



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.



Immunoinformatics prediction of overlapping CD8⁺ T-cell, IFN- γ and IL-4 inducer CD4⁺ T-cell and linear B-cell epitopes based vaccines against COVID-19 (SARS-CoV-2)



Abiodun J. Fatoba^a, Leah Maharaj^a, Victoria T. Adeleke^b, Moses Okpeku^a, Adebayo A Adeniyi^{c,d,*}, Matthew A. Adeleke^{a,*}

^a Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Westville, P/Bag X54001, Durban 4000, South Africa

^b Discipline of Chemical Engineering, University of KwaZulu-Natal, Howard Campus, Durban 4041, South Africa

^c Department of Chemistry, Faculty of Natural and Agricultural Sciences, University of the Free State, Bloemfontein, South Africa

^d Department of Industrial Chemistry, Federal University, Oye-Ekiti, Nigeria

ARTICLE INFO

Article history:

Received 7 August 2020

Received in revised form 11 December 2020

Accepted 2 January 2021

Available online 18 January 2021

Keywords:

Multi-epitope
Immunoinformatics
Immune-response
SARS-CoV-2
Vaccine

ABSTRACT

At the beginning of the year 2020, the world was struck with a global pandemic virus referred to as SARS-CoV-2 (COVID-19) which has left hundreds of thousands of people dead. To control this virus, vaccine design becomes imperative. In this study, potential epitopes-based vaccine candidates were explored. Six hundred (600) genomes of SARS-CoV-2 were retrieved from the viPR database to generate CD8⁺ T-cell, CD4⁺ T-cell and linear B-cell epitopes which were screened for antigenicity, immunogenicity and non-allergenicity. The results of this study provide 19 promising candidate CD8⁺ T-cell epitopes that strongly overlap with 8 promising B-cells epitopes. Another 19 CD4⁺ T-cell epitopes were also identified that can induce IFN- γ and IL-4 cytokines. The most conserved MHC-I and MHC-II for both CD8⁺ and CD4⁺ T-cell epitopes are HLA-A*02:06 and HLA-DRB1*01:01 respectively. These epitopes also bound to Toll-like receptor 3 (TLR3). The population coverage of the conserved Major Histocompatibility Complex Human Leukocyte Antigen (HLA) for both CD8⁺ T-cell and CD4⁺ T-cell ranged from 65.6% to 100%. The detailed analysis of the potential epitope-based vaccine and their mapping to the complete COVID-19 genome reveals that they are predominantly found in the location of the surface (S) and membrane (M) glycoproteins suggesting the potential involvement of these structural proteins in the immunogenic response and antigenicity of the virus. Since the majority of the potential epitopes are located on M protein, the design of multi-epitope vaccine with the structural protein is highly promising though the whole M protein could also serve as a viable epitope for the development of an attenuated vaccine. Our findings provide a baseline for the experimental design of a suitable vaccine against SARS-CoV-2.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

The recent emergence of the new coronavirus, referred to as 'COVID-19', is posing a global challenge to public health and has placed the global economy under financial burden. The virus, which was first reported in Wuhan, a city in the Hubei province of China [1] in late December 2019, has since spread to different continents of the world causing severe illness, which range from mild sicknesses to death. Globally, 16,523,815 confirmed cases and 655,112 death cases of COVID-19 have been reported across

216 countries as of 29 July 2020 (https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200729-covid-19-sitrep-191.pdf?sfvrsn=2c327e9e_2). The World Health Organization on January 30, 2020, declared COVID-19 outbreak as Public Health Emergency of International concern.

The epidemiological transmission of this virus from human to human has been attributed to sneezing and coughing, which release the droplet that transmits the virus [2]. Some of the symptoms associated with this viral infection include fever, sore throat and pneumonia [2]. To date, there is no certified cure [3]. Presently, treatment relies on symptomatic relief and self-isolation to prevent infection of other people.

The genetic component of this virus is a positive-sense single-stranded RNA with a similar genome to SARS-CoV and bat coronavirus, hence the name SARS-CoV-2 [4]. It belongs to the Coron-

* Corresponding authors at: Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Westville, P/Bag X54001, Durban 4000, South Africa.

E-mail addresses: adeniyiaa@ufs.ac.za (A.A. Adeniyi), adelekem@ukzn.ac.za (M.A. Adeleke).

aviridae family and classified as β -coronavirus with a 2B group [5]. Contrary to other human coronaviruses such as HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1 which are mostly associated with the common cold [6], the emergence of some human coronavirus such as SARS (Severe Acute Respiratory Syndrome), MERS (Middle East Respiratory Syndrome) and the recently COVID-19 have led to fatal epidemics and pandemic [7].

Similar to other coronaviruses, SARS-CoV-2 consists of both structural and non-structural proteins. The structural proteins include N-protein (Nucleocapsid), S-protein (Spike), E-protein (Envelope), and M-protein (Membrane) [8]. Based on the similarity of COVID-19 (SARS-CoV-2) to SARS-CoV, it has been suggested that the previous understanding of the protective immune response in SARS-CoV could be helpful in the design of effective vaccine for this novel SARS-CoV-2 [8]. Different reports have indicated the crucial role of the humoral and cell-mediated immune response in the control of SARS-CoV [9–11]. Although antibodies have been reported to be produced against N-protein of SARS-CoV, the responses were only for a short time [12]. On the other hand, T-cell response is highly immunogenic against all structural proteins of SARS-CoV and has been reported to be dominant and long-lasting specifically against N and S proteins [11].

Currently, immunoinformatics studies on different potential vaccine epitopes have been carried out by various researchers and several are under clinical trials [13–16]. Although most of these studies explored potential T-cell and B-cell epitopes from different proteins of the viral genome, the protein sequences retrieved were based on limited data of the viral genome that were available at different databases at the early onset of the pandemic. Due to the drastic impact of this virus and the emergence of updated genomic data, further study on the potential immunogenic T-cell and B-cell epitopes are crucial as this could guide the experimental design of vaccines against SARS-CoV-2. Therefore, this study provided details of the immunoinformatics approach towards the development of multi-epitope vaccine candidates against COVID-19 (SARS-CoV-2) with long-lasting immune response and effectiveness when validated.

2. Materials and methods

The flow chart of all the major steps in the design of the potential CD4⁺ and CD8⁺ T-cell, and B-cell epitopes and multi-epitope vaccines are shown in Figure S1. The details of the various steps are described in this section.

2.1. SARS-CoV-2 sequence retrieval and filtering

The genomic sequences of SARS-CoV-2 isolates were retrieved from the ViPR database (Virus Pathogen Database and Analysis Resource) (<http://www.viprbrc.org/>). All available sequences were used, including partial and complete. Human Leukocyte Antigen (HLA) CD8⁺ and CD4⁺ were also obtained from protein database (RCSB PDB (www.rcsb.org)) [17] with identifiers 3QZW and 3S4S respectively. All retrieved sequences were aligned by multi-sequence alignment using the CLUSTALW server [18] with the default parameters. The CLUSTAL W server uses the k-tuple method to generate a pairwise alignment, thereafter the server runs tree construction by the Neighbour-Joining method which will inform the final multiple sequence alignment [18]. From the alignment, conserved regions of 15 amino acids or more with the CD8⁺ and CD4⁺ epitopes were selected. The identified sequences were then subjected to antigenic (Vaxijen V2.0) [19] and transmembrane helix property prediction (based on a hidden Markov model) (TMHMM V2.0) [20]. For antigenicity prediction, the virus was the target organism and the threshold value was 0.4 as used

elsewhere in similar research [21]. The Vaxijen server uses auto cross-covariance transformation of protein sequences into uniform vectors and is alignment-independent (alignment to known antigens is not a reliable way to predict antigenicity) [19]. From these antigenic sequences, the transmembrane sequences as determined by TMHMM [20] were extracted for further analysis.

For this study, the definitions of ‘immunogenicity’ and ‘antigenicity’ were used as outlined by [22]. A substance is considered to be immunogenic if it can induce cellular and humoral responses. Conversely, a substance is antigenic if it can be recognized by antibodies that arise due to the immune response to the substance.

2.2. Prediction of CD8⁺ T-cell epitopes and MHC-I binding alleles

NetCTL v1.2 was employed to predict nonamers that can bind major histocompatibility (MHC) class I (HLA allele) molecules [23]. NetCTL incorporates aspects of proteasomal cleavage, TAP transport efficiency and MHC class I affinity to predict epitopes [23]. Within the NetCTL parameters, the A1 supertype was selected while the weight on C terminal cleavage, weight on TAP transport efficiency and a threshold for epitope identification were at 0.15, 0.05, and 0.75 respectively which were the default parameters. Nonamers above the threshold of 0.5 were input into the IEDB analysis tool (<http://tools.iedb.org/mhci/>) [24] to identify frequently and non-frequently occurring MHC class I binding alleles. The Stabilized Matrix Base Method (SMM) was used with an amino acid length of 9 and IC₅₀ value of < 250 to determine the CD8⁺ T-cell epitopes.

2.3. Prediction of CD4⁺ T-cell epitopes and MHC-II binding allele prediction

To predict CD4⁺ T-helper lymphocytes and the MHC class II alleles, the IEDB MHC II binding tool was used (<http://tools.iedb.org/mhcii/>) [25]. SMM method was utilized and peptides with an IC₅₀ < 250 were selected as the CD4⁺ T-cell epitopes.

2.4. Prediction of immunogenicity, antigenicity, conservancy and allergenicity

Immunogenicity of the CD8⁺ T-cell epitopes was deduced from the MHC I immunogenicity tool of IEDB (<http://tools.iedb.org/immunogenicity/>) [26]. This server describes immunogenicity as T-cell recognized peptide-MHC complexes [26]. Vaxijen (<http://www.ddg-pharmfac.net/vaxijen/Vaxijen/Vaxijen.html>) [15] was used to determine protective antigenicity scores for all epitopes, only those with scores above 0.5 were used as done in other studies [27,28]. The threshold was increased from the initial value of 0.4 to ensure higher accuracy at this stage of the protocol.

To assess conserved epitopes that correspond to a portion of protein sequence that will restrain the epitope [29], both the CD8⁺ and CD4⁺ T-cell epitopes were input into the IEDB conservation across antigens tool (<https://tools.iedb.org/conservancy/>) [30]. Following this, AllerTOP (<http://www.ddg-pharmfac.net/AllerTOP/>) was used to determine if the epitopes would be expected to cause an allergic reaction in patients or not. Thus the resulting epitopes were predicted to be immunogenic, conserved across antigens and had no allergenicity.

To identify IFN- γ inducer properties in CD4⁺ T-cells, the online tool IFNepitope (<https://crdd.osdd.net/raghava/ifnepitope>) tool was utilized (Dhanda et al., 2013). The CD4⁺ T-cell epitopes that were IFN- γ inducing then underwent IL-4 inducer prediction by the IL4pred tool (<http://crdd.osdd.net/raghava/il4pred/>) [31]. For IL4 prediction, the support vector machine (SVM) model was used with a threshold value of 0.1.

2.5. HLA-allele distribution

As it is known that different HLA alleles are differentially expressed in different ethnicities, it is important to determine the HLA-allele distribution globally to inform successful peptide-based vaccine development. To do this, the CD8⁺ and CD4⁺ T-cell epitopes were input into the IEDB population coverage analysis tool (<http://tools.iedb.org/population/>) [32].

2.6. Linear B-cell epitope prediction

In this, the initial sequences with an antigenicity score ≥ 0.4 from the Vaxijen server and identified as the outer membrane protein in TMHMM were input in the ABCpred server (<http://osddlinux.osdd.net/raghava/abcpred>) [33]. ABCpred predicts linear B-cell epitopes through the use of a recurrent neural network [33]. The window length and threshold values were hexadecamer and 0.51 respectively. The B-cell epitopes generated were later tested for their antigenicity and allergenicity. Overlapping B-cell epitopes with CD8⁺ T-cell epitopes were also filtered from the list of epitopes.

2.7. Multi-epitopes construction

All the selected promising CD4⁺ T-cells and B-cells epitopes were mapped to the surface of the COVID-19 complete genome (GenBank: MN908947.2) obtained from NCBI using the software Snapgene. The region of the epitopes that mapped well with the in house protein region as shown in Figs. 2 and S3 was selected as multi-epitopes. The multi-epitopes peptides located on surface (S) glycoprotein also called spike protein was cut from the pdb protein structure (6vxx), the ones located on the orf1ab polyprotein and nucleocapsid (N) phosphoprotein were generated using trRosetta tool (<https://yanglab.nankai.edu.cn/trRosetta/help/>) [34] while the long chains multi-epitopes peptides (Table 5) from the membrane (M) and ORF3a were obtained through the homology model.

2.8. Molecular docking

To determine the interaction between CD8⁺ T-cell epitopes with the chosen MHC-I HLA allele HLA-A*02:06 and CD4⁺ T-cell epitopes with MHC-II allele HLA-DRB1*01:01, the respective protein geometries (pdb: 3OXR and 1AQD respectively) were retrieved from the

RCSB Protein Data Bank in the PDB format [35]. Similarly, the interaction of multi-epitope with Toll-like receptor 3 (TLR) was also evaluated. The protein ‘2A0Z’ pdb was retrieved for TLR3 and docked with multi-epitope. Docking was conducted on pepATTRACT (<http://www.attract.ph.tum.de/services/ATTRACT/peptide.html>) [36] wherein 50 structures were added to the resulting pdb file and run on 12 processing cores. pepATTRACT is fully blind but performs similarly to local docking methods FlexPepDock and HADDOCK [36]. All other criteria were left as default. The docking file was retrieved from this and run on the Centre for High Performance Computing (CHPC) server, Lengau cluster to complete docking calculations. This was conducted for each CD8⁺ T-cell epitope and the HLA allele. From the 50 structures generated for each docking run, the structure with the lowest energy state was selected and viewed in Visual Molecular Dynamics software (VMD) [37]. The CD8⁺ T-cell epitope was identified from the HLA structure and indicated accordingly. The image was then rendered in VMD and subsequently Chimera [38] for publication.

Further docking of the multiepitopes to TLR was carried out using HADDOCK 2.2 [39–41]. The solvent accessibility (SA) of > 50 was used to define the active residues of the entire multi-epitopes that will interact with TLR3. The active site residues of the TLR was selected and the active residues define for residues with SA > 50 . Three docking results were generated from the Haddock dockings which were the results from the constrained docking, flexible docking and refinement in water solvent medium.

3. Results

3.1. Conservation, antigenicity and proteins localization

Out of the 600 genomes of SARS-CoV-2 that were subjected to multiple sequence alignment, 495 conserved protein sequences were generated. Antigenicity analysis with threshold ≥ 0.4 showed only 43 out of 495 conserved sequences to be antigenic and after being subjected to transmembrane topology analysis, only 38 of these antigenic sequences met exomembrane criteria (Table S1). Among these 38 protein sequences that met both antigenic and transmembrane criteria, the Vaxijen scores ranged from 0.4031 to 0.9547. These sequences were selected for further analysis.

3.2. CD8⁺ T-cell epitopes and their MHC-I binding allele identification

The 38 protein sequences obtained were used to generate the CD8⁺ T-cell epitopes using NetCTL v1.2 servers with a threshold of greater than or equal to 0.5 to ensure a high confidence in the prediction of the epitopes. A total of 975 nonamers were generated, but 330 were selected after being subjected to the IEDB MHC-I prediction tool under the criteria that the IC₅₀ value < 250 . Antigenicity and immunogenicity analysis were conducted on the 330 epitopes of CD8⁺ T-cell and only 18 CD8⁺ T-cell epitopes were both antigenic and immunogenic. Similarly, the 18 CD8⁺ T-cell epitopes interact with at least one different MHC-I HLA alleles. The epitopes IFLWLLWPV and YIIKLIFLW had the highest and lowest antigenic score of 1.4835 and 0.5220 respectively (Table 1). The highest and lowest immunogenic scores of 0.3785 and 0.0076 were also observed for IFLWLLWPV and AFLPFAMGI epitopes respectively. A total of 12 MHC class I binding alleles made up of 7 HLA-A (HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-A*23:01, HLA-A*30:01, HLA-A*31:01, HLA-A*68:02) and 5 HLA-B alleles (HLA-B*15:01, HLA-B*35:01, HLA-B*53:01, HLA-B*57:01, HLA-B*58:01) were observed for the 18 CD8⁺ T-cell epitopes. The epitope YIIKLIFLW with 5 HLA alleles had the maximum number of alleles. All the 18 epitopes were most conserved with CD4⁺ T-cell epitopes with a 100% conservancy score. The HLA-A*02:06 was the most

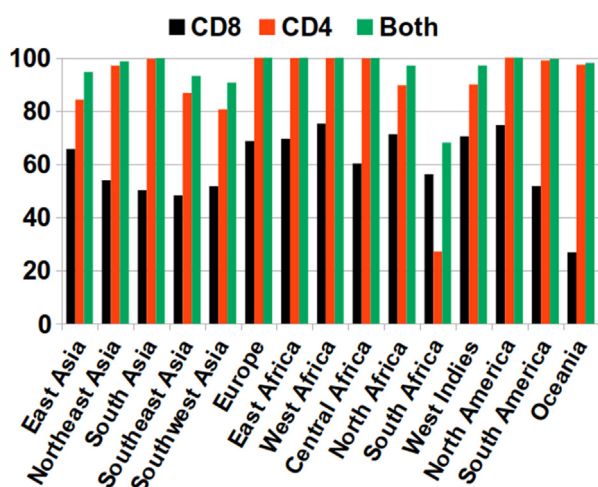


Fig. 1. Population coverage rate of overlapping CD8⁺ T cell epitopes and CD4⁺ T cell epitopes in SARS-CoV-19.

Table 1

The 18 final CD8⁺ T-cells epitopes of SARS-CoV-2 that are found to be antigenic, immunogenic, overlapping with CD4⁺ T-cell epitopes and interacting with MHC class I HLA-alleles.

CD8 ⁺ T-cells epitopes	Antigenicity score	Immunogenicity	MHC-I alleles	Position	Conservancy scores
IFLWLLWVW	1.4835	0.3785	HLA-A*02:01, HLA-A*02:06	80	100%
RNRFLYIIK	0.8364	0.3104	HLA-A*30:01,HLA-A*31:01	102	100%
ATRRIRGGD	0.6108	0.2956	HLA-A*30:01	57	100%
LWLLWVWTL	1.1590	0.2480	HLA-A*23:01	276	100%
ENAFLPFAM	1.0278	0.2309	HLA-A*68:02, HLA-B*35:01	151	100%
ATATIPIQA	0.7881	0.2155	HLA-A*68:02	54	100%
LAAYRINW	1.1825	0.2079	HLA-B*58:01, HLA-B*53:01, HLA-B*57:01	170	100%
LACFVLAAV	1.1343	0.1859	HLA-A*02:06, HLA-A*02:03	171	100%
FAMGIIAMS	1.1005	0.1624	HLA-A*68:02	160	100%
NLLLLFVTV	0.6912	0.1421	HLA-A*02:06,HLA-A*02:01,HLA-A*02:03	326	100%
TLACFVLA	0.5781	0.1248	HLA-A*02:03,HLA-A*02:01, HLA-A*02:06	240	100%
VTLACFVLA	0.5358	0.1218	HLA-A*02:06	80	100%
ELLHAPATV	1.0476	0.1123	HLA-A*02:03	149	100%
RVVVLSFEL	0.6783	0.0457	HLA-A*02:06	125	100%
YIIKLIFLW	0.5220	0.0330	HLA-B*58:01,HLA-B*53:01,HLA-A*02:01, HLA-A*02:06, HLA-B*57:01	130	100%
GIAIAMACL	0.5505	0.0209	HLA-A*02:03	13	100%
AMACLVGLM	0.9558	0.0086	HLA-A*02:03,HLA-B*15:01	31	100%
AFLPFAMGI	0.9095	0.0076	HLA-A*02:06	17	100%

conserved MHC-class 1 binding allele for all the 18 CD8⁺ T-cell epitopes (Table 1). Allergenicity screening of the 18 CD8⁺ T-cell epitopes revealed that they were non-allergens.

3.3. Identification of CD4⁺ T-cell epitopes and their MHC-II binding alleles from conserved sequences

The 495 conserved protein sequences were subjected to the IEDB MHC-II prediction tool under the following criteria: SMM-align based and IC₅₀ value < 250. A total of 200 pentadecamer peptides (epitopes) were generated and then subject to antigenicity test. Only 122 pentadecamers epitopes were antigenic and regarded as the potential CD4⁺ T-cell epitopes. The 122 pentadecamers were also subjected to overlapping test with the immunogenic CD8⁺ T-cell epitopes. The epitopes TLACFVLAAYRINW and RNRFLYIIKLIWL had the highest and lowest antigenic score of 1.3949 and 0.5114 respectively as shown in Table 2. A total of 13 MHC class II binding alleles (HLA-DPA1*01:03, HLA-DPA1*02:01, HLA-DPA1*03:01, HLA-DQA1*01:01, HLA-DRB1*01:01, HLA-DRB1*08:02, HLA-DRB1*11:01, HLA-DRB4*01:01, DPB1*01:01, DPB1*02:01, DPB1*04:02, DPB1*05:01, DQB1*05:01) were

observed for the 19 CD4⁺ T-cell epitopes. The most conserved MHC-class II binding allele for all the 19 CD4⁺ T-cell epitopes was HLA-DRB1*01:01 (Table 2).

3.4. Identification and analysis of overlapping CD8⁺ and CD4⁺ T-cell epitopes

Conservancy analysis was carried out on the 19 CD4⁺ T-cell epitopes with the immunogenic and antigenic 18 CD8⁺ T-cell epitopes (Table 1) and all the epitopes showed 100% conservancy. The overlapped CD4⁺ T-cell epitopes together with their antigenicity scores and interacting MHC-II alleles are shown in Table 2.

3.5. Analysis of the IFN-γ and IL-4 property of CD4⁺ T-cell epitopes

The overlapped CD4⁺ T-cells were also tested for their ability to induce interferon-gamma (IFN-γ) and interleukin 4 (IL-4) using IFNepitope and IL4pred prediction tool. The final 19 IFN-γ and IL-4 inducing CD4⁺ T-cell epitopes were overlapped 18 CD8⁺ T-cell epitopes indicated in Table 2. All these epitopes were also non-allergic.

Table 2

The 19 final CD4⁺ T-cells epitopes of SARS-CoV-2 that are found to be antigenic, IFN-γ and IL-4-inducer, overlapping with CD8⁺ T-cell epitopes and interacting with MHC class II HLA-alleles.

Overlapping CD4 ⁺ T-cells Epitopes	Antigenicity scores	MHC-II alleles	Conservancy (%)	IFN-γ	IL-4
ACFVLAAYRINWIT	1.1115	HLA-DRB4*01:01	100%	Inducer	Inducer
AFLPFAMGIIAMSAF	1.0499	HLA-DQA1*01:01/DQB1*05:01	100%	Inducer	Inducer
FFLYENAFLPFAMGI	0.8191	HLA-DPA1*02:01/DPB1*05:01	100%	Inducer	Inducer
FLPFAMGIIAMSAFA	1.0725	HLA-DRB4*01:01	100%	Inducer	Inducer
GIAIAMACLVGLMWL	0.9155	HLA-DRB4*01:01	100%	Inducer	inducer
IGYYRRATRRIRGGD	0.6649	HLA-DPA1*02:01/DPB1*01:01	100%	Inducer	Inducer
IKLIFLWLLWVWTLA	0.8704	HLA-DQA1*01:01/DQB1*05:01	100%	Inducer	Inducer
KLIFLWLLWVWTLAC	0.7344	HLA-DRB1*08:02	100%	Inducer	Inducer
LFPFAMGIIAMSAFAM	1.0879	HLA-DRB1*01:01	100%	Inducer	Inducer
NAFLPFAMGIIAMSA	0.9681	HLA-DPA1*03:01/DPB1*04:02	100%	Inducer	Inducer
PSDFVRATATIPIQA	0.5466	HLA-DPA1*03:01/DPB1*04:02	100%	Inducer	Inducer
PVTLACFVLAAYRYI	0.9703	HLA-DPA1*03:01/DPB1*04:02	100%	Inducer	Inducer
RNRFLYIIKLIWL	0.5114	HLA-DPA1*02:01/DPB1*01:01	100%	Inducer	Inducer
RVVVLSFELLHAPAT	0.7485	HLA-DQA1*01:01/DQB1*05:01	100%	Inducer	Inducer
SDFVRATATIPIQAS	0.5411	HLA-DQA1*01:01/DQB1*05:01	100%	Inducer	Inducer
TLACFVLAAYRINW	1.3949	HLA-DQA1*01:01/DQB1*05:01	100%	Inducer	Inducer
VHFVCLNLLLLFVTVY	1.3132	HLA-DPA1*03:01/DPB1*04:02	100%	Inducer	Inducer
VVLSFELLHAPATVC	0.8835	HLA-DRB1*11:01	100%	Inducer	Inducer
YRVVLSFELLHAPA	0.9711	HLA-DPA1*01:03/DPB1*02:01	100%	Inducer	Inducer

3.6. Distribution analysis of combined MHC-I and MHC-II HLA alleles

Evaluating the frequency of distribution of HLA alleles in different ethnicity is an important step that determines the possible effectiveness of a potential vaccine. Thus, the different HLA alleles that bound to the predicted CD8⁺ and CD4⁺ T-cell epitopes were subjected to population coverage analysis. Population coverage of combined MHC-I and MHC-II that is close to 100% indicates good coverage of a potential vaccine. The predicted T-cell epitopes indicated in Table 1 and 2 respectively were subjected to population coverage across 15 different geographical areas of the world as shown in Fig. 1. Among the 18 CD8⁺ T-cell epitopes, the highest coverage was observed in West Africa (75.16%) followed by North America (74.63%), North Africa (71.16%) and West Indies (70.37%) while the lowest coverage was found in Oceania (26.78%).

Better population coverage was obtained for CD4⁺ T-cell epitopes compared to CD8⁺ T-cell epitopes (Fig. 1). The highest coverage for CD4⁺ T-cell epitopes was found in North America (99.99%), Europe (99.94%), East Africa (99.84%), West Africa (99.84%), Central Africa (99.73%) and South Asia (99.52%) while the least coverage was in South Africa (27.07%). The combination of the two T-cell epitopes (CD8⁺ and CD4⁺ T-cell epitopes) resulted in improved population coverage. The best coverage for both was found in North America (100%) followed by Europe (99.98%), West Africa (99.96%), East Africa (99.95%), Central Africa (99.89%), South Asia (99.76%) and South America (99.49%) while the lowest coverage was still South Africa but of an improved percentage coverage (68.05%). The plots showing the percentage distribution of the epitopes hits for the countries South Africa, West Africa, Europe and North America are shown in Figure S2. A better distribution was observed in countries like West Africa, Europe and North America that had a higher percentage than South Africa where the percentage was small.

3.7. Prediction and selection of putative B-cell epitopes

B-cell epitopes were predicted from the 495 conserved sequences using the ABCpred server. A total of 187B-cell epitopes were generated. These were thereafter subjected to antigenicity and allergenicity tests of which only 21B-cell epitopes were selected. Further overlapping was then carried out with the CD8⁺ T-cell epitopes indicated in Table 1 resulting in 8B-cell epitopes that had 100% conservancy with CD8⁺ T-cell epitopes (Table 3).

To design target-specific vaccines, B-cells epitopes is important because of the ability of the extreme specificity of B-cell to neutralize pathogenic molecules through secretion of antibodies [42,43].

3.8. Multi-epitopes vaccines development

All the final 19 CD4⁺ T-cell and 8B-cell epitopes were mapped to the complete viral genomes as shown in Fig. 2 to know the position where they can be identified. The positions on the complete gen-

omes where the potential epitope-based vaccines candidates were identified are the orflab polyprotein, surface (S) glycoprotein, ORF3a protein, membrane (M) glycoprotein and nucleocapsid (N) phosphoprotein (Figure S3). All the 19 CD4⁺ T-cell epitopes and 8B-cell epitopes were identified in the complete genome. As shown in Fig. 2, the large protein in COVID-19 genome is orflab, and a total of 5 CD4⁺ T-cell epitopes (number E2, E3, E4, E9 and E10 corresponding to their numbering in Table 2) and 1B-cell epitope (E7 corresponding to their numbering in Table 3) were identified and were found within a very close range (11075 to 11,143 DNA base pair) as shown in Figure S3. The next big protein to orflab is the Spike glycoprotein where a total of 3 CD4⁺ T-cell epitopes (E14, E18 and E19) and 1B-cell (E6) where identified also within a very close range (23084 to 23,137 DNA base pair). All other proteins in COVID-19 are small proteins, 3 CD4⁺ T-cell epitopes (E11, E15 and E17) and 1B-cell epitope (E8) was identified but were found to be widely separated from each other (25465 to 25,872 DNA base pair). Therefore, we proposed three types of the multi-epitopes for the ORF3a; one is the whole region select called ORF3a, the second one is called ORF3a-1a, which comprised of the region of E11, E15 and E17 while the third one is called ORF3a-1b that encompassed all the residues after CD4⁺ T-cell epitopes E17 to B-cell E8 in Figure S3.

The highest number of epitopes are found in the membrane (M) protein of the genome even though it is a very short protein compared to the rest of the proteins where the potential vaccine epitopes are identified. A total of 7 CD4⁺ T-cell epitopes (E1, E5, E7, E8, E12, E13, E16) and 3B-cell epitopes (E1, E4, E5) were identified. All the identified epitopes are within a short-range (26646 to 26,804 DNA base pair). Since the whole M protein is also a small protein, we propose taking the whole protein as a potential vaccine (name M–whole in Table 5). This can be attenuated if necessary should there be any identified virulence in it. The M–protein is known to be the most structured protein that helps in the determination of the shape of the virus envelope [44]. One unique feature of the M–protein is the ability to bind to other structural proteins. The binding of the nucleocapsids or N-protein with M–protein helps to stabilize the N protein-RNA complex inside the internal virion and therefore promotes completion of viral assembly [44]. The N-protein from the genome is the last protein where our proposed epitopes are identified, and only one CD4⁺ T-cell epitope (E) was identified in it. All the proposed multi-epitopes potential vaccines candidates are shown in Table 5. The region was selected with some extra residues that corresponded to the in house protein identified using the Snapgene as shown in Figure S3.

3.9. Molecular docking simulation

3.9.1. The interaction of the CD8⁺ T-cell with MHC-I and CD4⁺ T-cell with MHC-II HLA alleles

The 3D modelling of all the promising epitopes for CD4⁺ and CD8⁺ T-cell and B-cell epitopes were modelled using online

Table 3
Putative linear B-cell epitopes of SARS-CoV-2 with their antigenicity and allergenicity.

B-cell epitopes	ABCpred score	Position	Vaxijen score	Conservancy (%)	Allergenicity
GIAIAMACLVGLMWLS	0.74	1206	0.9132	100%	Non-allergen
GKNWITGGIAIAMACL	0.78	1199	0.6906	100%	Non-allergen
LWPVTLACFVLAAYVN	0.59	1086	0.9305	100%	Non-allergen
LWPVTLACFVLAAYVR	0.56	1727	0.8961	100%	Non-allergen
RINWITGGIAIAMACL	0.77	1401	1.2392	100%	Non-allergen
VVLSFELLHAPATVCG	0.66	560	0.6184	100%	Non-allergen
YENAFLPFAMGIIAMS	0.67	1607	1.1003	100%	Non-allergen
YFLCWHNTNCYDYCIPI	0.84	773	0.8340	100%	Non-allergen

Table 4

The interaction energies of CD4⁺ T-cell epitopes with MHC-II HLA allele HLA-DRB1*01:01 and CD8⁺ T-cell epitopes with MHC-I HLA allele HLA-A*02:06 with the number of interacting residues, number of hydrogen bonds and number of non-bonded contacts.

Epitopes	Binding energy (kcal/mol)	Number of interacting residues	Number of interacting residues	Number of Hydrogen bond	Number of non-bonded contacts
CD4⁺ T-cell epitopes		HLA	Epitope		
ACFVLAAYRINWIT	-147.631	29	13	9	139
AFLPFAMGIIAMSAF	-121.093	24	12	8	114
FFLYENAFLPFAMGI	-110.855	26	11	4	117
FLPFAMGIIAMSAFA	-114.943	29	13	9	146
GIAIAMACLVGLMWL	-147.807	26	14	7	147
IGYYRRATRRIRGGD	-178.556	31	13	7	139
IKLIFLWLLWPVTLA	-124.515	24	11	3	134
KLIFLWLLWPVTLAC	-142.978	28	13	5	123
LPFAMGIIAMSAFAM	-128.490	27	13	12	128
NAFLPFAMGIIAMSA	-123.086	25	10	8	122
PSDFVRATATIPIQA	-90.283	23	12	3	126
PVTLACFVLAAYVRI	-129.133	28	13	9	140
RNRFLYIIKLIWL	-170.388	27	12	5	131
RVVVLSFELLHAPAT	-142.421	30	13	9	114
SDFVRATATIPIQAS	-158.042	21	14	16	183
TLACFVLAAYRINW	-131.079	23	12	10	103
VHFVCNLLLLFVTVY	-152.189	28	13	8	137
VVLSFELLHAPATVC	-137.146	28	11	11	139
YRVVLSFELLHAPA	-130.201	24	13	12	139
CD8⁺ T-cell epitopes					
AFLPFAMGI	-121.823	3	9	5	95
AMACLVGLM	-121.143	7	9	8	117
ATATIPIQA	-114.556	9	9	10	109
ATRRIRGGD	-137.499	10	9	12	122
ELLHAPATV	-136.211	6	9	9	113
ENAFLPFAM	-107.396	6	9	7	126
FAMGIIAMS	-114.161	6	9	6	120
GIAIAMACL	-85.360	8	9	9	86
IFLWLLWPV	-131.361	6	9	7	111
LAAYRINW	-106.038	6	9	8	103
LACFVLAAY	-113.552	7	9	7	117
LWLLWPVTL	-143.572	7	9	8	141
NLLLLFVTV	-130.618	6	9	6	144
RNRFLYIIK	-152.292	6	9	8	136
RVVVLSFEL	-131.722	4	9	4	111
TLACFVLAAY	-94.269	5	9	5	109
VTLACFVLA	-104.150	3	9	5	124
YIIKLIWL	-129.732	4	9	7	150

ATTRACT software. The most conserved MHC-I and MHC-II HLA-alleles that were used for the docking study are HLA-A*02:06 and HLA-DRB1*01:01 respectively. The promising 18 CD8⁺ T-cell epitopes were docked to HLA-A*02:06 and the 19 CD4⁺ T-cell epitopes were docked to HLA-DRB1*01:01. All the binding results of the CD8⁺ T-cell and CD4⁺ T-cell epitopes are shown in Table 4. The most promising CD8⁺ T-cell epitopes are RNRFLYIIK, LWLLWPVTL, ATRRIRGGD, ELLHAPATV, RVVLSFEL, and NLLLLFVTV with the binding energy of -152.292, -143.572, -137.499, -136.211, -131.722, and -130.618 respectively (Table 4). The binding site orientation of all the 18 final CD8⁺ T-cell epitopes binds in a similar to HLA-A*02:06 with coil structure except ATRRIRGGD (among the most promising list) and VTLACFVLA (-104.150) that assumed alpha-helix structure (Figure S4).

From Table 4, all the HLA-epitope complexes contained an appreciable number of hydrogen bonds which are important in promoting epitopes binding affinities. The overall interacting residues and non-bonded contacts are also presented. The binding site interaction of RNRFLYIIK that has the best binding energy among the CD8⁺ T-cell epitopes shows that it fits into many of the binding site pockets of HLA-A*02:06 (Fig. 3a). Many of the residues of HLA-A*02:06 have hydrophobic interactions with RNRFLYIIK while its specifically formed hydrogen bond interactions with Lys 66, Arg 97, Lys 146, Tyr 159, Glu 166, Glu 173 and a salt bridge with Glu 173 residues of HLA-A*02:06 (Fig. 3b). The binding site interaction

of the epitope IGYRRATRRIRGGD that was ranked the best among the 19 epitopes that bind to HLA-DRB1*01:01 is shown in Fig. 4a. The epitopes explored the binding pockets of the receptor and stretched across the binding site. Many of the receptor's residues participated in hydrophobic interaction with IGYRRATRRIRGGD. The epitope IGYRRATRRIRGGD forms many hydrophobic interactions and more specific hydrogen bonds interactions with GLN 7, SER 51, TRP 237, ASP 242, THR 253, HIS 257 and a salt bridge with ASP 64 of HLA-DRB1*01:01 (Fig. 4b).

3.9.2. The interaction of B-cell epitopes and multi-epitopes to Toll-like receptor 3 (TLR-3)

As Toll-Like Receptor 3 (TLR-3) is known to greatly facilitate the induction of immune response, the interaction of constructed multi-epitope with TLR3 was studied (Table 5). The minimum binding energies of the multi-epitope were best with water-flexible.

The multi-epitope binding to TLR-3 (Fig. 5) is that of the selected region (red colour) of membrane (M) protein and the whole M-protein. The whole M-protein did not use the epitopes selected region for the interaction (yellow) but rather used the unselected region. This resulted in better TLR-3 interaction of the whole M-protein compared to the selected epitopes part which is an indication that the other parts that are not found to be antigenic can help in improving the immunogenicity of the M-protein.

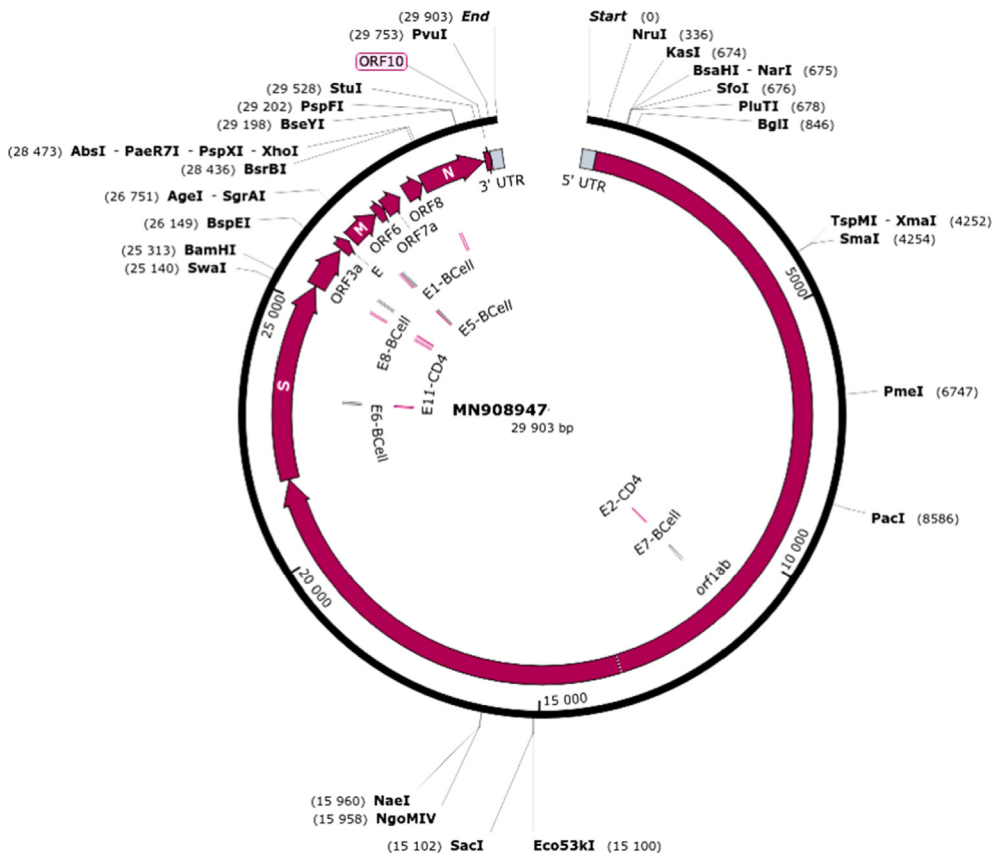


Fig. 2. The whole genome sequence (GenBank number MN908947.2) of COVID-19 built using Snapgene showing the position of the promising CD4⁺ T-cell epitopes and B-cell epitopes. The CD4⁺ T-cell epitopes were named from E1-CD4 to E9-CD4 representing alphabetically the epitopes as found in Table 2, the B-cell epitopes number E1-Bcell to E8-Bcell represented as found in Table 3.

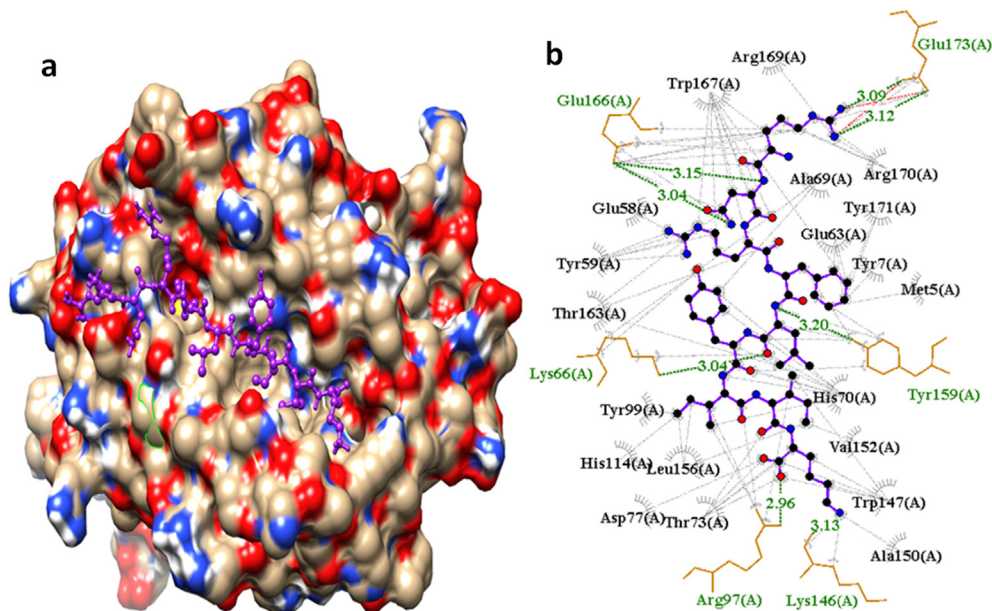


Fig. 3. The feature of the (a) binding interaction of the most promising CD8⁺ T-cell epitope RNRFLYIIK with HLA-A*02:06 and (b) the residue interaction showing the binding site residues that are contributing to the binding interactions. The residues of the receptor that are involved in hydrogen bonding are labeled in green while those involved in hydrophobic interaction are represented with red semicircle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

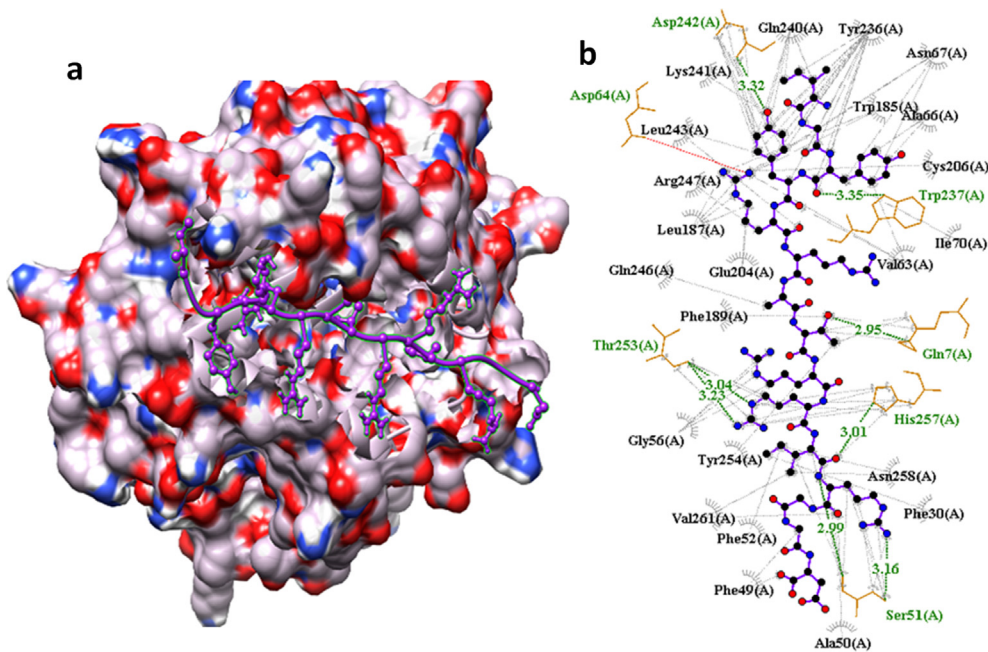


Fig. 4. The feature of the (a) binding interaction of the most promising CD4⁺ T-cell epitope IGYRRATRIRRGD with HLA-DRB1*01:0 and (b) the residue interaction showing the binding site residues that are contributing to the binding interactions. The residues of the receptor that are involve in hydrogen bonding are labeled in green while those involved in hydrophobic interaction are represented with red semicircle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

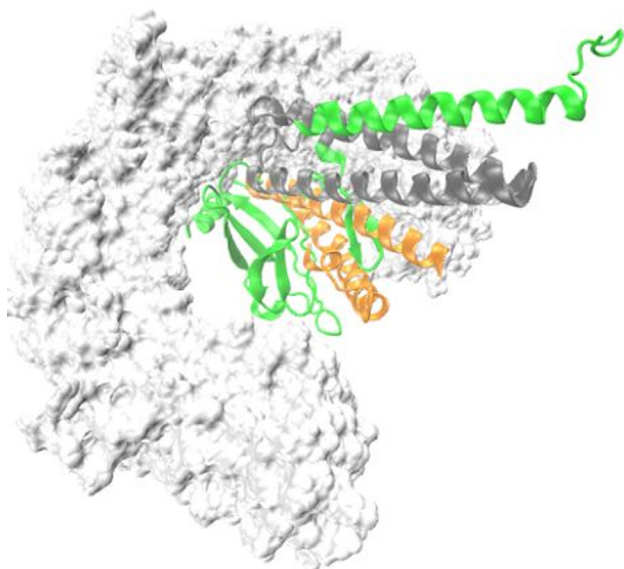


Fig. 5. The interaction of the whole M-protein and selected epitopes part (M-sel in orange and its position on the whole M-protein (green) colour in grey). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

The ongoing fight against the COVID-19 pandemic has become a global phenomenon and has placed demands on scientists for the search of a suitable vaccine against this virus. Although some vaccines are already at clinical and preclinical trials, there is still a need for extensive continuous research on COVID-19 vaccines as the solution is still yet to be found. Due to the limited knowledge of the immunological response of this virus and the urgency for the design of an effective vaccine, the computational design of

T-cell and B-cell immunological epitopes becomes important. Furthermore, vaccines developed through this mean does not contain live pathogen of the virus that can lead to pathogenicity reversal.

Presently, COVID-19 has been declared as a global issue with even first-world countries struggling to contain the disease outbreak. Different HLA type frequencies vary in different ethnicities due to high polymorphism of the MHC molecule, too much polymorphism may limit the proportion of the human population that responds to a particular antigen [45]. As such, the alleles considered in this study had to be proven to show sufficient population coverage on a large scale as shown in Fig. 1. The combined population coverage of both CD8⁺ CD4⁺ T-cell was highest in North America (100%) followed by Europe (99.98%), West Africa (99.96%), East Africa (99.95%), Central Africa (99.89%), South Asia (99.76%) and South America (99.49%), an indication of good qualities of potential vaccine epitopes. While the results showed that the selected alleles were sufficiently prevalent in most countries, South Africa showed the lowest coverage. This is similar to the separate study on the coverage of the epitopes based vaccine for Rift valley fever that also found the percentage coverage in South Africa to be small [29]. The reason for this can be traced to a low level of this type of data analysis in the area like South Africa [46,47] which may affect the population coverage results.

The present investigation presented epitopes-based vaccine candidates of 18 overlapping CD8⁺ T-cell epitopes and 19 IFN- γ and IL-4 inducer CD4⁺ T-cell epitopes that can bind to MHC-I and MHC-II respectively as represented in Tables 1 and 2 respectively. These epitope vaccine candidates contained cytotoxic T-lymphocyte that are 100% conserved in both T-cell and B-cell epitopes predicted. This gives uniqueness to these candidate vaccines in addition to their antigenic, non-toxic and immunogenic characteristics (Tables 1 and 2). A similar method has been used in the prediction of suitable vaccine epitopes for various infectious pathogens such as Zika virus glycoprotein, Dengue virus protein, Chikungunya virus protein, and Ebola virus [48–51]. All the predicted vaccine candidates in this study have strong potential to fight against COVID-19 when validated and tested.

Table 5

The multi-epitopes vaccine selected from the region of complete COVID-19 genomes that correspond with the proposed CD4 T-Cell epitopes and B-cell epitopes and the interaction energy with TLR.

Name	Sequence	Binding energy	Rigid	Flexible	Water-flexible
M-sel	ICLLQFAYANRNRFYIIKLIKLIWLL WPVTLACFVLAAYRINWITGGI AIAMACLVGLMWLSYFIASFRLFA	-110.839	-34.88032	-75.34169	-103.75328
M-complete	MADSNQTTVEELKLLQWNLV IGFLFTWICLLQFAYANRNRFY IIKLIKLIWLLWPVTLACFVLAAY RINWITGGIAIAMACLVGLMWLSYFI ASFRLFARTRSMWFSNPETNILLNVPL HGTILTRPLESELVIGAVILRGHLRIAG HHLGRCDIKDLPEITVATSRTLSYYK LGASQRVAGDSGFAAYSRYRIGNYKL NTDHSSSDNIALLVQ	-121.738	59.969969	14.51024	-68.86108
N-sel	DQJGYRRATRRIIRGGDGKMK	-104.797	-0.1492332	-29.95489	-55.4898
orf1ab-sel	TQWSLFFFLYENAFLPFAMGHIAMSAFAM	-94.365	-27.8293501	-53.685868	-89.3723
ORF3a-1a-sel	MDLFMRIFTIGTVTLKQGEIKD ATPSDFVRATAPIQASLPF GWLIVGVALLAVFQSASKII TLKKRWQLALSKGVHFVCN LLLLFVTVY	-88.954	-7.298504	-29.03831	-49.9221
ORF3a-1b-sel	SHLLVAAGLEAPFLY LYALVYFLQSFVRI IMRLWLCWKCRSKNP LLYDANYFLCWHTNC YDYCIPYNSVTSSIVIT SGDGTTSPISEHD	-104.512	-16.69614538	-47.02714	-58.1496
ORF3a-sel	MDLFMRIFTIGTVT LKQGEIKDATPSDFV RATAPIQASLPFGW LIVGVALLAVFQSASKI ITLKKRWQLALSKGVHF VCNLLLLFVTVYSHLLVA AGLEAPFLYLYALVYFLQSFVRI FVRIIMRLWLCWKCRSKNPLL YDANYFLCWHTNCYDYCIPY NSVTSSIVITSGDGTTSPISEHD	-109.959	46.812421	20.50033	-26.4873
S-sel	YQPYRVVLSFELLHAPATVCGPKK	-95.409	-19.41174	-61.47866	-77.1721

The tools used to design these epitopes were carefully selected by their accuracy. The IEDB T-cell epitope prediction servers are widely used and accepted in literature [52,53]. B-cell epitope prediction tools generally have low accuracy [54]. ABCpred has an accuracy of 65.93% [33] which can be considered to be average, but it is still considered to be one of the most effective epitope prediction servers available online [55]. This justifies the need for further analysis of results obtained from T- or B-cell prediction tools, including antigenicity testing and docking studies. The allerTOP server has been identified as the best allergenicity prediction server and has an accuracy of 88.7% [56]. While this is high, it must be acknowledged that there is still room for error. Therefore, the results presented here are predictive and must be further validated in the laboratory.

Cytotoxic T-cell (CD8⁺ T-cell) kills infected cells or secretes antiviral cytokine and as such restricts infection from tissues [57]. A robust immune response against most infectious pathogens such as viruses is elicited by CD8⁺ T-cell epitopes [58,59]. The immune response generated by T-cell is known to be long-lasting compared to B-cell immune response thus necessitating the importance of T-cell in vaccine design [60]. Similarly, CD8⁺ CD4⁺ T-cell responses play a crucial role in antiviral immunity [15]. According to the report of [11], T-cell response is highly immunogenic, dominant and long-lasting against N and S protein of SARS-CoV. A similar study by Grifoni et al [61] showed that T and B-cell epitopes from structural proteins of SARS-CoV-2 were conserved. One of

the T-cell epitopes ‘TLACFVLAAY’ from M-protein that was found in their study was also found in our study. The reason for the difference in other epitopes predicted by the Grifoni et al [61] from those found in this present study could be that the predicted epitopes in this study were found on M and Orf-3a proteins different from S and Orf-1ab proteins reported by the authors.

The study by Enayatkhani et al [7], found many similar T and B-cell epitopes (NLLLLFVTV, LLWPVTLAC, WPVTLACFV, ITGGIAIAM, IAIAMACLV) from N, M and Orf-3a proteins as found in our present study. Our study further presented other epitopes besides the similar epitopes that passed the screening test for immunogenicity and cytokines-inducing ability. Another similar study that was done by Bhattacharya et al [16] predicted T and B-cell epitopes from the Spike protein of the virus. However, the epitopes selected for multi-epitope vaccine construction differs from this present study. Although few of our predicted epitopes were found on the Spike (S) protein, the differences in the predicted epitopes on S-protein from those predicted by Bhattacharya et al could be due to differences in screening criteria. The selected epitopes from the present study were subjected to allergenicity test and cytokines-inducing ability which make it different from theirs that were subjected to only the antigenicity.

A recent study on samples from 20 convalescing COVID-19 patients showed that helper T-cell (CD4⁺ T-cell) induced a robust immune response against S, M and N protein [62]. The study also reported immunogenic epitopes of S, M and N proteins to induce

CD8⁺ T-cells though this was not specific to only recovered patients but highlights the promising potentials of epitopes located on these structural proteins of SARS-CoV-2. Studies on antibody screening among COVID-19 patient have also confirmed antibodies detection such as IgA, IgG and IgM against S and N- proteins of the SARS-CoV-2 virus [63,64]. Occasional detection of antibodies against M–proteins has also been reported in COVID-19 patient [65].

According to [66], antibodies detection in SARS-CoV-2 patient shows cross-reactivity with SARS-CoV with IgM and IgA detected 5 days after symptom onset while IgM was detected 14 days following the symptom onset. A similar study by [67] also shows that combined antigens from S and N protein-induced optimal antibody detection with high cross-reactivity of IgA and IgG in both SARS-CoV-2 and SARS-CoV. Based on the studies, predicted epitopes on S, M and N protein from SARS-CoV-2 as shown in this study could serve as a potential vaccine candidate and induce an appreciable level of antibodies when validated experimentally. It is important to note that this study is based on the present data of the virus, considering that new information is always being discovered. If any new mutations are observed, they are not expected to affect this proposed vaccine design given that they do not occur in the studied regions [8].

There was a strong binding affinity of CD8⁺ and CD4⁺ T-cell epitopes generated with both MHC-I and II respectively. Besides the immunogenicity and antigenicity of the 18 CD8⁺ and 19 CD4⁺ T-cell epitopes, six of the CD8⁺ epitopes (RNRFLYIIK, LWLLWPVTL, ATRRIRGGD, ELLHAPATV, RVVVLSFEL, and NLLLFVTV) and four CD4⁺ epitopes (IGYRRATRRIRGGD, RNRFLYIIKLIKLWL, SDFVRATA-TIPIQAS, and VHFVCNLLLFVTVY) demonstrate strong interaction with MHC-I and MHC-II.

TLRs are a group of transmembrane receptors that assist with detection of an invading pathogenic organism [68] by detecting dsRNA produced by viruses during DNA and RNA replication [69]. TLR activation results in a series of steps that lead to the regulation of the expression of cytokines, chemokines and type I IFNs [70], therefore characterization of this interaction is essential. Studies have shown that this interaction has been associated with the immune-activating roles of TLRs [71]. This allows the direct increase of IFN- γ production by antigen-specific CD8⁺ T-cells [72]. CD4⁺ T-cells are required for the induction of cytotoxic T lymphocyte activation to minor histocompatibility and some viruses. Interaction of TLR3 with multi-epitope generated in this study is expected to induce cytokines through TLR pathway that will promote and enhance the immune response against COVID-19 disease. Conclusion

Overall, this study provides potential 18 and 19 T-cell epitopes (each of CD8⁺ T-cell and CD4⁺ T-cell), 8B-cell epitopes and multi-epitope based vaccines that are highly antigenic, immunogenic and non-allergic with good population coverage that can assist in the experimental design of suitable vaccine for COVID-19 (SARS-CoV-2). They need to be validated, synthesized and tested to meet the urgent need for the prevention of COVID-19.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The College postdoctoral fellowship awarded to the first author by the College of Agriculture, Engineering and Science, University of KwaZulu-Natal is gratefully appreciated. The authors also wish to acknowledge Centre for High Performance Computing, South

Africa for the computer programs and facilities that were used in this project.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.vaccine.2021.01.003>.

References

- [1] Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* (London, England) 2020;395:470–3. [https://doi.org/10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9).
- [2] Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol* 2020;92:418–23. <https://doi.org/10.1002/jmv.25681>.
- [3] Igić R. Pharmacist's view of the new corona virus. *Scr Med (Brno)* 2020;51:6–8. <https://doi.org/10.5937/scriptamed51-25852>.
- [4] Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. *J Med Virol* 2020;92:522–8. <https://doi.org/10.1002/jmv.25700>.
- [5] Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis* 2020;91:264–6. <https://doi.org/10.1016/j.ijid.2020.01.009>.
- [6] Lim XY, Ng LY, Tam PJ, Liu XD. Human Coronaviruses: A Review of Virus-Host Interactions. *Dis* 2016;4. <https://doi.org/10.3390/diseases4030026>.
- [7] Enayatkhani M, Hasaniazad M, Faezi S, Guklani H, Davoodian P, Ahmadi N, et al. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *J Biomol Struct Dyn* 2020:1–19. <https://doi.org/10.1080/07391102.2020.1756411>.
- [8] Ahmed FS, Quadeer AA, McKay RM. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* 2020;12. <https://doi.org/10.3390/v12030254>.
- [9] Lin Y, Shen X, Yang RF, Li YX, Ji YY, He YY, et al. Identification of an epitope of SARS-coronavirus nucleocapsid protein. *Cell Res* 2003;13:141–5. <https://doi.org/10.1038/sj.cr.7290158>.
- [10] Wang J, Wen J, Li J, Yin J, Zhu Q, Wang H, et al. Assessment of immunoreactive synthetic peptides from the structural proteins of severe acute respiratory syndrome coronavirus. *Clin Chem* 2003;49:1989–96. <https://doi.org/10.1373/clinchem.2003.023184>.
- [11] Channappanavar R, Fett C, Zhao J, Meyerholz DK, Perlman S. Virus-specific memory CD8 T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. *J Virol* 2014;88:11034–44. <https://doi.org/10.1128/JVI.01505-14>.
- [12] Tang F, Quan Y, Xin Z-T, Wrammert J, Ma M-J, Lv H, et al. Lack of Peripheral Memory B Cell Responses in Recovered Patients with Severe Acute Respiratory Syndrome: A Six-Year Follow-Up Study. *J Immunol* 2011;186:7264 LP – 7268. <https://doi.org/10.4049/jimmunol.0903490>.
- [13] Joshi A, Joshi BC, Mannan MA ul, Kaushik V. Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. *Informatics Med Unlocked* 2020;19:100338. <https://doi.org/10.1016/j.imu.2020.100338>.
- [14] Tilocca B, Soggiu A, Sanguinetti M, Musella V, Britti D, Bonizzi L, et al. Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes Infect* 2020;22:188–94. <https://doi.org/10.1016/j.micinf.2020.04.002>.
- [15] Abdelmageed MI, Abdelmoneim AH, Mustafa MI, Elfadlol NM, Murshed NS, Shantier SW, et al. Design of multi epitope-based peptide vaccine against E protein of human 2019-nCoV: An immunoinformatics approach pp. 04.934232. *BioRxiv* 2020;2020(02). <https://doi.org/10.1101/2020.02.04.934232>.
- [16] Bhattacharya M, Sharma AR, Patra P, Ghosh P, Sharma G, Patra BC, et al. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. *J Med Virol* 2020;92:618–31. <https://doi.org/10.1002/jmv.25736>.
- [17] Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The protein data bank. *Acta Crystallogr Sect D Biol Crystallogr* 2002;58:899–907. <https://doi.org/10.1107/S0907444902003451>.
- [18] Thompson JD, Gibson TJ, Higgins DG. Multiple Sequence Alignment Using ClustalW and ClustalX pp. 3.1-2.3.22. *Curr Protoc Bioinforma* 2003;00(2). <https://doi.org/10.1002/0471250953.bi0203s00>.
- [19] Doytchinova IA, Flower DR. Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* 2007;8:4. <https://doi.org/10.1186/1471-2105-8-4>.
- [20] Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* 2001;305:567–80. <https://doi.org/10.1006/jmbi.2000.4315>.
- [21] Flower DR, Doytchinova I, Zaharieva N, Dimitrov I. Immunogenicity Prediction by Vaxijen: A Ten Year Overview. *J Proteomics Bioinform* 2017;10:298–310. <https://doi.org/10.4172/jpb.1000454>.

- [22] Ilinskaya AN, Dobrovolskaia MA. Understanding the immunogenicity and antigenicity of nanomaterials: Past, present and future. *Toxicol Appl Pharmacol* 2016;299:70–7. <https://doi.org/10.1016/j.taap.2016.01.005>.
- [23] Larsen MV, Lundegaard C, Lambeth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinform* 2007;8:424. <https://doi.org/10.1186/1471-2105-8-424>.
- [24] Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics* 2016;32:511–7. <https://doi.org/10.1093/bioinformatics/btv639>.
- [25] Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinform* 2007;8:238. <https://doi.org/10.1186/1471-2105-8-238>.
- [26] Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput Biol* 2013;9:e1003266.
- [27] Saha CK, Mahbub Hasan M, Saddam Hossain M, Asraf Jaham N, Azad AK. In silico identification and characterization of common epitope-based peptide vaccine for Nipah and Hendra viruses. *Asian Pac J Trop Med* 2017;10:529–38. <https://doi.org/10.1016/j.apit.2017.06.016>.
- [28] Rodr SD, Quiroz-castañeda RE, Aguilar-d H, Vara-pastrana E, Pescador-p D, Amaro-estrada I, et al. Immunoinformatic Analysis to Identify Proteins to Be Used as Potential Targets to Control Bovine Anaplasmosis 2020;2020.
- [29] Adhikari UK, Rahman MM. Overlapping CD8+ and CD4+ T-cell epitopes identification for the progression of epitope-based peptide vaccine from nucleocapsid and glycoprotein of emerging Rift Valley fever virus using immunoinformatics approach. *Infect Genet Evol* 2017;56:75–91. [10.1016/j.meegid.2017.10.022](https://doi.org/10.1016/j.meegid.2017.10.022).
- [30] Bui H-H, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinform* 2007;8:361. <https://doi.org/10.1186/1471-2105-8-361>.
- [31] Dhanda SK, Gupta S, Vir P, Raghava GPS. Prediction of IL4 Inducing Peptides. *Clin Dev Immunol* 2013;2013. <https://doi.org/10.1155/2013/26395263952>.
- [32] Bui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinform* 2006;7:153. <https://doi.org/10.1186/1471-2105-7-153>.
- [33] Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins Struct Funct Bioinforma* 2006;65:40–8. <https://doi.org/10.1002/prot.21078>.
- [34] Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci* 2020;117. <https://doi.org/10.1073/pnas.1914677117>. 1496 LP – 1503.
- [35] Boyington JC, Motyka SA, Schuck P, Brooks AG, Sun PD. Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature* 2000;405:537–43. <https://doi.org/10.1038/35014520>.
- [36] de Vries SJ, Rey J, Schindler CEM, Zacharias M, Tuffery P. The pepATTRACT web server for blind, large-scale peptide-protein docking. *Nucleic Acids Res* 2017;45:W361–4. <https://doi.org/10.1093/nar/gkx335>.
- [37] Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph* 1996;14:33–8. [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [38] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605–12. <https://doi.org/10.1002/jcc.20084>.
- [39] Van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastriis PL, Karaca E, et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol* 2016;428:720–5.
- [40] Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 2003;125:1731–7.
- [41] Van Dijk ADJ, Bonvin AMJJ. Solvated docking: introducing water into the modelling of biomolecular complexes. *Bioinformatics* 2006;22:2340–7.
- [42] Jespersen MC, Mahajan S, Peters B, Nielsen M, Marcatili P. Antibody specific B-cell epitope predictions: leveraging information from antibody-antigen protein complexes. *Front Immunol* 2019;10:298.
- [43] Zobayer N, Hossain ABMA, Rahman MA. A combined view of B-cell epitope features in antigens. *Bioinformatics* 2019;15:530.
- [44] Astuti I, Ysrafil. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. *Diabetes Metab Syndr Clin Res Rev* 2020;14:407–12. <https://doi.org/10.1016/j.dsx.2020.04.020>.
- [45] Misra N, Panda PK, Shah K, Sukla LB, Chaubey P. Population coverage analysis of T-Cell epitopes of Neisseria meningitidis serogroup B from Iron acquisition proteins for vaccine design. *Bioinformatics* 2011;6:255–61. <https://doi.org/10.6026/9732063006255>.
- [46] Tshabalala M, Mellet J, Pepper MS. Human leukocyte antigen diversity: A Southern African perspective. *J Immunol Res* 2015;2015. <https://doi.org/10.1155/2015/746151>.
- [47] Mellet J, Tshabalala M, Agbedare O, Meyer P, Gray CM, Pepper MS. Human leukocyte antigen (HLA) diversity and clinical applications in South Africa. *S Afr Med J* 2019;109:29–34. <https://doi.org/10.7196/SAMJ.2019.v109i8b.13825>.
- [48] Chakraborty S, Chakravorty R, Ahmed M, Rahman A, Waise TMZ, Hassan F, et al. A Computational Approach for Identification of Epitopes in Dengue Virus Envelope Protein: A Step Towards Designing a Universal Dengue Vaccine Targeting Endemic Regions. *Silico Biol* 2010;10:235–46. <https://doi.org/10.3233/ISB-2010-0435>.
- [49] Kori P, Sajjan SS, Madagi SB. In silico prediction of epitopes for Chikungunya viral strains. *J Pharm Investig* 2015;45:579–91. <https://doi.org/10.1007/s40005-015-0205-0>.
- [50] Alam A, Ali S, Ahamad S, Malik MZ, Ishrat R. From ZikV genome to vaccine: in silico approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. *Immunology* 2016;149:386–99. <https://doi.org/10.1111/imm.12656>.
- [51] Dash R, Das R, Junaid M, Akash MFC, Islam A, Hosen SZ. In silico-based vaccine design against Ebola virus glycoprotein. *Adv Appl Bioinform Chem* 2017;10:11–28. <https://doi.org/10.2147/AABC.S115859>.
- [52] Adhikari UK, Tayebi M, Rahman MM. Immunoinformatics Approach for Epitope-Based Peptide Vaccine Design and Active Site Prediction against Polyprotein of Emerging Oropouche Virus. *J Immunol Res* 2018;2018:6718083. <https://doi.org/10.1155/2018/6718083>.
- [53] Elhag M, Saadaldin MM, Abdelmoneim AH, Taha TS, Abdelrahman F, Hassan MA. Immunoinformatics Approach for Designing an Epitope-Based Peptide Vaccine against *Treponema pallidum* Outer Membrane Beta-Barrel Protein. *Immunome Res* 2020;16:1–12.
- [54] Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol* 2017;8:1–16. <https://doi.org/10.3389/fimmu.2017.00278>.
- [55] Wang X, Sun Q, Ye Z, Hua Y, Shao N, Du Y, et al. Computational approach for predicting the conserved B-cell epitopes of hemagglutinin H7 subtype influenza virus. *Exp Ther Med* 2016;12:2439–46. <https://doi.org/10.3892/etm.2016.3636>.
- [56] Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* 2014;30:846–51.
- [57] Garcia KC, Teyton L, Wilson IA. Structural Basis of T Cell Recognition. *Annu Rev Immunol* 1999;17:369–97. <https://doi.org/10.1146/annurev.immunol.17.1.369>.
- [58] Jin X, Newman MJ, De-Rosa S, Cooper C, Thomas E, Keefer M, et al. A novel HIV T helper epitope-based vaccine elicits cytokine-secreting HIV-specific CD4+ T cells in a Phase I clinical trial in HIV-uninfected adults. *Vaccine* 2009;27:7080–6. <https://doi.org/10.1016/j.vaccine.2009.09.060>.
- [59] He L, Cheng Y, Kong L, Azadnia P, Giang E, Kim J, et al. Approaching rational epitope vaccine design for hepatitis C virus with meta-server and multivalent scaffolding. *Sci Rep* 2015;5:12501. <https://doi.org/10.1038/srep12501>.
- [60] Panagiote E, Klenerman P, Lee LN, van der Burg SH, Arens R. Features of Effective T Cell-Inducing Vaccines against Chronic Viral Infections. *Front Immunol* 2018;9:276. <https://doi.org/10.3389/fimmu.2018.00276>.
- [61] Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* 2020;27(671–680). <https://doi.org/10.1016/j.chom.2020.03.002>.
- [62] Shah VK, Fimal P, Alam A, Ganguly D, Chattopadhyay S. Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past. *Front Immunol* 2020;11:1–17. <https://doi.org/10.3389/fimmu.2020.01949>.
- [63] Sun B, Feng Y, Mo X, Zheng P, Wang Q, Li P, et al. Kinetics of SARS-CoV-2 specific IgM and IgG responses in COVID-19 patients. *Emerg Microbes Infect* 2020;9:940–8. <https://doi.org/10.1080/22221751.2020.1762515>.
- [64] Iyer AS, Jones FK, Nodoushani A, Kelly M, Becker M, Slater D, et al. Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients. *Sci Immunol* 2020;5:1–13.
- [65] Shrock E, Fujimura E, Kula T, Timms RT, Lee I-H, Leng Y, et al. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* (80-) 2020.
- [66] Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling Early Humoral Response to Diagnose Novel Coronavirus Disease (COVID-19). *Clin Infect Dis* 2020;71:778–85. <https://doi.org/10.1093/cid/ciaa310>.
- [67] de Assis RR, Jain A, Nakajima R, Jasinskas A, Felgner J, Obiero JM, et al. Analysis of SARS-CoV-2 Antibodies in COVID-19 Convalescent Blood using a Coronavirus Antigen Microarray pp. 15.043364. *BioRxiv* 2020;2020(04). <https://doi.org/10.1101/2020.04.15.043364>.
- [68] Vercammen E, Staal J, Beyaert R. Sensing of viral infection and activation of innate immunity by toll-like receptor 3. *Clin Microbiol Rev* 2008;21:13–25. <https://doi.org/10.1128/CMR.00022-07>.
- [69] de Oliveira VL, Almeida SCP, Soares HR, Crespo A, Marshall-Clarke S, Parkhouse RME. A novel TLR3 inhibitor encoded by African swine fever virus (ASFV). *Arch Virol* 2011;156:597–609. <https://doi.org/10.1007/s00705-010-0894-7>.
- [70] Kawasaki T, Kawai T. Toll-like receptor signaling pathways. *Front Immunol* 2014;5:1–8. <https://doi.org/10.3389/fimmu.2014.00461>.
- [71] Dar AA, Patil RS, Chiplunkar SV. Insights into the relationship between toll like receptors and gamma delta T cell responses. *Front Immunol* 2014;5:1–12. <https://doi.org/10.3389/fimmu.2014.00366>.
- [72] Tabiasco J, Devèvre E, Rufer N, Salaun B, Cerottini J-C, Speiser D, et al. Human Effector CD8 + T Lymphocytes Express TLR3 as a Functional Coreceptor. *J Immunol* 2006;177:8708–13. <https://doi.org/10.4049/jimmunol.177.12.8708>.