In-silico design of a multi-epitope recombinant vaccine against SARS-CoV-2 targeting the receptor-binding domain

Harshawardhan Pande (✉ harshu.pande@slu.edu)
Saint Louis University

Research Article

Keywords: COVID-19, in-silico, reverse vaccinology, immunoinformatics, multi-epitope, vaccine

Posted Date: June 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1725962/v1

License: ☺️ ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

The COVID-19 pandemic caused by the SARS-CoV-2 virus has posed several challenges in the development of efficacious epitope-based vaccines. While multiple bioinformatic studies have attempted to predict epitopes, they have focused on the whole spike protein without considering antibody-mediated enhancement or Th-2 immunopathology, and some have missed some important but less antigenic epitopes in the receptor binding domain. Therefore, this study used in-silico methods to design and evaluate a potential multiepitope vaccine that specifically targets the receptor-binding domain due to its critical function in viral entry. Immunoinformatic tools were used to specifically examine the receptor-binding domain of the surface glycoprotein for suitable T cell and B cell epitopes. The selected 5 B cell and 8 T cell epitopes were then constructed into a subunit vaccine and appropriate adjuvants along with the universal immunogenic PADRE sequence were added to boost efficacy. The structure of the vaccine construct was predicted through a de-novo approach and molecular docking simulations were performed which demonstrated high-affinity binding to TLR 5 receptor and appropriate HLA proteins. Finally, the vaccine candidate was cloned into an expression vector for use as a recombinant vaccine. Similarities to some recent epitope mapping studies suggest a high potential for eliciting neutralizing antibodies and generating a favorable overall immune response.

Introduction

The COVID-19 pandemic has rapidly escalated to historic proportions, with over 500 million confirmed cases and 6 million deaths globally at the time of this writing. SARS-CoV-2, the virus responsible for this pandemic, is a 29,903-nucleotide single-stranded RNA virus belonging to the Coronaviridae family. It is part of the genus Beta-coronavirus and shares close genetic resemblance to the SARS-CoV virus responsible for the outbreak in 2003 [1]. The COVID-19 pandemic likely originated from a seafood market in Wuhan, China, and some studies predict that it was first transmitted to humans through zoonotic transfer from bats [2]. Like other coronaviruses, SARS-CoV-2 consists of 4 main structural proteins: Surface glycoprotein(S), Envelope protein(E), Membrane protein(M), and the Nucleocapsid protein(N) [3]. The Surface Glycoprotein (spike protein) mediates viral entry into the host cell through the binding of its S1 subunit with the human ACE2 receptor. However, the novel Coronavirus differs from other Beta-coronaviruses like SARS in terms of its significantly higher infectivity and low mortality rate. Some explanations point to the presence of its furin cleavage site and considerably stronger affinity between the human ACE2 receptor and the Receptor Binding Domain found in the S1 subunit of the spike protein [2].

Due to the novel nature of the virus, traditional vaccine approaches have been challenging. The first-generation vaccine approaches, which include inactivated and live-attenuated vaccines, prove ineffective against SARS-CoV-2 due to a high risk of infection and difficulty in production. Instead, several pharmaceuticals and researchers have shifted the focus to mRNA and DNA vaccines that eliminate the risk of any unwanted reactions. However, these novel methods may face logistical regulatory roadblocks, and inherently carry little antigenicity due to their purity [4]. Subunit vaccines offer a viable alternative
since they can elicit potent immune responses without causing unwanted allergic reactions and have been already tested successfully against a variety of pathogens. Peptide vaccines utilizing certain epitopes in the antigenic protein provide an even more specific and potent response. While they inherently carry low immunogenicity, they can be easily combined with suitable adjuvants to generate the desired response [5]. Conventional approaches in designing such vaccines are painstakingly slow and inefficient. However, the emergence of new bioinformatic tools can greatly speed up the otherwise laborious process of vaccine design. This method termed as reverse vaccinology relies on immunoinformatic tools such as epitope prediction to construct vaccines in-silico and then validate the results experimentally in animal models [6]. This in-silico method can be effectively applied to design a multi-epitope vaccine for SARS-CoV-2.

In regards to subunit and peptide vaccines for SARS-CoV-2, most efforts have relied on utilizing the entirety of the spike protein as an antigen or identifying immunogenic epitopes [7], [8]. However, challenge experiments with similar vaccines for SARS-CoV have demonstrated the occurrence of lethal Th2 immunopathology as well as antibody-dependent enhancement in some subjects which suggests that certain regions of the spike protein such as $S_{597-603}$ may cause hypersensitivity and ultimately prove unsafe [9]. Moreover, epitope predictions focusing on the whole spike protein have overlooked several potential epitopes in the Receptor binding domain (RBD) due to their comparatively lower predicted antigenic values. The RBD, however, has been shown to generate equally high concentrations of neutralizing antibodies without causing a Th2 response and has been proven as a successful candidate against SARS-CoV in previous subunit vaccine studies [10]. Recently it was discovered that the RBD of SARS-CoV-2 binds to the ACE2 receptor with a much greater affinity (EC50: 0.14 µg/ml) than the RDB of SARS-CoV (EC50: 1.32 µg/ml) and can generate antibodies in mice while preventing antibody mediated enhancement which further validates the efficiency of RBD of SARS-CoV-2 as a potential vaccine candidate [11][12]. Therefore, this study solely analyzed the Receptor Binding Domain for promising T cell and B cell epitopes to construct a multi-epitope vaccine.

**Methods**

**Antigenic protein sequence selection**

The amino acid sequence for the spike protein was accessed from the SARS-CoV-2 whole-genome sequence MN908947.3 found in the National Center for Biotechnological Information (NCBI) gene bank [13]. The results obtained by Tai *et al* helped identify the appropriate amino acid sequence for the RBD from the sequence of the spike protein [11]. The antigenicity of the RBD protein was predicted through the antigen prediction tool Vaxijen v2.0 [14].

**Prediction of B cell epitopes**

Linear B cell epitopes were predicted using the B cell epitope prediction tool found in the Immune Epitope Database and Analysis Resource (IEDB). A host of prediction methods including Bepipred [15], Bepipred
2.0 [16], Chou & Fasman Beta-turn [17], Emini Surface Accessibility [18], Karplus & Schulz Flexibility [19], Kolaskar & Tongaonkar Antigenicity [20], and Parker Hydrophilicity [21] were incorporated which base their predictions on Hidden Markov models, Random forest algorithms, and amino acid scales. Additionally, Bcepred, another linear B cell epitope prediction server based on physicochemical properties, was used in conjunction with IEDB to generate additional epitope candidates [22]. The collective results of both servers were analyzed to determine to most favorable linear B cell epitopes based on their IEDB scores and level of agreement between the two models. VaxiJen 2.0 with the threshold at 0.4 was used to further filter out epitopes with a low antigenicity.

**Prediction of T cell epitopes**

Potential epitopes for Cytotoxic T cells were predicted based on their binding to MHC class I molecules while Helper T cell epitopes were generated based on their binding to MHC class II molecules. In both cases, the Tepi tool in the IEDB server was chosen to predict epitopes with high MHC binding affinity, and the IEDB recommended prediction method was used which utilizes Artificial neural network (ANN), Stabilized matrix method (SMM), and Combinatorial Library to pick the best selection method based on the peptide sequence [23]. A set of 27 most frequently occurring Human Leukocyte Antigen (HLA) alleles was used for each MHC class, and a high threshold was set to select for the top 1% CTL and top 10% of HTL epitopes. Vaxijen 2.0 was again used to refine for epitopes with an antigenic tendency of greater than 0.4.

**Analysis and selection of appropriate T cell and B cell epitopes**

The generated epitopes were evaluated for safety by examining their toxicity and potential to act as an allergen. The toxicity of the candidates was determined through the ToxinPred server [24], and their allergenicity was predicted by Allergen FP1.0 which compares the similarity of the input peptide sequences with those of known allergens [25].

**Epitope conservancy and population analysis**

The epitope conservancy was determined by performing multiple sequence alignment of the RBD protein sequence with all the SARS-CoV-2 sequences found on China National Center for Bioinformation [26] to identify the genomic variation of the protein at different locations. The effectiveness of the CTL and HTL epitopes in different genetic population subsets was determined through the IEDB population analysis tool which finds the frequency of various HLA alleles expressed in people in different parts of the world and compares the data to the HLA restrictions for the provided T cell epitopes [27].

**Construction of vaccine and structural analysis**

Adjacent and/or overlapping epitopes were combined into a larger peptide sequence, and the peptide sequences were fused by the peptide linker AAY to construct a multi-peptide vaccine. The N and C terminals of bacterial flagellin were joined with rigid EAAAK linkers to either ends of the vaccine as an adjuvant. [28]. To further supplement the adjuvant and increase the construct’s immunogenicity, a
universal helper T cell epitope called Pan-DR epitope (PADRE) was added to the C terminal of the vaccine peptide sequence [29]. The physical and chemical properties of the vaccine were investigated by ProtParam to evaluate its stability [30]. PSIPRED predicted the secondary structure of the construct [31]. The 3D structure of the vaccine candidate was generated through SparksX server which predicts the structure of a protein through a de novo approach. [32], To improve the accuracy and stability of the structure, GalaxyRefine was used[33]. ProSA web [34] and RAMPAGE [35] servers were implemented to verify the structure by generating Ramachandran plots and z-score graphs.

**Protein-peptide docking and Immune system simulation**

The efficient docking of the T cell epitopes to their respective MHC molecules was modeled by Pepsite [36]. The molecular docking of the entire vaccine construct to the TLR-5 receptor was modeled using the PatchDock server [37]. The necessary structures of the MHC molecules and TLR-5 receptor were obtained from the Protein Data Bank (PDB). All protein-peptide interactions were displayed in iCn3D software [38].

**In-silico cloning and codon optimization**

The amino acid sequence of the vaccine construct along with an addition of a 6xHis tag was codon-optimized for expression in *E coli* K-12 through the JAVA Codon Adaption Tool [39]. The appropriate expression vector was chosen from Addgene[40], and the sequence was cloned into the plasmid through restriction-digest in Benchling [41].

**Results**

**Antigenic protein sequence selection**

The complete genome sequence of SARS-CoV-2 was taken from the NCBI database and downloaded in FASTA format for later use. The results of Tai *et al* indicate that amino acid residues 331-524 in the spike protein constitute the receptor binding domain of the virus, the neutralization of which can effectively prevent viral entry into the host cell, thereby serving as a useful candidate for a vaccine. The selected protein sequence was then run through Vaxijen 2.0 with a threshold of 0.4. The sever predicted a score of 0.5356 which indicates that the overall RBD protein has a strong potential for being an antigen (table 1).

**Prediction of B cell epitopes**

A preliminary list of potential B cell epitopes was generated through the IEDB B cell epitope prediction tool. Various prediction methods based on physiochemical properties and antigenicity were utilized. The parker hydrophilicity of the RBD protein was found to be 1.316(average), -3.743(minimum), and 5.371(maximum). The Chou and Fasman beta turn was found to be 1.040(average), 0.694(minimum) and 1.397(maximum). The Karplus and Schilz flexibility was found to be 0.989(average), 0.896(minimum), 1.112(maximum). Viable epitopes should be exposed on the surface for better contact and so the surface accessibility was evaluated. The Emini Surface Accessibility of the protein was found to be 1.000(average), 0.074(minimum), 3.839(maximum). The Kolaskar and Tongaonkar antigenicity was
found to be 1.042 (average), 0.907 (minimum), and 1.214 (maximum). IEDB’s Hidden Markov model based Bepipred and Random Forest algorithm based Bepipred 2.0 prediction was also used (figure 1). The epitopes that consistently matched through several prediction methods were chosen. Another server called BcePred that evaluated similar physiochemical methods also generated some additional epitopes while agreeing with most of the epitopes predicted by the IDEB tools. A total of 14 B cell epitopes were generated through both servers which met the criteria for having a Vaxijen antigenicity score of higher than 0.4 and length of greater than 4 amino acids (table S1).

**Prediction of T cell epitopes**

An initial list of Cytotoxic (CTL) and Helper T cell (HTL) epitopes was created with the help of the IEDB Tepi tool. Generally, T cell epitopes are predicted based on their binding affinity to different alleles of Human Leukocyte Antigen (HLA) molecules, a crucial step for antigen presentation to T cells. For selection of CTL epitopes, the 27 most frequently expressed HLA alleles of supertypes A and B were selected, and the threshold was set to select the epitopes of score of top 1% and the length of 8-11 peptides. Their antigenic potential was also evaluated and yielded a total of 23 potential CTL epitopes (table S2). Interestingly a few epitopes such as PYRVVVL SF and KLNDCFTNV were completely or nearly identical to their SARS counterparts which have been previously confirmed through MHC binding assays.

The IEDB Tepi tool was also used for the selection HTL epitopes, and the set of 26 most frequently expressed HLA alleles for MHC class II molecules was used. The threshold was set to generate the epitopes with a score in the top 10% and length of 8-11 peptides. The dataset was similarly filtered to only keep epitopes with high antigenicity and produced a total of 7 Helper T cell epitopes (table S3).

**Analysis and selection of appropriate T cell and B cell epitopes**

Evaluating the safety of any antigen or epitope remains of great importance since many epitopes can resemble potent allergens or toxins and therefore can result in adverse and unwanted immune responses. Accordingly, all the epitopes were tested for their resemblance to allergens and toxins using Allergen FP1.0 and ToxinPred respectively. Almost all of the epitopes were found to be nontoxic, but the majority of B cell and T cell epitopes were filtered out due to their high potential for acting as an allergen. Ultimately 3 CTL, 5 HTL, and 5 B cell epitopes were selected to constitute the multi-epitope vaccine based on their high antigenicity and high number of binding HLA alleles (Table 2). Although most of the selected HTL epitopes bound to several HLA alleles, only 2 CTL epitope demonstrated significant binding to at least 3 HLA alleles.

**Epitope conservancy and population analysis**

The final list of epitopes was evaluated for their conservancy across different SARS-CoV-2 isolates. The China National Center for Bioinformation was used because of its vast database of almost 4000 SARS-CoV-2 isolates from 154 locations worldwide. Multiple sequence alignment was performed on the RBD region across all the isolates as well as SARS-CoV, the results of which are shown in figure 2. Across all
the sequences of SARS-CoV-2, there were no mutations in the amino acid sequence despite 3 different nucleotide mutations. In contrast, significant variation was observed between SARS-CoV-2 and SARS-CoV protein sequences in the RBD region which may partially explain the failure of some SARS-CoV antibodies to cross neutralize SARS-CoV-2 infections.

Due to the fact that HLA alleles are expressed in very different frequencies in different genetic populations, it is important to ensure that the selected epitopes are able to bind to the HLA alleles across different segments of populations. Therefore, the IEDB population analysis tool was used to determine the population coverage of the T cell epitopes in populations across the world. The results of the population analysis are provided in figure 3. The HTL epitopes covered over 99% of the global population and had similar levels of coverage across all continents. However, the 3 CTL epitopes covered only 74% of the global population, in part because of the low number of corresponding alleles for each epitope. Nonetheless, the selected T cell epitopes for the vaccine component altogether cover 99.92% of the global population have over 99% population cover in all continents.

**Construction of multi-epitope vaccine**

The selected epitopes were fused into a single vaccine component along with an adjuvant and a universal immunogenic sequence to create the final vaccine construct. The epitopes were linked using AAY linkers to increase protein stability. The adjuvant and the universal immunogenic sequence were connected to the epitopes through rigid EAAAK liners. Rigid linkers provide the distinct advantage of separating the different proteins of interest and preventing any cross-interactions. Certain regions such as amino acid residues 174-193 contained several epitopes and so overlapping regions were combined into a single peptide chain encompassing all those epitopes (table 3). The N and C terminals of bacterial flagellin were added to the peptide chain as an adjuvant. These alpha helical regions of flagellin which are largely conserved across bacterial species act as a potent TLR 5 agonist and induce a mixed Th1 and 2 response. To further increase the immunogenicity of the vaccine, the pan-DR epitope (PADRE) with the sequence AKFVAAWTLKAATA was attached to the vaccine. This universal Helper T cell epitope has been shown to improve efficiency of several peptide vaccines by boosting both the helper T cell and cytotoxic T cell response. The final peptide sequence of the multi-epitope vaccine candidate along with the addition of flagellin and PADRE is shown in table 4.

**Structural Analysis**

The primary structure of the vaccine construct was analyzed for its physical and chemical properties. It was found to have the formula $C_{1720}H_{2758}N_{506}O_{555}S_3$ with a total of 5542 atoms and a molecular weight of 39.5 kDa. The estimated half-life of the vaccine was predicted to be 20 hours (in mammalian cells, in vitro), and 10 hours (in *E coli*, in vivo). Importantly, the instability index of the protein was determined to be 37.05 which classifies it as stable. The Aliphatic index of 87.86 reflected this observation. Furthermore, the Grand Average of hydropathicity was -0.288 which classified the protein as hydrophilic. The secondary structure was predicted using PSIPRED whose results are indicated in figure 4. The results
indicated that 77.53% of the protein consisted of alpha helices, 21.37% consisted of random coils, and 1.09% was made of extended strands.

The tertiary structure was constructed in SparksX and is shown in figure 5. Due to high levels of unfavorable residues in the initial structure, the model was refined in GalaxyRefine which generated 5 possible models. The top model was chosen and structurally evaluated in ProSA and PROCHECK. ProSA compares the structure to other experimentally determined protein structures and assigns it a z-score. The z-score for the vaccine was -5.06 which is well in range of z-scores for proteins of that size. The residue energy chart was also generated and the energies of most of the residues were negative, suggesting a stable structure. The structure was further evaluated in RAMPAGE and a Ramachandran plot was generated. 96.7% of residues were in the favored range, 2.5% in the allowed range, and 0.8% in the outlier range. The results of the tertiary structural analysis are illustrated in figure 5.

**Protein-peptide molecular docking**

The T cell epitope interactions with their respective HLA molecules was modeled in PepSite and the results are shown in figure 5. Effective T cells response necessitate strong bonding of the epitope with HLA molecules. For some T cell epitopes which were identical in SARS, the binding to HLA alleles has already been experimentally performed through MHC assays and so was not modeled here. For instance, PYRVVVLSF has already been proven to interact with HLA-A*01:01, HLA-A*23:01, HLA-A*24:02, and HLA-A*26:01. Nonetheless, all of the T cell epitopes chosen for modeling bound strongly to their corresponding HLA molecules at an average of 6 residues. For instance, the epitope ASVYAWNRK bonded to HLA-A*30:01 at ala-1, val-3, tyr-4, ala-5, asn-7, and arg-8. Due to the intrinsic lack of antigenicity of peptide vaccines, the interaction between the adjuvant and pattern recognition receptors such as Toll like receptors is critical for the generation of a potent immune response. Therefore, the molecular docking between the vaccine and Toll-like receptor 5 was modeled in Patch Dock (figure 6). As expected, the alpha helix rich flagellin interacted with TLR 5 at 14 key residues and demonstrated its function as an TLR 5 agonist.

**Codon optimization in silico cloning**

Finally, a 6xHis tag was added for purification purpose. For expression in bacteria, the genomic sequence of the vaccine construct has to be codon optimized to yield optimal expression since the same amino acids are coded using different nucleotides in different organisms. Therefore, the peptide sequence of the vaccine was codon optimized in JAVA Codon Adaption Tool and the generated sequence is shown in table 5.

The codon optimized sequence was then inserted into a vector for expression in *E coli* K12. The pET-28 a (+) expression vector was chosen from Addgene for its high expression rates [40]. The gene coding for
the vaccine construct was then inserted through restriction digestion which was modeled in Benchling (figure 7).

**Discussion**

The COVID-19 pandemic continues to infect millions and has already taken the lives of over 1 million people in the United States alone. In addition to the high infection rate, its symptoms characterized by fever and atypical pneumonia, are proving especially fatal to the elderly population and people with pre-existing health conditions such as heart disease, diabetes, and COPD. With the increasingly dire situation, there is an urgent need for an effective vaccine. While mRNA-based vaccine candidates by pharmaceuticals like Moderna and Pfizer have been successful so far, they require multiple boosters to maintain efficacy against new strains. More conventional methods like subunit vaccines offer a more feasible approach. Multi-epitope peptide vaccines offer great promise as they can target highly specific regions of the antigen to generate a more potent immune response while preventing any allergic reactions associated with traditional subunit vaccines. By incorporating t cell epitopes, in addition to b cell epitopes, epitope-based vaccines have the potential to provide sustained cellular immunity in case the humoral response wanes over time. Furthermore, multiepitope peptide vaccines can be generated at a much faster rate than conventional methods with the help of bioinformatic tools like epitope prediction servers and molecular docking simulations. This feature is of particular importance as new strains of the virus may be resistant to current vaccines, thus needing to update existing vaccines. Already, such epitope based approaches have been applied in several cancer vaccines in clinical trials as well as against pathogens Human Immunodeficiency Virus and Hepatitis B virus [5].

For a SARS-CoV-2 multi-epitope vaccine, the surface glycoprotein (spike protein) has been the focus of many researchers. The spike protein mediates viral entry into the cell by binding to the human ACE2 receptor with the aid of its protease which cleaves the Spike protein into S1 and S2 subunits [42]. The actual binding to the ACE2 receptor occurs through the RBD found in the S1 subunit of the virus [42]. Due to the role of the spike protein in viral pathogenesis, multiple studies have attempted to probe the spike protein for B and T cell epitopes. However, previous research into SARS-CoV vaccines has revealed that immunizations with whole spike proteins induced Antibody-dependent enhancement (ADE) when tested in mouse models. Wang et al later demonstrated through challenge experiments in Rhesus Macaques that the epitope S_{597−603} was responsible for ADE in some of their subjects [43]. In other studies, despite high titers of neutralizing antibodies, an adverse Th2 response characterized by eosinophil infiltration was seen in mice, and one post challenge examination revealed that immunization with a viral vector encoding the spike protein enhanced hepatitis in some ferrets [44] [45]. In contrast, immunization with the receptor binding domain has been used by several studies against SARS without any reports of ADE or unwanted immune responses. Indeed, several research groups such as Baylor College of Medicine are actively perusing receptor binding domain based subunit vaccine candidates for SARS-Cov-2 [46]. The records of dangerous side effects of some SARS vaccines and the high degree of similarity between the two viruses suggests that certain sections in the S2 subunit of the spike protein may prove unsafe for
vaccine development. Therefore, this study limited its scope to immunogenic epitopes in the receptor binding domain due to their high levels of safety.

Immunoinformatic tools from the IEDB predicted 5 linear B cell epitopes in addition to 3 Cytotoxic T cell and 5 Helper T cell epitopes that can constitute a multiepitope vaccine against the receptor binding domain. The B cell epitopes were predicted from both structure-based approaches and machine learning techniques such as Hidden Markov models which are trained on previously confirmed epitopes. This hybrid method of selection enabled the analysis of physicochemical properties of the epitopes such as beta turns, surface accessibility, and hydrophilicity as well as their similarities to other experimentally known epitopes. Meanwhile, the Tepi tool used binding affinity to MHC class I and II molecules to predict T cell epitopes which were then validated through simulations of molecular docking with their HLA alleles. All selected epitopes were also thoroughly examined for allergenicity, toxicity, and conservancy across different population groups.

Despite the highly targeted immune response created by peptide vaccines, epitopes by themselves are not immunogenic and quickly break down in the body. As a result, they don't usually activate pattern recognition receptors necessary for a sufficient immune response. However, their low immunogenicity can be overcome by conjugating them with suitable adjuvants that are highly immunogenic in nature [47]. Often, this is accomplished by incorporating certain pathogen derived molecules that are known to stimulate Toll Like Receptors (TLR). The N and C terminals of bacterial flagellin, a known TLR 5 agonist, are conserved across most species and are often used as adjuvants for peptide vaccines. The interaction of these domains with TLR 5 results in DC activation and maturation, promotes release of proinflammatory cytokines, and aids in recruiting B and T cells [48]. Due to the effectiveness of flagellin in generating a potent immune response and lack of serious side effects like a cytokine storm, flagellin has been successfully used as an adjuvant in multiple clinical trials [49]. The N and C domains of *E. coli* flagellin were therefore fused to the multiepitope vaccine, and in silico molecular docking did confirm the interaction between the adjuvant and TLR5. To further increase the potency of the multiepitope vaccine, the universal immunogenic sequence PADRE was also included. The pan DR epitope is a universal helper T cell epitope that boosts antigen specific immune responses by the recruitment of helper T cells. It binds to 15 most common HLA-DR supertypes and can generate a consistent Th response across population subsets without encountering the issue of polymorphism in HLA-DR receptors [29]. Since the humoral response as well as cellular response is dependent on effective Helper T cell stimulation, the inclusion of the Pan-DR was crucial in constructing an efficacious peptide vaccine. Ultimately, a multi-epitope vaccine candidate was designed which included both B cell and T cell epitopes along with suitable adjuvants to increase immunogenicity.

Interestingly, recent epitope mapping studies with the use of peptide microarrays have confirmed some of the epitopes predicted from immunoinformatic servers. Zang et al who mapped the Spike protein for linear immunodominant sites of 20–25 peptides each found a total of 9 such sites out of which 4 were in the RBD. 3 immunodominant sites ($S_{330-349}$, $S_{375-394}$, $S_{450-469}$) partially contained 3 predicted B cell epitopes while one immunodominant region encompassed both a predicted B cell epitope as well as 1
HTL epitope [50]. The same paper also mapped favorable regions for T cell epitopes using ELISpot assay. 3 predicted HTL epitopes were found in this region. Although none of the CTL epitopes lay in that region, 2 CTL epitopes PYRVVLSF and KLNDLCFTNV are identical to experimentally confirmed SARS-CoV HTL epitopes. The predicted B cell epitope FRKSNLKP holds the greatest potential among the predicted epitopes because it has been confirmed in several recent peptide microarray studies as well as previous papers using bioinformatic methods [51][52]. Despite some experimental evidence for the validity of epitopes predicted in this paper, in vitro and in vivo tests are required before any conclusions can be drawn regarding the safety and efficacy of the designed multi-epitope vaccine candidate. Subsequent studies should also consider optimal delivery systems and adjuvant formulations such as MPL A in order to generate a robust immune response. Future evaluations of the proposed multi-epitope construct will help create a vaccine candidate against the SARS-CoV-2 virus.

References


13. NCBI Resource Coordinators (2016). Database resources of the National Center for Biotechnology Information. *Nucleic acids research,* 44(D1), D7–D19. [https://doi.org/10.1093/nar/gkv1290](https://doi.org/10.1093/nar/gkv1290)


**Declarations**
Competing interests: The authors declare no competing interests.

Tables
Tables 1-5 are in the supplementary files section.

Figures

Figure 1

Results of various B cell epitope prediction methods are shown. (a) Bepipred prediction, (b) Bepipred 2.0 prediction, (c) Kolaskar and Tongaonkar antigenicity, (d) Parker hydrophilicity, (e) Chou and Fasman beta turn, (f) Karplus and Schulz flexibility, (g) Emini Surface Accessibility, (h) Bcepred prediction. In the IEDB
prediction graphs, the peaks shaded yellow represent favorable regions for B cell epitopes. In the Bcepred prediction methods, peptide sequences in blue represent potential B cell epitopes.

Figure 2

The results of the multiple sequence alignment of the RBD are shown. A small sample of the list of the SARS-CoV-2 isolates used to perform multiple sequence alignment is shown in the gray area and show no genetic variation. The comparison to bat coronavirus genome and SARS-CoV is shown at the bottom. Here a significant number of red lines are seen which indicate the presence of several amino acid mutations.

Figure 3

The population coverage of the epitope vaccine is shown based on the number T cell epitope and HLA allele matches. (a) World, (b) South Asia, (c) Northeast Asia, (d) North America, (e) Europe.
Figure 4

The predicted secondary structure using PSIPRED is shown. The prediction reflects a high percentage of Alpha helices at 77.53% due to the helical structure of the flagellin adjuvant while the rest is made of random coils (21.37%) and extended strands (1.09%).

Figure 5
The model and structural analysis of the vaccine construct is shown. (a) the 3-dimentional model of the vaccine designed by SparksX, (b) the knowledge-based residue energy which shows that the structure has mostly low residue energies, (c) Ramachandran plot indicating that 96.7% of residues ae in the favored region, (d) z score plot with a z score of -5.06 suggesting that the construct has similar z scores to other confirmed molecules of that size.

**Figure 6**

The molecular docking of the individual epitopes as well as the entire vaccine construct are shown. (a.) binding of KLNDLCFTNV to HLA-A*02:03, (b.) binding of KLNDLCFTNV to HLA-A*02:06, (c.) binding of ASVYAWNRK to HLA-A*03:01, (d.) binding of ASVYAWNRK to HLA-A*11:01, (e.) binding of ASVYAWNRK to HLA-A*68:01, (f.) binding of ASVYAWNRK to HLA-A*30:01, (g.) binding of the vaccine construct to TLR 5. The flagellin adjuvant interacted with TLR 5 at 14 residues.

**Figure 7**

The recombinant vaccine cloned into a pET-28 a (+) expression vector is shown. It can express the protein in *E coli* K-12

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx