Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Computers in Biology and Medicine 140 (2022) 105122

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Structure-guided design of multi-epitopes vaccine against variants of concern (VOCs) of SARS-CoV-2 and validation through In silico cloning and immune simulations

Fahad Humayun [a], Yutong Cai [b], Abbas Khan [a], Syed Ali Farhan [c], Fatima Khan [d], Usman Ishrat Rana [e], Anum binte Qamar [f], Nasim Fawad [g], Shazia Shamas [h], Dongqing-Wei [a,i,*]

[a] *State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint International Research Laboratory of Metabolic & Developmental Sciences and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200030, PR China*
[b] *Shenzhen College of International Education 3, Antuoshan 6 Road, Futian, Shenzhen, China*
[c] *Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan*
[d] *National Institute of Health, Islamabad, Pakistan*
[e] *Department of Radiology, General Hospital, Lahore, Pakistan*
[f] *Department of Biosciences and Bioinformatics, COMSATS University, Islamabad, Pakistan*
[g] *Poultry Research Institute, Rawalpindi, Pakistan*
[h] *Department of Zoology, University of Gujrat, Gujrat, Pakistan*
[i] *Peng Cheng National Laboratory, Vanke Cloud City Phase I Building 8, Xili Street, Nashan District, Shenzhen, Guangdong, 518055, PR China*

A R T I C L E I N F O

Keywords:
SARS-CoV-2
Multi-epitope subunit vaccine
Structural vaccinology
Immune simulation
In silico cloning

A B S T R A C T

Severe Acute Respiratory Syndrome Corovirus2 (SARS-CoV-2) has been determined to be the cause of the current pandemic. Typical symptoms of patient having COVID-19 are fever, runny nose, cough (dry or not) and dyspnea. Several vaccines are available in markets that are tackling current pandemic. Many different strains of SAR-CoV-2 have been evolved with the passage of time. The emergence of VOCs particularly the B.1.351 ("South African") variant of SARS-CoV-2 has been reported to be more resistant than other SARS-CoV-2 strains to the current vaccines. Thus, the current research is focused to design multi-epitope subunit Vaccine (MEV) using structural vaccinology techniques. As a result, the designed MEV exhibit antigenic properties and possess therapeutic features that can trigger an immunological response against COVID-19. Furthermore, validation of the MEV using immune simulation and in silico cloning revealed that the proposed vaccine candidate effectively triggered the immune response. Conclusively, the developed MEV needs further wet lab exploration and could be a viable vaccine to manage and prevent COVID-19.

1. Introduction

Coronaviridae, Roniviridae and Arteviridae families are the largest of RNA viruses, and Coronaviruses (CoVs) also belongs to these families. The structure of coronaviridae consists of 5′capped and unsegmented 3′ polyadenylated positive sense single-stranded RNA. It is responsible for causing infections that lead to human respiratory diseases [1,2]. COVs are categorized into four different classes as alpha, beta, delta, and gamma. Among all of the classes, beta and alpha cause infection in humans [3]. It has been considered that CoV strain is one of the most interesting strain for researchers since it has been recognized to cause coronavirus disease (COVID-19) [4] due to global pandemic in the year of 2019. Severe Acute Respiratory Syndrome Corovirus2 (SARS-CoV-2) has been determined to be the cause of pandemic [5] and genomic study revealed that SARS-CoV-2 and SARS virus responsible for disease outbreak in 2002–2003 are closely related to each other [6]. Diagnostic

* Corresponding author. State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint International Research Laboratory of Metabolic & Developmental Sciences and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200030, PR China.

E-mail addresses: fahadhumayun@outlook.com (F. Humayun), s19099.cai@stu.scie.com.cn (Y. Cai), abbaskhan@sjtu.edu.cn (A. Khan), jcsccfana@yahoo.com (S.A. Farhan), fatimanasrullahkhan@gmail.com (F. Khan), usman.ishrat@hotmail.com (U.I. Rana), hussain.anum94@gmail.com (A. Qamar), nasimfawad58@gmail.com (N. Fawad), shaziashams@yahoo.com (S. Shamas), dqwei@sjtu.edu.cn (Dongqing-Wei).

https://doi.org/10.1016/j.compbiomed.2021.105122
Received 25 September 2021; Received in revised form 30 November 2021; Accepted 3 December 2021
Available online 7 December 2021
0010-4825/© 2021 Elsevier Ltd. All rights reserved.

process initially determined primary spread of SARS-CoV-2 with respiratory droplets due to sneezing and coughing which later extended to the contact of body and somehow with fecal contact [7]. Based on clinical evidence, the time required for the symptoms of SARS-CoV-2 to present is 14 days after getting exposure, but some of the cases it may takes more than 14 days. Typical symptoms of patient having COVID-19 are fever, runny nose, cough (dry or not) and dyspnea [8].

The 33% of the genetic code of RNA based SARS CoV2 is comprised of structural or accessory proteins and 66% is made up of non-structural or essential proteins. The outer surface is covered by structural proteins which make up the system that initiates the infection and inactivates host defense mechanism. The study [9] described that SARS CoV-2 mutation proneness reflects genomic organization that exhibits complexity in developing a vaccine that is composed of 5′-leader-UTR-replicase-ORFab-S (Spike)-E (Envelope)-M(Membrane)-ORF6-ORF7a-ORF8-N(Nucleocapsid)-3′UTR-polyAtail [10,11] which included the accessory proteins as ORF3a, ORF8 and ORF7A for the pathogenesis of virus.

In treatment plan many medications have been subjected including hydroxychloroquine and dexamethasone for the clinical trials while subunit vaccines [12], nano-particle based vaccines along with viral vector vaccines are developed to deal with the pandemic. Some of the other vaccines as fusion-protein, recombinant proteins, DNA and live-attenuated vaccines are also considered for the trials and are available in market [13]. To deal with cost and time barriers, along with traditional approaches computational methods have also been used in which the virus is targeted for efficient results, [14]. In a recent study, it is reported that computational techniques are much more efficient against effects of MERS virus, Ebola virus chikungunya, and Zika [15]. Another approach known as reverse vaccinology is in practice that involves algorithms to evaluate the data based on immunological constraints including allergenicity, toxicity and epitopes binding efficiency which is more consistent in human leukocyte antigen (HLA) allelesto design the potential multiepitope subunit vaccine [16]. The consequent multiepitope vaccine is composed of multiple epitopes group that is integrated with respective linkers which increases the particular adaptive-immune response in the host cell [17].

In establishing the concept of novelty, studies reported that strains of SARS-CoV-2 in England, South Africa and some other countries made up the second wave of virus at the end of year 2020 which aggravated the lethality of the pandemic. Novel strains may inactivate immune system of host cell due to mutations thus vaccines that have been previously designed become useless in clinical trials [18]. The B.1.351 ("South African") variant of SARS-CoV-2 is significantly more resistant to neutralization by convalescent plasma (9.4 fold) and vaccine sera (10.3–12.4 fold) than other SARS-CoV-2 strains [19]. This raises the possibility of that particular variation to be resistant to the most widely used spike-based vaccinations [20]. Earlier lab tests revealed that Moderna's vaccine resulted in "a six-fold reduction in neutralizing titers with the B.1.351 variant compared to prior variants [21]. Pfizer-"neutralization BioNTech's of the B.1.351-spike virus" was around two-thirds less effective [22]. When evaluated in human volunteers, AstraZeneca's vaccine gave just 25% protection from mild to moderate illness caused by that variation, and its distribution in South Africa was halted [23]. In South Africa, Janssen's vaccine performed worse than it did elsewhere, owing to poor neutralization of infecting viruses [24].

To compete with this challenge, in this study an immunoinformatic approach is established which targets the spike protein of novel strain that is further investigated for development of immunogenic epitope (CTL, HTL and B cell). It has been considered that vaccine using spike protein for SARS-CoV-2 is more related to this approach. This approach has been recognized widely in vaccine design against number of pathogens as well as SARS-CoV-2 [25]. Another approach in vaccine design includes the role of molecular docking that constructs vaccine interactions by TLR4 receptor and determines complex stability by using molecular dynamics (MD) simulations. The previously discussed approaches are usually used to design vaccine against COVID-19 but the ultimate goal is to fight against the novel strain of SARS-CoV-2.

2. Material and methods

2.1. Protein sequence collection

It has been evident that new variants of SARS-CoV-2 have emerged in South Africa and some states/countries of the world that critically requires serious attention of researchers and medical practitioners. These variants have number of mutations as compared to the strain presented in Wuhan, China in the ending period of year 2019. To get structural evaluation a more targeted approach has been considered that calls for the antigenic vaccine to be utilized for different protein sequences that ultimately interact to fight against B.1.351 ("South African"). NCBI (https://www.ncbi.nlm.nih.gov/) was used to retrieve the dataset required for this approach. This dataset contains ORF3a (QWA53388), ORF6 protein (QWA53391), nucleocapsid phosphoprotein (QWA53395), ORF10 protein (QWA53396), ORF7a protein (QWA53392), envelop protein (QWA53389), ORFlab (QWA53385), ORF8 protein (QWA53394), membrane glycoprotein (QWA53390) and surface glycoprotein (QWA53387). The protein sequences were used in FASTA format as input for the immune-informatics analysis.

2.2. Investigation of trans-membrane helicityand antigenicity of protein

Protein antigenicity and transmembrane helicity played a key role in the development of a multi-epitope vaccine. Vaxijen server [26] with 4.0 threshold value was used for theevaluation of protein antigenicity while TMHMM [27] and Protter server [28] predicted the *trans*-membrane helicity.

2.3. Prediction of cytotoxic T – cell epitope

NetCTL pan version 1.1 shortlisted and evaluated seven extracellular protein peptides for the prediction of promiscuous epitopes that may increase immune response of host cell with interacting HLA-epitope binding. NetCTL pan evaluation used for all MHC (HLA) class I molecules super type representatives, allows prediction of 8, 9, 10 and 11 mer peptides, to lower the cost of experiment. Machine learning assistance including neural networks can predict high binding affinity of nonameric peptides more accurately [29]. Vaxijen, AllergenFP [30], Toxinpred server (http://crdd.osdd.net/raghava/toxinpred/) were used to confirm the antigenicity, allergenicity, and toxicity levels of the epitopes, respectively.

2.4. Immunogenecity prediction

Immunogenicity prediction (IEDB) refers to the ability of a viral infection to generate humoral and cellular response in the host cell. Sequence of NetCTLpan epitopes selected were in the form of fasta format to input peptides for predicting immunogenicity. For considering the positive immunogenicity epitopes and capacity of high binding epitopes IEDB immonogenecity tool server was selected. This server has the capacity to perform predictions at immunogenicity level and rely on epitope positions with expected peptide with respective amino acid physiochemical properties [29].

2.5. Prediction of helper T-cell epitope

In vaccine development, it is a necessary aspect that it must cover major of the world population. In concern of HLA alleles, the overcoming factors include panel of 27 alleles in reference to establish a population which is $>$99% [31]. The IEDB MHC II server is comprised upon combinatorial libraries that allow percentile rank to show the IC-50, in which high affinity of epitope-HLA complex is depicted by a lower percentile rank. Amino acids list was generated with resultant

epitopes that contain 15aa long epitopes [32]. These epitopes further considered for evaluation by Vaxigen, AllergenFP and TOXINPRED server that show analysis for epitope antigenicity, allergenicity and the toxicity reaction for host cell. IFN gamma induction is marked as important in the analysis that eliminate virus in host immune system response and efficiency of IFN gamma induction targeted in non-cross-reactive epitopes was predicted by IFNepitope server (http://crdd.osdd.net/raghava/ifnepitope/).

2.6. B-cell epitope prediction

B-cell epitopes uses an antibody-based immune response to detect viral infections. The B-cell interacting epitopes were analyzed using ABCpred (http://crdd.osdd.net/raghava/abcpred/). As an input file, FASTA files of all seven proteins were employed. These epitopes were then tested in the host cell using Vaxijen, AllergenFP, and the TOXINPRED service to determine antigenicity, allergenicity and toxicity.

2.7. Multi-epitope vaccine design

In general consideration it has been observed that these characteristics are preferred to design vaccine subunit as highly antigenic, immunogenic, non-allergic, non-toxic. As a result, only those epitopes were selected for future investigation that made MEV with the above-mentioned characteristics included. The adjuvant for the EAAAK linker focuses on the first cytotoxic T-lymphocytes (CTL) epitope and improves the immunological response. The epitopes connected with AAY, GPGPG and KK linkers are being tested for interaction and compatibility in terms of maintaining separate immunogenic activity. These linkers are essential for achieving prolonged conformation (flexibility), protein folding, and functional domain separation, all of which contribute to a more stable protein structure [33]. The AAY linker was used to connect CTL epitopes, the GPGPG linker was used to connect HTL epitopes, and the KK linker was used to connect B cells [34]. In this work, β-defensin was used as an adjuvant, and it is a simple 45 amino acid long peptide that acts as both an immunomodulator and an antibacterial agent [35].

2.8. Vaccine construct analysis

The construction of binding affinity for vaccine design by B and T-cell, post-injection toxicity and IgE antibody-associated allergic reaction with reference to host cell are common challenges that arise in vaccine development. To overcome these issues, investigations were performed using antigenicity by Vaxigen, AlgPred for allergenicity and for toxicity Toxinpred. For the post-injection symptoms in vaccine design, physiochemical properties analysis plays a major role. To recognize the higher antigenic vaccine, lower allergic MEV was evaluated by using Protparam server [36]. The half-life, stability index, theoretical isoelectric point (pI), molecular weight (KDa), extinction coefficient, grand average of hydropathicity (GRAVY) and aliphatic index are included. For secondary structure prediction along with tertiary structure, MEV PESIPRED [37] and Phyre2 modeling mode [38] were used respectively. RaptorX (http://raptorx.uchicago.edu) was used to produce cross-validation of model structure in ab-initio vaccine modelling. The vaccine model was refined using GalaxyRefine server and validated through highly recognized server ProSa [39] and QMEAN [40].

2.9. Molecular docking

The interaction between antigenic molecule and immune receptor is an essential element in vaccine development and in order to find these interactions Molecular Docking was performed for the analysis with human receptors and MEV. Human Receptor TLR3 (Toll-like Receptors-3) has been extensively researched, and investigations have revealed that it plays a critical role in the production of antiviral immune responses. Hawkdock (http://cadd.zju.edu.cn/hawkdock/) was used for the docking of MEV with TLR3 [41]. Furthermore, the online server PDBsum (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html) and Chimera [42] were used to create a traditional sketch of interactions among docked proteins.

2.10. MD simulation

The Molecular Dynamics Simulation (MDS) analysis was targeted by using Amber20 [43] with TIP3P solvent along with FF19SB force field. MDS is launched for 100ns to duplicate the methods that were published [44]. In conclusion, a system model was built that selected TIP3P water as solvent and the system charge was neutralized by adding twenty-nine sodium ions. This system was constructed to be close to natural system. The system is relaxed before simulation and MDS is applied for 100 ns with trajectories record to analyze RMSD and RMSF.

2.11. Immune simulations

C-ImmSim is an online server (http://kraken.iac.rm.cnr.it/C-IMMSIM/) that was used for multi epitope vaccine immune simulation profile and it comprises of Celada-Seiden model which is used to predict humoral and immune cellular stimulation for targeted vaccine development as reverse (against) factor [45]. The standard parameters were used for simulation including simulation steps 672, random seed 12,345, vaccine injection without LPS and simulation volume 10. The selected parameters were applied for vaccine injection in selecting B5802, MHCs A0101, DRB3_0101, B4403, DRB1_0101) and 1000 antigens. The MEV injection was induced for one first day for 4 weeks and 16 weeks that equals 1,84,336-time steps respectively. To evaluate the virus' clearance capacity, an 8-h gap was observed between two-time steps. With an active immune system, similar antigen was incorporated, and the elimination of virus was observed. This rotation was based on injected virus for 24 weeks as 504 steps while the multiplication factor was 0.2, infectivity 0.6 and number of viruses turned to 1000. Different plots were used to analyze the immune responses after the injection of MEV.

2.12. In silico cloning and codon adaptation

Reverse translation was applied on the amino acid sequence of MEV, and codon optimization is also necessary for the expression of cloning of MEV genes. Java Codon Adaptation tool (http://www.jcat.de/) was used to perform the optimization of code. Prokaryotic organism such as *E. coli* K12 strain was used to express the enhancement of expression as efficiency with MEV. Bacterial ribosome binding sites, rho-independent transcription terminators, and certain cleavage sites of restriction enzyme were excluded. By using snap gene tool (https:/snapgene.com/), significant effects have been covered in restriction endonuclease *Hin*dIII and *Bam*HI included at N and C terminals for MEV gene respectively.

3. Results and discussion

3.1. Result

3.1.1. Antigenicity and transmembrane helicity of proteins analysis

To determine antigenicity and *trans*-membrane helicity, amino acid sequences in FASTA format of target protein were collected. All proteins had antigenicity scores that indicated their ability to outperform the immune response in the host cell. The viral protein transmembrane helicity was tested to trigger an immunological response from the host cell. We identified extracellular, cytosolic peptides and transmembrane using the TMHMM and Protter servers. According to the protein Protter findings, ORF6, ORF10 and ORF8 protein along with nucleoprotein phosphoprotein are all fully cytosolic. As a result, these three proteins were removed from future research, and the remaining seven proteins

were evaluated to be included in the chimeric SARS CoV2 vaccine. These extracellular peptides of these proteins might aid in interacting with host PAMPs and increasing the proposed vaccinations' solubility.

3.1.2. Prediction of cytotoxic T-cell epitopes, including immunogenicity, allergenicity, toxicity and antigenicity

NetCTLpan 1.1 was used to predict Cytotoxic T-lymphocyte epitopes for all seven proteins using the identical HLA supertypes. Peptide binders were found in all proteins with various HLA supertypes. We discovered various epitope binders from the server findings, including membrane glycoprotein, envelope protein, ORF7a protein, ORF3a protein, ORF1ab polyprotein and surface glycoprotein. Different epitopes of ORF3a protein (2 epitopes), ORF1ab polyprotein (50 epitopes), membrane glycoprotein (4 epitopes) and surface glycoprotein (17 epitopes) were further shortlisted based on toxicity, immunogenicity, allergenicity and antigenicity. After analysis of peptide, the epitopes that potentially boost immune response by activating HLA were chosen.

3.1.3. IFN gamma induction analysis and prediction of helper T-cell epitopes

The HLA allele reference set was used to construct 15 amino acid extended epitopes mediated by helper T cells. Consequently, 0.1 percentile rank was selected to be the threshold for filtering binding affinity of HLA interacting epitopes. Antigenicity, allergenicity, and toxicity levels in the low percentile epitopes were used to shortlist the epitopes. ORF1ab polyprotein (3 epitopes), membrane glycoprotein (7 epitopes), ORF3a (7 epitopes) and surface glycoprotein (8 epitopes) were found as HLA-II binders with low allergenicity and toxicity along with high antigenicity booster. According to the combined findings of all investigations, IFN gamma appears to have a significant role in the immunological response of human cells to viruses. Virus clearance from human cells is aided by high IFN gamma expression. The epitope's ability to trigger IFN secretion via T-cells (TC) was determined by the IFNepitope server's positive score which is given in Table 1.

3.1.4. B cell epitope prediction

The ABCpred server was used for B cell epitope prediction. Various epitopes of selected proteins including membrane glycoprotein, ORF3a protein, surface glycoproteinand ORF1ab protein were examined which were then chosen for further antigenicity, allergenicity and toxicity investigation. Membrane glycoprotein (8 epitopes), and ORF1ab protein (15 epitopes), ORF3a protein (6 epitopes) and surface glycoprotein (12 epitopes) were shortlisted for further investigation.

3.1.5. Investigation of physicochemical characteristics of vaccine design MEV

Top score CTL, HTL and B cell epitopes from each protein were selected to construct vaccine MEV. All selected epitopes are shown in Table 2, Table 3, and Table 4. MEV vaccine was shown to be antigenic and free of allergenicity and toxicity. The highly antigenic vaccine design MEV was found to have 30 kDa molecular weight, 283and 34.03 instability index, all of which indicates the protein's stability. The aliphatic and GRAVY index values in the MEV construct are 77.67 and −0.091, respectively. The Ramachandran plot revealed 96% residue in the favored region, which was used to model, improve, and confirm the MEV tertiary structure. The model's Z-score is −6.15, and its QMEAN is −2.19. RaptorX was used to create an ab initio model to validate the Phyre 2 model. Both are similar models are similar, and the Ramachandran plot of RaptoX model revealed 99% residue in the favored region. This serves as proof that the phyre2 model is accurate and can be utilized for further research. MESV was compared to the proteome of *Homo sapiens*, and the results indicated that it is unrelated to any human protein.

The AAY linker was used to connect CTL epitopes, the GPGPG linker was used to connect HTL epitopes, and the KK linker was used to connect B cell. β-defensin was used as an adjuvant, and it is a simple 45 amino acid long peptide that acts as both an immunomodulator and an antibacterial agent. The amino acid sequence of MEV is shown in Fig. 1.

3.1.6. MEV and TLR3 molecular docking

Contact between the immune receptor and antigenic molecule is necessary to initiate the immune response. Molecular docking of a representative immune receptor TLR3 with the proposed MEV was

Table 1
Helper T cell epitopes selected for vaccine design using IEDB MHC II.

Epitope sequence	Method	Result	Score
TLSYYKLGASQRVAG	SVM	POSITIVE	0.16141022
RTLSYYKLGASQRVA	SVM	POSITIVE	0.3791622
ITVATSRTLSYYKLG	SVM	POSITIVE	0.15373236
NMLRIMASLVLARKH	SVM	POSITIVE	0.13368254
PNMLRIMASLVLARK	SVM	POSITIVE	0.39136883
SFLGRYMSALNHTKK	SVM	POSITIVE	0.40123996
KKRWQLALSKGVHFV	SVM	POSITIVE	0.30805221
KRWQLALSKGVHFVC	SVM	POSITIVE	0.26929309
LKKRWQLALSKGVHF	SVM	POSITIVE	0.091300556
ITLKKRWQLALSKGV	SVM	POSITIVE	0.26323932
GINITRFQTLLALHR	SVM	POSITIVE	0.10772613
TRFASVYAWNRKRIS	SVM	POSITIVE	0.73155567
RFASVYAWNRKRISN	SVM	POSITIVE	0.53166283

Table 2
CTL epitopes short listed for vaccine design using NetCTLpan 1.1. Vaxijen, AllergenFP, Toxinpred server were further used to confirm the antigenicity, allergenicity, and toxicity levels of the shortlisted epitopes, respectively.

Protein Name	Epitope Sequence	Length	Score
Membrane glycoprotein	AGDSGFAAY	9	0.03981
Membrane glycoprotein	LVGLMWLSY	9	−0.06867
Orf1ab	VSDIDITFL	9	0.38916
Orf1ab	FLFVAAIFY	9	0.33519
ORF3a	STDTGVEHV	9	0.22152
ORF3a	TSPISEHDY	9	0.10373
Spike glycoprotein	QLTPTWRVY	9	0.31555
Spike glycoprotein	VLPFNDGVY	9	0.1815

Table 3
HTL epitopes short listed for vaccine design using IEDB MHC II. These epitopes further considered for evaluation by Vaxigen, AllergenFP and TOXINPRED server that show analysis for epitope antigenicity, allergenicity and the toxicity reaction.

Protein Name	Epitope Sequence	Length	Score
Membrane glycoprotein	TLSYYKLGASQRVAG	15	0.16141022
Membrane glycoprotein	ITVATSRTLSYYKLG	15	0.15373236
ORF1ab	NMLRIMASLVLARKH	15	0.13368254
ORF1ab	SFLGRYMSALNHTKK	15	0.40123996
ORF3a	KKRWQLALSKGVHFV	15	0.30805221
Spike glycoprotein	GINITRFQTLLALHR	15	0.10772613
Spike glycoprotein	TRFASVYAWNRKRIS	15	0.73155567

Table 4
B cell epitopes shortlisted for vaccine design using ABCpred. Vaxijen, AllergenFP, and the TOXINPRED service to determine antigenicity, allergenicity and toxicity.

Protein Name	Epitope Sequence	Length	Score
Membrane glycoprotein	VAGDSGFAAYSRYRIGNYKL	20	0.73
Membrane glycoprotein	LWPVTLACFVLAAVYRINWI	20	0.72
ORF1ab	LEILQKEKVNINIVGDFKLN	20	0.96
ORF1ab	DLVPNQPYPNASFDNFKFVC	20	0.95
ORF3a	KQGEIKDATPSDFVRATATI	20	0.84
ORF3a	LKKRWQLALSKGVHFVCNLL	20	0.79
Spike glycoprotein	ADSFVIRGDEVRQIAPGQTG	20	0.93
Spike glycoprotein	TKLNDLCFTNVYADSFVIRG	20	0.90

GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKKAAAKVSDIDITFLAAYFLFVAAIFYAAYSTDTGVEH
VAAYTSPISEHDYAAYAGDSGFAAYAAYLVGLMWLSYAAYQLTPTWRVYAAYVLPFNDGVYGPGPGNMLRIMASLVL
ARKHGPGPGSFLGRYMSALNHTKKGPGPGKKRWQLALSKGVHFVGPGPGTLSYYKLGASQRVAGGPGPGITVATSRT
LSYYKLGGPGPGGINITRFQTLLALHRGPGPGTRFASVYAWNRKRISKKNVPLHGTILTRPLLESELVIKKLWPVTLACFVL
AAVYRINWIKKLEILQKEKVNINIVGDFKLNKKDLVPNQPYPNASFDNFKFVCKKKQGEIKDATPSDFVRATATIKKLKKR
WQLALSKGVHFVCNLL KKADSFVIRGDEVRQIAPGQTG KKTKLNDLCFTNVYADSFVIRG

Fig. 1. Amino acid sequence of MEV. Ajduvant is shown in Green color, CTL epitopes are shown in Golden color, B cell epitopes are shown in Red color, HTL epitopes are shown in Blue color while linkers are represented in Black color. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

performed to determine the binding capability of MEV to innate immune receptors. The complex having best net global energy of 363.18 kJ/mol, was predicted by the docking evaluation. According to a visual examination of the complex, the MEV binds deeply in the core of TLR3 and promotes weak van Dar Waals interactions along with rigorous hydrogen bond with TLR3 residues (see Fig. 2). PDBsum and Chimera were used to have an insight of the docking results.

3.1.7. Molecular dynamic simulation analysis

The complex was subjected to MDS for 100 ns, and the results were evaluated for RMSD, RMSF, and other factors. According to RMSD calculations, the MEV-TLR3 complex is stable during simulation with an RMSD of about 3 Å while less stable before the binding of MEV to TLR3. Similarly, according to the RMSF study with an RMSF of about 3 Å, the MEV-TLR3 complex exhibits less volatility and compared to the RMSF study before the binding of MEV to TLR3. According to the MDS results, the MEV multiepitope vaccine design had a stronger and more consistent interaction with the TLR3 complex (Fig. 4). These findings proposed that this vaccine construct elicited an immunological response which is capable of clearing SARS CoV2 (see Fig. 5).

3.1.8. Immunogenicity evaluation of MEV

Every primary and secondary immune responses contribute to the infection in some way and possibly consistent through real immune response. The response of in-silico host immune system response to the antigen is depicted in Fig. 3. High IgG + IgG and IgM concentrations

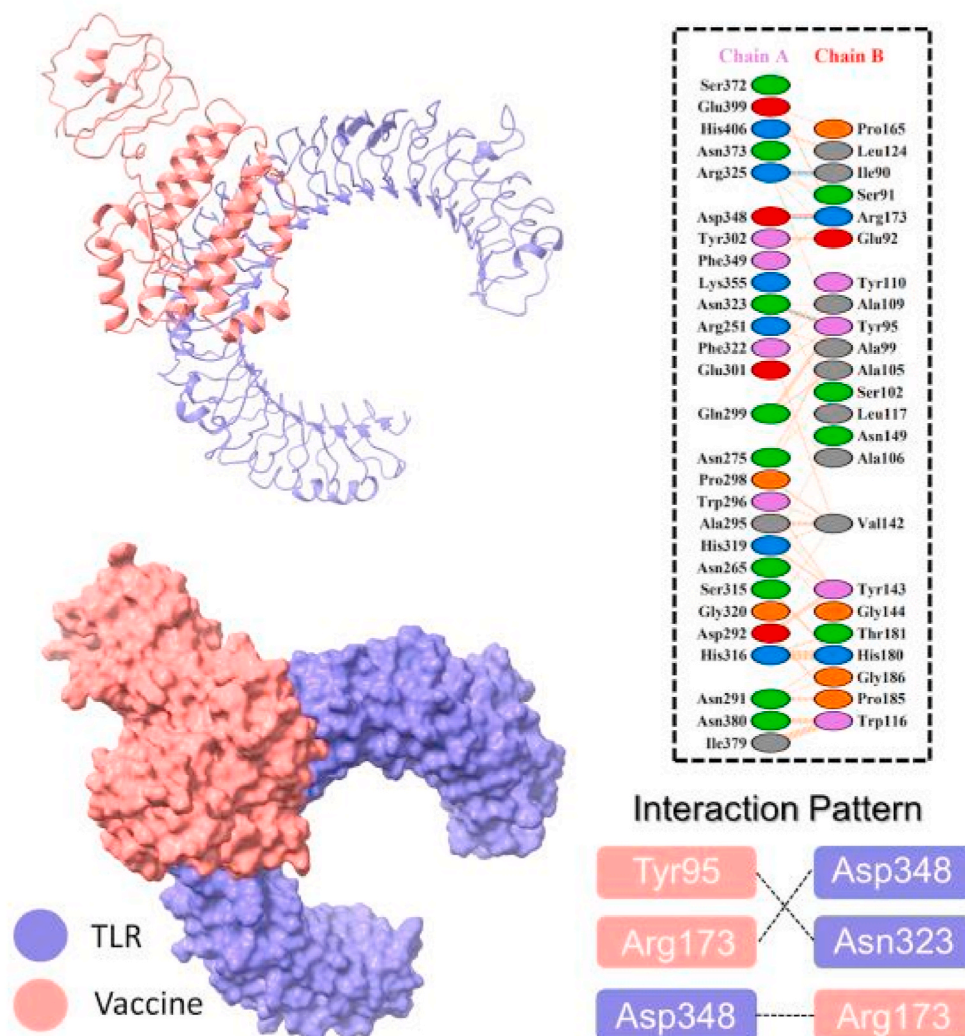

Fig. 2. Dock pose of MEV and TLR3 complex and the 2D interaction of TLR3 and MEV.

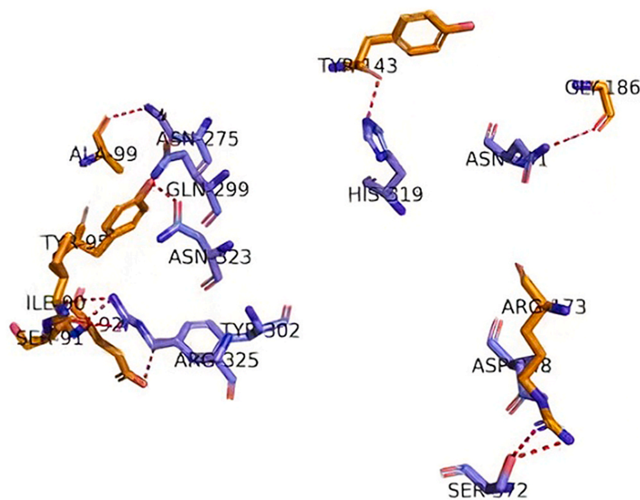

Fig. 3. 3D interaction pattern of TLR3 and MEV.

defined the first reaction, which was followed by IgG1, IgG1 + IgG2, and IgM at both the primary and secondary phases, with decrease in antigen. All these results points to the MEV's immunological response and eventual elimination following successive contacts.

3.1.9. *In silico cloning within E. coli system*

To ensure the expression of MEV generated from SARS-CoV-2 in commonly utilized *E. coli* hosts, in silico cloning was performed. First, the MEV codons were altered to accommodate the use of expression system codons of *E. coli* through K12 strain. The optimized MEV construct has 849 nucleotides, a CAI of 0.87, and a GC content range of 53.7%, indicating that it has a high potential for reproducibility and good expression of protein. To aid the purification/cloning process, the buffer compatible restriction enzymes *Hin*dIII and *Bam*HI restriction sites were connected to both ends of the MEV optimized nucleotide sequence. Finally, between the restriction sites, the revised MEV sequence was cloned to the various cloning sites of the = pET30a (+) vector. The clone had a length of 6196 base pairs. (Fig. 6).

4. Discussion

Recent advances in the field of immune-informatics have led to development of a number of tools and servers that can help minimize

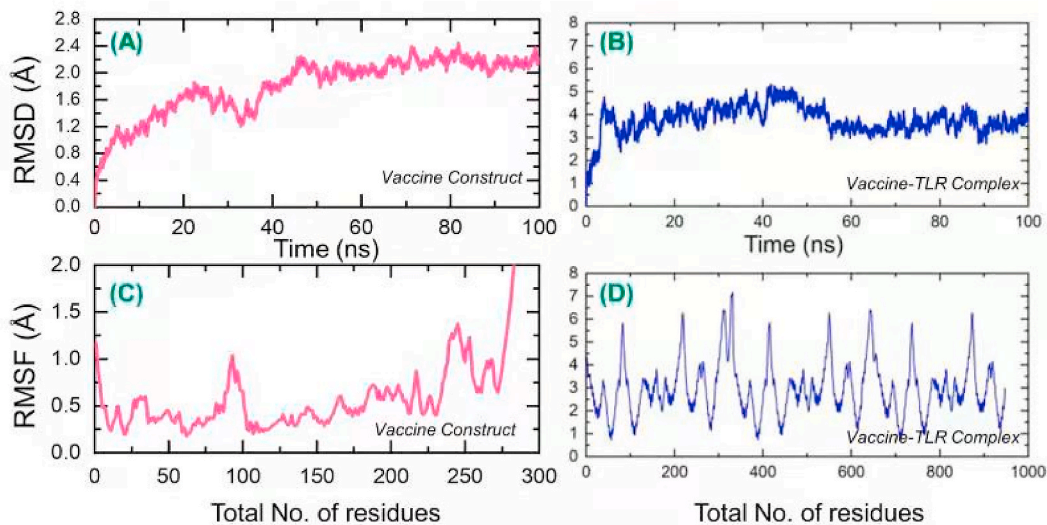

Fig. 4. Molecular dynamics simulation analysis. (A) Root-mean-square deviation before binding of MEV and TLR3, (B) Root-mean-square deviation after binding of MEV and TLR3, (C) Root mean square fluctuations before MEV binding to TLR3, (D) Root mean square fluctuations after MEV binding to TLR3.

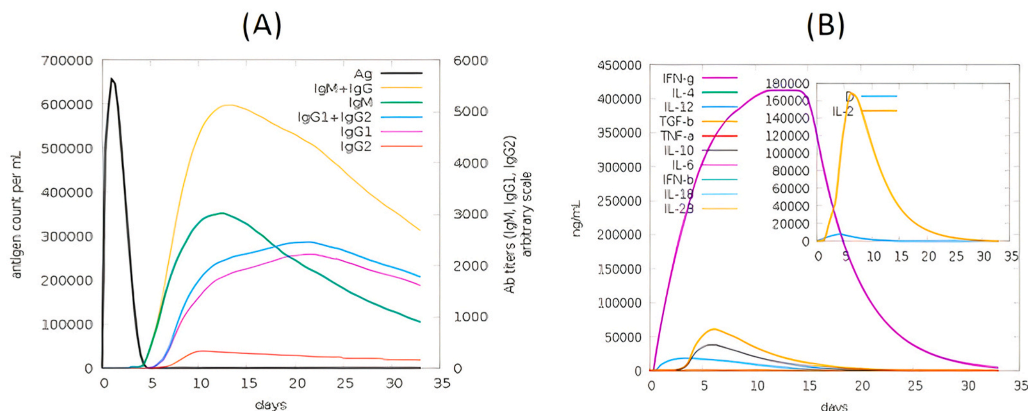

Fig. 5. (A) The virus, immunoglobulins and immunocomplexes. (B) Cytokines and interleukins concentrations. The danger signal is shown in the inset plot together with the leukocyte growth factor IL-2.

Fig. 6. Codon optimized MEV was cloned in silico into the *E. coli* K12 expression system. Black color depicts the backbone of the plasmid and red indicates the inserted DNA sequence. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

time and expense of developing conventional vaccines. The creation of successful multiepitope vaccines is tricky due to difficulties in selecting acceptable antigen candidates and immune-dominant epitopes. As a result, immune-informatics methods for predicting suitable antigenic epitopes of a targeted protein are critical for developing a MEV [46].

With the aid of reverse vaccinology, a multiepitope vaccine called MEV for SARS CoV2 has been suggested, which combines structural and non-structural proteins; outer surface-exposed epitopes. Although various groups have attempted to develop a vaccine against SARS-CoV2, the current study describes a detailed protocol to developing a vaccine construct that addresses all potential design flaws such as the absence of surface epitopes, the presence of cross-reactivity with human proteins, and the lack of an immune response. Considering South African strain, that is highly transferable and pathogenic while containing important mutations, so that MEV can be effective for the new strains emerging in different parts of the world. Envelope and membrane proteins along with surface glycoprotein all have a role in internalization and surface adhesion in SARS CoV2. Whereas, non-structural proteins are virulence-associated components, that induce immune-pathogenesis. Surface glycoprotein (spike protein) interaction with the ACE2 receptor aids SARS CoV2 host cell internalization.

B-cell, HTL, and CTL epitopes were included in MEV that causes successful responses to a specific virus [47]. Few groups have created SARS-CoV-2 subunit vaccines, but they have only utilized a single protein for vaccine design [48,49], and they have only used CTL epitopes without considering the relevance of HTL or B cell epitopes. CTL prevents pathogen transmission by secreting specific antiviral cytokines and identifying and destroying infected cells [10]. The discovery, creation, and optimization of effective adjuvants are closely related to increasing the efficacy of a designed vaccination [50]. Adjuvants are molecular complexes that, when handled inside a vaccination design, increase an immunological response [51]. Linkers, also known as's-pacers,' are important components in the development of protein vaccines. They play a crucial role in the structural stability, interdomain interactions, and vaccination functionality [52]. Spacers connect the three key components of a vaccine construct: B Cell epitopes, T Cell epitopes, and intramolecular adjuvants. Lack of quantity in synthesis, misfolded protein structures, and/or diminished bioactivity can all result from the fusing of functional domains without/with an unsuitable linker [53].

In this study, we used a next-generation vaccine design method to generate a MEV construct that can elicit immune responses against SARS-CoV-2. The MEV design is expected to induce both cell-mediated and humoral immune responses. The receptor's interaction and binding patterns with the vaccination protein remained constant and improved. Furthermore, effective immune responses were found in real life during immunological simulation. MEV, which was carefully created using such an approach, might thus become an asset in the fight against viral diseases.

Experimental techniques are used to provide initial raw data for computational/immunoinformatics approaches. The accuracy of immune-informatics predictions might be limited by the quality of the

data and the effectiveness of the computer methods used. To guarantee the true potential of developed MESV to prevent COVID-19, more in vivo and in vitro research is necessary.

5. Conclusion

Due to the unavailability of vaccine for new strains, COVID 19 has a high death and morbidity rate; thus, effective treatment methods, fast development, testing, and production are required immediately for this worldwide pandemic illness. New SARS-CoV-2 variants have been detected in South Africa and a few other states/countries across the world, indicating a severe situation that requires the attention of scientist and doctors. Currently, the B.1.351 (South Africa) form is of greater concern, as it is less susceptible to sera from persons who have been vaccinated.

Using different immunoinformatics and computational methods, we analyzed SARS-CoV-2 structural proteins for antigenic epitopes and proposed a possible MEV. The outcomes of this study might help researchers save time and money when studying experimental epitope targets. The MEV has sufficient physicochemical and structural characteristics to activate all components of the host immune system. It also appears to have a stable interaction with the innate immune receptor TLR3, which makes it more effective to attack immune system of the host. However, more in-vitro and in-vivo tests are needed to determine its efficacy in the battle against COVID-19.

Funding information

This research is supported by grants from the National Science Foundation of China (Grant No. 32070662,61832019, 32030063) , the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the Science and Technology Commission of Shanghai Municipality (Grant No.: 19430750600), as well as SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2021ZD02). The computations were partially performed at the Pengcheng Lab and the Center for High-Performance Computing, Shanghai Jiao Tong University.

Declaration of competing interest

The authors declare that there is no conflict of interest.

References

[1] M.A. Alamri, M. Tahir Ul Qamar, M.U. Mirza, R. Bhadane, S.M. Alqahtani, I. Muneer, M. Froeyen, O.M. Salo-Ahen, Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro, J. Biomol. Struct. Dyn. (2020) 1–13.

[2] M.T. Ul Qamar, S. Saleem, U.A. Ashfaq, A. Bari, F. Anwar, S. Alqahtani, Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study, J. Transl. Med. 17 (2019) 1–14.

[3] A.H. De Wilde, E.J. Snijder, M. Kikkert, M.J. van Hemert, Host factors in coronavirus replication. Roles of Host Gene and Non-coding RNA Expression in Virus Infection, 2017, pp. 1–42.

[4] C. Wang, P.W. Horby, F.G. Hayden, G.F. Gao, A novel coronavirus outbreak of global health concern, Lancet 395 (2020) 470–473.

[5] L.E. Gralinski, V.D. Menachery, Return of the coronavirus: 2019-nCoV, Viruses 12 (2020) 135.

[6] J.S. Peiris, K.Y. Yuen, A.D. Osterhaus, K. Stöhr, The severe acute respiratory syndrome, N. Engl. J. Med. 349 (2003) 2431–2441.

[7] M. Cotten, S.J. Watson, A.I. Zumla, H.Q. Makhdoom, A.L. Palser, S.H. Ong, A.A. Al Rabeeah, R.F. Alhakeem, A. Assiri, J.A. Al-Tawfiq, Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus, mBio 5 (2014) e01062-13.

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

[9] J. Chen, J. Shi, T. Yau, C. Liu, X. Li, Q. Zhao, J. Ruan, S. Gao, Bioinformatics analysis of the 2019 novel coronavirus genome, Chinese J. Bioinf. 18 (2) (2020) 96–102.

[10] S. Mishra, T Cell Epitope-Based Vaccine Design for Pandemic Novel Coronavirus 2019-nCoV, 2020.

[11] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, A Novel Coronavirus from Patients with Pneumonia in China, New England journal of medicine, 2019, 2020.

[12] J. Liu, R. Cao, M. Xu, X. Wang, H. Zhang, H. Hu, Y. Li, Z. Hu, W. Zhong, M. Wang, Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro, Cell Discov. 6 (2020) 1–4.

[13] C. Wen-Hsiang, S. Ulrich, J. Hotez Peter, B.M. Elena, The SARS-CoV-2 Vaccine Pipeline: an Overview. Current Tropical Medicine Reports, 2020 [Internet] 20.

[14] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: genome structure, replication, and pathogenesis, J. Med. Virol. 92 (2020) 418–423.

[15] B. Ahmad, U.A. Ashfaq, M.-u. Rahman, M.S. Masoud, M.Z. Yousaf, Conserved B and T cell epitopes prediction of ebola virus glycoprotein for vaccine development: an immuno-informatics approach, Microb. Pathog. 132 (2019) 243–253.

[16] E. De Gregorio, R. Rappuoli, Vaccines for the future: learning from human immunology, Microb. Biotechnol. 5 (2012) 149–155.

[17] K.K. Jensen, M. Andreatta, P. Marcatili, S. Buus, J.A. Greenbaum, Z. Yan, A. Sette, B. Peters, M. Nielsen, Improved methods for predicting peptide binding affinity to MHC class II molecules, Immunology 154 (2018) 394–406.

[18] V. Baruah, S. Bose, Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV, J. Med. Virol. 92 (2020) 495–500.

[19] I. Ahmed, G. Jeon, Enabling artificial intelligence for genome sequence analysis of COVID-19 and alike viruses, Interdiscipl. Sci. Comput. Life Sci. (2021) 1–16.

[20] C.K. Wibmer, F. Ayres, T. Hermanus, M. Madzivhandila, P. Kgagudi, B. Oosthuysen, B.E. Lambson, T. De Oliveira, M. Vermeulen, K. Van der Berg, SARS-CoV-2 501Y. V2 escapes neutralization by South African COVID-19 donor plasma, Nat. Med. 27 (2021) 622–625.

[21] Moderna, Moderna COVID-19 Vaccine Retains Neutralizing Activity against Emerging Variants First Identified in the UK and the Republic of South Africa, 2021.

[22] Y. Liu, J. Liu, H. Xia, X. Zhang, C.R. Fontes-Garfias, K.A. Swanson, H. Cai, R. Sarkar, W. Chen, M. Cutler, Neutralizing activity of BNT162b2-elicited serum, N. Engl. J. Med. 384 (2021) 1466–1468.

[23] J. Cohen, South Africa suspends use of AstraZeneca's COVID-19 vaccine after it fails to clearly stop virus variant, Science 10 (2021).

[24] U. Food, D. Administration, Fact sheet for healthcare providers administering vaccine (vaccination providers): emergency use authorization (EUA) of the Pfizer-Biontech COVID-19 vaccine to prevent coronavirus disease 2019 (COVID-19), 2020. December 30.

[25] A.S. Cockrell, J.C. Johnson, I.N. Moore, D.X. Liu, K.W. Bock, M.G. Douglas, R. L. Graham, J. Solomon, L. Torzewski, C. Bartos, A spike-modified Middle East respiratory syndrome coronavirus (MERS-CoV) infectious clone elicits mild respiratory disease in infected rhesus macaques, Sci. Rep. 8 (2018) 1–9.

[26] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinf. 8 (2007) 1–7.

[27] A. Krogh, B. Larsson, G. Von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580.

[28] U. Omasits, C.H. Ahrens, S. Müller, B. Wollscheid, Protter: interactive protein feature visualization and integration with experimental proteomic data, Bioinformatics 30 (2014) 884–886.

[29] T. Stranzl, M.V. Larsen, C. Lundegaard, M. Nielsen, NetCTLpan: pan-specific MHC class I pathway epitope predictions, Immunogenetics 62 (2010) 357–368.

[30] I. Dimitrov, L. Naneva, I. Doytchinova, I. Bangov, AllergenFP: allergenicity prediction by descriptor fingerprints, Bioinformatics 30 (2014) 846–851.

[31] J. Greenbaum, J. Sidney, J. Chung, C. Brander, B. Peters, A. Sette, Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes, Immunogenetics 63 (2011) 325–335.

[32] P. Wang, J. Sidney, C. Dow, B. Mothé, A. Sette, B. Peters, A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach, PLoS Comput. Biol. 4 (2008), e1000048.

[33] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, Nucleic Acids Res. 45 (2017) W24–W29.

[34] N. Nezafat, Y. Ghasemi, G. Javadi, M.J. Khoshnoud, E. Omidinia, A novel multi-epitope peptide vaccine against cancer: an in silico approach, J. Theor. Biol. 349 (2014) 121–134.

[35] V. Solanki, V. Tiwari, Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against Acinetobacter baumannii, Sci. Rep. 8 (2018) 1–19.

[36] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31 (2003) 3784–3788.

[37] D.W. Buchan, D.T. Jones, The PSIPRED protein analysis workbench: 20 years on, Nucleic Acids Res. 47 (2019) W402–W407.

[38] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (2015) 845–858.

[39] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007) W407–W410.

[40] G. Studer, C. Rempfer, A.M. Waterhouse, R. Gumienny, J. Haas, T. Schwede, QMEANDisCo—distance constraints applied on model quality estimation, Bioinformatics 36 (2020) 1765–1771.
[41] G. Weng, E. Wang, Z. Wang, H. Liu, F. Zhu, D. Li, T. Hou, HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA, Nucleic Acids Res. 47 (2019) W322–W330.
[42] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[43] R. Salomon-Ferrer, D.A. Case, R.C. Walker, An overview of the Amber biomolecular simulation package, Wiley Interdiscipl. Rev.: Comput. Mol. Sci. 3 (2013) 198–210.
[44] V. Tiwari, Denovo designing, retro-combinatorial synthesis, and molecular dynamics analysis identify novel antiviral VTRM1. 1 against RNA-dependent RNA polymerase of SARS CoV2 virus, Int. J. Biol. Macromol. 171 (2021) 358–365.
[45] N. Rapin, O. Lund, M. Bernaschi, F. Castiglione, Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system, PLoS One 5 (2010), e9862.
[46] Z. Nain, M.M. Karim, M.K. Sen, U.K. Adhikari, Structural basis and designing of peptide vaccine using PE-PGRS family protein of Mycobacterium ulcerans—an integrated vaccinomics approach, Mol. Immunol. 120 (2020) 146–163.
[47] L. Zhang, Multi-epitope vaccines: a promising strategy against tumors and viral infections, Cell. Mol. Immunol. 15 (2018) 182–184.
[48] M.I. Abdelmageed, A.H. Abdelmoneim, M.I. Mustafa, N.M. Elfadol, N.S. Murshed, S.W. Shantier, A.M. Makhawi, Design of a multiepitope-based peptide vaccine against the E protein of human COVID-19: an immunoinformatics approach, BioMed Res. Int. 2020 (2020).
[49] M. Bhattacharya, A.R. Sharma, P. Patra, P. Ghosh, G. Sharma, B.C. Patra, S.S. Lee, C. Chakraborty, Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): immunoinformatics approach, J. Med. Virol. 92 (2020) 618–631.
[50] S. Parvizpour, M.M. Pourseif, J. Razmara, M.A. Rafi, Y. Omidi, Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches, Drug Discov. Today 25 (2020) 1034–1042.
[51] J. Dietrich, L.V. Andreasen, P. Andersen, E.M. Agger, Inducing dose sparing with inactivated polio virus formulated in adjuvant CAF01, PLoS One 9 (2014), e100879.
[52] N. Nezafat, M. Sadraeian, M.R. Rahbar, M.J. Khoshnoud, M. Mohkam, A. Gholami, M. Banihashemi, Y. Ghasemi, Production of a novel multi-epitope peptide vaccine for cancer immunotherapy in TC-1 tumor-bearing mice, Biologicals 43 (2015) 11–17.
[53] H. Guo, Y. Yang, F. Xue, H. Zhang, T. Huang, W. Liu, H. Liu, F. Zhang, M. Yang, C. Liu, Effect of flexible linker length on the activity of fusion protein 4-coumaroyl-CoA ligase:: stilbene synthase, Mol. Biosyst. 13 (2017) 598–606.