
Immunoinformatics Design of Novel Multi-Epitope Subunit Vaccine for SARS-CoV-2 by Exploring Virus Conserved Sequences of the Spike Glycoproteins

Itemobong Ekaidem^{1,2,*}, Anietie Moses^{2,3}, Youtchou Tattfeng^{2,4}

¹Department of Chemical Pathology, University of Uyo, Uyo, Nigeria

²Institute of Biomedical Research and Innovations, University of Uyo, Uyo, Nigeria

³Department of Medical Microbiology and Parasitology, University of Uyo, Uyo, Nigeria

⁴Department of Medical Laboratory Science, Niger Delta University, Wilberforce Island, Nigeria

Email address:

seityjen@gmail.com (I. Ekaidem), itemobongekaidem@uniuyo.edu.ng (I. Ekaidem)

*Corresponding author

To cite this article:

Itemobong Ekaidem, Anietie Moses, Youtchou Tattfeng. Immunoinformatics Design of Novel Multi-Epitope Subunit Vaccine for SARS-CoV-2 by Exploring Virus Conserved Sequences of the Spike Glycoproteins. *International Journal of Immunology*.

Vol. 8, No. 3, 2020, pp. 42-52. doi: 10.11648/j.iji.20200803.12

Received: July 6, 2020; **Accepted:** July 21, 2020; **Published:** October 27, 2020

Abstract: Sars-CoV-2 infection also called COVID-19 is characterized by fever and signs of acute respiratory distress syndrome (ARDS). It is currently a global pandemic with high mortality rate in those with severe disease. Lack of effective vaccine and approved drug for treatment created a disastrous condition among the global communities. This study was designed as a step ahead in the path of protein-based subunit vaccine development. The primary amino acid sequence of SARS-CoV-2 spike glycoprotein was used to design a protein subunit vaccine construct. The molecular weight of vaccine protein was 58.4 kDa with a total number of 8170 atoms and 584 amino acid residues. The theoretical pI was found to be 8.54 showing its slightly basic nature while the total number of negative and positive charged residues were 29 and 33 respectively. The peptide vaccine construct has 147 (25.17%) polar residues and 375 (64.21%) hydrophobic residues. The vaccine construct has cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL) and B cell epitopes of varying lengths having great potential to stimulate high levels of IFN- γ production. It has potent antigenic properties but lacked allergenicity. It is stable and have a good binding affinity for the TLR-4 receptor. In general, this modelling applied a series of immunoinformatics tools in a sequential manner to find an effective vaccine that may be used effectively in fighting against the COVID-19 pandemic. This modelling, however, needs real life experimental validation to prove the workability of the computational work.

Keywords: Sars-CoV-2, COVID 19, Subunit Vaccine, Immunoinformatics, Multi-Epitopes Vaccine

1. Introduction

The Chinese Centre for Disease Control (China CDC) reported on December 31, 2019, a cluster of severe pneumonia cases of unknown aetiology in the city of Wuhan in the Hubei province of China. The disease was thereafter identified to be caused by a viral agent belonging to the family of corona viruses and resembling the severely acute respiratory syndrome corona virus-1 (SARS-CoV-1) that emerged in China in 2003 and Middle East respiratory syndrome corona virus (MERS-CoV) virus of 2012

epidemic. The viral agent was later called severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) and the disease named COVID-19 and declared a pandemic by the World Health Organization (WHO). The current outbreak of COVID-19 has spread to about 213 countries and territories of the world and about 8,242,999 confirmed cases have been reported globally with 430,806 deaths, as at June 18, 2020, according to the World Health Organization (WHO) Covid 19 Worldwide dashboard in collaboration with the China CDC and public health centres in other countries [1]. Currently all Africa countries have been affected and in Nigeria, the number of confirmed cases is increasing daily.

As at June 18, 2020, the Nigeria Centre for Disease Control (NCDC) reported a total of 18,480 confirmed cases with 475 death cases in Nigeria. However, this region of the world is reported to be least hit by the pandemic compared to others when considering the case fatality rate.

SARS-CoV-2 is a member of the *Coronaviridae* family, which are named after their crown-like appearance under the electron microscope [2, 3]. The family includes two subfamilies: *Letovirinae* and *Orthocoronavirinae*. The *Orthocoronavirinae* includes the genera *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus*. Betacoronaviruses and Alphacoronaviruses are known to typically infect only mammals, whereas gammacoronaviruses and deltacoronaviruses typically infect avian species and sometimes mammals as well. These viruses are common human pathogens and two types of alphacoronaviruses (229E, NL63) and two types of betacoronaviruses (OC43, HKU1) cause common cold in humans. Other pathogenic betacoronaviruses for humans include SARS-CoV-1, MERS-CoV and now SARS-CoV-2. [2, 4].

Coronaviruses have a very large (30kb+) single stranded positive sense RNA genome encoding for a number of open reading frames (orf). One frame encodes the spike or S protein, a class I fusion protein that mediates attachment of the virus to cell surface receptors to facilitate uptake into the endosomes. Proteolytic cleavage of the spike protein and fusion of viral and endosomal membranes trigger release of viral RNA into the cytosol. The RNA contains a 5' cap structure and a 3' poly (A) tail that allows expression of the replicase, which is encoded by approximately two thirds of the genome. The other third codes for the structural and accessory proteins [5]. In many coronaviruses, including SARS-CoV-2, the S protein is also cleaved into two subunits, S1 and S2, often by furin-like proteases, before entry into host cells [6].

Technology in vaccine development has significantly evolved in the last decade, making it possible for the production of several types of candidate vaccines viz RNA and DNA vaccines, recombinant protein vaccines and cell culture-based vaccines. Following the SARS-CoV-2 pandemic, there has been growing interest and race for coronavirus vaccine development to stem the tide of COVID-19. Scientists are currently trying to find out the best corona vaccine approach against the novel coronavirus using different technologies such as the use of viral nucleic acid (DNA and RNA molecules), whole virus (live attenuated and inactivated viruses), protein-based (surface protein subunits and virus-like particles) and viral vectors (replicating and non-replicating) [7]. Studies on SARS-CoV-1 and the related MERS-CoV vaccines have also shown that the spike protein on the surface of the virus could serve as an ideal target for a vaccine [8]. The genomic sequence of SARS-CoV-2 was recently identified and made widely available by Chinese researchers [9, 10]. Currently, there is paucity of information on which parts of the SARS-CoV-2 sequence are recognized by human immune responses. Such information could be of immediate relevance in vaccine design and evaluation of

vaccine candidate immunogenicity as well as monitoring of the potential consequences of mutational events and epitope escape as the virus is transmitted through human populations. Epitope data, yet available for SARS-CoV-2 are scanty in the literature, but there is a significant body of information about epitopes for coronaviruses in general, and in particular for Betacoronaviruses like SARS-CoV-1 and MERS-CoV, which cause acute respiratory disease in humans [11].

Although the structure of the spike protein of SARS-CoV-2 had been demonstrated at high resolution, further contributing to our understanding of this vaccine target [12], virus genetic polymorphism resulting in rapidly changing protein structures may affect the potency and efficacy of a protein-based candidate vaccine. Here, we adopt immunoinformatic approach to design a multi-epitope polypeptide subunit vaccine for SARS-CoV-2 using conserved sequences of virus spike glycoproteins. We also used validated bioinformatic tools to predict B and T cell epitopes of the conserved sequences of spike glycoproteins that are likely to be recognized in humans.

2. Methodology

2.1. Sequence Retrieval and Preparation

The NCBI Protein Database was used to retrieve the amino acid sequences of the spike glycoproteins of SARS-CoV-2 with protein id AF344187.1. The conserved domains of the spike glycoprotein sequence was determined from the Conserved Domain Database (CDD) of NCBI and blasted on the UniProtKB/Swiss-Prot (Swissprot) to retrieve conserved sequences using Conserved Domain Architecture Retrieval Tool (CDART). Four protein sequences with similarity score of 2 namely: Feline infectious peritonitis virus (strain 79-1146), P10033; Human coronavirus 229E, P15423 (query sequence); Avian infectious bronchitis virus (strain Beaudette), P11223 and SARS coronavirus, P59594 were selected for alignment and construction of a consensus sequence using MEGA 7 alignment tool. The consensus peptide sequence so constructed was then used as a query sequence on NCBI Blast to retrieve ten (10) protein sequences which were subjected to antigenicity prediction using AntigenPro. Sequences with antigenicity score of approximately 0.8 [13] were selected for use in the peptide construct.

2.2. Sequence Analysis by Immunoinformatics Tools

2.2.1. MHC Class I Binding Peptides Prediction

Three selected protein sequences of human corona virus spike glycoprotein were submitted in the FASTA format to predict the cytolytic T-Lymphocyte (CTL) epitope at the threshold score of 0.75 (default) using NetCTL 1.2 server [14]. Selected epitopes having a combined score of greater than 1.0 were selected as CTL epitope and further subjected to the Immune Epitope Database (IEDB) MHC class I immunogenicity prediction module. Predicted CTL epitopes for each human coronavirus spike glycoprotein selected as

input sequences were those epitopes with higher IEDB MHC class I immunogenicity score since they are known to have greater probability of eliciting an immune response [15].

2.2.2. Helper T-lymphocyte (HTL) Epitope Prediction

The Helper T-Lymphocyte (HTL) epitope prediction was performed using IEDB MHC-II epitope prediction module for all the three (3) selected human coronavirus spike glycoprotein sequences [16]. The available parameters were kept default except for MHC II allele selection where H2-IAad was the nominated alleles. Output epitopes were ranked and selected based on their percentile rank score where lower percentile rank represented greater binding affinity for HTL receptor.

2.2.3. Designing of Multi-epitope Subunit Vaccine

A total of three linkers namely EAAK, AAY, and GPGPG, were used to construct the final vaccine peptide. AAY and GPGPG linkers were added at the intra-epitope position to link the CTL and HTL epitopes, respectively. The EAAK linker was used to attach an adjuvant, APPHALS, which is also a Toll-like receptor agonist to the N-terminal of the peptide vaccine construct. A total of 13 CTL and 23 HTL epitopes were involved in the construct. These were to ensure that the end product will effectively activate both the innate and adaptive immune responses.

2.2.4. IFNepitope Prediction Scan for Interferon-gamma Producing Epitopes

To prove our work that the peptide vaccine construct will have ability to activate Th1 type immune response followed by the IFN- γ production, the peptide sequence of the vaccine was subjected to the IFNepitope server using the scan option. The sequence was submitted in the FASTA format followed by the selection of Scan prediction. Motif and SVM hybrid were selected as the approach and IFN-gamma versus other cytokines as model of prediction.

2.2.5. B-cell Epitope Prediction

ABCPREDS (default method) server was used to predict the linear B-cell epitopes of 20 amino acids long [17, 18]. The amino acid sequence of final vaccine construct was used as an input sequence in plain format followed by the selection of fixed length epitope prediction method and length of the epitope. The specificity threshold was set to be by default at 75% to obtain the result in a user-friendly format. While conformational epitopes were predicted using ElliPro server for the input of tertiary protein structure of the vaccine construct.

2.2.6. Antigenicity and Allergenicity prediction of the Peptide Vaccine CONSTRUCT

The antigenicity of the predicted peptide vaccine construct was determined using ANTIGENpro server, a sequence based, pathogen independent and alignment-free prediction method that is used to check the antigenic behaviour of vaccine proteins [19]. AllerTOP v. 2.0 was used to check the allergenicity of the vaccine construct based on the method that uses auto cross-covariance (ACC) transformation of

protein sequences into uniform equal-length vectors. Input protein sequence of vaccine protein was classified by the k-nearest neighbour algorithm (kNN, k = 1) which is based on the training set of 2427 known allergen from different species and 2427 nonallergen from similar species [20].

2.3. Assessment of Physicochemical Properties of the Peptide Vaccine Construct

ProtParam web server is a part of Expert Protein Analysis System (EXPASY) and was used to define various physicochemical properties of the predicted vaccine construct. The primary protein sequence of the vaccine was used to predict the various physicochemical parameters including molecular weight (kDa), estimated half-life, theoretical pI, aliphatic index, grand average of hydropathy (GRAVY) and molar extinction coefficient.

2.3.1. Tertiary Structure Prediction

Protein molecule achieves maximum stability in its lowest energy state by proper bending and twisting to form a tertiary structure. It is the interaction between the amino acids side chain residue which is responsible for stabilizing the protein structure. The 3-dimensional structure of the predicted vaccine construct was obtained by utilizing Phyre2 structure prediction server. Phyre2 is a pure *ab initio* method that can be used to build a 3D model in a template-free manner [21]

2.3.2. Refinement of 3D Vaccine Model and Validation

It is the degree of likeness between the target and available template structure that determines the quality of protein model structure created by contemporary protein structure prediction techniques [22, 23]. Therefore, it was necessary to improve the template-based prediction model beyond the accuracy by utilizing the template information. To fulfil this thought, output model of Phyre2 server was subjected to the GalaxyRefine web server [24], which is based on the CASP10 tested refinement method. GalaxyRefine performs reshaped structure perturbation followed by overall structural relaxation by performing molecular dynamics simulation.

2.3.3. Disulfide Engineering for Vaccine stability

Disulfide bonds are covalent Interactions that provide considerable stability to protein model by maintaining precise geometric conformations. Disulfide engineering is a novel approach for creating disulfide bonds into the target protein structure. Therefore, the refined model of final the vaccine construct was subjected to disulfide engineering using the Disulfide by Design 2.0 [25]. Initially, the refined protein model was uploaded and run for the residue pair search that can be used for the disulfide engineering purpose. A total of 3 residue pairs were selected to mutate them with cysteine residue using create mutate function of the Disulfide by Design 2.0 server.

2.4. Codon Adaptation and in Silico Cloning

The primary sequence of vaccine protein was submitted to the Java Codon Adaptation Tool (JCAT) to adapt their codon

usage to most sequenced prokaryotic organisms (*E. coli* K12) [13]. The CAI value and GC content of the adapted sequence were also noted. The adapted nucleotide sequence corresponding to the designed vaccine construct was cloned into the *E. coli* pET30 (+) vector using the restriction cloning module of SnapGene tool [26].

2.5. Molecular Docking of Vaccine Constructs with TLR4

TLR-4 was selected as receptor and its PDB file (PDB id: 4G8A) was obtained from RCSB-Protein Data Bank while the refined model of vaccine protein was used as a ligand. Protein-protein docking was performed using the ClusPro 2.0: protein-protein docking server, to check the binding affinity between them [27].

3. Results and Discussion

3.1. Sequence Retrieval and Preparation

Three (3) sequences with antigenicity score of approximately 0.8 which were selected for use in the peptide construct are shown in table 1. The use of conserved domain consensus sequence to construct the peptide vaccine was to ensure that vaccine remain active and efficacious even in the event of viral protein polymorphism. Although the specificity of the vaccine may not be 100%, its sensitivity could be enhanced by this novel approach. Provide justification for scores that signify selection/non-selection!!

Table 1. Retrieval of Coronaviruses Spike glycoprotein sequences from the NCBI database and their antigenicity scores for the peptide vaccine construct.

Serial No.	Protein Accession No.	Protein Name	Antigenicity Scores	Selected/ Non-selected
1	AAR07628.1	spike glycoprotein [SARS coronavirus BJ302]	0.8305	Selected
2	AAR07629.1	spike glycoprotein [SARS coronavirus BJ302]	0.7793	Non-selected
3	ADC35483.1	spike glycoprotein precursor [SARS coronavirus HKU-39849]	0.7795	Non-selected
4	AAR07625.1	spike glycoprotein [SARS coronavirus BJ302]	0.7781	Non-selected
5	AAR07627.1	spike glycoprotein [SARS coronavirus BJ302]	0.7761	Non-selected
6	AAU81608.1	S protein [SARS Coronavirus CDC#200301157]	0.7915	Selected
7	ABD72984.1	spike glycoprotein [Severe acute respiratory syndrome-related coronavirus]	0.7893	Selected
8	ACZ72093.1	spike glycoprotein precursor [SARS coronavirus wtic-MB]	0.7802	Non-selected

3.2. CTL Epitope Prediction and Immunogenicity Assessment

The prediction the Cytotoxic T Lymphocyte receptor specific immunogenic epitopes was performed using the NetCTL 1.2 server and total 132 CTL epitopes having COMB Score ≥ 0.75 and 9mer length sequences were obtained from the 3 selected proteins (14). The NetCTL server used for this experiment combines predictors of proteasomal processing, TAP transport, and MHC I binding to produce an overall score for each peptide's intrinsic potential of being a T cell epitope. Thirteen of these 9mer epitopes with higher IEDB immunogenicity score were used

as input sequences for CTL immune responses as shown in table 2. The immunogenicity of epitopes was determined as per the instruction of IEDB (28), higher scores indicate greater probability to elicit an immune response. Cytotoxic T-lymphocytes are a CD8+ subset of T-cell responses to kill those target cells having intracellular viral, bacterial or protozoan infection [29]. During infection, whenever cytotoxic T lymphocytes encounter the MHC-I mounted antigen specific to their receptor, they enter the cell cycle and perform several mitotic divisions followed by differentiation into the effector and memory cells [30]. Delete what is repeated as in methodology section! Provide justification for scores that signify selection/non-selection!!

Table 2. Cytotoxic T-Lymphocytes (CTL) prediction and immunogenicity by NetCTL 1.2).

S/N	Accession Number	Epitope	Start position	Length	Comb Score	IEDB Immunogenicity Score	Selected/ Non-selected
1	AAR07628.1	ATSTGNYNY	430	9	3.3053	0.00537	Selected
2	AAR07628.1	STGNYNYKY	432	9	3.0280	-0.13432	Not Selected
3	AAR07628.1	CTPPALNCY	467	9	1.7583	-0.02406	Not Selected
4	AAR07628.1	NATKFPSVY	330	9	1.1637	-0.21634	Not Selected
5	AAR07628.1	CVADYSVLY	348	9	2.5753	-0.09595	Not Selected
6	AAR07628.1	TQDLFLPFY	55	9	1.8726	0.15864	Selected
7	AAR07628.1	NAFNCTFEY	155	9	1.0697	0.17283	Selected
8	AAR07628.1	NKDGFLVYV	189	9	0.9973	0.17283	Selected
9	AAU81608.1	HTSSMRGVY	33	9	3.0674	-0.28973	Not Selected
10	AAU81608.1	SAAAYFVGY	248	9	1.4644	0.21331	Selected
11	AAU81608.1	LTDDMIAAY	847	9	3.7201	0.03968	Selected
12	AAR07628.1	TSQKSIVAY	669	9	2.1228	-0.23272	Not Selected
13	AAR07628.1	KTSVDCNMY	715	9	2.4743	-0.14937	Not Selected
14	AAR07628.1	GTATAGWTF	862	9	1.8825	0.33138	Selected
15	AAR07628.1	TTGIGYQPY	486	9	2.0576	0.0702	Selected
16	AAR07628.1	ASSEVAVLY	590	9	3.0826	0.15244	Selected
17	AAR07628.1	GAEHVDTSY	638	9	1.5677	0.06223	Selected

S/N	Accession Number	Epitope	Start position	Length	Comb Score	IEDB Immunogenicity Score	Selected/ Non-selected
18	ABD72984.1	TSQKSIVAY	669	9	2.1228	-0.23272	Not Selected
19	ABD72984.1	TQDLFLPFY	774	9	3.1247	0.21023	Selected
20	ABD72984.1	NAFNCTFEY	820	9	2.7523	0.14875	Selected
21	ABD72984.1	NKDGFLYVY	998	9	2.6432	0.21268	Selected

3.3. HTL Epitope Prediction

Helper T-lymphocyte is the key player of both humoral and cell-mediated immune response and hence are crucial part of the prophylactic and immunotherapeutic vaccines [31]. The 3 selected spike glycoprotein sequences subjected to IEDB MHC-II epitope prediction module and 1947 epitopes of 15mer length were obtained for H2-1Ad allele. In

order to express high immunogenic epitopes, they must have a lower percentile rank and IC50 value [26]. Only 23 epitopes out of the 1947 that have lowest percentile rank ranging from 0.03 – 3.0 were selected for the vaccine construct (Table 3). Their *simm* align IC50 value ranged from 483 – 800 denoting that the 23 epitopes with lowest percentile rank, also have enormous capacity for the HTL immune responses.

Table 3. Predicted Helper T-lymphocyte (HTL) specific epitopes and their percentile rank obtained from the immune epitope database.

S/N	Allele	Sequence ID	Start	End	Peptide	Method	Percentile Rank
1	H2-1Ad	2	996	1010	RAAEIRASANLAATK	Consensus (simm/nn)	1.46
2	H2-1Ad	1	848	862	TVDMIAAYTAALVSG	Consensus (simm/nn)	1.52
3	H2-1Ad	1	845	859	PLLTVDMIAAYTAAL	Consensus (simm/nn)	1.60
4	H2-1Ad	3	785	800	AIPFAMQMQLRLNYV	Consensus (simm/nn)	0.73
5	H2-1Ad	1	994	1008	LIRAAEIRASANLAA	Consensus (simm/nn)	0.74
6	H2-1Ad	1	990	1004	VTQQLIRAAEIRASA	Consensus (simm/nn)	0.79
7	H2-1Ad	1	989	1003	YVTQQLIRAAEIRAS	Consensus (simm/nn)	1.04
8	H2-1Ad	1	995	1009	IRAAEIRASANLAAT	Consensus (simm/nn)	1.10
9	H2-1Ad	1	849	863	VDMIAAYTAALVSGT	Consensus (simm/nn)	1.28
10	H2-1Ad	1	850	864	DMIAAYTAALVSGTA	Consensus (simm/nn)	1.29
11	H2-1Ad	1	846	860	LLTVDMIAAYTAALV	Consensus (simm/nn)	1.31
12	H2-1Ad	1	666	680	LRSTSQKSIVAYTMS	Consensus (simm/nn)	1.45
13	H2-1Ad	1	991	1005	TQQLIRAAEIRASAN	Consensus (simm/nn)	0.81
14	H2-1Ad	2	992	1006	QQLIRAAEIRASANL	Consensus (simm/nn)	0.82
15	H2-1Ad	2	993	1007	QLIRAAEIRASANLA	Consensus (simm/nn)	1.02
16	H2-1Ad	2	998	1012	AEIRASANLAATKMS	Consensus (simm/nn)	1.65
17	H2-1Ad	3	849	863	DDMIAAYTAALVSGT	Consensus (simm/nn)	1.68
18	H2-1Ad	1	844	858	PPLLTVDMIAAYTAA	Consensus (simm/nn)	1.75
19	H2-1Ad	3	665	679	LLRSTSQKSIVAYTM	Consensus (simm/nn)	1.80
20	H2-1Ad	2	997	1011	AAEIRASANLAATKM	Consensus (simm/nn)	1.83
21	H2-1Ad	1	743	757	TQLNRALSGIAAEQD	Consensus (simm/nn)	2.25
22	H2-1Ad	2	851	865	MIAAYTAALVSGTAT	Consensus (simm/nn)	2.30
23	H2-1Ad	3	848	862	TDDMIAAYTAALVSG	Consensus (simm/nn)	2.35

3.4. Construction of Multi-epitope Subunit Vaccine

A peptide vaccine construct of 584 amino acid residues was designed using the 13 CTL and 23 HTL epitopes [13]. In order to attain maximum immune response TLR-4 agonist (RS09) was used as an adjuvant at the N-terminal site of the vaccine construct [32]. Each joint was occupied by the suitable linkers as described by Nezafat and colleagues [31] for example, adjuvant and CTL epitopes were combined together by EAAAK linker, intra-CTL and intra-HTL epitopes joint by AAY and GPGPG linker, respectively. Final vaccine construct was obtained having adjuvant, linkers, CTL and HTL epitopes in a sequence moving from N-terminal to C-terminal. As this designed subunit vaccine consisting of immunogenic CTL and HTL epitopes along with suitable adjuvant and linkers, it may have the ability to evoke adequate immune response against SARS-CoV-2 within the human host [26]. One of the mechanisms by which the innate immune system senses the invasion of pathogenic

microorganisms is through the Toll-like receptors (TLRs), which recognize specific molecular patterns that are present in microbial components. Stimulation of different TLRs induces distinct patterns of gene expression, which not only leads to the activation of innate immunity but also instructs the development of antigen-specific acquired immunity.

3.5. Antigenicity and Allergenicity Prediction of Designed Vaccine

The antigenicity Probability score of the designed vaccine construct was 0.81 using an alignment-free ANTIGENpro server. The server is a sequence based, pathogen independent and alignment-free prediction method that is used to check the antigenic behaviour of vaccine proteins. It uses SVM classifier to summarize the probable antigenic or non-antigenic nature of proteins. ANTIGENpro uses the existing protein antigenicity microarray files of eight feature sets for five pathogens to construct two-stage architecture, the first one is multiple representations of the primary protein

sequence and the second one is five machine learning algorithms [19]. The antigenicity probability score recorded here was acceptable and represents the antigenic nature of a peptide vaccine construct. The antigenicity score for this vaccine construct is comparable with the antigenicity of subunit vaccine reported elsewhere [26]. A vaccine given to human host must be immunogenic and capable of triggering significant humoral immune response which ultimately leads to the memory cell formation against the pathogenic epitopes. Similarly, allergenicity prediction of the vaccine construct was determined using AllerTOP online server and found that the vaccine protein is nonallergic in nature and safe for the human use [19]. Allergy is an overreaction by our immune system to the previously encountered, ordinarily harmless substance that may manifest as sneezing, wheezing, skin rash, and swelling of the mucous membrane. It is associated with the IgE antibody response. Hence, the predicted vaccine construct must be free from being allergenic in nature

3.6. B-cell Epitope Mapping for Peptide Construct

The identification of B-cell epitopes is an essential part in vaccine designing. Using BCPREDS server, a total 565 B-

cell linear epitopes of 20mer length were predicted from the primary input sequence of final vaccine construct at a default threshold of 0.50. Among them, only 17 epitopes were observed to have high probability score of ≥ 0.80 (Table 4). B-cells are a key player of humoral immunity. An epitope corresponding to the B-cell receptor plays an important role in vaccine design for antibody production [33]. Therefore, due to the selection of highest scoring B-cell epitope among designed subunit vaccine, our vaccine may have the ability to enhance humoral immunity as well as cell mediated immunity [13]. ElliPro- Epitope Prediction based on structural protrusion was performed using IEDB Analysis Resource tools for continuous and discontinuous B cell epitope. This method predicts epitopes based on solvent accessibility and flexibility. Continuous and discontinuous epitopes of 57 amino acids long were also predicted from the final 3D model of vaccine construct. A total of 6 continuous and 6 discontinuous B cell epitopes were generated at a default threshold probability score of 0.50. Among these, two (2) continuous and one (1) discontinuous epitopes had a probability score of ≥ 0.70 (Figure 1B). The obtained probability scores also confirming the immunogenic behaviour of the designed subunit vaccine [18].

Table 4. B Cell epitopes prediction for SARS-CoV-2 spike glycoprotein vaccine construct using (BCPREDS server).

S/N	Sequence	Start position	Score
1	AAVASSEVAVLYAAYGAEHV	69	0.89
2	ANLAATKMSGPGDMMIAA	476	0.88
3	GPGPGTVDMIAAYTAALVSG	425	0.87
4	TAALVSGTAGPGPGLLTVDM	358	0.86
5	GIGYQPYAAVASSEVAVLYA	62	0.85
6	AVLYAAYGAEHVDTSYAAYG	77	0.84
7	IVAYTMGPGPGAAEIRASAN	539	0.84
8	LVSGTGPGPGDMIAAYTAAL	342	0.83
9	RASGPGGIRAAEIRASANL	304	0.83
10	GPGVDMIAAYTAALVSGTGP	329	0.82
11	AGPGPGVTQQLIRAAEIRAS	206	0.82
12	ATAGWTFAAYSAAAYFVGYA	98	0.81
13	IAAYTAALGPGPGAEIRASA	457	0.81
14	YVYAAAYATSTGNYNYYAAYTT	42	0.81
15	PGPGLLTVDMIAAYTAALVG	368	0.81
16	GYVTQQLIRAAEIRASGPGP	291	0.80
17	ANLGPMPGQLIRAAEIRASA	264	0.80

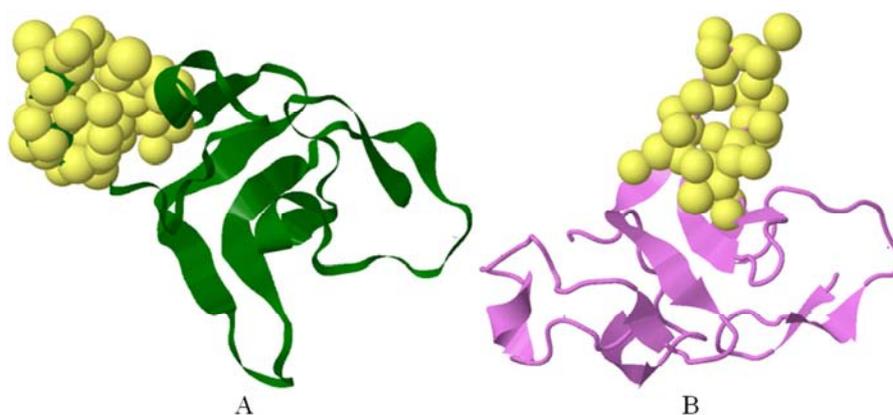


Figure 1. Humoral epitope predictions for subunit vaccine. (A) Showing the continuous B-cell epitopes (yellow colour) among the 3D structure of final vaccine construct (green colour). (B) Discontinuous conformational B-cell epitopes (yellow colour) showing the sequence subunits composed of antigenic epitopes that will come in direct contact with immune receptor.

3.7. IFNepitope Prediction Scan for Interferon-Gamma Producing Epitopes

The peptide vaccine construct was tested for its ability to activate Th1 type immune response followed by the IFN- γ production. The peptide sequence of the vaccine was subjected to the IFNepitope server using the scan option. About 210 positive epitopes were generated, out of which 75 epitopes had a probability score of ≥ 0.50 . Interferon gamma (IFN γ) is a dimerized soluble cytokine that is produced by activated or antigen-stimulated lymphocytes. It is a cytokine that is critical for innate and adaptive immunity against viral, some bacterial and protozoal infections. IFN γ is an important activator of macrophages and inducer of Class II major histocompatibility complex (MHC) molecule expression.

3.8. Physicochemical Properties Assessment

The physicochemical parameters of vaccine construct were determined by using ProtParam server and evaluated for seven parameters. The molecular weight of vaccine protein was found to be 58.4 kDa with a total number of 8170 atoms and 584 amino acid residues. The high molecular weight protein will favour the antigenicity of the vaccine construct [13]. The theoretical pI was found to be 8.54 showing its slightly basic nature while the total numbers of negative and positive charge residues were 29 and 33 respectively. The extinction coefficient was found to be and 1mg/ml (17.113 μ M) solution of the protein has absorbance of 1.06 at 280nm measured in water, assuming that all cysteine residues are reduced. The score obtained for instability index was 32.46, showing the stable nature of vaccine construct. The peptide vaccine construct has 147 (25.17%) polar residues and 375 (64.21%) hydrophobic residues. The value of the aliphatic index and Grand average of hydropathicity (GRAVY) were 79.59 and 0.159, respectively. The estimated value of aliphatic index represents the thermostable nature of

designed subunit vaccine because higher the value of aliphatic index, greater will be associated with thermo stability (26). A positive value of GRAVY for the input subunit vaccine represents the hydrophobic nature of vaccine (26). Conclusively, the designed vaccine is immunogenic, thermostable and hydrophobic in nature.

3.9. Tertiary Structure Prediction, Refinement, and Validation

The tertiary structure was predicted by using the Phyre2 server and 3D model was obtained and shown in Figure 2A. The best template used for the homology modelling was structure of the rhs-repeat containing BC component of the secreted ABC toxin complex from *Yersinia entomophaga*, chain B (PDB ID: 4IGL) with 95.3% confidence and 13% coverage. The B chain of toxin complex is not toxic on its own but requires complete assembly of the A, B and C subunits to express insecticidal properties [34]. Total 76 amino acid residues were modelled as a single domain with 84% disorder. Secondary structure information resulting in the presence of 3% helix, 7% Beta sheet, and 90% coiled structure. Further, protein refinement using GalaxyRefine leads to the increase in a number of residues in the favoured region [26]. Initially, 85.2% of residues were in the Rama-favoured region while after refinement the number of residues in the Rama-favoured region reached to 93.2%. The refinement output was also validated by plotting Ramachandran plot and found that 93.8% residue in Rama-favoured region, 2.5% in the generously allowed region and only 3.7% residues in outlier region (Figure 2B). Ramachandran plot is based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

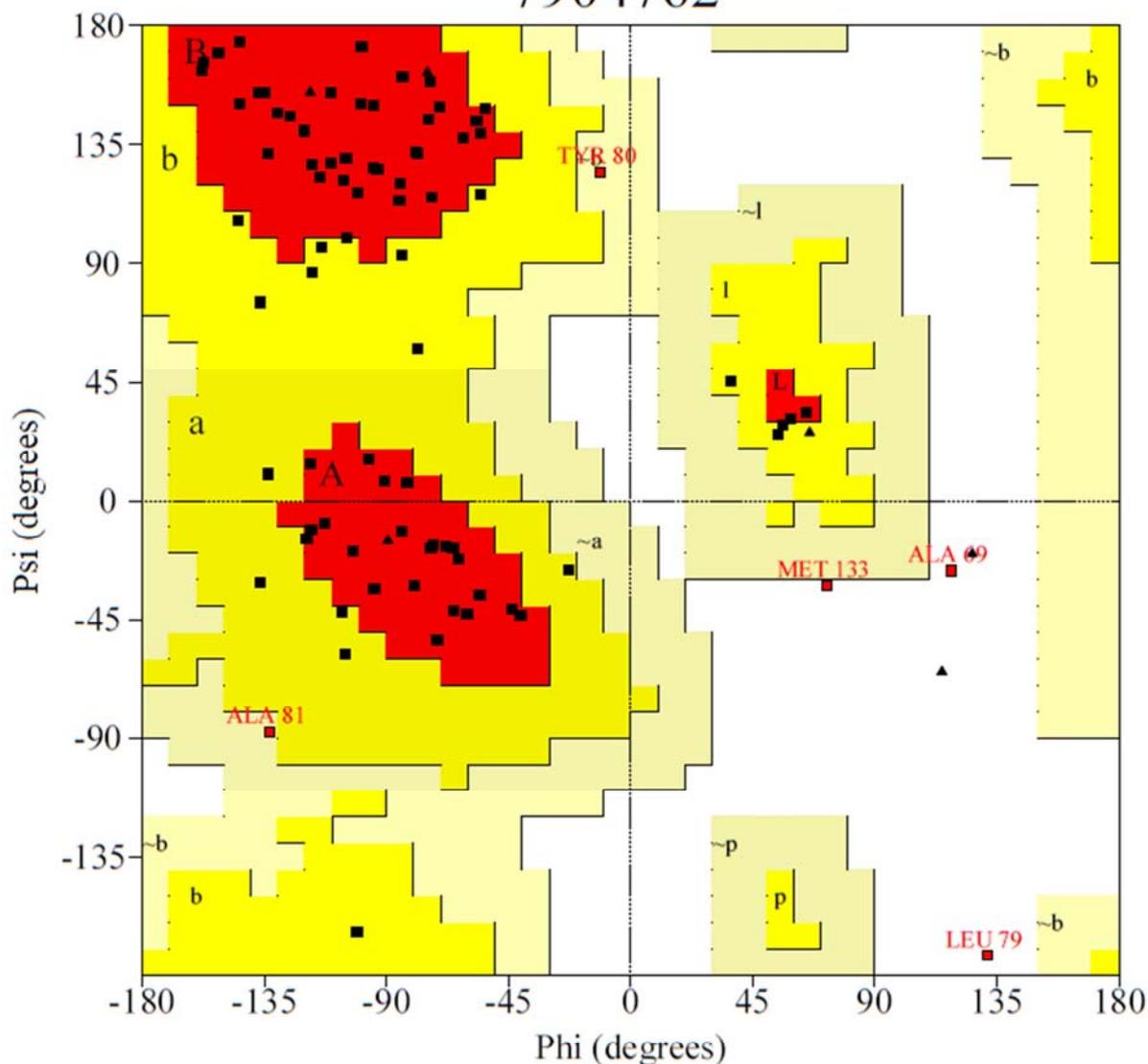


A

PROCHECK

Ramachandran Plot

7904762



Plot statistics

Residues in most favoured regions [A,B,L]	59	72.8%
Residues in additional allowed regions [a,b,l,p]	17	21.0%
Residues in generously allowed regions [~a,~b,~l,~p]	2	2.5%
Residues in disallowed regions	3	3.7%

Number of non-glycine and non-proline residues	81	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	6	
Number of proline residues	2	

Total number of residues	90	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

B

Figure 2. Tertiary structure prediction and validation of vaccine construct. (A) Tertiary structure predicted for the subunit vaccine construct showing helix, sheet and coiled region. (B) Ramachandran plot formation to validate the 3D modelled structure.



Figure 3. Disulfide engineering to improve protein stability. Showing total 4 mutated residue pairs and disulfide bonds in green colour. These residues were selected based on their energy, chi3 value, and B-factor.

3.10. Disulfide Engineering for Vaccine Stability

In order to stabilize the modelled structure of final vaccine constructs disulfide engineering was performed using Disulfide by design v2.0 [25] and found that there are total of 12 pairs of residues that can be used for the purpose of disulfide engineering. But after evaluation on other parameters like energy and Chi3 value, only four pairs of residues were finalized because their value comes under the allowed range i.e. the value of energy should be less than 2.2 and Chi3 should be in between -87 and +97 degree [35]. Therefore, total 8 mutations were created at the residue pairs namely Ala69-Val114, Ala85-Tyr95, Phe113-Tyr116, and Ala153-Gln106 (Figure 3).

3.11. Codon Adaptation and in Silico Cloning

The sequence of peptide vaccine was translated to a DNA

sequence and adapted the codons for *E. coli* K12 strain using JCAT server. It was observed that the GC-content of the improved sequence was 58.34% while the codon adaptive index (CAI) value of the codons was 1.0 which is quite satisfactory [13]. The main purpose of *in silico* cloning was to express the vaccine protein epitope of SARS-CoV-2 spike glycoprotein origin into *E. coli* expression system. Therefore, it was necessary to adapt the codon respective to subunit vaccine construct as per the codon usage of *E. coli* expression system. Thereafter, BstBI and BanII restriction sites were created on the adapted DNA sequence and cloned into the pQE30 vector (Figure 4). The target sequence in the clone is represented in red colour in between aforementioned restriction sites [23]. The target sequence was also tagged with 6-histidine residues on one end that will be helpful in the purification process. The total length of the clone was 5047bp.

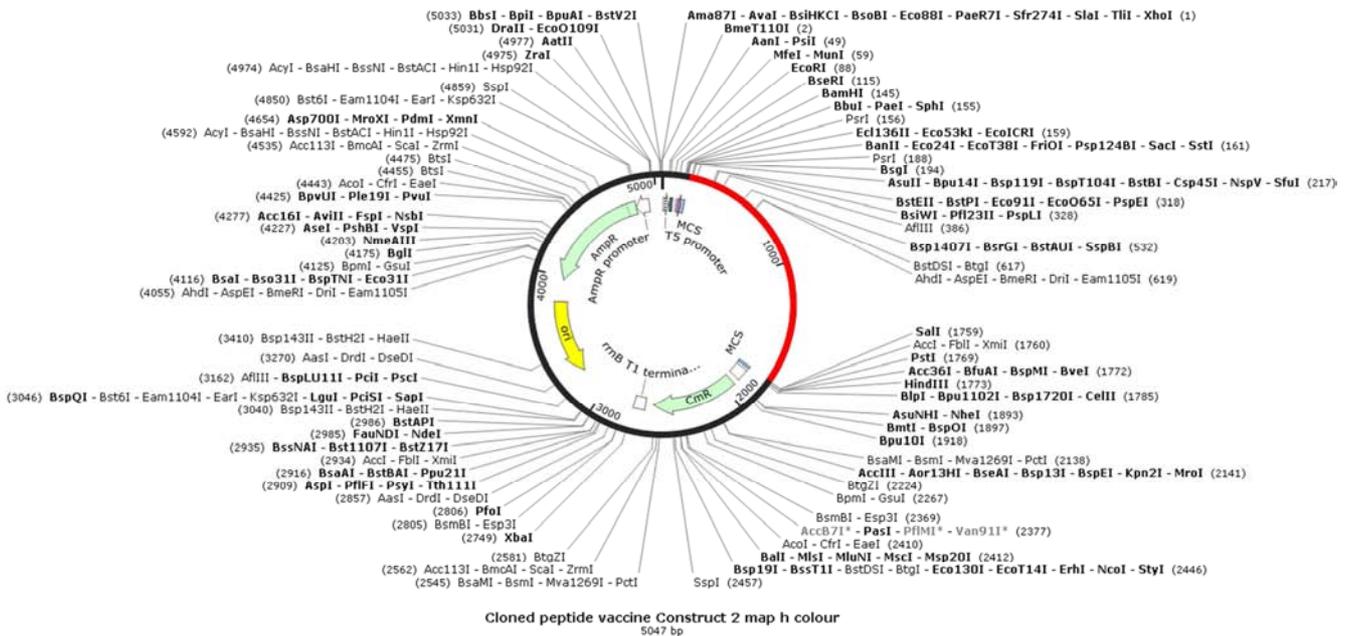


Figure 4. In silico cloning of the adapted vaccine sequence into pQE30 vector showing the region of choice in red colour inserted between BstBI (217) and BanII (1759) while the vector has shown in black lines.

3.12. Molecular Docking of Vaccine Constructs with TLR4

Molecular docking of subunit vaccine protein and TLR-4 receptor was performed using the ClusPro 2.0 and total 30 models were generated [23]. Among them, only that model was selected which properly occupied the receptor and having lowest energy score and found that model number 0.00 fulfil the desired criteria that's why selected as the best-docked complex (Figure 5). The energy score obtained for the model 0.00 was found to be -1252 which is lowest among all other predicted docked complex showing highest binding affinity.

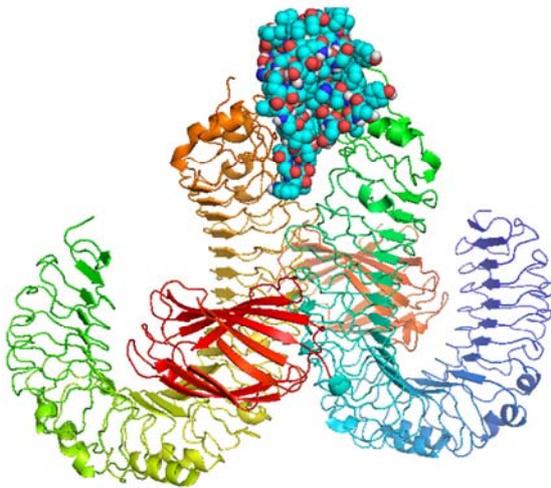


Figure 5. Docked complex of vaccine protein and TLR-4 receptor here vaccine construct is shown in the sphere form, docked on the TLR-4 receptor.

4. Conclusion

Sars-CoV-2 infection also called COVID-19 is characterized by fever and signs of acute respiratory distress syndrome (ARDS). It is currently a global pandemic with high mortality rate in those with severe disease. Lack of effective vaccine and approved drug for treatment created a disastrous condition among the global communities. Therefore, it's the need of time to search for new options to tackle this global pandemic. This study was designed as a step ahead in the path of protein-based vaccine development. The primary amino acid sequence of SARS-CoV-2 spike glycoprotein was used to design a protein subunit vaccine construct. The vaccine construct has CTL, HTL and B cell epitopes of varying lengths having great potential to stimulate high levels of IFN- γ production. It has potent antigenic properties but lacked allergenicity. It is stable and have a good binding affinity for the TLR-4 receptor. In general, this modelling applied a series of immunoinformatics tools in a sequential manner to find an effective vaccine that may be used effectively in fighting against the COVID-19 pandemic. This modelling, however, needs real life experimental validation to prove the workability of the computational work.

References

- [1] WHO (2020). Novel Coronavirus (2019-nCoV) Situation Report 23. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200212-sitrep-23-ncov.pdf?sfvrsn=equals;41e9fb78_4.
- [2] Chan, J. F., Lau, S. K., To, K. K., Cheng, V. C., Woo, P. C., Yuen, K. Y. (2015) Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like dis-ease. *Clin Microbiol Rev* 28 (2): 465–522.
- [3] Elfiky, A. A., Mahdy, S. M., Elshemey, W. M. (2017) Quantitative structure-activity relation-ship and molecular docking revealed a potency of anti-hepatitis C virus drugs against human corona viruses. *J Med Virol*; 89 (6): 1040–1047.
- [4] Chan, J. F. W., Kok, K. H., Zhu, Z., Chu, H., To, K. K. W., Yuan, S. F., Yuen, K. Y. (2020) Genomic characterization of the 2019 novel humanpathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, *Emerging Microbes & Infections*, 9: 1, 221-236, <http://dx.doi.org/10.1080/22221751.2020.1719902>.
- [5] Jiang, S., Du, L. & Shi, Z. (2020) An emerging coronavirus causing pneumonia outbreak in Wuhan, China: calling for developing therapeutic and prophylactic strategies. *Emerg. Microbes Infect.* 9, 275–277.
- [6] Hoffman, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Muller, M. A., Drosten, C., Pohlman, S. (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 1–10.
- [7] Callaway, E. (2020) The race for coronavirus vaccines: a graphical guide. *Nature* 2020; 580: 576-577 doi: 10.1038/d41586-020-01221-y.
- [8] Krammer, F. & Palese, P. (2015). Advances in the development of influenza virus vaccines. *Nat Rev Drug Discov* 14, 167-182.
- [9] Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., *et al.* (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 382, 727-733.
- [10] Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., Fouchier, R. A. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367, 1814–1820.
- [11] Walls, A. C., Park, Y. J., Tortorici, A. M., Wall, A., Andrew T. McGuire, A. T., Veesler, D. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein *Cell* 180, 1–12 <https://doi.org/10.1016/j.cell.2020.02.058>.
- [12] Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020a). Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the ACE2 receptor. *bioRxiv*, 2020.2002.2019.956235.

- [13] Khatoon, N., Pandey, R. K. & Prajapati, V. K. (2017) Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci. Rep.* 7, 8285, <https://doi.org/10.1038/s41598-017-08842-w>.
- [14] Larsen, M. V., Laundegaard, C., Lamberth, L., Buus, S., Lund, O., Nyelsen, M. (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 8, 424, <https://doi.org/10.1186/1471-2105-8-424>.
- [15] Pradhan, D., Yadav, M., Verma, R., Khan, N. S., Jena, L., Jain, A. K. (2017) Discovery of T-cell Driven Subunit Vaccines from Zika Virus Genome: An Immunoinformatics Approach. *Interdiscip. Sci.* 9, 468–477, <https://doi.org/10.1007/s12539-017-0238-3>.
- [16] Black, M., Trent, A., Tirrell, M. & Olive, C. (2010) Advances in the design and delivery of peptide subunit vaccines with a focus on toll-like receptor agonists. *Expert Rev. Vaccines* 9, 157–173, <https://doi.org/10.1586/erv.09.160>.
- [17] EL - Manzalawy, Y., Dobbs, D. & Honavar, V. (2008) Predicting linear B - cell epitopes using string kernels. *J. Mol. Recogn.* 21, 243–255.
- [18] Shi, J. Zhang, J., Li, S., Sun, J., Teng, Y., Wu, M., Li, J., Li, Y., Hu, N., Wang, H., Hu, Y. (2015) Epitope-Based Vaccine Target Screening against Highly Pathogenic MERS-CoV: An In Silico Approach Applied to Emerging Infectious Diseases. *PLoS One* 10, e0144475, <https://doi.org/10.1371/journal.pone.0144475>.
- [19] Magnan, C. N., Zeller, M., Kayala, M. A., Vigil, A., Randall, A., Felgner, P. L., Baldi, P. (2010) High-throughput prediction of protein antigenicity using protein microarray data *Bioinformatics*, 26 (23): 2936–2943, <https://doi.org/10.1093/bioinformatics/btq551>.
- [20] Dimitrov, I., Flower, D. R. & Doytchinova, I. (2013) AllerTOP - a server for in silico prediction of allergens. *BMC Bioinformatics* 14, S4–S4, <https://doi.org/10.1186/1471-2105-14-S6-S4>.
- [21] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., Stemberg, M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10: 845–858. doi: 10.1038/nprot.2015.053.
- [22] Chander, S., Pandey, R. K., Penta, A., Choudhary, B. S., Sharma, M., Malik, R., Prajapati, V. K., Murugesan, S. (2017) Molecular docking and molecular dynamics simulation-based approach to explore the dual inhibitor against HIV-1 reverse transcriptase and Integrase. *Comb. Chem. High Throughput Screen.* 20, 1–13, <https://doi.org/10.2174/1386207320666170615104703>.
- [23] Pandey, R. K., Prajapati, P., Goyal, S., Grover, A. & Prajapati, V. K. (2016) Molecular Modeling and Virtual Screening Approach to Discover Potential Antileishmanial Inhibitors Against Ornithine Decarboxylase. *Comb. Chem. High Throughput Screen.* 19, 813–823. <http://doi:10.2174/1386207319666160907100134>.
- [24] Heo, L., Park, H. & Seok, C. (2013) GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* 41, W384–388.
- [25] Craig, D. B. & Dombkowski, A. A. (2013) Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* 14, 346, <https://doi.org/10.1186/1471-2105-14-346>.
- [26] Ali, M., Pandey, R. K., Khatoon, N., Narula, A. & Mishra, A. (2017) Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci. Rep.* 7, 9232, <https://doi.org/10.1038/s41598-017-09199-w>.
- [27] Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., Vajda, S. (2017) The ClusPro web server for protein-protein docking. *Nature Protocols.* 2017 Feb; 12 (2): 255-278.
- [28] Calis, J. J. A., Maybeno, M., Greenbaum, J. A., Weiskopf, D., DeSilva, A. D., Ste, A., Kesmir, C., Peters, B. (2013) Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput. Biol.* 9, e1003266, <https://doi.org/10.1371/journal.pcbi.1003266>.
- [29] Jordan, K. A. & Hunter, C. A. (2010) Regulation of CD8 (+) T Cell Responses to Infection With Parasitic Protozoa. *Exp. Parasitol.* 126, 318–325.
- [30] Moseman, E. A. & McGavern, D. B. (2013) The great balancing act: regulation and fate of antiviral T-cell interactions. *Immunol. Rev.* 255, 110–124.
- [31] Nezafat, N., Ghasemi, Y., Javadi, G., Khoshnoud, M. J. & Omidinia, E. (2014) A novel multi-epitope peptide vaccine against cancer: an *in silico* approach. *J. Theor. Biol.* 349, 121–134.
- [32] Li, M., Jiang, Y., Gong, T., Zhang, Z. & Sun, X. (2016) Intranasal Vaccination against HIV-1 with Adenoviral Vector-Based Nanocomplex Using Synthetic TLR-4 Agonist Peptide as Adjuvant. *Mol. Pharm.* 13, 885–894.
- [33] Chan, J., Mehta, S., Bharrhan, S., Chen, Y., Achka, J. M., Casadewall, A., Flynn, J. (2014) The role of B cells and humoral immunity in Mycobacterium tuberculosis infection. *Semin. Immunol.* 26, 588–600.
- [34] Busby, J. N., Panjekar, S., Landsberg, M. J., Hurs, M. R., Lott, J. S. (2013) The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. *Nature.* 501: 547-550.
- [35] Rana, A. & Akhter, Y. (2016) A multi-subunit based, thermodynamically stable model vaccine using combined immunoinformatics and protein structure-based approach. *Immunobiology* 221, 544–557, <https://doi.org/10.1016/j.imbio.2015.12.004>.