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^{1*}

¹Department of Bioinformatics, Colors Medical Laboratories, Cairo, Egypt

Article Info

Article type:
Research

Article History:

Received: 2021-02-23
Accepted: 2021-03-15
Published: 2021-03-17

* Corresponding author:

Peter T. Habib

Department of Bioinformatics,
Colors Medical Laboratories, Cairo,
Egypt

Email: p.habib911@gmail.com

Keywords:

In Silico Vaccine Design
SARS-CoV-2
Reverse Vaccinology
Immunoinformatics

ABSTRACT

Introduction: 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.

Material and Methods: The material of SARS-COV2 Surface Glycoprotein (S), Membrane Protein (M), and Envelope Protein (E) were downloaded from the NCBI protein database. Each protein has undergone epitopes prediction for MHC class I epitopes, MHC class II epitopes, and Antibody of B-cell epitopes. Selected epitopes according to their antigenicity score was tested for allergenicity and toxicity. Finally, filtered epitopes were used in vaccine construction. Vaccines were constructed, docked against Toll-like receptor 3, and undergone Molecular Dynamic simulation. The vaccine with the best scores, subjected to immune stimulation and cloning design.

Results: Three vaccines were constructed, COVac-1, COVac-2, and COVac-3. Each vaccine was submitted into a deep investigation. The molecular dynamic simulation determines the stability and physical movement of protein atoms and molecules. After Molecular dynamics simulation, COVac-1 was having the best scores. COVac-1 was then subjected to immune simulation analysis to insure the stimulation of innate and adaptive immunity. After passing the immune simulation, COVac-1 was integrated into E.coli pET-30b plasmid using in silico cloning design.

Conclusion: Viral pandemics are threatened to face humanity today. The best scenario to fight against any pandemic is utilizing the full power of computational biology, especially immune-informatics, to design and discover in silico new vaccines or molecules that may stimulate the immune system against the invader pathogens or inhibit the pathogen life cycle.

Cite this paper as:

Habib PT. 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. Front Health Inform. 2021; 10: 67. DOI: [10.30699/fhi.v10i1.279](https://doi.org/10.30699/fhi.v10i1.279)

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 kilo bases. 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, and

avian [1-3]. So far, seven distinct human coronaviruses (HCoVs) have been identified. Four HCoVs strains, i.e., HCoV-OC43, HCV-229E, HCV-NL63, and HCV-HKU1, could induce common cold in immune-depressed people and two other HCoVs strains, i.e., SARS-CoV, and MERS-CoV, might cause severe acute respiratory syndrome [4-7]. 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 [8]. 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 [9]. The most common symptoms of COVID-19 onset include fever, cough, tiredness, diarrhea, and patients experience trouble breathing in extreme circumstances [10]. 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. 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. Interferon are a little powerful in combination with ribavirin. There are many efforts to use natural compounds and molecules to inhibit SARS-CoV-2 binding to ACE2, the receptor responsible for viral penetration [11] and classifier of proteins or coronavirus from CT scan using machine learning [12, 13]. However, we must further assess the efficacy of the combination solution [14]. This analysis was conducted in order to design a new epitope-based vaccines against three SARS-CoV-2 proteins. Those proteins are namely, surface glycoprotein (S) responsible for viral fusion events in the human cell during viral penetration [15, 16]; the Envelope protein (E) which is the cover protein for the virus components. It is also involved in many aspects of the virus' life cycle [17] and membrane glycoprotein (M) that mediates the interaction between virions and cellular receptors [18]. 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. [19, 20].

MATERIAL AND METHODS

The current experiment has been performed to improve possible SARS-CoV-2 vaccines by using reverse vaccination strategies, which is the using the pathogen's proteins to design a vaccine from it. In this experiment, the materials imported from and the

methods used were adapted from [21] work, which constructs vaccine by prediction of the epitope for B-cells, T- cells, and MHC class I and II. Those predicted epitopes have undergone investigation to filter the epitope that is non-allergenic, non-toxic, and potent antigenic assemble the filtered epitopes to build the vaccine construct. The potential vaccine constructs were analyzed using molecular docking and dynamic simulation to determine the strength and intactness of the vaccine. Finally, the vaccine passed the dynamic simulation and molecular docking analysis was chosen to design a cloning plasmid based on the E. coli plasmid system.

RESULTS

Viral protein sequence identification, selection, and retrieval

Proteins downloaded from NCBI database. Proteins retrieved in FASTA format was selected for the potential vaccine design. Those proteins are Membrane Glycoprotein (accession no: YP_009724393.1), 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: Selected proteins for candidate vaccine design

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

Antigenicity prediction and physicochemical property analysis of the protein sequences

Vaxijen webserver reports the three proteins as potent antigens that may stimulate the immune system (Table 2). For these three selected proteins, the physicochemical property analysis was conducted to determine some of the protein characteristics. 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.

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).

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

MHC class I and II molecules are two primary classes of major histocompatibility complex (MHC) molecules. They are found on the surface of the cell of all nucleated cells in the bodies of humans. The class I MHC function is to presenting intracellular proteins to cytotoxic T cells (CTLs). 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 [22]. Based on the antigenicity values however, eight MHC class I, 6 MHC class II, and four 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 [22]. Subsequently, vaccine construction was selected from epitopes with high antigenic, non-allergic, and non-toxicity.

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, GGGG, GPVPG, 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 [23]. The newly built vaccines were: COVac-1, COVac-2, and COVac-3 and further analyzed their physical properties (Table 7).

Table 2: The physicochemical properties of viral proteins

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

Table 3: The predicted epitopes for the Envelope protein

Protein	MHC Class /Antibody	Epitope	AS	Percentile 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
		LVTAILTALRLCAY	0.4070	0.38	ANTIGEN	NON-ALLERGEN	Non-Toxin
		VLLVTLAILTALRL	0.7218	0.38	ANTIGEN	NON-ALLERGEN	Non-Toxin
		VTLAILTALRLCAYC	0.8599	0.38	ANTIGEN	NON-ALLERGEN	Non-Toxin
	Antibody	YVYSRVKNLNSSRPV	0.4492	-	ANTIGEN	NON-ALLERGEN	Non-Toxin

Table 4: The predicted epitopes for the Membrane protein

Protein	MHC Class/ Antibody	Epitope	AS	Scores	Antigenicity (Virus model, threshold = 0.4)	Allergenicity	Toxicity
Membrane Protein	MHC Class I	ATSRTLSTYY	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	NGTITVEELKKLEQW	-0.1969	-	NON-ANTIGEN	ALLERGEN	Non-Toxin
		YRIGNYKLNTDHSSSDNIA	0.2216	-	NON-ANTIGEN	ALLERGEN	Non-Toxin

Table 5: The predicted epitopes for the Surface Glycoprotein

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
		VVLSFELLHAPATV	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
		RVVLSFELLHAPAT	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
		YNSASFSTFKCYGVSP TKLNDLCFT	1.4031	-	ANTIGEN	NON-ALLERGEN	Non-Toxin
		GDEV RQIAPGQTGKIADYNYKLP	1.1017	-	ANTIGEN	NON-ALLERGEN	Non-Toxin
		ELLHAPATVCGPKSTNLVKN	0.0029	-	NON-ANTIGEN	ALLERGEN	Non-Toxin
		NCTEVPVAIHADQLTPT	0.3987	-	NON-ANTIGEN	NON-ALLERGEN	Non-Toxin
		ASYQTQTNSPRRARSVASQ	0.2556	-	NON-ANTIGEN	NON-ALLERGEN	Non-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: The selected and filtered out epitope for the three protein

Protein Name	MHC Class I	MHC Class II	Antibody
Surface Glycoprotein	TLDSKTQSL	VLSFELLHAPATVCG	KHTPINLVRDLPQGFS
	WTAGAAAYY	VVLSFELLHAPATVC	GDEV RQIAPGQTGKIADYNYKLP
	LLFNKVTLA	RVVLSFELLHAPAT	YNSASFSTFKCYGVSP TKLNDLCFT
	-	VVLSFELLHAPATV	-
	-	FVFLVLLPLVSSQCV	-
	-	MFVFLVLLPLVSSQC	-
Membrane Protein	ATSRTLSTYY	-	-
Envelope Protein	FLLVTLAILTALRLC	-	YVYSRVKNLNSSRVP
	LVTAILTALRLCAY	-	-
	VFLLVTLAILTALRL	-	-
	VTLAILTALRLCAYC	-	-

Table 7: The vaccine constructs and their physicochemical properties

Vaccine name	The protein sequence of Constructed Vaccines	Length	GRAVY Score	The instability index (II)	Molecular weight	Theoretical pI
COVac-1	EAAAKGIINTLQKYYCRVRGGRCVAVL SCLPKKEEQIGKCSRGRKCCRRKKEA AAKAKFVAAWTLKAAAAGGGLDSDK TQSLGGGSATSRTLSYGGGSLFNK VTLAAGGSWTAGAAAYYGGPGVVV SFELLHAPATVCGPGPGFLLVTLAILT ALRLCGPGPGVVVLSFELLHAPATV PGPGVTLAILTALRLCAYCGPGPGVF LLVTLAILTALRLKKYVYSRVKNLNS SRVPKKKHTPINLVRDLPQGFSKKG DEVQRQAPGQTGKIADYNYKPKKY NSASFSTFKCYGVSPTKLNDLCFTGG GS	311	0.075	29.99	32451.97	9.90
COVac-2	EAAAKMAKLSTDELLDAFKEMTLLE LSDFVKKFEETFEVTAAPVAVAAA GAAPAGAAVEAAEEQSEFDVILEAAG DKKIGVIVVREIVSGLGLKEAKDLV DGAPKPLEKVAKEAADEAKAKLEA AGATVTVKEAAAKAKFVAAWTLKA AAGGSWTAGAAAYYGGGSLDSTKT QSLGGGSATSRTLSYGGGSLFNKV TLAAGPGPGVVVLSFELLHAPATV PGVFLVTLAILTALRLGPGPGRVV LSFELLHAPATGPGPGFLLVTLAILTA LRLCGPGPGVVLSFELLHAPATVCGP GPGVTLAILTALRLCAYCKYNSASF STFKCYGVSPTKLNDLCFTKKYVYSR VKNLNSSRVPKKKHTPINLVRDLPQ GFSKKGDEVQRQAPGQTGKIADYNYK LPGGGS	416	0.216	25.49	42730.60	9.07
COVac-3	EAAAKMAENPNIDDLPAPLAALGA ADLALATVNDLIANLRERAEETRAE TRTRVEERRARLTQFQEDLPEQFIEL RDKFTTEELRKAEGYLEAATNRYN ELVERGEAALQRLRSQTAFEDASAR AEGYVDQAVELTQEALGTVASQTRA VGERAAKLVGIELEAAAKAKFVAAW TLKAAAAGGSLFNKVTLAAGGSWT AGAAAYYGGGSLDSTKTQSLGGGSAT SRTLSYGGPGPGRVVLSFELLHAPA TGPGPVTLAILTALRLCAYCGPGPG VVLSFELLHAPATVCGPGPGVFLVT LAILTALRLGPGPGVVLSFELLHAP ATVGGPGPGFLLVTLAILTALRLCKG DEVQRQAPGQTGKIADYNYKPKKY NSASFSTFKCYGVSPTKLNDLCFTKK YVYSRVKNLNSSRVPKKKHTPINLVR DLPQGFSGGGS	445	-0.028	33.64	46918.76	9.05

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: Analysis of antigenicity, allergy, and physicochemical

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

Secondary and tertiary structure prediction of the vaccine constructs

The secondary structure refers to recurring arrangements within a chain of polypeptides in adjacent amino acid residue space. The amide hydrogen and carbonyl oxygen bonds in the peptide backbone maintain it. α -helix and β -structures are the main secondary structures. 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 1, 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 root mean square deviation in Ångström (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 (Fig 4, 5, 6). The RaptorX server used the templates [24] 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

Fig 1: COVac-1 secondary structure prediction

Fig 2: COVac-2 secondary structure prediction

Fig 3: COVac-3 secondary structure prediction

Fig 4: The 3D structure of COVac-1

Fig 5: The 3D structure of COVac-2

Fig 6: The 3D structure of COVac-3

3D structure refinement and validation

The three vaccine constructs were refined and validated in the 3D structure refinement and validation step to ensure that the protein structure is in the right orientation and order. 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 [25-27] 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).

Fig 7: Ramachandran plot for the three vaccines. COVac-1 (top right), COVac-2 (top left), and COVac-3 (bottom).

Vaccine protein disulfide engineering

A disulfide protein bond is a covalent connection in two cysteine residues between sulfur atoms (–SH) in the Thiol group. The disulfide is formed by oxidation of the two thiols (also called the SS-binding, Disulfide Bridge, or crosslink). It thus connects the two cysteines and their respective major peptide chains with the covalent disulfide binding. Conversely, a reduction reaction can disrupt a disulfide bond (e.g.,

using dithiothreitol). 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 [28]. The COVac-1 generated 43 amino acid pairs that could form disulfide bonds. However, only five 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.

Protein-protein docking study

The prediction of the complex structure given the structures of each protein is the protein-protein docking. The notion of a physical and steric complementarity at the protein-protein interface is at the heart of the docking methodology. 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 had 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 the best vaccine construct among the three constructed vaccines (Fig 8). Later, in silico immune simulation, the molecular dynamics simulation and in silico codon adaptation studies were conducted only on the COVac-2 vaccine.

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 The (helper) and T C (cytotoxic) cells with memory growth (Fig 8e, F). 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.

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 [29]. 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 Fig 9 docked complexes COVac-2 and TLR8 generated good 3.315510×10^{-6} 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 9g) [30].

Fig 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).

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 [31]. 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.

Fig 10: The results of the codon adaptation study of the best constructed vaccine, COVac-1.

Fig 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 bp 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.

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 [32].

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, GGS, GPGPG, and KK linkers after antigenicity and allergenicity check. The EAAK 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.

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.

AUTHOR'S CONTRIBUTION

The author agree on this final form of the manuscript, and attested that all authors contributed in the final draft of the manuscript.

CONFLICTS OF INTEREST

The author declare no conflicts of interest regarding the publication of this study.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

REFERENCES

1. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* 2016; 24(6): 490-502. PMID: 27012512 DOI: 10.1016/j.tim.2016.03.003 [PubMed]
2. Weiss SR, Navas-Martin S. Coronavirus pathogenesis

and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev.* 2005; 69(4): 635-64. PMID: 16339739 DOI: 10.1128/MMBR.69.4.635-664.2005 [PubMed]

3. Masters PS, Perlman S. Coronaviridae. In: Fields B, Knipe DM, Howley PM. *Fields virology.* 6th ed. Lippincott Williams & Wilkins; 2013.

4. van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJM, Wolthers KC, et al. Identification of a new human coronavirus. *Nat Med*. 2004; 10(4): 368-73. PMID: 15034574 DOI: 10.1038/nm1024 [\[PubMed\]](#)
5. Hamre D, Procknow JJ. A new virus isolated from the human respiratory tract. *Proc Soc Exp Biol Med*. 1966; 121(1): 190-3. PMID: 4285768 DOI: 10.3181/00379727-121-30734 [\[PubMed\]](#)
6. Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med*. 2003; 348(20): 1967-76. PMID: 12690091 DOI: 10.1056/NEJMoa030747 [\[PubMed\]](#)
7. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*. 2012; 367(19): 1814-20. PMID: 23075143 DOI: 10.1056/NEJMoa1211721 [\[PubMed\]](#)
8. Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, et al. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: What lessons have we learned? *Int J Epidemiol*. 2020; 49(3): 717-26. PMID: 32086938 DOI: 10.1093/ije/dyaa033 [\[PubMed\]](#)
9. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet*. 2020; 395(10223): 470-3. PMID: 31986257 DOI: 10.1016/S0140-6736(20)30185-9 [\[PubMed\]](#)
10. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020; 395(10223): 497-506. PMID: 31986264 DOI: 10.1016/S0140-6736(20)30183-5 [\[PubMed\]](#)
11. Habib PT, Saber-Ayad M, Hassanein SE. In silico analysis of 716 natural bioactive molecules from atlantic ocean reveals candidate molecule to inhibit spike protein. *Research Square*. 2021; Preprint.
12. Randhawa GS, Soltysiak MPM, el Roz H, de Souza CPE, Hill KA, Kari L. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *PLoS One*. 2020; 15(4): e0232391. PMID: 32330208 DOI: 10.1371/journal.pone.0232391 [\[PubMed\]](#)
13. Habib PT. COVATOR: A software for chimeric coronavirus identification. *bioRxiv*. 2020; Preprint.
14. Fehr AR, Perlman S. Coronaviruses: An overview of their replication and pathogenesis. *Methods Mol Biol*. 2015; 1282: 1-23. PMID: 25720466 DOI: 10.1007/978-1-4939-2438-7_1 [\[PubMed\]](#)
15. Petit CM, Melancon JM, Chouljenko VN, Colgrove R, Farzan M, Knipe DM, et al. Genetic analysis of the SARS-coronavirus spike glycoprotein functional domains involved in cell-surface expression and cell-to-cell fusion. *Virology*. 2005; 341(2): 215-30. PMID: 16099010 DOI: 10.1016/j.virol.2005.06.046 [\[PubMed\]](#)
16. Cavanagh D. The coronavirus surface glycoprotein. In: Siddell SG (eds). *The coronaviridae*. Springer; 1995.
17. Schoeman D, Fielding BC. Coronavirus envelope protein: Current knowledge. *Virol J*. 2019; 16(1): 1-22. PMID: 31133031 DOI: 10.1186/s12985-019-1182-0 [\[PubMed\]](#)
18. Rottier PJM. The coronavirus membrane glycoprotein. In: Siddell SG (eds). *The coronaviridae*. Springer; 1995.
19. Chong LC, Khan AM. Vaccine target discovery. In: Ranganathan S, Nakai K, Schonbach C (eds). *Encyclopedia of bioinformatics and computational biology*. Elsevier; 2019.
20. Mar'ia RR, Arturo CJ, Alicia J-A, Paulina MG, Gerardo A-O. The impact of bioinformatics on vaccine design and development. In: Afrin F (eds). *Vaccine*. INTECH; 2017.
21. Ullah MA, Sarkar B, Islam SS. Exploiting the reverse vaccinology approach to design novel subunit vaccines against Ebola virus. *Immunobiology*. 2020; 225(3): 151949. PMID: 32444135 DOI: 10.1016/j.imbio.2020.151949 [\[PubMed\]](#)
22. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res*. 2019; 47(D1): D339-43. PMID: 30357391 DOI: 10.1093/nar/gky1006 [\[PubMed\]](#)
23. Wu C-Y, Monie A, Pang X, Hung C-F, Wu TC. Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. *J Biomed Sci*. 2010; 17(1): 1-10. PMID: 21092195 DOI: 10.1186/1423-0127-17-88 [\[PubMed\]](#)
24. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*. 2012; 7(8): 1511-22. PMID: 22814390 DOI: 10.1038/nprot.2012.085 [\[PubMed\]](#)
25. Sateesh P, Rao A, Sangeeta SK, Babu MN, Grandhi RS. Homology modeling and sequence analysis of anx3. *International Journal of Engineering Science and Technology*. 2010; 2(5): 1125-30.
26. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*. 1993; 26(2): 283-91.
27. Zobayer M, Aowald Hossain ABM. In silico characterization and homology modeling of histamine receptors. *Journal of Biological Sciences*. 2018; 18(4): 178-91.
28. Craig DB, Dombkowski AA. Disulfide by design 2.0: A web-based tool for disulfide engineering in proteins. *BMC Bioinformatics*. 2013; 14: 1-7. PMID: 24289175 DOI: 10.1186/1471-2105-14-346 [\[PubMed\]](#)
29. Chauhan V, Rungta T, Goyal K, Singh MP. Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Sci Rep*. 2019; 9(1): 2517. PMID: 30792446 DOI: 10.1038/s41598-019-39299-8 [\[PubMed\]](#)
30. López-Blanco JR, Aliaga JI, Quintana-Ort'i ES, Chacón P. iMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Res*. 2014; 42(Web Server

issue): W271-6. PMID: 24771341 DOI: 10.1093/nar/gku339 [\[PubMed\]](#)

31. Solanki V, Tiwari V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*. *Sci Rep*. 2018; 8(1): 9044. PMID: 29899345 DOI: 10.1038/s41598-018-26689-7

[\[PubMed\]](#)

32. Demkowicz WE, Maa JS, Esteban M. Identification and characterization of vaccinia virus genes encoding proteins that are highly antigenic in animals and are immunodominant in vaccinated humans. *J Virol*. 1992; 66(1): 386-98. PMID: 1727494 DOI: 10.1128/JVI.66.1.386-398.1992 [\[PubMed\]](#)