Designing a multi-epitope peptide-based vaccine against SARS-CoV-2

Abhishek Singh[1, 2], Mukesh Thakur[1]*, Lalit Kumar Sharma[1] and Kailash Chandra[1]

[1]Zoological Survey of India, New Alipore, Kolkata- 700053, India

[2]Gujarat Forensic Sciences University, Sector 9, Gandhinagar, Gujarat 382007, India

*** Corresponding authors:**
Dr. Mukesh Thakur
Scientist C - Centre for DNA Taxonomy &
Coordinator - Centre for Forensic Sciences
Zoological Survey of India, New Alipore
Kolkata-700053, West Bengal, India
Mobile: +91-8171051282; FAX : 91-33-24008595
E. mail: thamukesh@gmail.com

Abstract

COVID-19 pandemic has resulted so far 14,395,16 confirmed cases with 85,711 deaths from the 212 countries, or territories. Due to multifacet issues and challenges in implementation of the safety & preventive measures, inconsistent coordination between societies-governments and most importanly lack of specific vaccine to SARS-CoV-2, the spread of Wuhan originated virus is still uprising after taking a heavy toll on human life. In the present study, we mapped several immunogenic epitopes (B-cell, T-cell, and IFN-gamma) over the entire structural proteins of SARS-CoV-2 and by applying various computational and immunoinformatics approaches, we designed a multi-epitope peptide based vaccine that predicted high immunogenic response in the largest proportion of world's human population. To ensure high expression of the recombinant vaccine in *E. coli*, codon optimization and in-silico cloning were also carried out. The designed vaccine with high molecular affinity to TLR3 and TLR4, was found capable to initiate effective innate and adaptive immune response. The immune simulation also suggested uprising high levels of both B-cell and T-cell mediated immunity which on subsequent exposure cleared antigen from the system. The proposed vaccine found promising by yielding desired results and hence, should be tested by practical experimentations for its functioning and efficacy to neutralize SARS-CoV-2.

Keywords: SARS-CoV-2; structural proteins; multi-epitope vaccine; immunoinformatics; computational modelling

Introduction

Coronavirus disease 2019 (COVID-19) is one of the most potent pandemics characterized by respiratory infections in Human. As per the latest report of World Health Organization (WHO) dated 10th of April 2020, the worldwide coverage of COVID-19 exceeded 14,395,16 confirmed cases with mortality of 85,711 deaths from 212 countries, areas or territories [1]. The outbreak of COVID-19 emerged at Wuhan city of Hubei Province of P.R. China in the early December of 2019 [2], but the report of severe pneumonic patients with unknown etiology of symptoms was released by the Chinese Center for Disease Control on December 31^{st,} 2019 [3]. Later on, the novel Betacoronavirus (SARS-CoV-2) was confirmed as the causative agent for the clusters of pneumonic cases by the health professionals. Investigations on the origin of infections linked the SARS-CoV-2 virus to a seafood market in Wuhan city, China which got spread to a large extent of human population in China in a very short period of time. By the time, scientists/ health professionals paid a serious note to understand the mode of transmission of COVID-19, it had reached and infected several people in different countries with an insight into the escalated transmission rate. This led the WHO to declare COVID-19 as a Public Health Emergency of International Concern (PHEIC) on 30th of January 2020 [4].

Until the outbreak of Severe Acute Respiratory Syndrome (SARS) in 2002 and 2003, coronaviruses were considered as a mild pathogenic virus causing respiratory and intestinal infections in animals and humans [5-9]. With a decade long pandemic of SARS, another potent pathogenic coronavirus, i.e., Middle East Respiratory Syndrome Coronavirus (MERS-CoV) appeared specifically in the Middle Eastern countries [10]. Extensive studies on these two coronaviruses have led to an understanding that they are genetically diverse but likely to be originated from bats [11, 12]. Apart from these two viruses, four other coronaviruses are commonly detected in Humans, viz., HKU1, 229E, OC43 and NL63 [13, 14]. Among all the coronaviruses reported so far, SARS-CoV, MERS-CoV, and SARS-CoV-2 have relatively high pathogenicity and escalated mortality rates. SARS CoV alone was responsible for 8422 positive cases and 919 probable fatalities over a vast population of 32 countries [15]. With the first mortality case reported in 2012, 2496 cases were found positive for MERS-CoV and 868 deaths

were reported from 27 countries [15]. Both of these pandemics were considered as highly pathogenic in the last two decades until the outbreak of SARS-CoV-2 in December 2019. SARS-CoV-2 has already spread in 212 countries and caused 85,522 deaths in just last three and half months [1]. As in the current scenario, several countries are facing the most fatal problem of community spread of SARS-CoV-2 e.g. China, Italy, Spain, France, Iran, USA, India, Germany etc. Though, the estimated mortality rate due to SARS-CoV-2 is about 3.6% in China and about 1·5% outside China [16], the problem is grave and need to be respond urgently to curb the rapid loss of human life.

Coronaviruses are broadly divided into four genera, i.e., Alpha-CoV, Beta-CoV, Gamma-CoV and Delta-CoV, where Alpha-CoV and Beta-CoV are only reported to infect mammals but Gamma-CoV and Delta-CoV are also capable to infect both mammals and birds [17]. Extensive study of their biology suggests that they are non-segmented enveloped viruses with a positive-sense single-stranded RNA of a size range of 26-32 kb [18]. Some of the recent research suggested that SARS-CoV-2 has an identical genomic characterization as of Beta-CoV and thus the genomic structure of SARS-CoV-2 follows 5'-leader-UTR-replicase-S (Spike)–E (Envelope)-M(Membrane)-N (Nucleocapsid)-3' UTRpoly (A) tail with the subsequent species-specific accessory gene at 3' terminal of the viral genome [19]. The structural protein includes Spike/Surface protein (S), Envelop protein (E), Membrane protein (M) and Nucleocapsid protein (N). These Spike/Surface proteins are known to have crucial roles in survival and viral pathogenesis as the receptor binding capability and entry to the host cell is regulated by the influence of Surface protein [20]. The Envelop and Membrane proteins are responsible for viral assembly and Nucleocapsid protein is necessary for RNA genome synthesis [20]. This complex genetic makeup and moderate mutation rate of SARS-CoV-2 requires the strategic development of vaccine by targeting all the structural proteins. Unfortunately, there is no approved vaccine available as on date against SARS-CoV-2. However, a few studies have predicted peptide-based vaccine targets for specific glycoproteins [21, 22] but no study has covered all the structural proteins for vaccine designing. With the recent development of integrated bioinformatics and immunoinformatics approaches, vaccine designing and its specific application has become rapid and cost effective as with the computational simulations and predictions, the targeted immunogenic peptide can be designed and validated for its efficacy to work on human, even before having the vaccine in hand for practical exposure and experiments. However,

conventional methods have simultaneously proven effective in vaccine designing with some limitation, e.g. sacrificing the whole organism or large protein residues which unnecessarily increases the antigenic load and probability of allergenicity [23]. We believe that an ideal multi-epitope vaccine candidate can enhance immune response and subsequently lower down the risk of re-infection by upraising the host immunogenicity. In this perspective, here we applied integrated approaches of vaccine designing by mapping B-cell, T-cell and IFN-gamma epitopes (humoral and adaptive immune response) of humans on the respective structural proteins based on several essential criteria.

Methodology

Retrieval of SARS-CoV-2 protein sequences and antigenicity prediction

All the protein sequences of SARS-CoV-2 available on NCBI (GenBank: MN908947.3) was retrieved and structural protein sequences were extracted for further analysis. We predicted antigenicity from each protein using the online server VaxiJen v2.0 [24] and proteins that showed antigenicity ≥ 0.4 in the virus category were subjected for further analysis.

Prediction of physiochemical and secondary structural properties of the target proteins

The physicochemical properties of target proteins were examined using Expasy Protpram online server [25] and the conformational states, e.g. Helix, sheet, turn and coil were predicted using the online Secondary Structure Analysis Tool (SOPMA) [26] with default parameters.

3 Dimensional homology modeling and validation

Three-dimensional structure of target proteins was modeled using three homology modeling tools,i.e. I Tasser, Raptor-X and Phyre2 [27-29] and post processing to reduce any distortions in the modeled structure was done using the Galaxy refine server [30].The modeled structures were subjected to RAMPAGE server for Ramachandran plot analysis for quality check and validation[31 and visualized using Chimera 1.10.1 visualization system [32].

Prediction of T-cell epitopes

Cytotoxic T-cell (CTL) epitopes

Nine residues long CTL epitopes recognized by HLA class-I supertypes, i.e., A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, B62 were predicted using NetCTL.1.2 server [33] considering the default values of weight on C terminal cleavage, weight on TAP transport efficiency and threshold for epitope identification. Further, another set of CTL epitopes recognized by HLA class I alleles like A*02:02, B*07:02, C*04:01, E*01:01, etc were identified

by consensus method of Immune Epitope Database (IEDB) tool [34]. The identified epitopes were scrutinized based on the consensus score i.e. ≤2 and only those epitopes were considered as strong binders and selected for further analysis which were predicted by more than one allele of both the method used.

Helper T-cell (HTL) epitopes

Fifteen residue long HTL epitopes recognized by HLA Class II DRB1 alleles were predicted using Net MHC II pan 3.2 server [35]. The thresholds for strong binder and weak binder were set as default. Further another set of HTL epitopes of 15 residues length recognized by HLA DR alleles were identified using the consensus method as implemented in the IEDB server [34] The epitopes with a percentile rank of ≤2 and predicted by more than one allele of both the methods were considered as a strong binder and scrutinized for further analysis.

Identification of overlapping T-cell epitopes

Considering the fact that epitopes with affinity for multiple HLA alleles tend to induce relatively more immune response in the host cell, we scrutinized overlapping epitopes with affinity to both HLA class I and class II alleles for immunogenicity and allergen prediction using VaxiJen v2.0 server [24] and AllerTop v.2.0 [36] tool with an immunogenic threshold of 0.4. These epitopes would have high potential to activate both CTL and HTL cells.

Identification of Continuous and Discontinuous B-cell epitopes

B-cell epitopes are short amino acid sequences recognized by the surface-bound receptors of B-lymphocyte. Their identification plays a major role in vaccine designing, and thus two types- Continuous and Discontinuous B- cell epitopes were predicted over the corresponding structural protein sequences. The continuous B-cell epitopes, present on the surface of proteins, were predicted using BCpred 2.0 [37] server and the discontinuous epitopes which were fairly large were predicted using Ellipro server [38]. The prediction parameters like Minimum score and Maximum distance (Anstrom) were set as 0.8 and 6, respectively.

Identification of IFN- gamma epitopes

Since, IFN gamma is the signature cytokine of both the innate and adaptive immune system and epitopes with potency to induce IFN gamma could boost the immunogenic capacity of any vaccine. Thus, IFN gamma epitopes were predicted for each structural protein by the IFNepitope server [39].

Characterization of predicted epitopes

Conservation analysis

Predicted T-cell and B-cell epitopes were submitted to the IEDB conservancy analysis tool [40] to identify the degree of conservancy in the structural protein sequences and the epitopes with 100% conservancy were selected for further analysis.

Population coverage and autoimmunity identification

Immunogenetic response of human population towards the selected overlapping CTL and HTL epitopes against their respective HLA genotype frequencies was predicted using IEDB population coverage analysis tool [41]. Those epitopes exhibiting 50% or more population coverage were scrutinized and to reduce the probability of autoimmunity, all the selected epitopes were subjected to BlastP search analysis against the Human proteome. Epitopes showing similarity to any human protein were excluded from the further analysis.

Interaction analysis of epitopes and their HLA alleles

The sequence of scrutinized epitopes was submitted to an online server PEPFOLD 3 [42] for 3D structure modeling. The structure of the most common HLA alleles i.e. HLA-DRB1 *01:01 (HLA class II) and HLA-A* 02:01 (HLA class I) in human population were retrieved from Protein data bank (PDB) with a PDB ID of 2g9h and 1QEW. Any ligand associated with the HLA allele's structure was removed and energy-minimization was carried out. Thereafter, the modeled epitopes were docked with the corresponding HLA allele using a web server Patchdock [43], to study the interaction pattern of receptor and ligand. The parameters like clustering RMSD value and the complex type was kept as 0.4 and default. The result obtained from Patchdock was forwarded to the Firedock server [44] for the refinement of the best 10 models. Based on Global energy value, the complex with the lowest global energy was scrutinized and the corresponding CTL and HTL epitopes were selected for the final vaccine construct.

Construction and quality control of final multi-epitope vaccine

For the construction of a multi-epitope vaccine against SARS-CoV-2, we followed Chauhan et al. [23] with some modifications. The epitopes were finally selected based on meeting the following criteria: 1). epitopes must be immunogenic, non-allergic and overlapping with affinity to both HLA class I and class II alleles; 2).epitopes must be capable to activate both CTL and HTL cells and must have a minimum 50% of the population coverage; 3).epitopes should not be overlapping with any human gene and the predicted B-Cell epitopes should overlap with finalized CTL and HTL epitopes and should be present on the surface of the target protein.

Based on these criteria, the HTL, CTL and IFN gamma epitopes were included in the final construct of the multi-epitope vaccine. The HTL and IFN gamma epitopes were linked by GPGPG linkers, whereas the CTL epitopes were linked by AAY linkers [23]. An adjuvant, 50S ribosomal protein L7/L12 (Locus RL7_MYCTU) with an NCBI accession no: P9WHE3 was added to the N terminal to enhance the immunogenicity of the constructed vaccine.

Antigenicity, allergenicity and physiochemical properties of vaccine

The antigenicity of the proposed vaccine was predicted using VaxiJen v2.0 server [24] and allergenicity was predicted using Allertop v.2.0 prediction tool [36]. The physicochemical property of the designed vaccine was predicted by submitting the final sequence of the vaccine to the Expasy Protpram online server [25].

Prediction of secondary structure

The secondary structure of the constructed vaccine was predicted using the SOPMA tool [26]. With the input of vaccine sequence and output width 70, the parameters like the number of conformational states, similarity threshold, and window width were set as 4 (Helix, Sheet, Turn, Coil), 8 and 17.

Modeling, refinement, and validation of vaccine construct

The final model of the constructed vaccine was prepared using a homology modeling tool SPARKS-X [45]. Among the top 10 predicted models, the model with the highest Z-score was selected and subjected for refinement using the Galaxy refine server [30]. Among all the refinement tools, Galaxy refine is one of the most efficient server according to CASP10 evaluation. Then after, we utilized ProSA-webtool [46] to assess the overall and local quality of the predicted model based on the Z-score. If the Z-score lies outside the range of native protein of similar size, a chance of error increases in the predicted structure. Further to check the overall quality of the refined structure of vaccine, Ramachandran plot analysis was carried out using the RAMPAGE server [31]. The final model of the vaccine construct was visualized using Chimera 1.10.1 visulaization system [32].

Prediction of Continuous and Discontinuous epitopes

Continuous and Discontinuous B- cell epitopes in the vaccine construct were predicted using BCpred 2.0 [37] and IEDB server [34], respectively. For effective designing of vaccine, the constructed vaccine should posses' B-cell epitopes for potential enhancement of immunogenic reactions.

Molecular docking of constructed vaccine with TLR3 and TLR4 receptor

The 3D structures of human TLR3 and TLR4 were retrieved from protein data bank (PDB ID: 2A0Z and 2z63). Any ligand attached to the retrieved structures was removed and the receptor was subjected for model refinement using Galaxy refine server [30]. Molecular docking analysis was performed using the Cluspro v.2 protein-protein docking server [47] to analyze the interaction pattern of vaccine with TLR3 and TLR4. The server provides cluster scores based on rigid docking by sampling billions of conformation, pairwise RMSD energy minimization. Based on the lowest energy weight score and members, the final vaccine-TLR3 complex and TLR4 complex model was selected and visualized using Chimera 1.10.1 visulaization system [32] .

***In silico* cloning and vaccine optimization**

Efficiency of the multi-epitope vaccine construct in cloning and expression is of the utmost importance for vaccine designing. The final multi-epitope vaccine construct was submitted to the Java Codon Adaptation Tool (JCat) [48], where codon optimization was performed in *E. coli* strain K12. In addition to this, the parameters like Avoid rho-independent transcription terminators, Avoid prokaryotic ribosome binding sites, Avoid Cleavage Sites of Restriction Enzymes were selected for the final submission. The output of the JCat tool provided CAI value and GC content of the optimized sequence. Ideally, the CAI value should be greater than 0.8 or nearer to 1 [49], and the GC content should be in a range of 30 to 70% [50] to ensure that the vaccine has potential translation, stability and transcription efficiency. Further, an in silico cloning of a multi-epitope vaccine construct in *E. coli* pET– 28(+) vector was performed using the SnapGene 4.2 tool [51].

Immune simulations of vaccine construct

To characterize the real-life immunogenic profiles and immune response of the multi-epitope vaccine, C-ImmSim server [52] was utilized. Position-specific scoring matrix (PSSM) and machine learning are the two methods, based on which C-ImmSim predicts immune epitopes and immune interactions. It simultaneously simulates three different anatomical regions of mammals, i.e. Bone marrow, Thymus and tertiary lymphatic organs. To conduct the immune simulation, three injections at an interval of four weeks were administered with 1000 vaccine molecules per injection. The parameters like Random seed, simulation volume, and simulation step were kept as 12345, 10µl and 1050. Several studies suggested that the minimum interval between two injections should be kept four weeks and thus the time steps followed for three injections were 1,

84 and 168 where each time step is equal to 8 hours of real-life [53]. Since, several patients wordwide got recurrent infection of SARS-CoV-2, 12 consecutive injections were administered four weeks apart for the assessment of after effects of vaccine exposure on SARS-CoV-2.

Results

Target proteins and their antigenicity

We scrutinized target structural protein sequences, i.e. Surface, Membrane, Envelop and Nucleocapsid proteins from the whole-genome of SARS-CoV-2, 29903 bp ss- RNA (GenBank: MN908947.3). Antigenicity prediction from the structural protein sequences revealed that Envelop proteins has the highest antigenicity score 0.6025, followed by Membrane, Nucleocapsid and Surface protein (table 1).

Physiochemical and secondary structural properties

Theoretical estimates of isoelectric point (PI) suggested that all the structural proteins were basic in nature and their predicted instability indices were below 40 indicating their stable nature except the Nucleocapsid protein with an instability index of 55.09 (table S1). The predicted secondary structure characteristics of all the structural proteins showed varaible percentage of Alpha Helix, Extended Strand, Beta Turn and Random Coil (table S2).

3 Dimensional structure of target proteins

The refined finalized 3D structures obtained from Raptor-X was better than the models generated from I Tasser and Phyre2 as Raptor-X models showed least percentage of outliers and therefore selected for further analysis (figure 1). Details of various models of the structural proteins through Ramachandran plot analysis are given in table S3.

Prediction of CTL and HTL epitopes

A total of 29 CTL epitopes in Envelop protein, 77 in Nucleocapsid, 340 in Surface and 89 epitopes in Membrane protein were predicted with strong binding affinity for multiple alleles. Similarly, 35 HTL epitopes in Envelop protein, 51 in Nucleocapsid, 252 in surface and 23 epitopes in Membrane protein were predicted. All these epitopes were strong binders and predicted percentile rank of ≤ 2.

Identification of overlapping T cell epitopes

The CTL epitopes overlapping with HTL epitopes were scrutinized and subjected to immunogenicity and allergenicity prediction. In total, 16 CTL overlapped epitopes were predicted in Envelop protein, 8 in Nucleocapsid protein, 25 in Surface protein and 9 epitopes in Membrane protein (Table S4-S11). All these selected epitopes were predicted to be non-allergic with high antigenic scores.

Identification of B-cell epitopes

The identified continuous B cell epitopes present on the surface of their respective proteins were selected based on their antigenic, non-allergic and overlapping nature in the finalized CTL and HTL epitopes. With these criteria, five epitopes were obtained in Surface protein, three in Nucleocapsid and one in Membrane protein (table S12). No epitope with such criteria was identified in Envelop protein. Further, one discontinuous epitope in Envelop protein, seven in Nucleocapsid, nine in Surface and four discontinuous epitopes were predicted in Membrane protein (table S13).

Identification of IFN- gamma epitopes

The server IFNepitope predicted eight epitopes in Envelop protein, 39 in Membrane, 66 in Nucleocapsid and 281 in Surface protein. Based on antigenicity and non-allergenicity, eight epitopes were scrutinized in Envelop protein, 13 in Membrane, 43 in Nucleocapsid and 81 in Surface protein. To obtain the highest immunogenicity and reduce the overload of epitopes in vaccine construct, we selected one epitope from each protein with the highest antigenicity score and non-allergenicity (table S14).

Characterization of predicted epitopes

Conservation analysis, Population coverage, and autoimmunity identification

The conservation prediction resulted that all the selected T cell and B cell epitopes were 100% conserved among the target structural proteins and population coverage prediction showed that five epitopes in Envelop protein, two in Membrane, two in Nucleocapsid and six epitopes in Surface protein covered more than 50% of worldwide population. No selected epitope with >50% population coverage showed homology with any human protein. Thus, altogether 30 epitopes, 15 each of CTL and HTL, met the above criteria were processed for the interaction analysis with the commonly occurring HLA alleles in the Human population (table S15).

Interaction analysis of epitopes and their HLA alleles

Molecular interaction analysis of each selected CTL and HTL epitopes with their respective HLA alleles resulted in 10 predicted docked complex. Each complex showed different Global energy value. Out of the characterized 30 epitopes, the epitopes with highest negative value of Global energy (set threshold -40) were considered as effectively interacting epitopes with the corresponding HLA allele, and thus selected for final vaccine construct (table 2). This resulted in two CTL and HTL epitopes in Envelop protein, one in Membrane, two in Nucleocapsid and two in Surface protein. For an ideal vaccine, the epitopes included should have potential affinity towards their respective alleles and thus greater the interaction between epitopes and their alleles, more is the affinity.

Construction of final multi-epitope vaccine

We selected seven CTL, seven HTL and four IFN gamma epitopes for construction of the multi-epitope vaccine (figure S1). The adjuvant was coupled by EAAAK linker with CTL epitope and subsequently, AAY linker was used to couple CTL epitopes and GPGPG linker was used to couple HTL and IFN gamma epitopes (figure 2). The final multi-epitope vaccine construct was composed of 430 amino acid residues which was then validated for antigenic, allergenic and physiochemical properties.

Evaluation of antigenicity, allergenicity, physiochemical properties and secondary structure of the vaccine construct

The multi-epitope vaccine construct found to be immunogenic (Ag score- 0.627), non-allergic with 45.131 KDa predicted molecular weight. The prediction suggested that vaccine construct was stable and basic in nature with instability index 23.46 and theoretical PI 9.01. The half-life of the vaccine in mammalian reticulocytes (in vitro) was 30 hrs, while it was about 20 hrs in yeast (in-vivo) and 10 hrs in *E. coli* (in vivo).

The predicted secondary structure of the vaccine constructed consisted 42.33% alpha-helix, 22.09% extended strand, 4.42% beta-turn and 31.16% random coil.

Modeling, refinement, and validation of vaccine construct

Of the top 10 predicted models, the model with the Z-score of 7.61 with template 1dd3A was selected for refinement. The ProSA-web tool assessed the overall and local quality of the refined model with a Z-score of -0.47 which was close to the range of native protein of similar size (figure 3). Further, the Ramachandran plot analysis revealed 92% of residues in the favored

region, 6.1% in the allowed region and 1.9% outliers. This signified that the quality of the predicted vaccine construct was adequate for further analysis.

Prediction of Continuous and Discontinuous epitopes

BCpred 2.0 server predicted four continuous epitopes i.e. AKILKEKYGLD, EILDKSKEKTSFD, LKESKDLV and VPKHLKKGLSKEEAESLKKQLEEV on the surface of predicted vaccine (figure S2), while IEDB predicted six discontinuous epitopes (figure S3) with a score higher than 0.8 (Table S16).

Molecular docking of constructed vaccine with TLR3 and TLR4 receptor

Cluspro v.2 predicted 30 models each of vaccine receptor TLR3 complex and TLR4 complex with their corresponding cluster scores (Table S17-S18). Among these models, the model number 2 in TLR3 complex and model number 1 in TLR4 complex were selected as a best-docked complex with the lowest energy score of -1199.1 with 46 members (TLR3) and lowest energy score of -1229.9 with 79 members (TLR4). This signifies potential molecular interaction between predicted vaccine construct with TLR3 and TLR 4 receptors (figure 4).

***In silico* cloning and vaccine optimization**

The Java Codon Adaptation Tool optimized the codon usage of the vaccine and produced an optimized codon sequence of length 1290 nucleotides. The CAI value of optimized sequences was 0.95, close to 1 and the GC content was found to be 54.41 %, fall between the range of 30 to 70%. Therefore, the adaptation was satisfactory indicating potential expression of vaccine construct in host *E.coli*. Later on, the restriction sequence of Xho I and Not I restriction enzymes were added to N and C terminal of adapted codon sequence. Further, the adapted sequence was also cloned in pET– 28(+) vector using SnapGene tool (figure 5).

Immune simulations of vaccine construct

C-ImmSim server-generated immune response was consistent with the actual immune response as depicted by the results (figure 6 and S4). The primary response was generated as marked by the increase in the level of IgM. Similarly, the secondary response was characterized by an increased level of IgM+IgG, IgG1+IgG2, IgG1, IgG2, and B-cell populations. On the subsequent exposure of vaccine, a decrease in the level of antigens was observed indicating the development of immunogenic response in the form of immune memory. Both the T cell populations i.e. CTL

and HTL developed increased response corresponding to the memory cells signifying the immunogenicity of T cell epitopes included in the vaccine construct. Increased activity of macrophages was also observed at each exposure, whereas the activity of NK cells was observed consistent throughout the period. A significant increase in the level of IFN gamma, IL-10, IL-23, and IL-12 was also observed at subsequent exposure (figure 6). After the repeated exposure of vaccine construct through 12 injections after a regular interval, a remarkable increase in the level of IgM+IgG, IgG1+IgG2. Though, IgG1 increased slightly but IgM and IgG2 increased consistently. A surge in the memory corresponding to B cell and T cells was observed throughout the exposure while the level of IFN gamma was consistently high from first to last exposure (figure S4). This signifies that the vaccine proposed in this study has generated a strong immune response in case of short exposure and immunity increases even on subsequent repeated exposure.

Discussion

The recent outbreak of SARS-CoV-2 in Wuhan city of China raised several questions regarding the susceptibility of Humans against these novel pathogens. As predicted earlier about the re-emergence of novel variants of SARS-CoV [54], SARS-CoV-2 proved to be more lethal as compared to the previous variants in terms of wide spread of virus in a short period of time. This escalated rate of transmission initiated a pursuit for a vaccine development against SARS-CoV-2, which is currently a worldwide pandemic with over 1.43 million cases and 85, 711 deaths [1]. With strict and comprehensive measures, vaccine development and its application could play a key role in eliminating the virus from the Human population or to restrain the spread among individuals and different populations. Thus several efforts are being made to address the challenge that appeared in the current scenario with appreciable advancements in the understanding of virus biology and its etiology. The lack of knowledge regarding the response of the immune system against viral infection is one of the major limitations in the path of vaccine development for SARS-CoV-2.

This study is the piolt attempt in describing the potential immunogenic target over the structural proteins and proposes a novel multi-epitope vaccine construct, by providing new rays of hope in the initial phase of vaccine development. Certain criteria like poor antigenicity, allergenicity, low affinity towards the immune cells, autoimmunity and oversize that could influence the

effectiveness, have been evaluated on the proposed vaccine construct following various computational and immunoinformatics approach. The retrieved structural proteins and their antigenicity score suggested that the Envelop protein is the most potent protein to generate immune response considering its role in viral assembly. Since, immunity against any pathogen is prominently dependent, how it gets recognized by B cells and T cells. We identified epitopes corresponding to B cells and T cells in each structural protein so that both humoral and cellular immunity can be induced with the exposure of vaccine construct. The proposed vaccine construct is designed based on the epitopes that have been selected with the most robust criteria,e.g. their nature of antigenic, non-allergic, 100% conserved among the target proteins, their affinity for multiple alleles, no homology with any of the human proteins, their worldwide coverage of human population (>50% of worldwide population covered) and effective molecular interaction with their respective HLA alleles. Further,, IFN gamma epitopes which are equally effective in immune regulatory, antiviral and antitumor activities, the final vaccine construct was designed with also due considering of the IFN gamma epitopes to increase the immune recognition ability. The designed vaccine of 430 amino acid residues has a molecular weight of 45.131K Da, also falls in the defined range of average molecular weight of a mullti-epitope vaccine. The instability index and theoretical PI also suggested that the vaccine is stable and basic in nature and the estimated half-life suggested that the recognized peptide does not possess a short half-life and would remain viable for a span adequate to generate a potential immune response. Vaccine model refinement and validation indicated that quality of the predicted model was good as more than 90% residues were in the favored region [55]. The B-cell continuous epitopes predicted over the surface of the vaccine suggested that the vaccine is capable to get recognized by B-cells indicating its effectiveness. TLR 3 and TLR 4 have proven recognition capability in both SARS-CoV and MERSCoV. Considering the similar genome characterization of SARS-CoV and SARS- CoV-2, the molecular interaction of vaccine with TLR 3 and TLR 4 through docking analysis suggested that the constructed vaccine has a significant affinity towards the toll-like receptors to act as sensor for recognizing molecular patterns of pathogen and initiating immune response. Thus the vaccine TLR complex is capable of generating an effective innate immune response against SARS-CoV-2. Further, Codon adaptation improved the expression of the recombinant vaccine in *E coli*. Strain K12 with significant codon adaptation index and GC content indicating elevated expression level. After successful cloning of the recombinant vaccine

in the pET– 28(+) vector, the simulation based generated immune response suggested that the primary and secondary immune response will be coherent with the actual expected response. Subsequent exposure for three injections indicated the potential capability of vaccine for generating memory cells and a higher probability of significant immune response on later exposure to coronavirus. The consistent high level of IFN gamma also supported the activation of humoral immunity. To understand the immunogenic potential of vaccine, repeated exposure in the form of 12 injections was administered and the result suggested the consistent generation of a strong immune response. With all these immunoinformatics approaches, the vaccine designed against SARS-CoV-2 showed promising results in inducing immune response.

Without proper control methods or vaccines, it is difficult to curb coronavirus pandemic given the current situation. However, several medications have been tested against SARS-CoV-2, but none of them showed complete effectiveness. In this study, a vaccinomics approach was carried out to design a multi-epitope vaccine against SARS-CoV-2 using several in silico and immunoinformatic approaches. Based on the computational and immunoinformatics approaches, we propose that the designed vaccine has all the potential to induce both the innate and adaptive immune systems and can neutralize the SARS-CoV-2, and therefore, an experimental validation of the designed vaccine must be undertaken by the professional for reaching a conclusion for its safety, efficacy and success.

Acknowledgements

We thank Zoological Survey of India, Kolkata for providing access to the computational support, working space and facilities to undertake the present study.

References

1. World Coronavirus disease 2019 (COVID-19) Situation Report – 77, 9th April 2020.
2. Zhang, Y.Z., Holmes, E.C. A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. *Cell*. 2020.
3. World-Health-Organization Coronavirus disease (COVID-19) outbreak. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed on 3rd April 2020).

4. World-Health-Organization Statement on the second meeting of the International Health Regulations(2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). Available online: https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-ofthe-international-health-regulations-(2005) emergency-committee-regarding-the-outbreak-of-novelcoronavirus-(2019-ncov) (accessed on 29 March 2020).
5. Masters, P. S., Perlman, S. In Fields Virology **Vol. 2** (eds Knipe, D. M. &Howley, P. M.) 825–858 (Lippincott Williams & Wilkins, 2013).
6. Zhong, N. S. et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet*. **362**, 1353–1358 (2003).
7. Drosten, C. et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med*. **348**, 1967–1976 (2003).
8. Fouchier, R. A. et al. Aetiology: Koch's postulates fulfilled for SARS virus. *Nature*. **423**, 240 (2003).
9. Ksiazek, T. G. et al. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med*. **348**, 1953–1966 (2003).
10. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med*. **367**, 1814–1820 (2012).
11. Lau, S. K. et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl Acad. Sci*. **102**, 14040–14045 (2005).
12. Ithete, N. L. et al. Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerg. Infect. Dis*. **19**, 1697–1699 (2013).
13. Su, S. et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol*. **24**, 490–502 (2016).
14. Forni, D., Cagliani, R., Clerici, M., Sironi, M. Molecular evolution of human coronavirus genomes. *Trends Microbiol*. **25**, 35–48 (2017).
15. Meo, S A., et al. Novel Coronavirus 2019-nCoV: Prevalence, Biological and Clinical Characteristics Comparison With SARS-CoV and MERS-CoV. *European Review for Medical and Pharmacological Sciences*. **24**, 4, pp. 2012-2019 (2020).

16. Baud, D. et al. Real estimates of mortality following COVID-19 infection**,** *The Lancet Infectious diseases*. (2020).
17. Woo, P. C. et al. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J. Virol*. **86**, 3995–4008 (2012).
18. Su, S. et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol*. **24**, 490-502 (2016).
19. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*. (2020).
20. Schoeman, D., Fielding, B.C. Coronavirus envelope protein: current knowledge. *Virol J*.**16**, 69 (2019).
21. Abdelmageed, M.I. et al. Design of multi epitope-based peptide vaccine against E protein of human 2019-nCoV: An immunoinformatics approach. *bioRxiv*. (2020).
22. Bojin, F., Gavriliuc, O., Margineanu, M., Paunescu, V. Design of an Epitope-Based Synthetic Long Peptide Vaccine to Counteract the Novel China Coronavirus (2019-nCoV). *Preprints*. (2020).
23. Chauhan, V. et al. Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Sci Rep*. **9**, 2517 (2019).
24. Doytchinova, I. A., Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. **8**, 4 (2007).
25. Wilkins, M. R. et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol*. **112**, 531–552 (1999).
26. Geourjon, C., Deleage, G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*. **11**, 681–684 (1995).
27. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nature methods*. **12**, 7 (2015).
28. Källberg, M. et al. Template-based protein structure modeling using the RaptorX web server. *Nature protocols*. **7**, 1511 (2012).

29. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*. **10**, 845 (2015).
30. Heo, L., Park, H., Seok, C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic acids research*. **41,** 384-388 (2013).
31. Lovell, S. C. et al. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*. **50**, 437–450 (2003).
32. Pettersen, E. F. et al. UCSF Chimera- a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. **13**, 1605–1612 (2004).
33. Larsen, M. V. et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*. **8**, 424 (2007).
34. Kim, Y. et al. Immune epitope database analysis resource. *Nucleic acids research*. **40**, 525–530 (2012).
35. Jensen, K. K. et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*. **154(3),** 394-406 (2018).
36. Bangov, I., Doytchinova, I., Dimitrov, I., Flower, D. AllerTOP v.2 - A server for in silico prediction of allergens. *J. Mol. Model*. **20**, 2278-2284 (2014)..
37. Manzalawy, Y.EL., Dobbs, D., Honavar, V. Predicting linear B- cell epitopes using string kernels. *Journal of Molecular Recognition: An Interdisciplinary Journal*. **21**, 243–255 (2008).
38. Ponomarenko, J. et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC bioinformatics*. **9**, 514 (2008).
39. Dhanda, S. K., Vir, P., Raghava, G. P. Designing of interferon-gamma inducing MHC class-II binders. *Biology direct*. **8**, 30 (2013).
40. Bui, H. H., Sidney, J., Li, W., Fusseder, N., Sette, A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC bioinformatics*. **8**, 361 (2007).
41. Bui, H. H. et al. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC bioinformatics*. **7,** 153 (2006).
42. Lamiable, A. et al. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic acids research*. **44**, 449–454 (2016).

43. Schneidman, D., Inbar, Y., Nussinov, R., Wolfson, H. PatchDock and SymmDock: Servers for Rigid and Symmetric Docking. *Nucleic acids research*. **33**. 363-7 (2005).
44. Mashiach, E., Schneidman, D., Andrusier, N., Nussinov, R., Wolfson, Haim. FireDock: A Web Server for Fast Interaction Refinement in Molecular Docking. *Nucleic acids research*. **36**. 229-32 (2008).
45. Y. Yang., E. Faraggi., H. Zhao., Y. Zhou. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics*. **27(15),** 2076–2082 (2011).
46. Wiederstein, M., Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*. **35**, 407–410 (2007).
47. Kozakov, D. The ClusPro web server for protein-protein docking. *Nat. Protoc.* 12, 255–278 (2017).
48. Grote, A. et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research*. **33**, 526–531 (2005).
49. Morla, S., Makhija, A., Kumar, S. Synonymous codon usage pattern in glycoprotein gene of rabies virus. *Gene*. **584**, 1–6 (2016).
50. Ali, M. et al. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci Rep*. **7**, 9232 (2017).
51. SnapGene software (from Insightful Science; available at snapgene.com).
52. Rapin, N., Lund, O., Bernaschi, M., Castiglione, F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One*. **5**, e9862 (2010).
53. Castiglione, F., Mantile, F., De Berardinis, P., Prisco, A. How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Comput. Math. Methods Med*. 1–9 (2012).
54. Cui, J., Li, F., Shi, Z. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*. **17**, 181–192 (2019).

574 55. Wlodawer, A. Stereochemistry and validation of macromolecular structures. *Methods*
575 *Mol. Biol.* **1607**, 595–610 (2017).

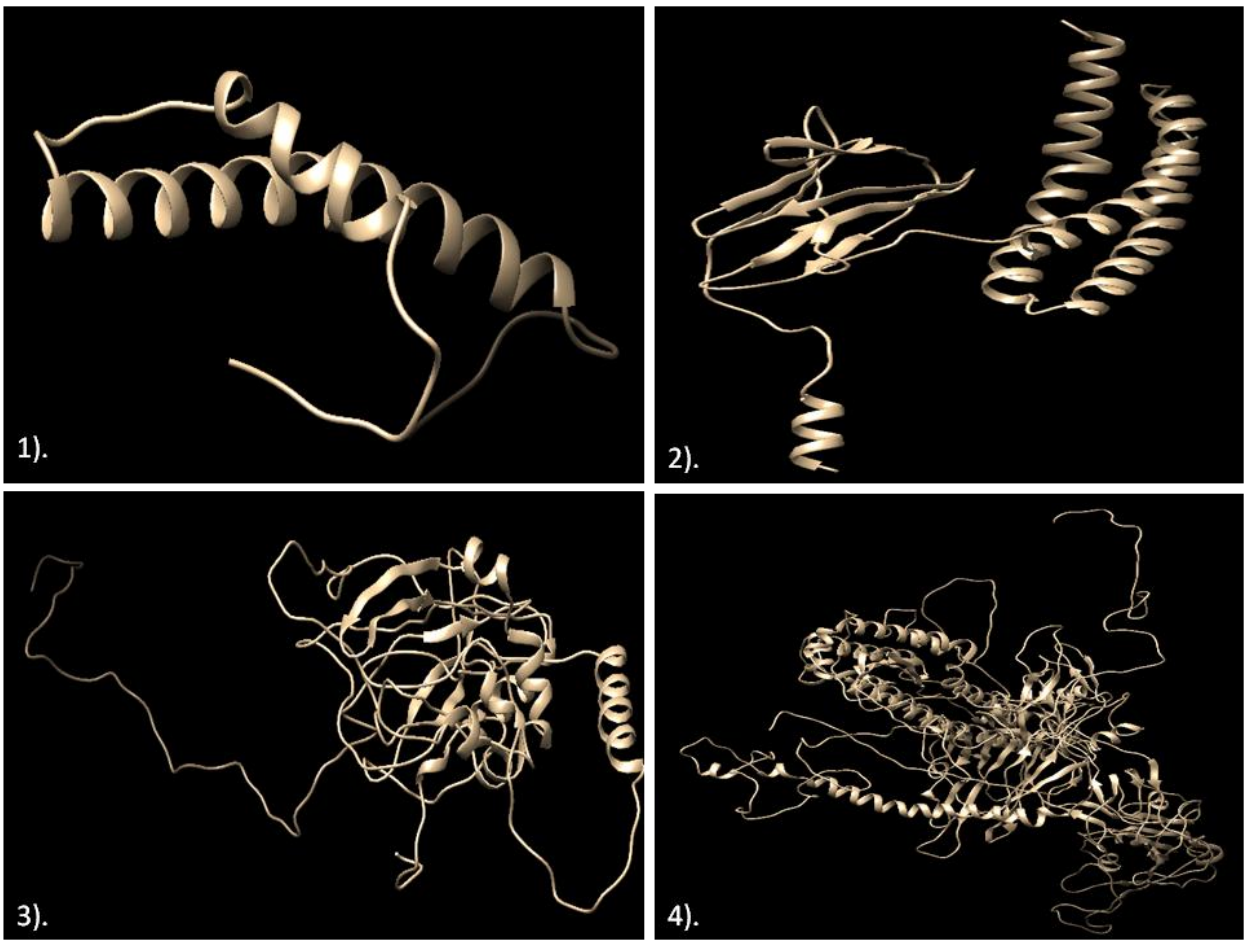

Figure 1). The 3D models of structural proteins of SARS- CoV-2.1). Envelop protein; 2). Membrane protein; 3). Nucleocapsid protein; 4). Surface protein

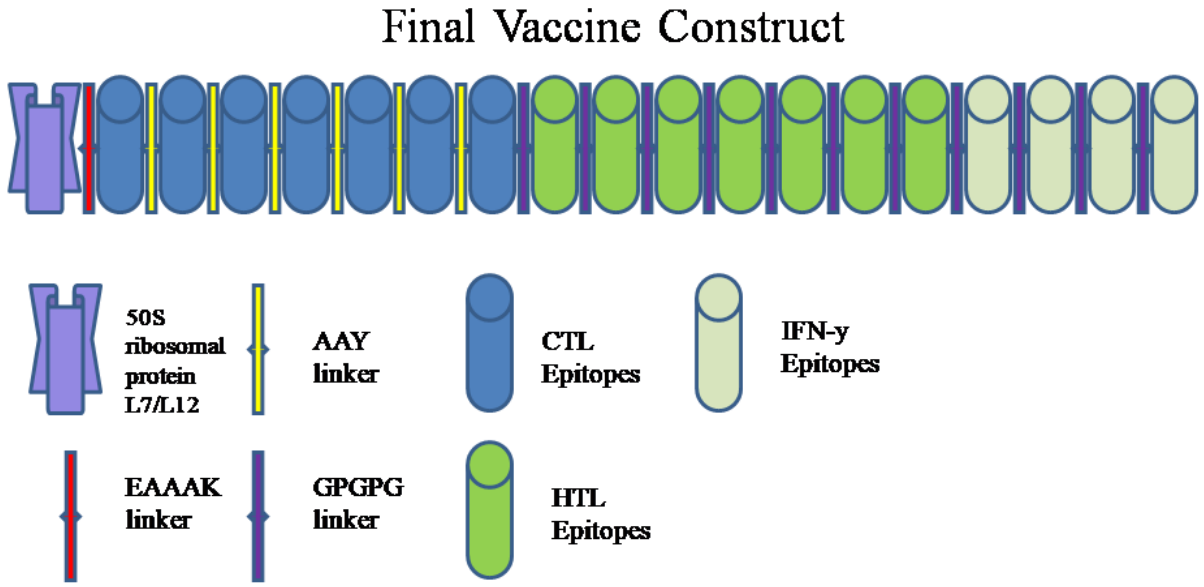

Figure 2. Schematic diagram of final vaccine construct

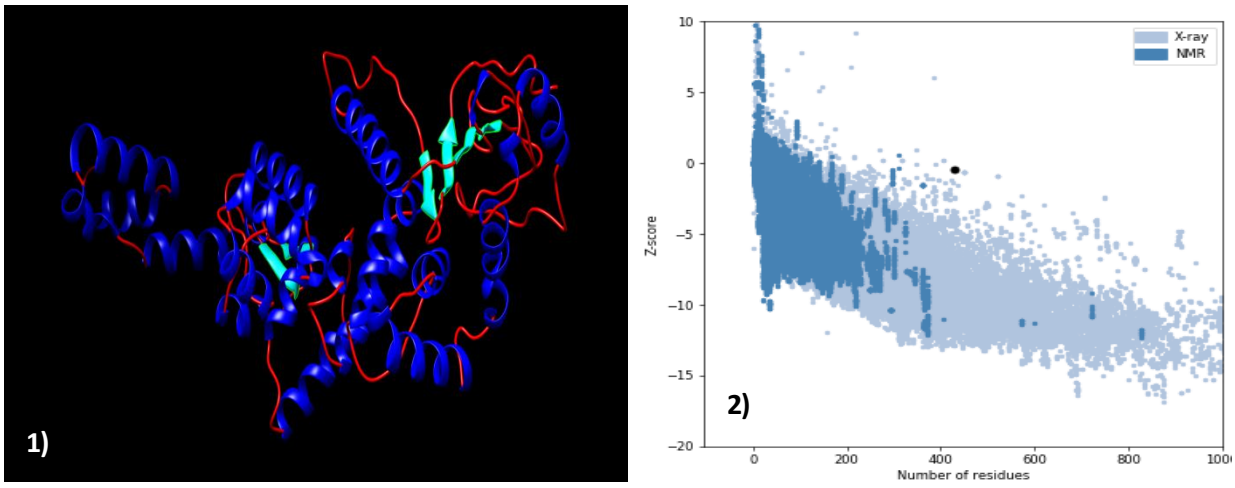

Figure 3. 1). 3 dimensional structure of final vaccine construct, 2). ProSA validation of predicted structure with Z score of -0.47.

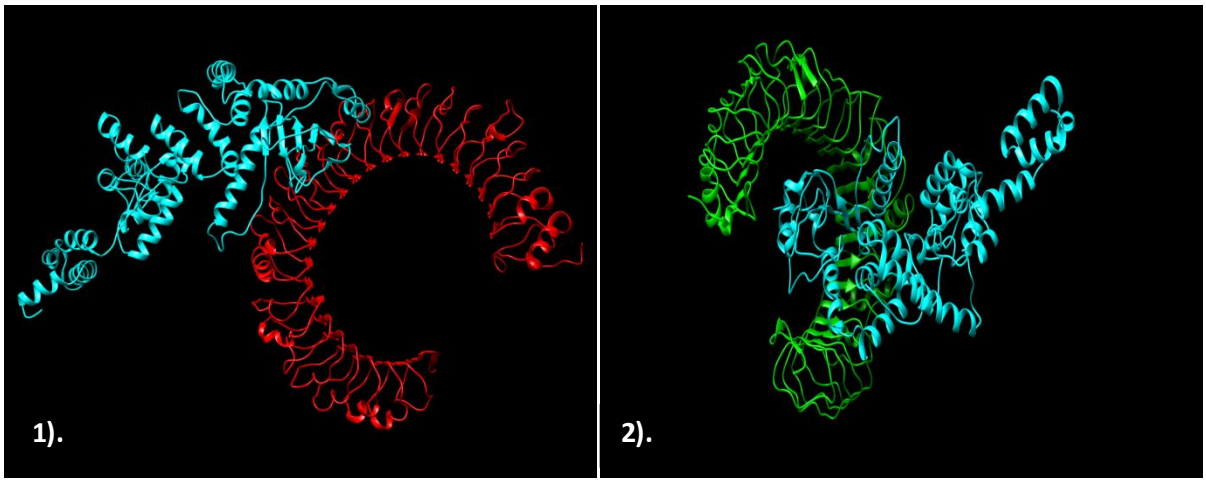

Figure 4. The interaction pattern of designed vaccine with TLR3 and TLR4. **1).** Vaccine (Cyan) docked with receptor TLR3 (Red). **2).** Vaccine (Cyan) docked with receptor TLR4 (Green).

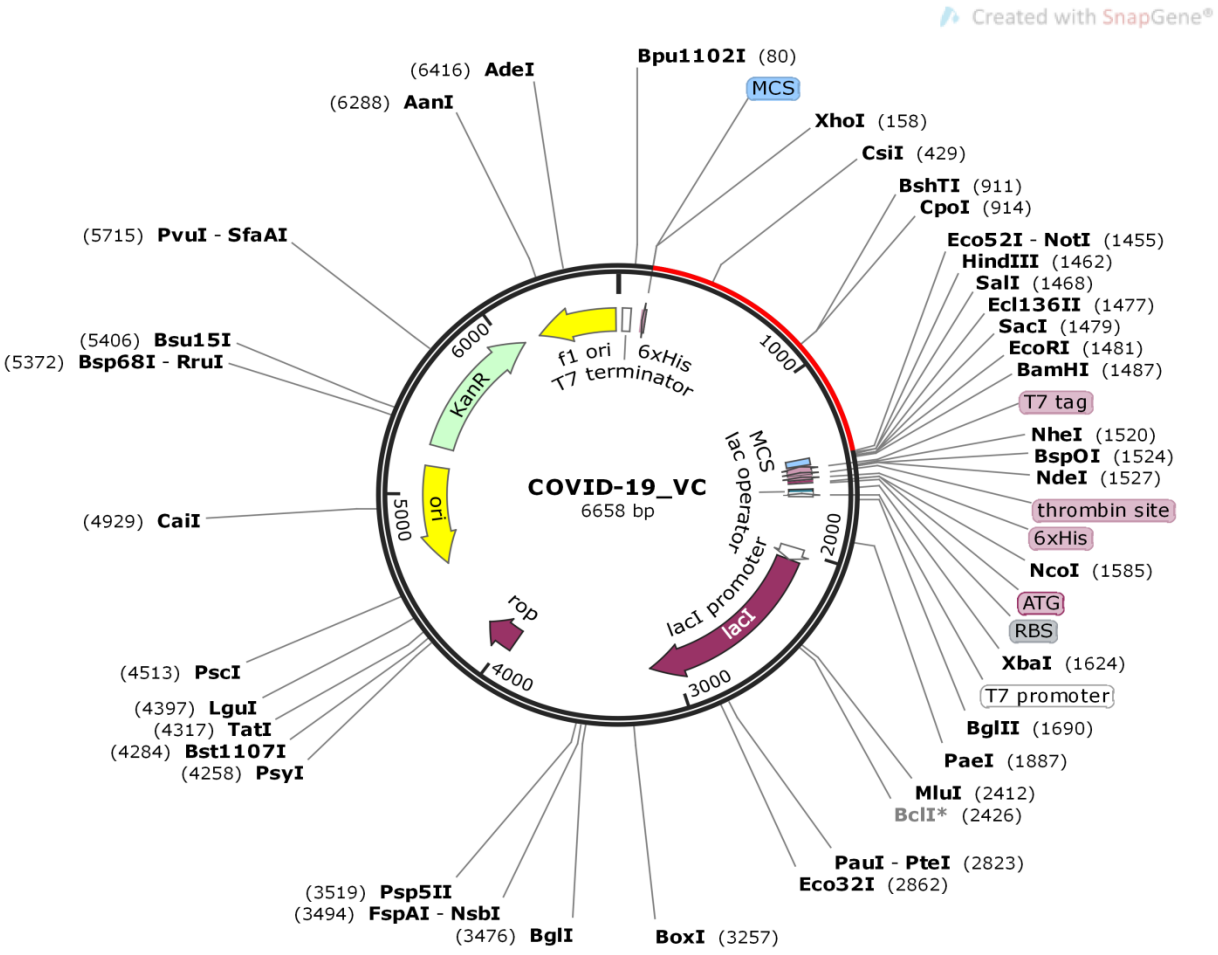

Figure 5. In silico cloning of vaccine. The segment represented in red is the multi-epitope vaccine insert in pET– 28(+) expression vector.

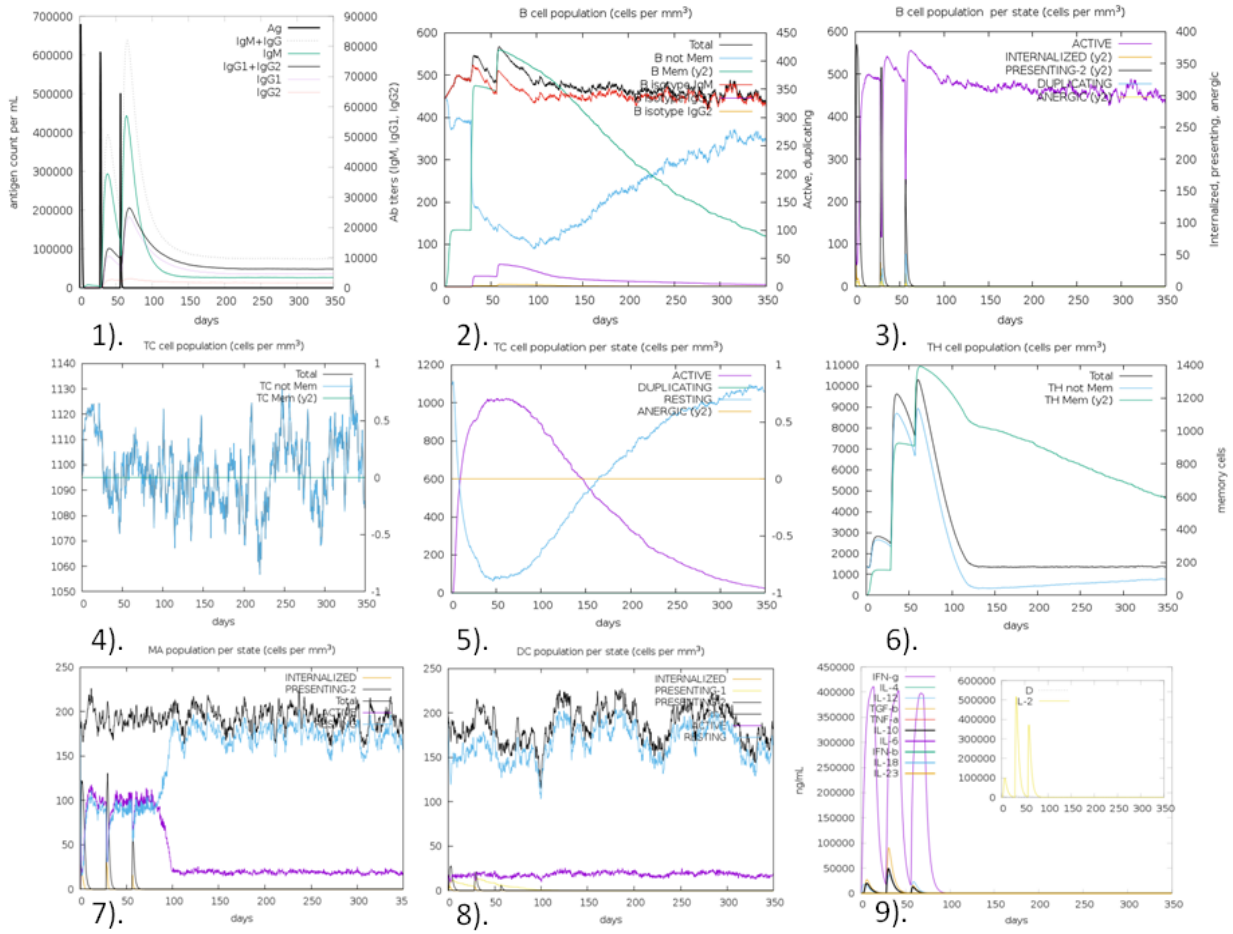

Figure. 6. In silico simulation of immune response using vaccine as an antigen after subsequent three injections. 1). Antigen and Immunoglobins. 2). B-cell population. 3). B-cell population per state. 4). Cytotoxic T-cell population. 5). Cytotoxic T-cell population per state. 6). Helper T-cell population. 7). Macrophages population per state. 8). Dendritic cell population per state. 9). Cytokine production.

Table 1. Details of target protein and their antigenicity score

Structural Protein	Protein ID (NCBI)	Size (aa)	Antigenicity score
Surface protein	QHD43416.1	1273	0.4646
Membrane protein	QHD43419.1	222	0.5102
Envelop protein	QHD43418.1	75	0.6025
Nucleocapsid protein	QHD43423.2	419	0.5059

Table 2. Final epitopes selected for multi-epitope vaccine construct after docking analysis with respective HLA alleles

S.No	Protein	HTL	Global_Energy	CTL	Global_Energy
1	Envelop	LLFLAFVVFLLVTLA	-71.44	LLFLAFVVF	-41.45
2	Envelop	LAFVVFLLVTLAILT	-64.39	FLLVTLAIL	-43.31
3	Membrane	TLACFVLAAVYRINW	-53.12	TLACFVLAA	-54.85
4	Nucleocapsid	GDAALALLLLDRLNQ	-87.95	GDAALALLL	-40.49
5	Nucleocapsid	AQFAPSASAFFGMSR	-58.74	AQFAPSASA	-43.29
6	Surface	IPFAMQMAYRFNGIG	-46.33	FAMQMAYRF	-41.37
7	Surface	FVFLVLLPLVSSQCV	-74.92	FVFLVLLPL	-55.15