Vaccine Design, Adaptation, and Cloning Design for Multiple Epitope-Based Vaccine Derived From SARS-CoV-2 Surface Glycoprotein (S), Membrane Protein (M) and Envelope Protein (E): In silico approach

Peter T. Habib (p.habib911@gmail.com)
Colors Medical Laboratories

Research Article

Keywords: In silico vaccine design, SARS-CoV-2, reverse vaccinology, immunoinformatics

Posted Date: February 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-241638/v1

Abstract

The SARS Coronavirus-2 (SARS-CoV-2) pandemic has become a global epidemic that has increased the scientific community's concern about developing and finding a counteraction against this lethal virus. So far, hundreds of thousands of people have been infected by the pandemic due to contamination and spread. This research was therefore carried out to develop potential epitope-based vaccines against the SARS-CoV-2 virus using reverse vaccinology and immunoinformatics approaches. Three potential vaccine constructs were designed after intensive computational experimentation, and one vaccine model was chosen as the best vaccine based on a molecular docking analysis that is intended to work efficiently against SARS-CoV-2. In order to verify biological stability and find an appropriate mass production technique for the chosen vaccine, molecular dynamics simulation, and silico codon adaptation studies were subsequently carried out. This study should help to maintain current research efforts to secure a definitive preventive measure against this contagious disease.

1. Introduction

Coronaviruses are a family of viruses belonging to the Coronaviridae family and the Nidovirales order. These viruses are single-stranded, positive-sense RNA viruses with a genome size ranging from 26 to 32 kilobases. Coronaviruses are known to cause acute upper respiratory tract infections and significant respiratory infections in children and adults and infect humans as well as certain other species such as murine, porcine, feline, bovine, avian. (Masters & Perlman, 2013; Su et al., 2016; Weiss & Navas-Martin, 2005) So far, seven distinct human coronaviruses (HCoVs) have been identified. Four HCoVs, i.e., HCoV-OC43, HCV-229E, HCV-NL63, and HCV-HKU1, induce common cold in immunocompromised people and two other HCoVs, i.e., coronavirus, that cause severe acute respiratory syndrome (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), cause severe respiratory diseases (Drosten et al., 2003; Hamre & Procknow, 1966; van der Hoek et al., 2004; Zaki et al., 2012). The seventh strain believed to infect humans is the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), which is responsible for the latest pandemic worldwide, causing the deadly coronavirus disease-2019 (COVID-19). COVID-19 was first detected in a group of pneumonia patients in Wuhan, China, in December 2019 (Peeri et al., 2020). The first COVID-19 fatality case was identified in Wuhan, China, on 11 January 2020, and the first affected case outside China was reported in Thailand on 13 January 2020. (Wang et al., 2020). The most common symptoms of COVID-19 onset include fever, cough, tiredness, diarrhea, and patients experience trouble breathing in extreme circumstances (Huang et al., 2020). On 11 March 2020, the World Health Organization (WHO) proclaimed COVID-19 a pandemic, as affected cases outside China rose 13 times by the end of February 2020, and more than 4000 deaths were registered globally (World Health Organization., 2021). At this time, 108,879,471 cases were recorded on 13 March 2020, 2,397,946 cases of mortality,80,912,198 cases of recovery in 177 countries were registered globally (Hopkins, 2020). Interferons are a little powerful in combination with ribavirin. There is many efforts to use bioactive compounds in order to inhibit SARS-CoV-2 binding to ACE2 (P. T. Habib et al., 2021) and classification of viral proteins or corona virus from CT scan using machine learning (P. Habib, 2020; Randhawa et al., 2020). However, we must further assess the efficacy of the combination solution (Fehr & Perlman, 2015). This experiment was conducted in order to design a new epitope-based vaccination against three SARS-CoV-2 proteins, namely, surface glycoprotein responsible for membrane fusion events during viral inlet (Cavanagh, 1995; Petit et al., 2005); the Envelope protein (E) which is a tiny envelope, integral protein, involving in assembly, buddings, forming envelopes, and pathogenesis. it is also involved in many aspects of the virus' life cycle (Schoeman & Fielding, 2019) and membrane glycoprotein that mediates the interaction between virions and cellular receptors (Rottier, 1995). Reverse vaccinology and immuno-informatics are used to examine the genome and genetic material of the specific virus in which novel antigens of a virus or microorganism or a disease organism are identified. The methods of bioinformatics are used in reverse vaccinology to classify and analyze certain novel antigens. These techniques are used to dissect the genome and genetic structure of a pathogen in order to improve the future vaccine. The reverse vaccinology approach also makes it possible for scientists, during the vaccine process, to understand the antigenic segments of a virus or pathogen. These methods for developing vaccines are fast, inexpensive, reliable, simple, and cost-effective. The methods were successfully employed in developing vaccines to combat many viruses, such as Zika, Chikungunya, etc. (Chong & Khan, 2019; Mar\'\ia et al., 2017).

2. Material And Methods

The current experiment has been performed to improve possible SARS-CoV-2 vaccines by using reverse vaccination strategies. In this experiment, the materials imported from and the methods used were adapted from (Ullah et al., 2020) work

3. Results

3.1. Viral protein sequence identification, selection, and retrieval

Proteins downloaded from NCBI database (HTTPS:// www.ncbi.nlm.nih.gov/). Proteins that retrieved in FASTA format was selected for the potential vaccine design. Those proteins are Membrane Glycoprotein (accession no: YP 09724393.1), Envelope Protein (accession no: YP 009724392.1), and surface Glycoprotein (accession no: YP 009724390.1). Table (1) lists the NCBI accession number of the protein sequences.

Table 1

ID	*Protein*	*Location*
YP_009724393.1	Membrane Glycoprotein	Virus Membrane
YP_009724392.1	Envelope Protein	Virus Membrane
YP_009724390.1	Surface Glycoprotein	Virus Membrane

3.2. Antigenicity prediction and physicochemical property analysis of the protein sequences

Vaxigen webserver reports the three proteins as potent antigens (Table 2). For these three selected proteins, the physicochemical property analysis was conducted. Surface glycoprotein has the highest molecular weight, but it has the lowest pI of them. Membrane glycoprotein has the highest pI, but the GRAVY score is still positive.

Table 2

Protein Name	Antigenicity	Length	Half-life time	GRAVY Score	The instability index (II)	Molecular weight	Theoretical pI
Envelope Protein	Antigenic	75	30 hours	1.128	38.68	8365.04	8.57
Membrane Glycoprotein	Antigenic	222	30 hours	0.446	39.14	25146.62	9.51
Surface Glycoprotein	Antigenic	1273	30 hours	-0.079	33.01	141178.47	6.24

Surface Glycoprotein's theoretical pI of 6.24 was the highest forecast. A similar half-life of 30 h was found to be predicted for the three proteins. However, the highest predicted instability index and positive hydropathic average were for Membrane glycoprotein (GRAVY)

3.3. T-cell and B-cell epitope prediction and their antigenicity, allergenicity, and topology determination

For possible vaccine construction using the server IEDB (https://www.iedb.org/), which generated numerous epitopes, class I and class II MHC epitopes were predicted (Table 3,4,5). The server contains data on human, non-human primates, and other animal species related to allergy, infectious diseases, self-immunity, and transplantation, which has been confirmed experimentally and approved for antibody and T-cell epitopes. By analyzing these experimental data and examining the input protein, the server predicts epitopes (Vita et al., 2019)(Vita et al., 2019)Based on the antigenicity values, however, eight MHC class I, 6 MHC class II, and 4 antibody epitopes have been selected (Table 6) after filtering the epitopes generated by AS and by percentiles from the top 12 MHC class I, sixteen MHC class II, and nineteen antibody epitopes (Table 3,4, 5). The percentile values are the forecast binding affinity and fewer percentile values represent a higher binding affinity (Vita et al., 2019). Subsequently, vaccine construction was selected from epitopes with high antigenic, non-allergic, and non-toxicity.

Table 3

Protein	MHC Class / Antibody	Epitope	AS	scores	Antigenicity (Virus model, threshold = 0.4)	Allergenicity	Toxicity
Envelope Protein	MHC Class I	SLVKPSFYV	0.4140	0.06	ANTIGEN	ALLERGEN	Non-Toxin
	MHC Class II	FLLVTLAILTALRLC	0.6311	0.38	ANTIGEN	NON-ALLERGEN	Non-Toxin
		LLVTLAILTALRLCA	0.3840	0.38	NON-ANTIGE	NON-ALLERGEN	Non-Toxin
		LVTLAILTALRLCAY	0.4070	0.38	ANTIGEN	NON-ALLERGEN	Non-Toxin
		VFLLVTLAILTALRL	0.7218	0.38	ANTIGEN	NON-ALLERGEN	Non-Toxin
		VTLAILTALRLCAYC	0.8599	0.38	ANTIGEN	NON-ALLERGEN	Non-Toxin
	Antibody	YVYSRVKNLNSSRVP	0.4492	-	ANTIGEN	NON-ALLERGEN	Non-Toxin

Table 4

Protein	MHC Class / Antibody	Epitope	AS	scores	Antigenicity (Virus model, threshold = 0.4)	Allergenicity	Toxicity
Membrane Protein	MHC Class I	ATSRTLSYY	0.6108	0.03	ANTIGEN	NON-ALLERGEN	Non-Toxin
	MHC Class II	LSYYKLGASQRVAGD	0.2247	0.67	NON-ANTIGEN	ALLERGEN	Non-Toxin
		RTLSYYKLGASQRVA	0.5644	0.67	ANTIGEN	ALLERGEN	Non-Toxin
		TLSYYKLGASQRVAG	0.4376	0.67	ANTIGEN	ALLERGEN	Non-Toxin
	Antibody	NGTITVEELKKLLEQW	-0.1969	-	NON-ANTIGEN	ALLERGEN	Non-Toxin
		YRIGNYKLNTDHSSSSDNIA	0.2216	-	NON-ANTIGEN	ALLERGEN	Non-Toxin

Table 5

Protein	MHC Class / Antibody	Epitope	AS	Percentile scores	Antigenicity (Virus model, threshold = 0.4)	Allergenicity	Toxicity
Surface Glycoprotein	MHC Class I	LTDEMIAQY	0.1043	0.01	NON-ANTIGEN	NON-ALLERGEN	Non-Toxin
		YLQPRTFLL	0.4532	0.02	ANTIGEN	ALLERGEN	Non-Toxin
		VLNDILSRL	0.8524	0.03	NON-ANTIGEN	NON-ALLERGEN	Non-Toxin
		TSNQVAVLY	0.4387	0.03	ANTIGEN	ALLERGEN	Non-Toxin
		TLDSKTQSL	1.0685	0.03	ANTIGEN	NON-ALLERGEN	Non-Toxin
		RLQSLQTYV	-0.2167	0.05	NON-ANTIGEN	NON-ALLERGEN	Non-Toxin
		KIADYNYKL	1.6639	0.05	ANTIGEN	ALLERGEN	Non-Toxin
		RLDKVEAEV	0.0765	0.06	NON-ANTIGEN	ALLERGEN	Non-Toxin
		WTAGAAAYY	0.6306	0.06	ANTIGEN	NON-ALLERGEN	Non-Toxin
		LLFNKVTLA	0.6150	0.08	ANTIGEN	NON-ALLERGEN	Non-Toxin
	MHC Class II	LSFELLHAPATVCGP	0.5062	0.03	ANTIGEN	ALLERGEN	Non-Toxin
		VLSFELLHAPATVCG	0.4784	0.03	ANTIGEN	NON-ALLERGEN	Non-Toxin
		VVLSFELLHAPATVC	0.8618	0.03	ANTIGEN	NON-ALLERGEN	Non-Toxin
		SFELLHAPATVCGPK	0.2085	0.09	NON-ANTIGEN	ALLERGEN	Non-Toxin
		VVVLSFELLHAPATV	0.8083	0.09	ANTIGEN	NON-ALLERGEN	Non-Toxin
		FVFLVLLPLVSSQCV	0.7185	0.24	ANTIGEN	NON-ALLERGEN	Non-Toxin
		MFVFLVLLPLVSSQC	0.5741	0.24	ANTIGEN	NON-ALLERGEN	Non-Toxin
		RVVVLSFELLHAPAT	0.7485	0.24	ANTIGEN	NON-ALLERGEN	Non-Toxin
	Antibody	SQCVNLTTRTQLPPAYTNSFTRGVY	0.6860	-	ANTIGEN	ALLERGEN	Non-Toxin
		FSNVTWFHAIHVSGTNGTKRFDN	0.6767	-	ANTIGEN	ALLERGEN	Non-Toxin
		DPFLGVYYHKNNKSWME	0.5821	-	ANTIGEN	ALLERGEN	Non-Toxin
		MDLEGKQGNFKNL	1.2592	-	ANTIGEN	ALLERGEN	Non-Toxin
		KHTPINLVRDLPQGFS	0.6403	-	ANTIGEN	NON-ALLERGEN	Non-Toxin
		KSFTVEKGIYQTSNFRVQP	0.5729	-	ANTIGEN	ALLERGEN	Non-Toxin
		YNSASFSTFKCYGVSPTKLNDLCFT	1.4031	-	ANTIGEN	NON-ALLERGEN	Non-Toxin
		GDEVRQIAPGQTGKIADYNYKLP	1.1017	-	ANTIGEN	NON-ALLERGEN	Non-Toxin
		ELLHAPATVCGPKKSTNLVKN	0.0029	-	NON-ANTIGEN	ALLERGEN	Non-Toxin
		NCTEVPVAIHADQLTPT	0.3987	-	NON-ANTIGEN	NON-ALLERGEN	Non-Toxin
		ASYQTQTNSPRRARSVASQ	0.2556	-	NON-ANTIGEN	NON-	Non-

				ALLERGEN	Toxin
YTMSLGAENSVAYSNN	0.6434	-	ANTIGEN	ALLERGEN	Non-Toxin
KQIYKTPPIKDFGGF	-0.3896	-	NON-ANTIGEN	NON-ALLERGEN	Non-Toxin
LADAGFIKQYGDCLG	0.2071	-	NON-ANTIGEN	ALLERGEN	Non-Toxin
RNFYEPQIITTD	0.3529	-	NON-ANTIGEN	ALLERGEN	Non-Toxin
SCCKFDEDDSEPVLKG	0.4347	-	ANTIGEN	ALLERGEN	Toxin

Table 6

Protein Name	MHC Class I	MHC Class II	Antibody
Surface Glycoprotein	TLDSKTQSL	VLSFELLHAPATVCG	KHTPINLVRDLPQGFS
	WTAGAAAYY	VVLSFELLHAPATVC	GDEVRQIAPGQTGKIADYNYKLP
	LLFNKVTLA	RVVVLSFELLHAPAT	YNSASFSTFKCYGVSPTKLNDLCFT
	-	VVVLSFELLHAPATV	-
	-	FVFLVLLPLVSSQCV	-
	-	MFVFLVLLPLVSSQC	-
Membrane Protein	ATSRTLSYY	-	-
Envelope Protein	FLLVTLAILTALRLC	-	YVYSRVKNLNSSRVP
	LVTLAILTALRLCAY	-	-
	VFLLVTLAILTALRL	-	-
	VTLAILTALRLCAYC	-	-

3.6. Vaccine construction

Three vaccines were selected using the selected epitopes intended to fight SARS-CoV-2. Three different adjuvants have been used for vaccines: beta-defensin, ribosomal protein, L7/L12 protein, and HABA protein and different linkers such as EAAK, GGGS, GPGPG, and KK linkers. A vaccine construction sequence PADRE is an important sequence. It can increase the vaccine's strength with minimum toxicity. Moreover, the PADRE sequence also improves the CTL response, thus ensuring a potent immune response (Wu et al., 2010). The newly built vaccines were: COVac-1, COVac-2, and COVac-3 and further analyzed their physical properties (Table 7).

Table 7

Vaccine name	Constructed Sequence
COVac-1	**EAAAK**GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKK**EAAAK**AKFVAAWTLKAAA**GGGSTLDSKTQSLGGGSATSRTLSYYGGGSLLFNKV**
COVac-2	**EAAAK**MAKLSTDELLDAFKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAAGDKKIGVIKVVREIVSGLGLKEAKDLVDGAP
COVac-3	**EAAAK**MAENPNIDDLPAPLLAALGAADLALATVNDLIANLRERAEETRAETRTRVEERRARLTKFQEDLPEQFIELRDKFTTEELRKAAEGYLEAATNRYNELVER

3.7. Antigenicity and allergenicity of the vaccine constructs

Table 8 lists the results of the analysis of antigenicity, allergy, and physicochemical property. All three vaccine buildings are both antigenic and non-allergenic.

Table 8

Vaccine name	Antigenicity score	Antigenicity	Allergenicity
COVac-1	**0.5547**	**ANTIGEN**	**NON-ALLERGEN**
COVac-2	**0.5248**	**ANTIGEN**	**NON-ALLERGEN**
COVac-3	**0.5633**	**ANTIGEN**	**NON-ALLERGEN**

3.8. Secondary and tertiary structure prediction of the vaccine constructs

In the secondary structure analysis, COVac-2 was shown to be the highest percentage of 138 amino acid coils and the highest percentage of 33 amino acids forming an alpha helix (Fig. 2). However, CV-3 had the highest percentage of 91 amino acids in extended strand formation (Fig. 3). COVac-2 estimated at 9.9785 was the lowest RMSD(Å). In Å is estimated the average deviation from the experimental structure. The less the quality of the 3D model is, the better. COVac-3 thus has the largest RMSD: 12,475 showed the worst 3D structure results from the experiment. Three different templates were also used for the 3D structures of the three different vaccines. The RaptorX server used the templates (Källberg et al., 2012) for the 3D structures of query vaccine construction. The outcome of the 3D structural analysis is presented in Table 9.

Table 9: Calculated of RMSD for the three construct

Name of the vaccine	RMSD(Å)
COVac-1	10.359
COVac-2	9.9785
COVac-3	12.475

3.9. 3D structure refinement and validation

The three vaccine constructs were refined and validated in the 3D structure refinement and validation step. The server PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK/) divides Ramachandran into four regions: the most favored (red), additional allowable (yellow) region, generous allowable region (light yellow), and disallowed area (represented by white color). According to the server, over 90% of their amino acids in this most favored region and additionally permitted regions should have a valid protein (the best quality protein). A few percent of the amino acids of the protein may also occur in the additional region and generously allowed region. However, there should be no amino acid in the unauthorized region (Laskowski et al., n.d.; Sateesh et al., 2010; Zobayer & others, 2019) For further analysis and validation, the 3D protein structures created in the previous step have been refined. With the help of the Ramachandran Plots, the refined structures were validated. The analysis showed that COVac-2 vaccines had an outstanding 91.8% of the amino acids in the allowed region, 7.3% of the amino acids in the additional areas, and 0.9% of the amino acids in the disregarded regions. In the most allowed areas, the COVac-3 vaccine contained 90.6% of amino acids, 8.9% of amino acids in the additional allowed regions, 0.5% of amino acids in the generously allowed regions, and 0.0% of the amino acids in the regions that are disallowed. With 89,2% of amino acids in allowed regions, 9,6% of amino acids in additional allowed regions, 0,8% of amino acids in the generously allowed regions, and 0, 4% of amino acids in disallowed regions, COV-1 vaccines showed the worst result. (Fig. 7).

3.10. Vaccine protein disulfide engineering

In protein disulfide engineering, The amino acid pairs with a bond value below 2.00 kcal/mol were selected in the experiment. In this study, the amino acid pairs that had bond energy values less than 2.2 kcal/ mol, were selected (Craig & Dombkowski, 2013). The COVac-1 generated 43 amino acid pairs that could form disulfide bonds. However, only 5 pairs were selected because they had the bond energy, less than 2.00 kcal/mol: 47-CYS 61-CYS, 268-CYS 271-CYS, 287-PRO 331-CYS, 325-CYS 331-CYS, and 21-CYS 25-CYS. Although COVac-2 and COVac-3 generated 44 and 37 pairs of amino acids, respectively, that might form disulfide bonds, where 5 amino acids selected in both COVac-2 and COVac-3 showed bond energy less than 2.00 Kcal/mol. The selected amino acid pairs of COVac-1, COVac-2, and COVac-3 formed the mutant version of the original vaccines.

3.11. Protein-protein docking study

The docking study for protein-protein has been conducted to find the best-constructed vaccine for COVID-19. The vaccine constructs with the best result in the molecular docking were considered as the best vaccine construct. Based on the docking results, the best constructed vaccine was found to be COVac-2 according to ClusPro 2.0, COVac-2 showed the best weighted score (Center: -863.0 and Lowest Energy: -1069.3) and the largest number of members (58). COVac-2 comes in the second rank with 57 members and the Lowest energy (-1090.6). COVac-3 was the worst one with 56 members and (-1241.1). when analyzed with PatchDock and FireDock servers, CV-2 showed the best and lowest global energy (-0.29), attractive VdW (-22.45), and repulsive VdW (15.40). again, COVac-1 has the worst global energy (-7.11) and COVac-3 -1.17. Since COVac-2 showed the best results in the protein-protein docking study with almost all the targets by all the servers and with the TLR-8, it was considered as the best vaccine construct among the three constructed vaccines (Fig. 8 and Table 13). Later, the molecular dynamics simulation, in silico immune simulation, and in silico codon adaptation studies were conducted only on the COVac-2 vaccine.

3.12. In silico immune simulation

C-ImmSim studies mechanisms of the successive and effective immune responses of the cell condition and the memory of immune cells. The effect is that few cells increase their half-life substantially and live longer than other cells. ImmSim server immune simulation results confirmed consistency with true immune responses. The response was illustrated by high IgM levels. Also, an increase in the B-cell population was characterized as an increase in immunoglobulin expression (IgG1+IgG2, IgM, and IgG+IgM), resulting in a decrease in antigen concentration (Fig. 8A, C). There is also a clear increase in the population of Th (helper) and T C (cytotoxic) cells with memory growth (Fig. 8F, E). IFN-γ production was also identified to have been stimulated after immunization (Fig. 8D). The T-cell population results have been approachable significantly with the growth of the memory and consistent exposure for all other immune cell populations.

3.12. A molecular dynamics simulation study

The results of molecular dynamics simulation of COVac-2-TLR-8 docked complex is illustrated in Fig. 9. Protein dynamic simulation determines the stability and physical movement of protein atoms and molecules (Chauhan et al., 2019). The simulation was thus conducted to determine the vaccine protein's relative stability. The deformability graph shows the peaks representing the protein regions with a moderate deformation rate (Fig. 9b). The complex's B-factor chart makes it easy to visualize and compare NMA and the PDB field of the docked complex (Fig. 9c). The value of the docked complex is shown in the figure. 9 docked complexes COVac-2 and TLR8 generated good 3.315510e-06 values. The graph shows the variance with red bars and the cumulative variance with green colored bars. (Fig. 9e). The co-variance map of the complex, where red color represents the correlated motion between a pair of residues, uncorrelated motion is indicated by white color as well as the anti-correlated motion is marked by blue color. The elastic map of the complex refers to the connection between the atoms and darker gray regions indicate stiffer regions (Fig. 9 g) (López-Blanco et al., 2014).

3.13. Codon adaptation and in silico cloning study

The number of nucleotides in a probable COVac-2 sequence would be 1255 as the COVac-2 protein had 416 amino acids before reverse translation. The codon adaptation index (CAI) value of 0.94 of COVac-2 indicated that the DNA sequences have a higher percentage of the codons that should be used by the cellular machinery of the target organism E. coli strain K12 (codon bias). For this reason, the production of the COVac-2 vaccine should be carried out successfully (Solanki & Tiwari, 2018). The GC content of the improved sequence was 52.80% (Fig. 10). The predicted DNA sequence of COVac-2 has inserted into the pET-30b(+) vector plasmid between the EcoRI and BamHI restriction sites and since the vaccine DNA sequence did not have restriction sites for EcoRI and BamHI restriction enzymes, EcoRI and BamHI restriction sites were conjugated at the N-terminal and C-terminal sites, respectively. The newly constructed vector is illustrated in Fig. 11.

4. Discussion

The current study has been conceived to develop potential SARS-CoV-2 vaccines, which are the cause of the recent COVID-19 pandemic worldwide. Tens of thousands of people worldwide have already been killed by pneumonia. Therefore, potential vaccines to fight this lethal virus were predicted in this study. Three candidate virus proteins were identified and selected from the NCBI database to perform the vaccine construction. For further analysis, only highly antigenic sequences have been chosen because highly antigenic proteins can produce better immunogenic responses (Demkowicz et al., 1992).

We have predicted linear B and T-cell epitopes using immunoinformatics tools that may promote cell and humoral immunity. These epitopes of B-cells and T-cells may theoretically be used to produce vaccines targeting the viral protein and maybe reliable for stimulating both humoral and cell-mediated immunity. In the present research, T-cell and B-cell epitopes were predicted via the IEDB server. For adaptive immune stimulation, T-cell epitopes are necessary and are sufficient to cooperate with MHC molecules. To build the epitope-based vaccine, we predicted B and T cell epitopes for nominated antigens and joined them with EAAK, GGGS, GPGPG, and KK linkers after antigenicity and allergenicity check. The EAAAK linker was also fused between the adjuvant and the epitopes sequences for the best expression and bioactivity improvement of the vaccine. The constructed multi-epitope vaccine showed higher scores of antigenicity both on the Vaxijen v2.0 server. Multi-epitopic vaccines have less immunogenicity and need adjuvants. Molecular docking and MD simulation were implemented, and the RMSD plot representing the steady binding of the complex. Immune simulation results showed corresponding to typical immune

responses. The generated immune responses increased generally after repeated exposure to the antigen. The development of memory B-cells and T-cells was apparent, with several months of memory in B-cells. Particularly stimulated were helper T cells. Another interesting finding was the increase in IFN-μ and IL-2 levels following the initial injection and the peak following the repeated exposure to antigen. This shows that T H cells are high and have therefore an efficient and humorous response to the Ig production. This must be expressed in a suitable host by recombinant protein. The preferred option for recombinant protein expression is the E.coli Systems for expression. Codon optimization has been performed to ensure that the recombinant E.coli vaccine has high levels of expression. System of E.coli (K12 strain). Both the GC content and the CAI score were beneficial for high-level protein expression in bacteria. The next step is the expression and numerous immunology analyses necessary to confirm the results obtained by immunoinformatic analyses of this peptide within a bacterial system.

5. Conclusion

One of the most deadly pandemics has recently occurred due to the SARS-CoV-2. Prevention of the new infection is both very difficult and obligatory. The potential of in silico methods can be used to find demanded solutions with fewer tests and mistakes to save the scientists both time and cost. Potential subunit vaccines against SARS-CoV-2 have been designed using various reverse vaccinology and immunoinformatics techniques in this study. The highly antigenic viral proteins and epitopes were employed to design the vaccines. Various types of computational studies with the vaccine constructs suggested show the possibility of a good immunogenic response. Consequently, these suggested vaccine constructs could be used effectively for vaccinations for preventing and spreading SARS-CoV-2 if satisfactory results are achieved in numerous in vivo and in vitro tests. Our present study should therefore help scientists to develop possible SARS-CoV2 vaccines and therapeutics.

Declarations

Competing interests: The authors have declared that no competing interests exist.

References

Cavanagh, D. (1995). The coronavirus surface glycoprotein. In *The coronaviridae* (pp. 73–113). Springer.

Chauhan, V., Rungta, T., Goyal, K., & Singh, M. P. (2019). Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Scientific Reports*, *9*(1), 1–15.

Chong, L. C., & Khan, A. M. (2019). Vaccine target discovery. *Encyclopedia of Bioinformatics and Computational Biology*, 241.

Craig, D. B., & Dombkowski, A. A. (2013). Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics*, *14*(1), 1–7.

Demkowicz, W. E., Maa, J. S., & Esteban, M. (1992). Identification and characterization of vaccinia virus genes encoding proteins that are highly antigenic in animals and are immunodominant in vaccinated humans. *Journal of Virology*, *66*(1), 386–398.

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A. M., & others. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *New England Journal of Medicine*, *348*(20), 1967–1976.

Fehr, A. R., & Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. *Coronaviruses*, 1–23.

Habib, P. (2020). COVATOR: A Software for Chimeric Coronavirus Identification. *BioRxiv*.

Habib, P. T., Saber-Ayad, M., & Hassanein, S. E. (2021). *In Silico Analysis of 716 Natural Bioactive Molecules Form Atlantic Ocean Reveals Candidate Molecule to Inhibit Spike Protein*.

Hamre, D., & Procknow, J. J. (1966). A new virus isolated from the human respiratory tract. *Proceedings of the Society for Experimental Biology and Medicine*, *121*(1), 190–193.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., & others. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*(10223), 497–506.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature Protocols*, *7*(8), 1511–1522.

Laskowski, R. A., MacArthur, M. W., & David, S. M. (n.d.). DS and Thornton, JM (1993) Procheck–a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr*, *26*, 283.

López-Blanco, J. R., Aliaga, J. I., Quintana-Ort\'\i, E. S., & Chacón, P. (2014). iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Research*, *42*(W1), W271–W276.

Mar\'\ia, R. R., Arturo, C. J., Alicia, J.-A., Paulina, M. G., & Gerardo, A.-O. (2017). *The impact of bioinformatics on vaccine design and development*. InTech, Rijeka, Croatia.

Masters, P. S., & Perlman, S. (2013). Coronaviridae, p 825–858. *Fields Virology, 1*.

Peeri, N. C., Shrestha, N., Rahman, M. S., Zaki, R., Tan, Z., Bibi, S., Baghbanzadeh, M., Aghamohammadi, N., Zhang, W., & Haque, U. (2020). The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *International Journal of Epidemiology*, *49*(3), 717–726.

Petit, C. M., Melancon, J. M., Chouljenko, V. N., Colgrove, R., Farzan, M., Knipe, D. M., & Kousoulas, K. G. (2005). Genetic analysis of the SARS-coronavirus spike glycoprotein functional domains involved in cell-surface expression and cell-to-cell fusion. *Virology*, *341*(2), 215–230.

Randhawa, G. S., Soltysiak, M. P. M., el Roz, H., de Souza, C. P. E., Hill, K. A., & Kari, L. (2020). Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *Plos One*, *15*(4), e0232391.

Rottier, P. J. M. (1995). The coronavirus membrane glycoprotein. In *The Coronaviridae* (pp. 115–139). Springer.

Sateesh, P., Rao, A., Sangeeta, S. K., Babu, M. N., & Grandhi, R. S. (2010). Homology modeling and sequence analysis of anxC3. 1. *International Journal of Engineering Science and Technology*, *2*, 1125–1130.

Schoeman, D., & Fielding, B. C. (2019). Coronavirus envelope protein: current knowledge. *Virology Journal*, *16*(1), 1–22.

Solanki, V., & Tiwari, V. (2018). Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against Acinetobacter baumannii. *Scientific Reports*, *8*(1), 1–19.

Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C. K., Zhou, J., Liu, W., Bi, Y., & Gao, G. F. (2016). Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends in Microbiology*, *24*(6), 490–502.

Ullah, M. A., Sarkar, B., & Islam, S. S. (2020). Exploiting the reverse vaccinology approach to design novel subunit vaccines against Ebola virus. *Immunobiology*, *225*(3), 151949.

van der Hoek, L., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J. M., Wolthers, K. C., Wertheim-van Dillen, P. M. E., Kaandorp, J., Spaargaren, J., & Berkhout, B. (2004). Identification of a new human coronavirus. *Nature Medicine*, *10*(4), 368–373.

Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A., & Peters, B. (2019). The immune epitope database (IEDB): 2018 update. *Nucleic Acids Research*, *47*(D1), D339–D343.

Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, *395*(10223), 470–473.

Weiss, S. R., & Navas-Martin, S. (2005). Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiology and Molecular Biology Reviews*, *69*(4), 635–664.

Wu, C.-Y., Monie, A., Pang, X., Hung, C.-F., & Wu, T. C. (2010). Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. *Journal of Biomedical Science*, *17*(1), 1–10.

Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E., & Fouchier, R. A. M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine*, *367*(19), 1814–1820.

Zobayer, M., & others. (2019). *In silico Characterization and Homology Modeling of Histamine Receptors*.

Figures

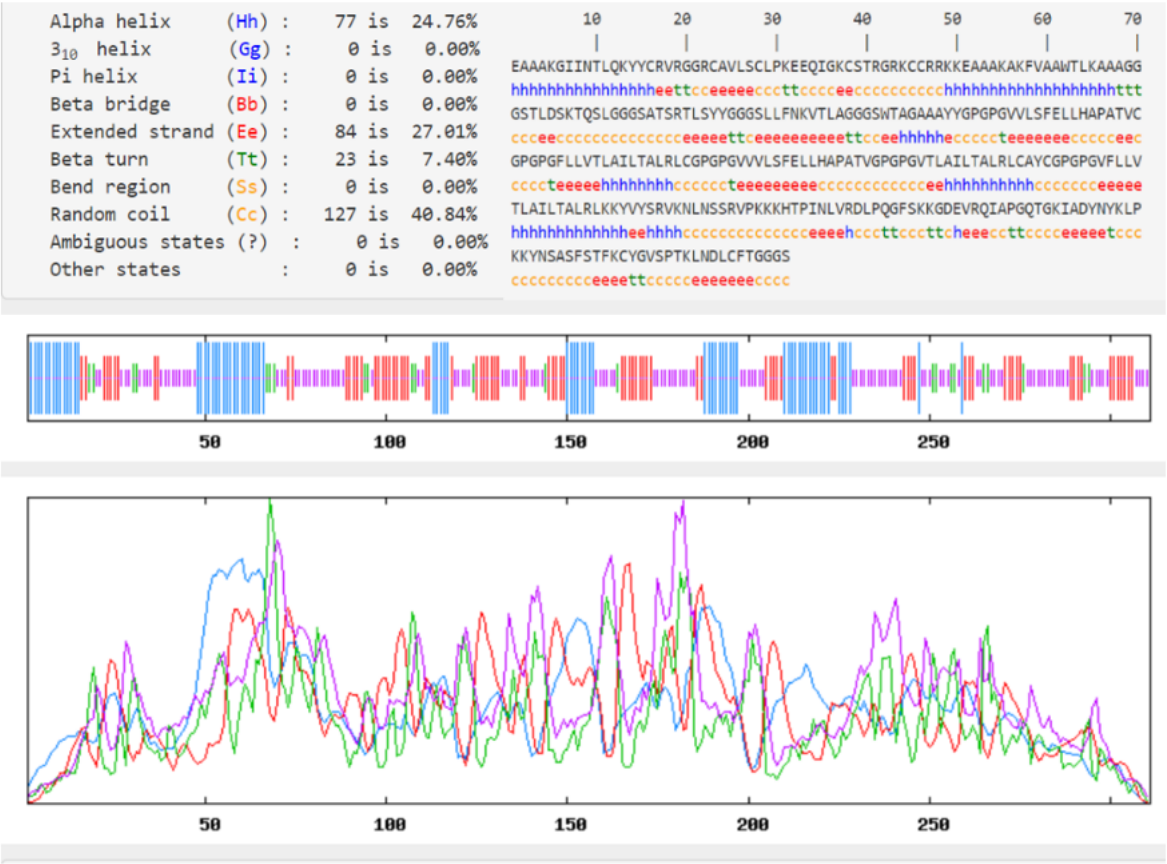

Figure 1

COVac-1 secondry structure prediction

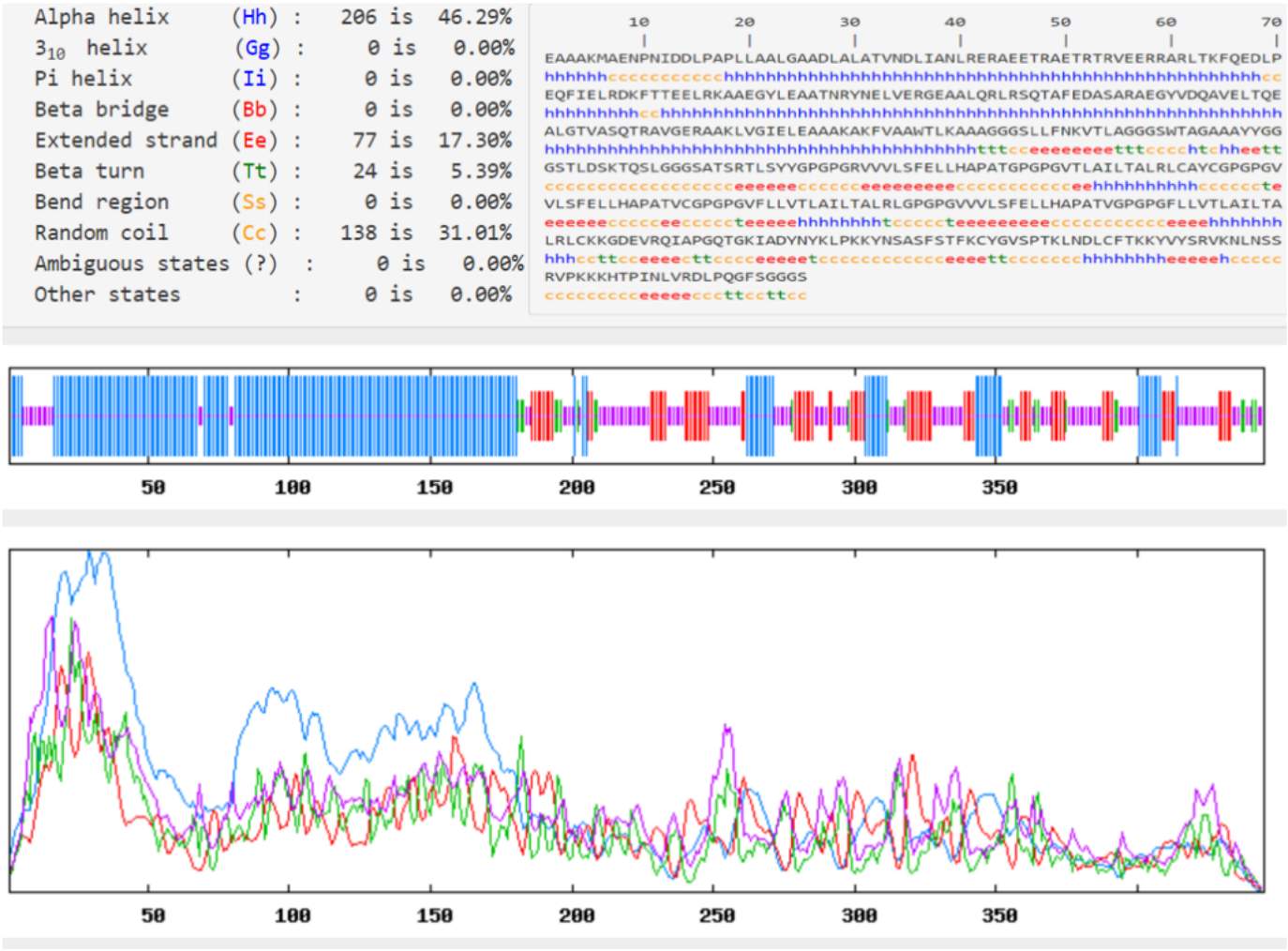

Figure 2

COVac-2 secondry structure prediction

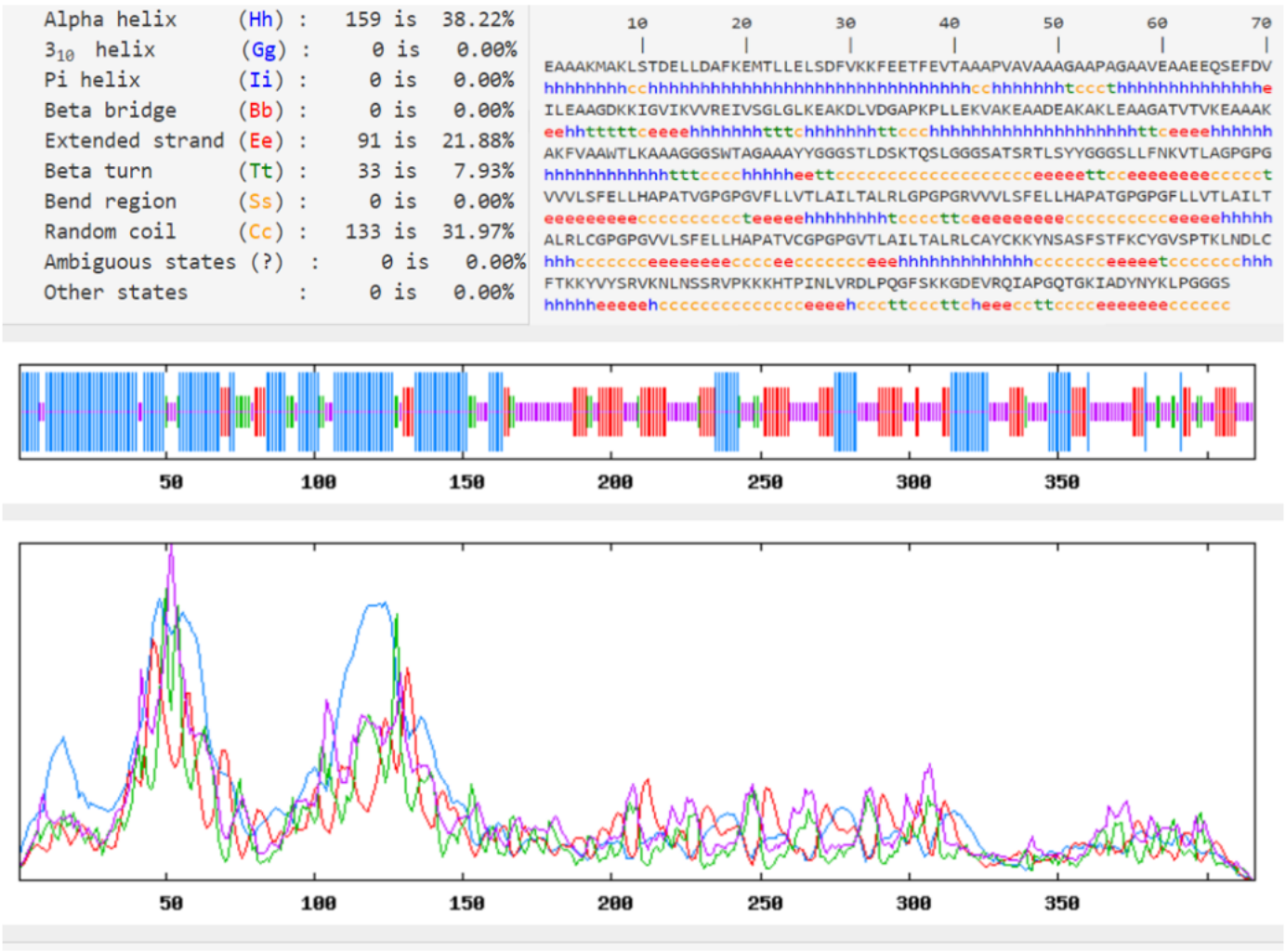

Figure 3

COVac-3 secondry structure prediction

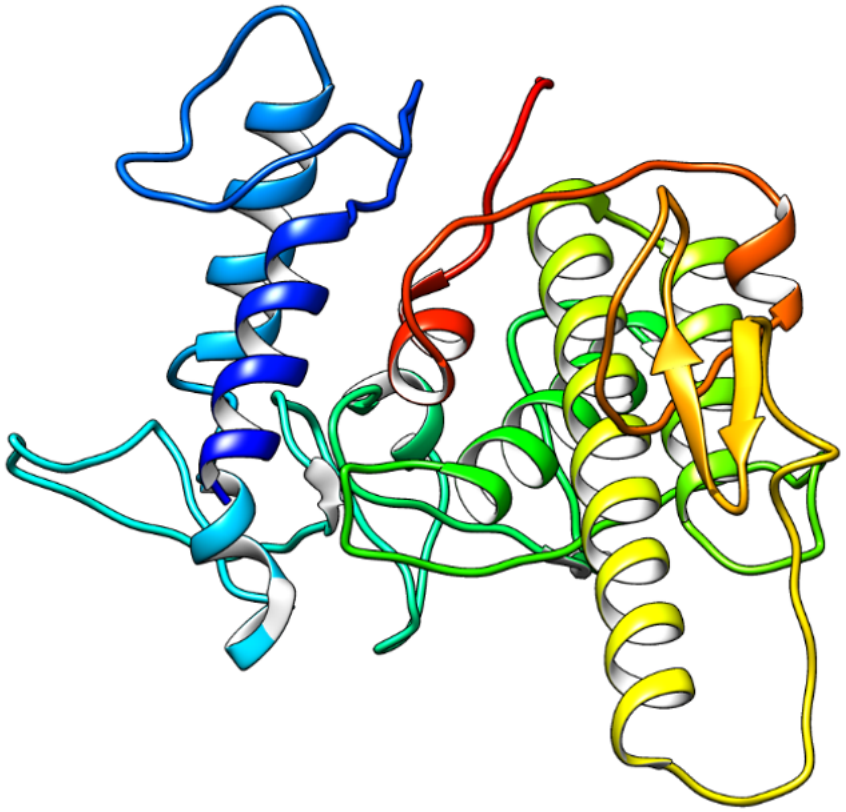

Figure 4

The 3D structure of COVac-1

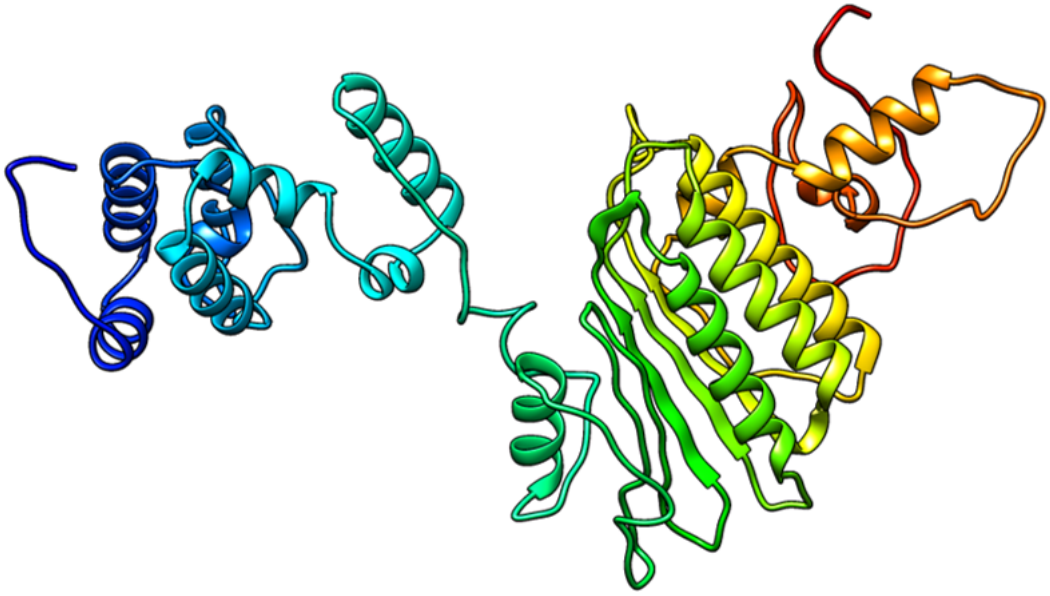

Figure 5

The 3D structure of COVac-2

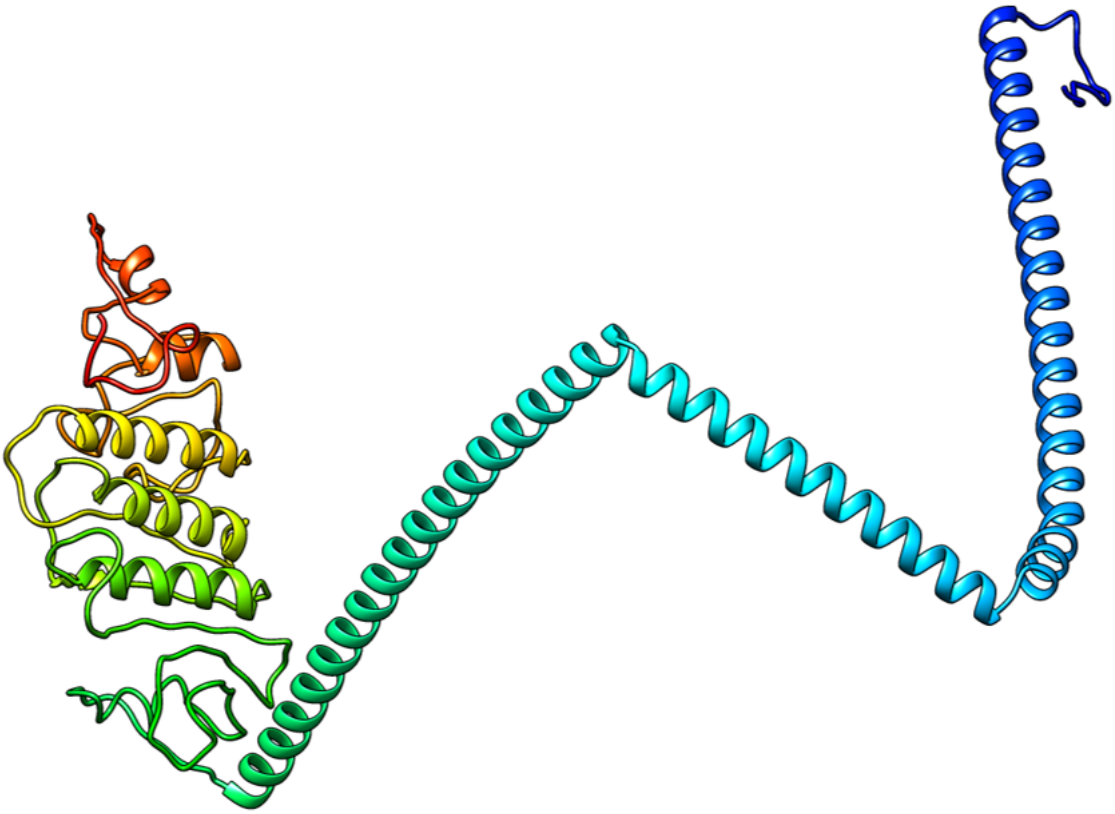

Figure 6

The 3D structure of COVac-3

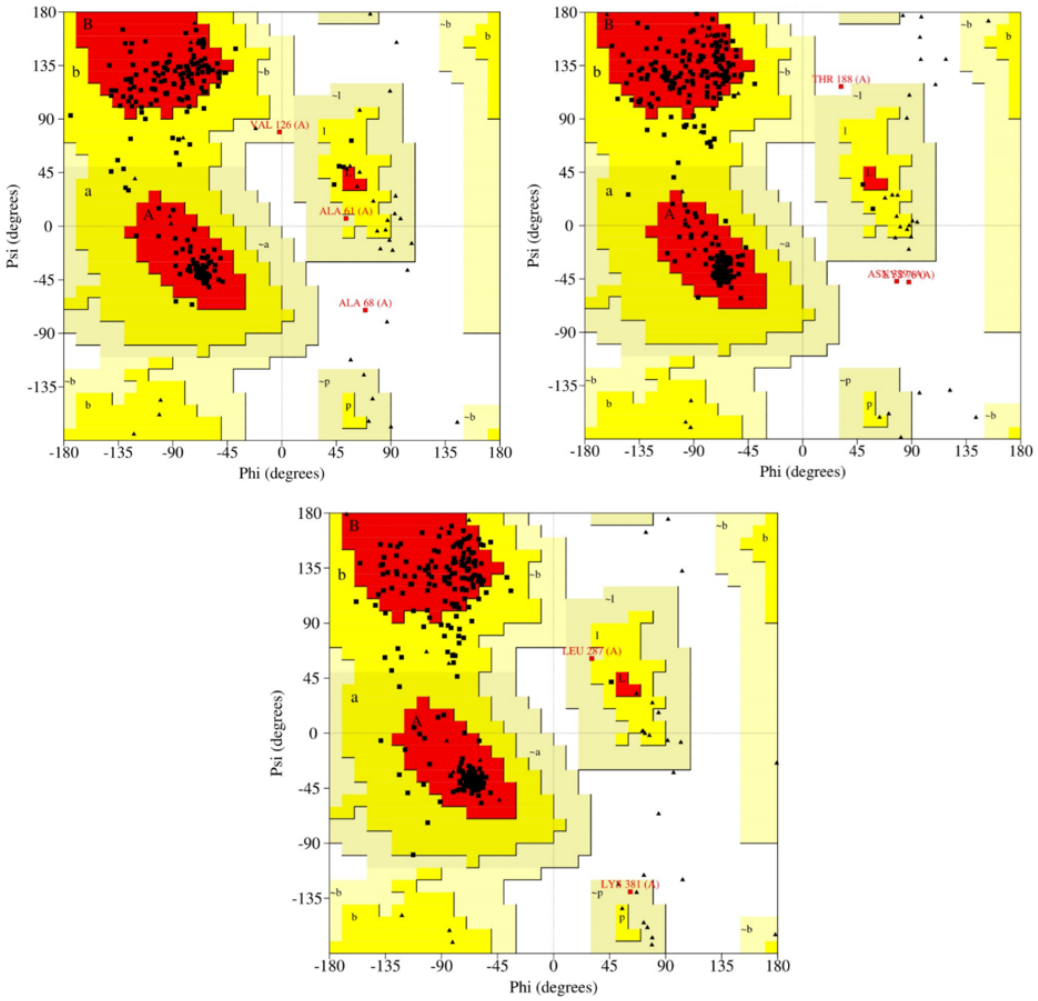

Figure 7

Ramachandran plot for the three vaccine. COVac-1 (top right), COVac-2 (top left), and COVac-3 (bottom).

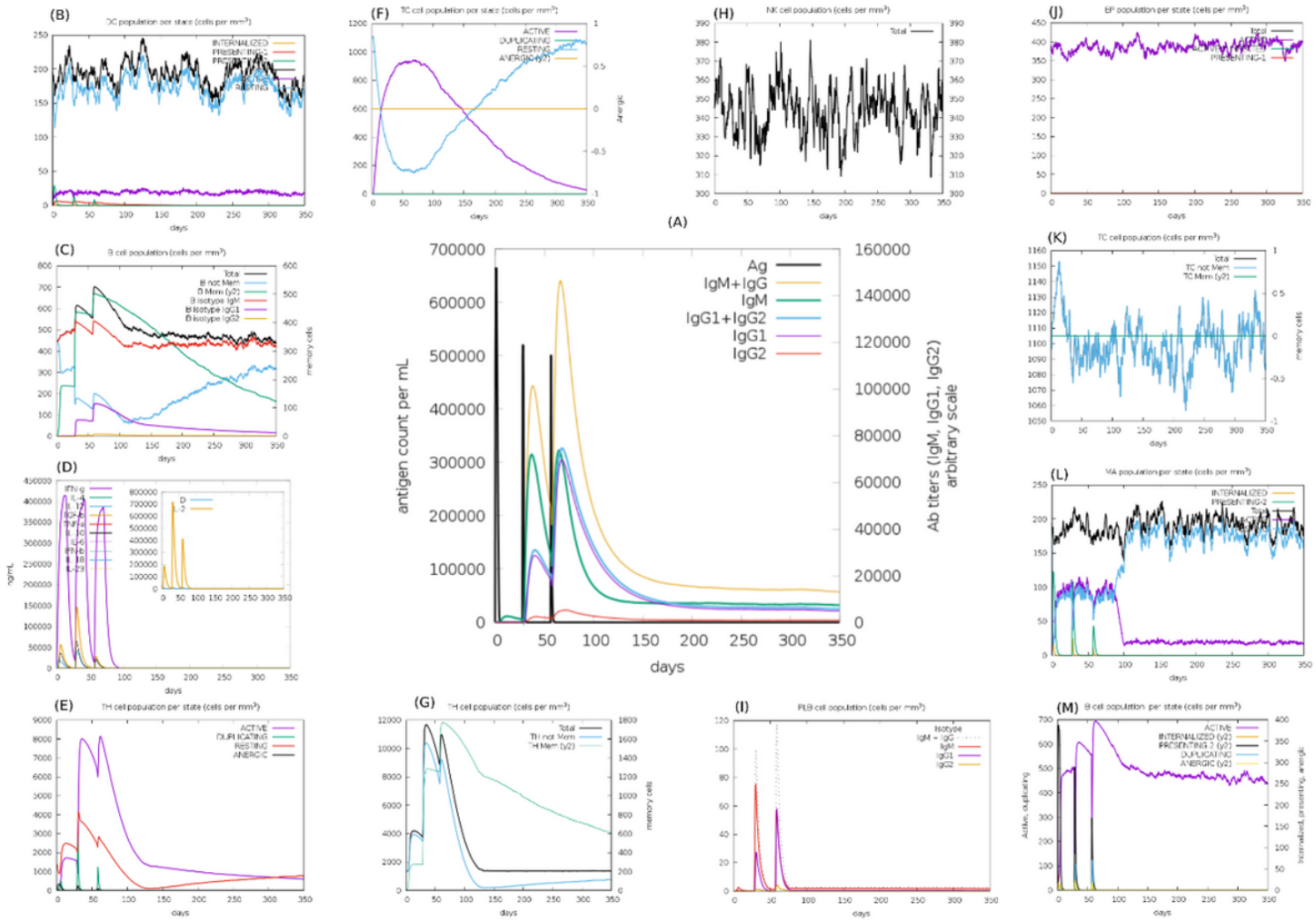

Figure 8

(A) Antigen, immunoglobulins, and Antibodies that sub-divided per isotype. (B) Dendritic cells. (C) B lymphocytes. (D) Cytokines. (E) CD4 T-helper lymphocytes. (F) CD8 T-cytotoxic lymphocytes (G) CD4 T-helper lymphocytes. (H) Natural Killer cells. (I) Plasma B lymphocytes. (J) Epithelial cells. (K)CD8 T-cytotoxic lymphocytes. (L) Macrophages. (M) B lymphocytes.

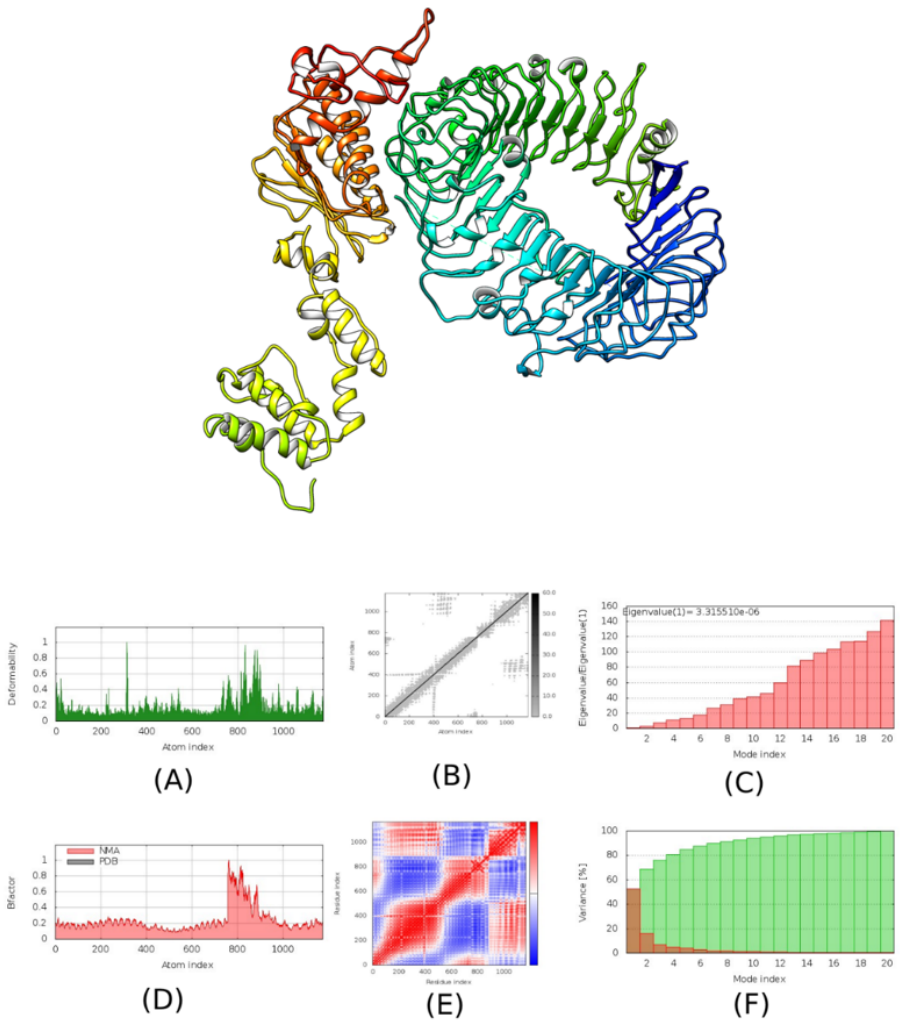

Figure 9

The results of a molecular dynamics simulation study of COVac-2 and TLR-8 docked complex. Here, (A) deformability, (B) elastic network (darker gray regions indicate stiffer regions), (C) eigenvalues, (D) B-factor, (E) co-variance map (correlated (red), uncorrelated (white), or anti-correlated (blue) motions) and (F) variance (red color indicates individual variances and green color indicates cumulative variances).

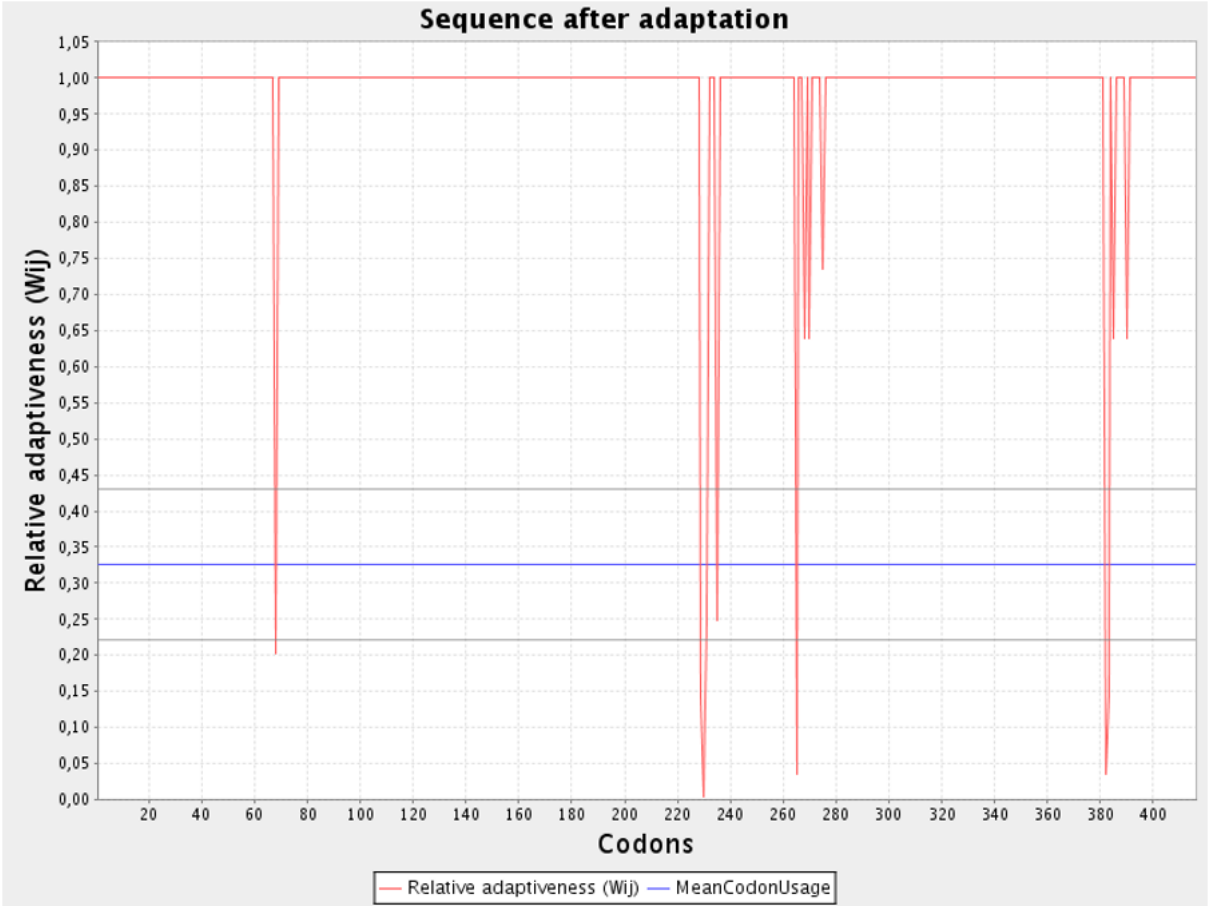

Figure 10

The results of the codon adaptation study of the best constructed vaccine, COVac-1.

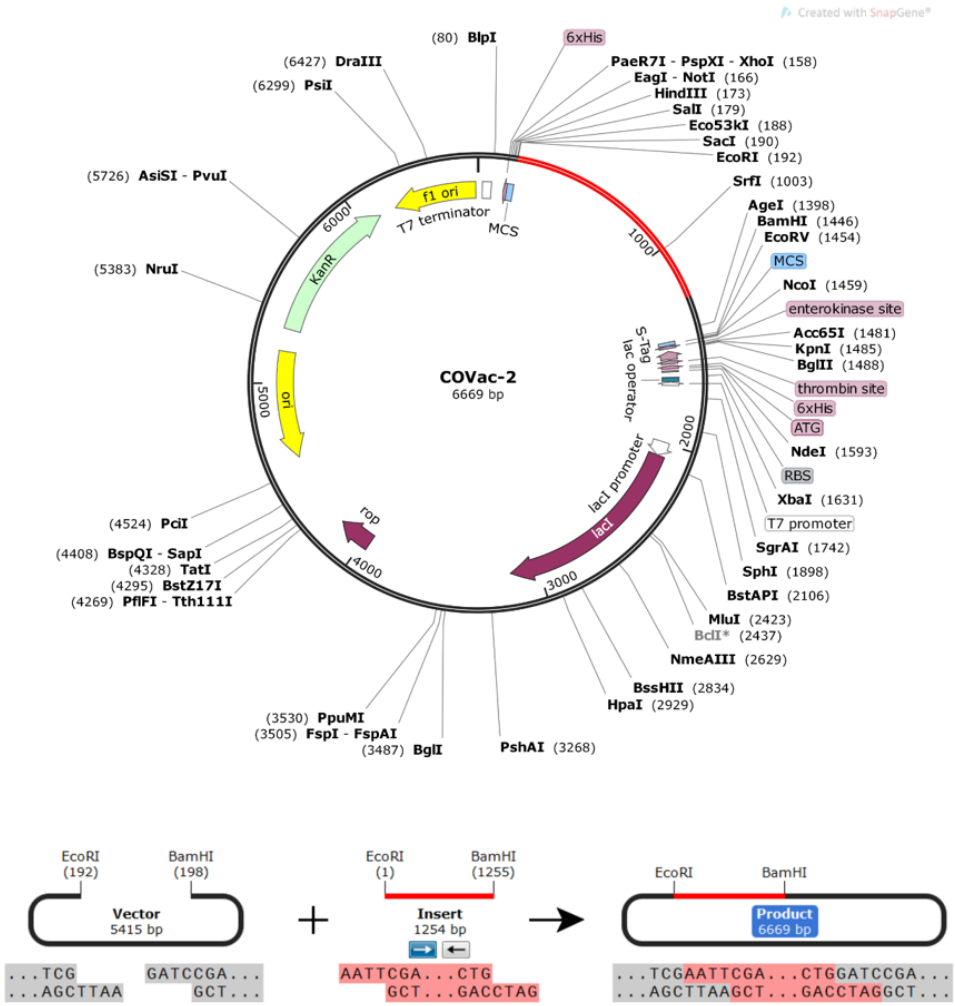

Figure 11

Constructed pET-30b(+) vector with the COVac-2 insert (marked in red color). In the plasmid, the larger purple-colored arrow represents the lacI gene (from 2073 bp to 3155 bp), the smaller purple-colored arrow represents the rop gene (from 3964 bp to 4155 bp), the yellow-colored arrow represents the origin of replication (from 4585 bp to 5173 bp), the light green colored arrow represents the AmpR (ampicillin resistance) gene (from 5295 bp to 6110 bp), the white rectangle represents the T7 terminator (from 26 bp to 73 bp), the light blue colored arrow represents the multiple cloning site (from 158 bd to 192 bp) and the desired gene has been inserted (marked by red color) between the 195 bp and 1255 bp nucleotide. Various restriction enzyme sites are mentioned in the plasmid structure.