



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.



Bioinformatics-based SARS-CoV-2 epitopes design and the impact of spike protein mutants on epitope humoral immunities

Qi Sun^{a,b,1}, Zhuanqing Huang^{a,b,1}, Sen Yang^{b,c}, Yuanyuan Li^{a,b}, Yue Ma^{a,b}, Fei Yang^{a,b}, Ying Zhang^a, Fenghua Xu^{a,*}

^a Pharmaceutical Sciences Research Division, Department of Pharmacy, Medical Supplies Centre of PLA General Hospital, Beijing, China

^b Medical School of Chinese PLA, Beijing, China

^c Chinese People's Armed Police Force Hospital of Beijing, Beijing, China

ARTICLE INFO

Keywords:

SARS-CoV-2

Epitope

Antigen

Bioinformatics

Immune effect

ABSTRACT

Background: Epitope selection is the key to peptide vaccines development. Bioinformatics tools can efficiently improve the screening of antigenic epitopes and help to choose the right ones.

Objective: To predict, synthesize and testify peptide epitopes at spike protein, assess the effect of mutations on epitope humoral immunity, thus provide clues for the design and development of epitope peptide vaccines against SARS-CoV-2.

Methods: Bioinformatics servers and immunological tools were used to identify the helper T lymphocyte, cytotoxic T lymphocyte, and linear B lymphocyte epitopes on the S protein of SARS-CoV-2. Physicochemical properties of candidate epitopes were analyzed using IEDB, VaxiJen, and AllerTOP online software. Three candidate epitopes were synthesized and their antigenic responses were evaluated by binding antibody detection.

Results: A total of 20 antigenic, non-toxic and non-allergenic candidate epitopes were identified from 1502 epitopes, including 6 helper T-cell epitopes, 13 cytotoxic T-cell epitopes, and 1 linear B cell epitope. After immunization with antigen containing candidate epitopes S₂₀₆₋₂₂₁, S₄₀₃₋₄₂₅, and S₁₁₅₇₋₁₁₇₀ in rabbits, the binding titers of serum antibody to the corresponding peptide, S protein, receptor-binding domain protein were (415044, 2582, 209.3), (852819, 45238, 457767) and (357897, 10528, 13.79), respectively. The binding titers to Omicron S protein were 642, 12,878 and 7750, respectively, showing that N211L, DEL212 and K417N mutations cause the reduction of the antibody binding activity.

Conclusions: Bioinformatic methods are effective in peptide epitopes design. Certain mutations of the Omicron would lead to the loss of antibody affinity to Omicron S protein.

1. Introduction

As the COVID-19 pandemic continues, the evolving mutation of SARS-CoV-2 is ongoing and has created great challenges in blocking the transmission of the virus and brought about a global public health crisis. Currently, the Omicron variant has quickly raised serious concerns globally, and the efficacy of current vaccines based on the original strains deserves further study. Coronavirus genome is composed of 30,000 nucleotides and encodes four major structural proteins: spike protein (S), membrane protein (M), nucleocapsid protein (N), envelope protein (E) (Boopathi et al., 2021). The S protein is a type I transmembrane glycoprotein including 1273 amino acids and can be

hydrolyzed by proteinase to form subunits S1 and S2. The receptor-binding domain (RBD) on the S1 fragment is responsible for interaction with the cellular receptor angiotensin-converting enzyme-2 (ACE-2), and the S2 fragment is in charge of the fusion of virus and host cells. M protein, the most abundant structural protein of coronaviruses consisting of 222 amino acids, is generally regarded as one of the most conserved candidate antigens. N protein is a structural protein with 419 highly conserved amino acid sequence. It performs many functions, including nucleocapsid formation, signal transduction, RNA replication, and transcription of mRNAs (Mcbride et al., 2014). Coronavirus E proteins, composed of 76–109 amino acids, have channel activity (Zhang et al., 2020). With limited immunogenicity, E proteins cannot be used as

* Corresponding author at: Pharmaceutical Sciences Research Division, PLA General Hospital, 28 Fu Xing Road, Beijing 100853, China.

E-mail address: xufh@301hospital.com.cn (F. Xu).

¹ These authors contributed equally to this work.

immunogens.

At present, the vaccine remains the most economical and effective method to prevent COVID-19 infection. There are different technical platforms for COVID-19 vaccine development. Inactivated virus vaccines (Sinopharm (Al Kaabi et al., 2021)); Sinovac (Zhang et al., 2021), mRNA-based vaccines (Pfizer (Sahin et al., 2020)); Moderna (Corbett et al., 2020), viral vector vaccines (Cansino (Zhu et al., 2020), Johnson (Jurazek et al., 2020), and recombinant protein subunit vaccines (Novavax (Keech et al., 2020)) against SARS-CoV-2 have been approved in succession for emergency clinical use and are being rolled out worldwide. S protein is the antigen of choice for all the vaccines approved and most of the vaccines under research by now. It could stimulate the immune response of both B and T lymphocytes and induce neutralizing antibodies. Even so, B and T cells usually recognize small epitope regions of antigens, which made peptide vaccine possible.

Peptide vaccines are made up of small peptide segments originated from pathogen protein. The peptide segments can be fully chemically synthesized. Compared to vaccines from other technical platforms, peptide vaccines are more suitable for virus variation and can be produced more rapidly, efficiently and less costly. The immunization effect can be enhanced by a combination of multiple peptides from different epitopes and even different viral strains.

It is crucial for individual antigens to effectively stimulate the protective immune response in the development of vaccines. A significant challenge in peptide vaccines development is to screen and design immunogens of high efficiency because short peptides have small molecular weight and usually are weak in arousing immunogenic effect. Meanwhile, the harmful immune responses should be balanced. Thus, in this study, we have screened the B cell and T cell epitopes on the S protein of SARS-CoV-2 for further peptide vaccines development. The allergenicity and toxicity of the epitopes were assessed at the same time.

Several studies have reported the prediction of the epitopes of SARS-CoV-2 (Grifoni et al., 2020; Kiyotani et al., 2020; Bhattacharya et al., 2020; Safavi et al., 2020). The results of different predictions were not quite the same because different prediction tools employed different

algorithms. In addition, since the predictions were based on only part of the epitope features, such as the amino acid structure, the surface area, spatial distribution and intermolecular contact, etc, each prediction has its own limitation. Besides, most predictions were not certificated by experimental data. In this study, we integrated the main bioinformatics servers and immunoinformatic tools popular used to improve the accuracy of the calculation, and evaluate the antigenic responses of candidate epitopes against original and Omicron S protein. The schematic procedure of this research was shown in Fig. 1.

2. Method

2.1. Protein sequence and alignment of mutant strains

Upon discovery and isolation of the first strain of the novel coronavirus, China shared the viral sequence with World Health Organization and registered the gene sequence of the original strain on the National Center for Biotechnology Information (NCBI) as GenBank-MN908947.3 (<https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3>). In pace with the prevalence of COVID-19, mutations continue to occur and new coronavirus variant strains have been detected consecutively. The main variants include B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta), B.1.429 (Epsilon) and B.1.1.529 (Omicron).

Outbreak.info (<https://outbreak.info/>) is a standardized, searchable platform for investigating and analyzing SARS-CoV-2 and COVID-19 data from Scripps Institute's Center for viral Systems Biology. The website provides daily monitoring reports on pedigree and mutation of the virus, including data from state, county, and country, based on >2.6 million genomes compiled by the GISAID Initiative. We compared sequence mutations via the Outbreak.info database, which was used to compare the S protein sequence mutation sites of six mutant strains, with the mutation rate being set at 10%. 1,115,216 sequences of the B.1.1.7 strain, 34,787 sequences of the B.1.351 strain, 65,649 sequences of P.1 strain, 46,074 sequences of the B.1.429 strain, 142,021 sequences of the B.1.617.2 strain and 2134 sequences of the B.1.1.529 strain were

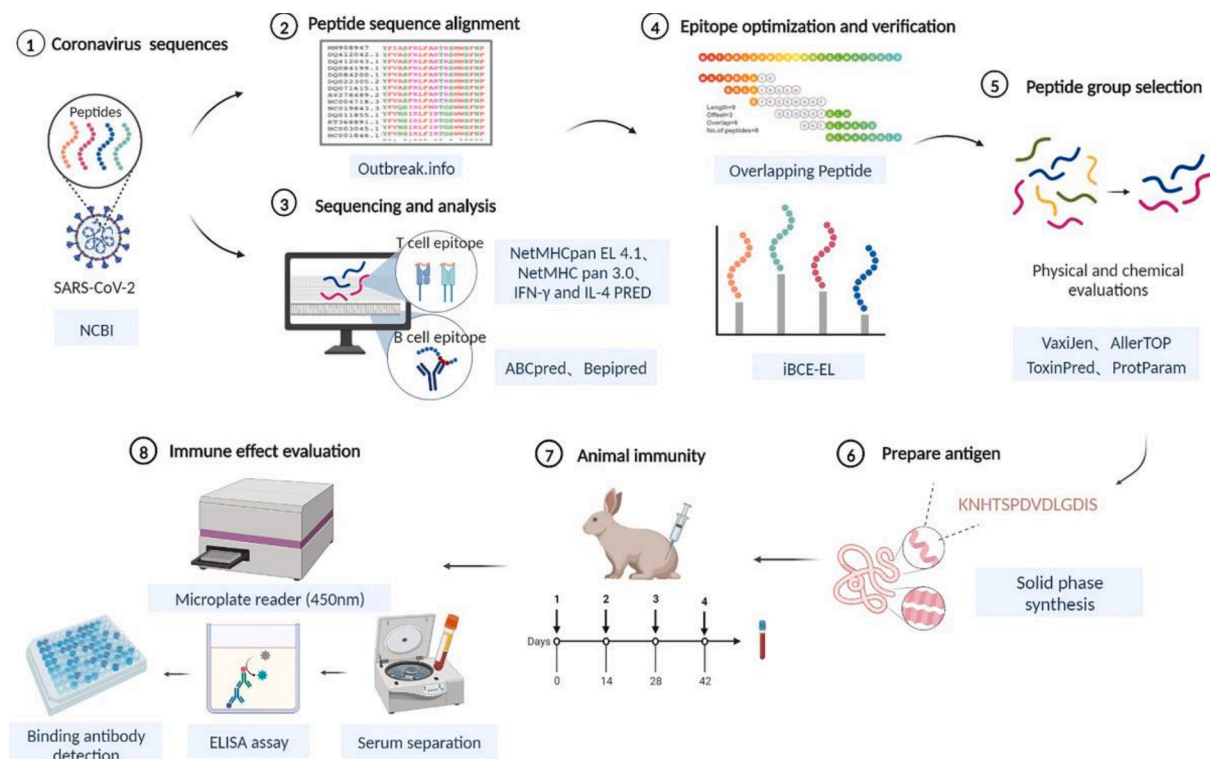


Fig. 1. Research Procedure.

analyzed.

2.2. Prediction of T cell epitopes

Twelve human leukocyte antigen (HLA) alleles, including 5 HLA-A alleles, 4 HLA-B alleles and 3 HLA-C alleles, each of which were reported to be present in the Chinese population at a frequency of >12 %, were selected to predict HLA-class I epitopes (He et al., 2018). Similarly, 8 HLA alleles with >12 % existing frequencies, containing 5 HLA-DRB1 and 3 HLA-DQB1 haplotypes, were chosen for HLA-class II epitope prediction. The genotypic frequency of the HLA allele was shown in Fig. 2.

2.2.1. Prediction and assessment of CTL epitopes

Bioinformatics tools NetMHCpan 4.1 EL (<http://tools.iedb.org/mhci/>) (Reynisson et al., 2020) and NetMHCpan 3.0 (<https://tools.iedb.org/main/tcell/>) (Nielsen and Andreatta, 2016) were used to assess the binding affinity of peptide segments with 8 to 14 amino acids in the S protein sequence to HLA-I molecules and thus to identify potential cytotoxic T lymphocyte (CTL) epitopes. Accordingly, based on the calculated affinity between the peptide and MHC in the “antigen peptide-MHC” complex, the top 1 % epitopes with IC₅₀ values less than 500 nM were picked up as potential epitope candidates for peptide vaccines.

IEDB database were employed to assess the immunogenicity of the candidate epitopes by figuring out their MHC-I Immunogenicity score (<http://tools.iedb.org/immunogenicity/>) (Calis et al., 2013). It is generally believed that the higher the epitope score, the greater the likelihood of inducing antibody response to it. In order to reduce the false-positive rate, we set 0.2 as the standard threshold for screening (Calis et al., 2013). Therefore, in this study, only epitopes with scores >0.2 in the IEDB calculation were picked out and continued to the follow-up investigations.

2.2.2. Prediction and assessment of HTL epitopes

The binding capacity of the 12–18 amino acid epitopes, originated from the novel coronavirus S protein, to HLA-II molecules was calculated with the use of bioinformatics tool NetMHCIIpan 4.0 (<http://tools.iedb.org/mhcii/>) (Reynisson et al., 2020); and those epitopes with top 0.2 % percentile (strong binders) were assessed the potential as candidate epitopes for peptide vaccines.

Helper T lymphocyte (HTL) cells polarize into diverse T-cell populations like Th1, Th2, Th17, or iTregs (Nielsen and Andreatta, 2016). Th1 cells release interferon-gamma (IFN- γ) which helps macrophages identify and eliminate viruses within cells. In Th2 cell subsets, interleukin-4 (IL-4) is the main cytokine secreted, which appears to

promote the proliferation and differentiation of antigen-presenting cells. Therefore, we take advantage of servers IFNepitope (<https://webs.iiitd.edu.in/raghava/ifnepitope/index.php>) (Dhanda et al., 2013b) and IL4pred (<https://webs.iiitd.edu.in/raghava/il4pred/index.php>) (Dhanda et al., 2013a) to infer the latency of the epitopes to induce interferon- γ (IFN- γ) and interleukin-4 (IL-4) respectively, with default parameters.

2.3. B cell epitope prediction

B-cell epitopes can be classified into two types: linear epitopes and conformational epitopes (El-Manzalawy et al., 2008). Linear B lymphocyte (LBL) epitopes are composed of sequential amino acids that participate in antibody binding and their interaction is based on the primary structure of the epitope (Nevagi et al., 2018). Conformational epitopes are composed of amino acids that are far apart in primary sequence but are in close proximity in the folded structures. In this study, we primarily concentrated on the identification of LBL epitopes.

ABCpred (https://webs.iiitd.edu.in/raghava/abcpred/ABC_submission.html) (Saha and Raghava, 2006) is an artificial neural network tool that predicts LBL epitopes with the prediction accuracy being about 66 %. In computing, the threshold for an active LBL epitope was set at 0.85, resulting in the sensitivity of epitope predictions from 95.5 % to 99.5 %. LBL epitopes usually consist of 5 ~ 30 amino acids, while in this study were set to 18 amino acids (Saha and Raghava, 2006). As to Bepipred server (<http://tools.iedb.org/bcell/>) (Jespersen et al., 2017), to predict LBL epitopes, the random forest regression algorithm was trained by a fivefold cross-validation method. In our prediction, the probability threshold is set at 0.35. Peptide series of probability >0.35 were considered as candidate epitopes, except for those less than 5-amino acid-long (Xu et al., 2020). Epitope candidates reckoned by both of the two prediction methods were more likely the effective B cell epitopes.

The overlapping peptide library tools (https://www.genscript.com.cn/overlapping_library.html) (Gershoni et al., 2007) aim to design short peptide sequences based on the target protein or long peptide. This library provides information on protein bioactivity, immune response specificity and antibody binding activity, turning out an ideal tool for screening linear epitopes. Peptide design in overlapping peptide library primarily depends on two parameters: the peptide chain length and offset number. Choosing the peptide length and appropriate step size would reduce experimental cost and increase data value. In our study, the peptide length was set to 14 and the number of amino acid offset was set to 5 to optimize the LBL candidate epitopes obtained above. The LBL epitopes obtained above were examined by IBCE-EL server (<https://theleelab.org/iBCE-EL/>) (Manavalan et al., 2018), only active epitopes were selected for further analysis.

2.4. Physiological and physicochemical properties analysis of candidate epitopes

Epitope antigenicity can be estimated by VaxiJen sever (<http://www.ddg-pharmfac.net/vaxijjen/VaxiJen/VaxiJen.html>) (Doytchinova and Flower, 2007) according to auto cross covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties. The prediction accuracy of VaxiJen sever ranged from 70 % to 89 % according to different organisms. In our exploration, the antigenicity of the candidate epitopes were calculated on virus model with the default threshold being set at 0.4 (Doytchinova and Flower, 2007; Doytchinova and Flower, 2007; Irini, 2008). Then AllerTOP (<https://www.ddg-pharmfac.net/AllerTOP/>) (Dimitrov et al., 2014) and ToxinPred (<https://webs.iiitd.edu.in/raghava/toxinpred/index.html>) (Gupta et al., 2013) servers were utilized to define whether the candidate epitopes were allergenic or toxic. ProtParam tool in ExPASy sever (<https://web.expasy.org/protparam/>) was employed to analyze the molecular weight, theoretical electronic point(pI), aliphatic index

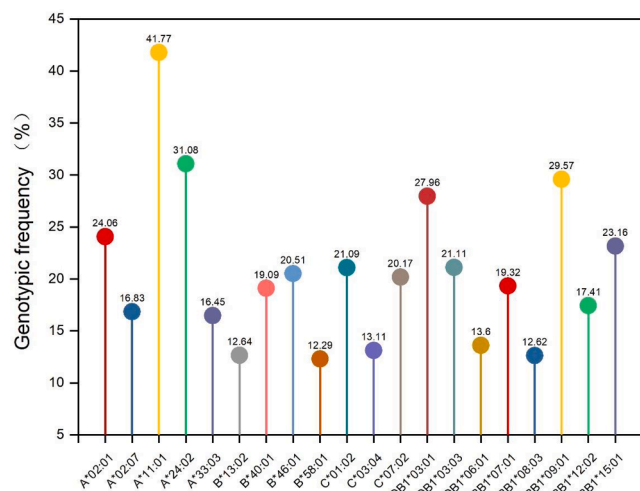


Fig. 2. Common HLA-allele distribution frequencies in the Chinese population.

(AI), instability index (II), total hydrophilic average, and the stable epitopes were sketched based on their instability index and half-time (Wilkins et al., 1999).

2.5. Prediction of population coverage

A given epitope will elicit a response only in individuals that express an MHC molecule capable of binding that particular epitope (Jain et al., 2021). The frequencies of specific HLA alleles vary dramatically between ethnic groups. A web based tool, IEDB population coverage (https://tools.iedb.org/tools/population/iedb_input) (Bui et al., 2006), has been used for population coverage analysis. COVID-19 has affected all over the world, in this study, Europe, East Asia, North America, China, United States, India and the whole world have been taken as target populations. By inputting the epitopes and the corresponding MHC I and II alleles, the coverage of our predicted T cell epitopes was analyzed.

2.6. Evaluation of humoral immune response of the predicted epitopes

2.6.1. Peptide synthesis and purification

Epitope peptides were synthesized by using a three-channel peptide automatic synthesizer (CS360). The peptide was synthesized by solid-phase peptide synthesis (SPPS) as stepwise addition of amino acid from carboxyl end to amino end. After the desired amino acid chain was built, the peptide was cleaved from the resin by to get a crude product. Then the crude product was separated and purified using high-performance liquid chromatography (HPLC) with C18 reverse phase chromatographic column packing. This peptide epitope was then coupled with the carrier protein to obtain peptide-hemocyanin conjugate (immunogen). Peptide epitopes synthesized here include S₂₀₆₋₂₂₁ (KHTPINLVRDLPQGFS), S₄₀₃₋₄₂₅ (RGDVRQIAPGQTGKIADYNYKL) and S₁₁₅₇₋₁₁₇₀ (KNHTSPDVLGDIS).

2.6.2. Immunogenicity evaluation

The immunization was carried out with New Zealand white rabbits at a four-dose immunization program. The rabbit was initially intradermally injected with the epitope immunogen plus Freund's complete adjuvant at multiple sites on the back. The boosters were performed with the immunogen plus Freund's incomplete adjuvant at a 14-day interval. Serum was collected from immunized animals before immunization and 10 days after each of the three boost injections. Peptide epitope-specific, spike-specific and RBD-specific antibody responses were evaluated by enzyme-linked immunosorbent assays (ELISA).

Briefly, 96-well plates were coated with 2 µg/ml peptide epitope, and 1 µg/ml recombinant original SARS-CoV-2 S protein (Sino Biological, Cat:40589-V08H9), the variant Omicron S protein (Sino Biological, Cat:40589-V08H26) or RBD protein (Sino Biological, Cat:40592-V08H) in 0.01 M carbonate-bicarbonate buffer solution and incubated overnight at 4 °C. Plates were then washed three times with PBS-0.05 % Tween 20 (PBST) and blocked for 2 h with block buffer at 37 °C. After block, serial 4-fold dilutions of inactivated serum, starting at 1:500 (rabbit), were added to wells and the plates were incubated for 1 h at room temperature. After three washes with wash buffer, the plates were added with Horseradish peroxidase (HRP) conjugated goat anti-rabbit IgG (1:20,000, ZSGB-BIO) and incubated for 1 h at 37 °C. The plates were then washed three times with wash buffer and added with TMB Chromogen Solution A 50 µl and then TMB Chromogen Solution B 50 µl to each well followed by 15 min of incubation at 37 °C. The reaction was stopped with 100 µl/well 2 M sulfuric acid and the absorbance at 450 nm (A450nm) was measured by the ELISA plate reader (Spectra Max M2). The absorbance values were plotted as a function of the reciprocal dilution of serum samples. Reciprocal plasma dilutions corresponding to 50 % maximal binding (i.e EC50) were computed using the Prism software (GraphPad Software v. 8.02).

3. Results

3.1. Variant strain alignments

The Outbreak.info database was used to compare sequence mutation sites, we analyzed 1,115,216 sequences of the B.1.1.7 strain, 34,787 sequences of the B.1.351 strain, 65,649 sequences of P.1 strain, 46,074 sequences of the B.1.429 strain, 14,2021 sequences of the B.1.617.2 strain and 2134 sequences of the B.1.1. 529 strain. The result is shown in Fig. 3.

Mutations such as L452R, E484K, N501Y, D614G, and P681R/H are widely considered to enhance the ability of the virus to enhance transmissibility, escape immune protection, and aggravate disease (Deng et al., 2021; Starr et al., 2020; Lopez Bernal et al., 2021). The D614G mutation in particular has attracted attention since it has quickly become the dominant strain of SARS-CoV-2 circulating worldwide (Korber et al., 2020). D614 is a surface residue in the vicinity of the furin cleavage site, like P681R/H mutation, may increase the cleavage efficiency of the spike protein S1-S2, facilitate the fusion of the virus and host cells, and thus promote the virus entering into the host cells, which consequently makes the virus more infectious. In United Kingdom, Brazil, South Africa and Botswana, fast-spreading strains share a mutation called N501Y at the RBD protein. Studies in cell and animal model systems have shown that the mutation of N501Y may enable novel coronaviruses to bind more tightly to the ACE2 receptor and thus enhance their infectious properties (Hongjing et al., 2020). Both L452R and E484K occur at the receptor-binding motif (RBM) on RBD. The latest variant Omicron (B.1.1.529) has 15 mutations on the RBD, much more than those on the other variants, and might have enhanced transmissibility and immune evasion.

3.2. Prediction of T cell epitopes

CTL could directly kill virus-infected cells and damage the infected cells via releasing cytotoxic proteins (Kalita et al., 2020). Calculation with NetMHCpan 4.1 EL and NetMHCpan3.0 resulted in 722 and 612 peptides with the top 1 % of "antigen peptide-MHC" affinity score and with IC50 below 500 nM, respectively. After eliminating redundant and nested peptides, we got 337 peptide segments. After further immunogenicity test with MHC-I Immunogenicity server, 45 candidate epitopes left (Fig. 4). The length of these CTL epitopes ranged from 8 to 13 amino acids.

Helper T cells are further divided into different subtypes, secreting different cytokines and chemokines via different transcription factors, thus enhancing humoral immunity as well as promoting cellular immunity. As a result of calculation by the NetMHCIIpan 4.0 EL tool, 127 HTL predicted epitopes are located primarily at 11 regions on the S protein. The epitopes regions vary between 12 and 18 amino acids (Table 1). In those 127 predicted peptide epitopes, 77 epitopes could stimulate IFN-γ activity, among which 59 epitopes could simultaneously induce IL-4 release, according to IFNepitope and IL4pred computation. (Fig. 4).

3.3. B cell epitope prediction and optimization

Based on ABCPred, we got 15 linear B cell epitopes of the new coronavirus S protein, while Bepipred gave 26 epitopes (Fig. 4). Some of the epitopes from the two predictions shared common segments (5 consensus segments in total, listed in Table S1). S₁₁₅₇₋₁₁₇₆ epitope has the longest common segment (9 amino acids) in both predictions. It has been reported S₁₁₅₇₋₁₁₇₃ could bind to the serum samples of SARS patients, (He et al., 2004) indicating its activity as an epitope. Therefore, we optimized the S₁₁₅₇₋₁₁₇₆ segment by using overlapping peptide library design software. On peptide length being 14 and offset being 5, two peptide epitopes were redesigned as S₁₁₅₇₋₁₁₇₀(KNHTSPDVLGDIS) and S₁₁₆₆₋₁₁₇₆(LGDISGINASV). Similarly, the other four consensus

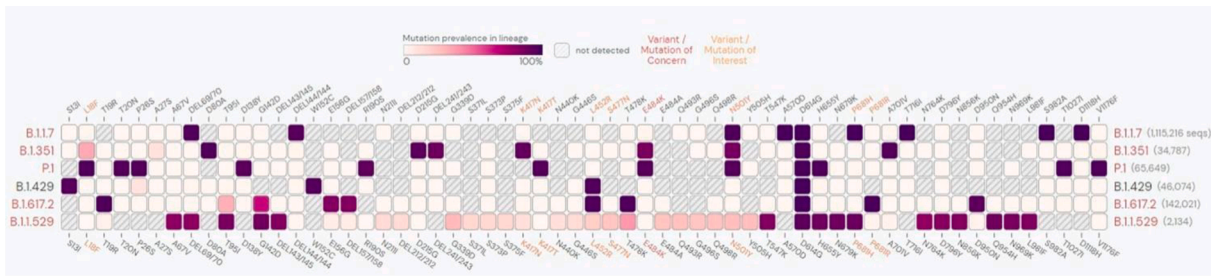


Fig. 3. Sequence alignment of variant strains.

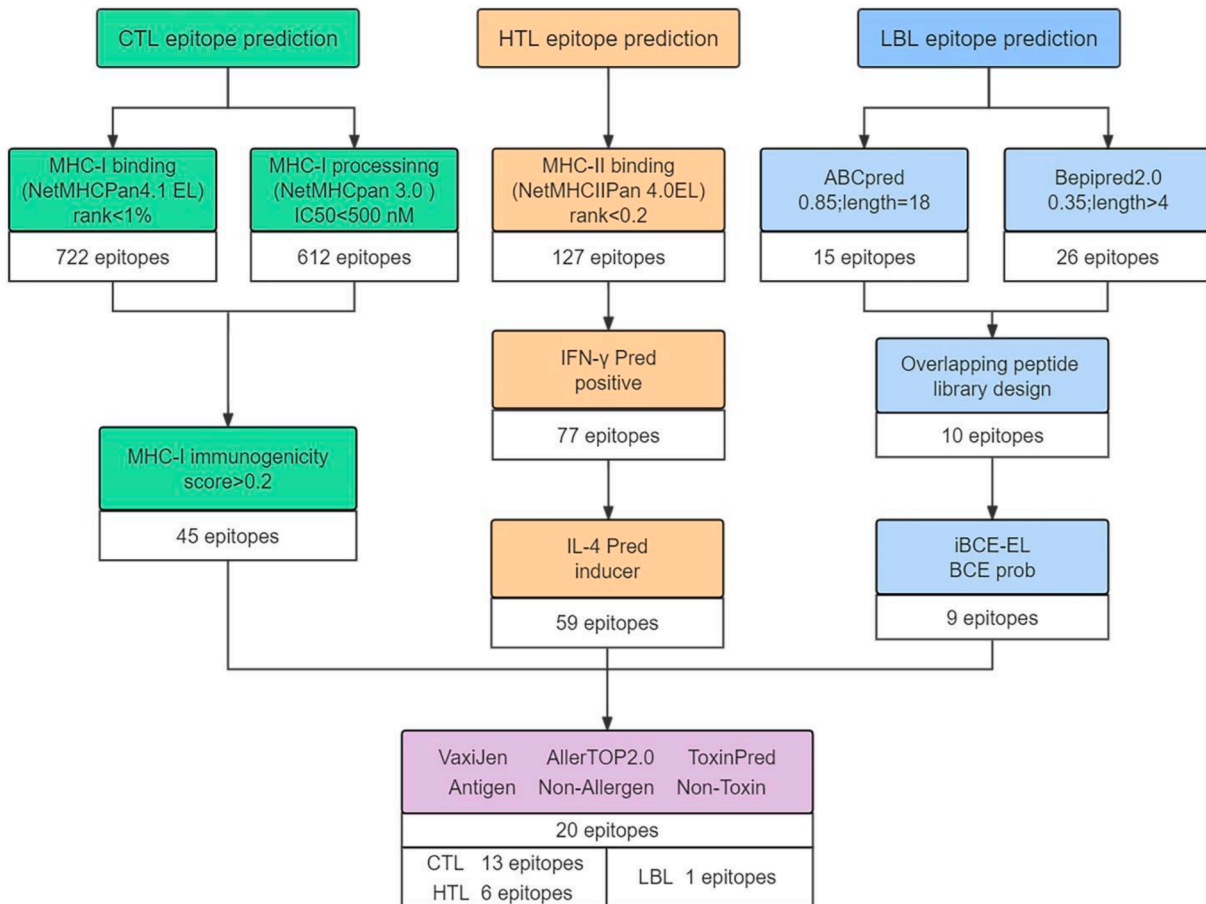


Fig. 4. Illustration of the epitope prediction and screening process.

Table 1
The HTL epitopes sequences, amino acid position of S protein by cluster analysis (NetMHCIIpanv4.0 EL).

Cluster Number	Epitope Number	Alignment	Amino-acid Position
1.1	Consensus	NDGVYFASTEKSN	87–99
2.1	Consensus	KHTPINLVRDLPPQGS	206–221
3.1	Consensus	FTVEKGIYQTSNFRVQPTE	306–325
4.1	Consensus	DDFTGCVIAWNSNLDKSVG	427–446
5.1	Consensus	IPTNFTISVTTEILPV	714–729
6.1	Consensus	PLLTDEMQYTSALLAGTTS	863–884
7.1	Consensus	QTYVTQLIRAAEIRASANLAATKM	1005–1029
8.1	Consensus	ISGINASVVNIQKEIDRLN	1169–1187
9.1	Singleton	DKVFRSSVLHSTQD	40–53
10.1	Singleton	SNVTWFHAIHVS	60–71
11.1	Singleton	ESIVRFPNITNL	324–335

segments were also optimized, and 10 peptide fragments in total were obtained. Only S₁₁₅₇₋₁₁₇₀(KNHTSPDVLGDIS) is a probable antigen without allergenicity and toxicity (Table S2).

3.4. Assessment of epitope antigenicity, allergenicity and toxicity

By VaxiJen, AllerTOP (Dimitrov et al., 2014) and ToxinPred (Gupta et al., 2013) server, a panel of 13 CTL epitopes, 6 HTL epitopes, and 1 linear B cell epitopes were finally obtained with attributes of being relatively highly antigenic, non-allergic and non-toxic (Table 2).

3.5. Physiological and physicochemical properties of candidate epitopes

According to ExPASy ProtParam, the peptide S₁₁₅₇₋₁₁₇₀(KNHTSPDVLGDIS) has a molecular weight (MW) of 1497. Its theoretical isoelectric point value is 4.41 and was considered to be acidic. Its in vitro half-life was predicted to be 1.3 h in mammalian

Table 2
Screening of candidate epitopes and their physicochemical properties.

Category	Peptide Sequence	Start	End	Length	VaxiJen Score	AllerTOP	ToxinPred	Molecular weight	Theoretical pI	The estimated half-life	The instability index (II)	Aliphatic index	Grand average of hydrophobicity (GRAVY)
CTL	DLPIGINITR	228	237	10	1.8171	Non-Allergen	Non-Toxin	1111.31	5.84	1.1 h (mammalian reticulocytes, in vitro)0.3 min (yeast, in vivo). > 10 h (Escherichia coli, in vivo).	63.71 unstable	156	0.31
CTL	VTWFHAIHV	62	70	9	0.5426	Non-Allergen	Non-Toxin	1109.3	6.89	100 h (mammalian reticulocytes, in vitro). > 20 h (yeast, in vivo). > 10 h (Escherichia coli, in vivo).	-3.53 stable	118.89	1.056
CTL	EQYIKWPWYI	1207	1216	10	1.1122	Non-Allergen	Non-Toxin	1425.65	6.1	1 h (mammalian reticulocytes, in vitro)0.30 min (yeast, in vivo). > 10 h (Escherichia coli, in vivo).	9.35 stable	78	-0.79
CTL	KVTLADAGFIK	825	835	11	0.8594	Non-Allergen	Non-Toxin	1162.39	8.59	1.3 h (mammalian reticulocytes, in vitro)0.3 min (yeast, in vivo)0.3 min (Escherichia coli, in vivo).	-21.78 stable	115.45	0.591
CTL	VTLADAGFIK	826	835	10	0.8702	Non-Allergen	Non-Toxin	1034.22	5.81	100 h (mammalian reticulocytes, in vitro). > 20 h (yeast, in vivo). > 10 h (Escherichia coli, in vivo)	-16.47 stable	127	1.04
CTL	IAGLIAIVM	1221	1229	9	0.4716	Non-Allergen	Non-Toxin	900.19	5.52	20 h (mammalian reticulocytes, in vitro)0.30 min (yeast, in vivo). > 10 h (Escherichia coli, in vivo).	-0.54 stable	227.78	2.956
CTL	TLADAGFIK	827	835	9	0.5781	Non-Allergen	Non-Toxin	935.09	5.5	7.2 h (mammalian reticulocytes, in vitro). > 20 h (yeast, in vivo). > 10 h (Escherichia coli, in vivo).	-9.98 stable	108.89	0.689
CTL	FYEPQITDNTF	1109	1121	13	0.4578	Non-Allergen	Non-Toxin	1588.73	3.67	1.1 h (mammalian reticulocytes, in vitro)0.3 min (yeast, in vivo)0.2 min (Escherichia coli, in vivo).	61.1 unstable	60	-0.338
CTL	FFSNVTWFH	58	66	9	0.5951	Non-Allergen	Non-Toxin	1184.32	6.74	1.1 h (mammalian reticulocytes, in vitro)0.3 min (yeast, in vivo)0.2 min (Escherichia coli, in vivo).	-17.24 stable	32.22	0.389
CTL	YEYQIKWPWYI	1206	1216	11	0.9881	Non-Allergen	Non-Toxin	1588.83	6	2.8 h (mammalian reticulocytes, in vitro). 10 min (yeast, in vivo). 2 min (Escherichia coli, in vivo).	2.55 stable	70.91	-0.836
CTL	ADQLTPTWRV	626	635	10	0.5883	Non-Allergen	Non-Toxin	1186.33	5.88	4.4 h (mammalian reticulocytes, in vitro). > 20 h (yeast, in vivo). > 10 h (Escherichia coli, in vivo).	-14.52 stable	78	-0.56
CTL	KYEYQIKWPWYI	1205	1216	12	1.1033	Non-Allergen	Non-Toxin	1717	8.38	1.3 h (mammalian reticulocytes, in vitro)0.3 min (yeast, in vivo)0.3 min (Escherichia coli, in vivo)	3.18 stable	65	-1.092
CTL	FFSNVTWF	57	65	8	0.4403	Non-Allergen	Non-Toxin	1047.18	5.52	1.1 h (mammalian reticulocytes, in vitro)0.3 min (yeast, in vivo)0.2 min (Escherichia coli, in vivo).	-20.65 stable	36.25	0.838

(continued on next page)

Table 2 (continued)

Category	Peptide Sequence	Start	End	Length	VaxiJen Score	AllerTOP	ToxinPred	Molecular weight	Theoretical pI	The estimated half-life	The instability index (II)	Aliphatic index	Grand average of hydropathicity (GRAVY)
HTL	PTNFTISVTTEILPV	715	729	15	1.1349	Non-Allergen	Non-Toxin	1631.89	4	>20 h (mammalian reticulocytes, in vitro). > 20 h (yeast, in vivo).	47.83 unstable	116.67	0.68
HTL	IPTNFTISVTTEILPV	714	729	16	0.9295	Non-Allergen	Non-Toxin	1745.05	4	20 h (mammalian reticulocytes, in vitro)0.30 min (yeast, in vivo). > 10 h (Escherichia coli, in vivo)	43.67 unstable	133.75	0.919
HTL	EKGiyQTSNFRVQPTE	309	324	16	0.8559	Non-Allergen	Non-Toxin	1897.07	6.24	1 h (mammalian reticulocytes, in vitro)0.30 min (yeast, in vivo). > 10 h (Escherichia coli, in vivo).	13.44 stable	42.5	-1.244
HTL	VEKGIYQTSNFRVQPTE	308	324	17	0.8296	Non-Allergen	Non-Toxin	1996.21	6.11	100 h (mammalian reticulocytes, in vitro). > 20 h (yeast, in vivo). > 10 h (Escherichia coli, in vivo)	13.24 stable	57.06	-0.924
HTL	VEKGIYQTSNFRVQPTEs	308	325	18	0.7311	Non-Allergen	Non-Toxin	2083.29	6.11	100 h (mammalian reticulocytes, in vitro). > 20 h (yeast, in vivo). > 10 h (Escherichia coli, in vivo)	23.76 stable	53.89	-0.917
HTL	RAAEIRASANLAATK	1013	1028	16	0.5709	Non-Allergen	Non-Toxin	1542.76	10.84	1 h (mammalian reticulocytes, in vitro). 2 min (yeast, in vivo). 2 min (Escherichia coli, in vivo)	22.17 stable	92	-0.153
LBL	KNHTSPDVLGDIS	1157	1170	14	1.5175	NON-ALLERGEN	Non-Toxin	1497.58	4.41	1.3 h (mammalian reticulocytes, in vitro)0.3 min (yeast, in vivo)0.3 min (Escherichia coli, in vivo).	19.16 stable	76.43	-0.921

reticulocytes, whereas only about 3 min in yeast and escherichia coli. The instability index (II) of this S₁₁₅₇₋₁₁₇₀ peptide is 19.16, and since a value >40 implies instability, it is considered as stable. Its aliphatic index was estimated 76.63, confirming its thermal stability (Ikai, a., 1980). The average hydrophilicity score of S₁₁₅₇₋₁₁₇₀ is -0.921, meaning that it is soluble in water and can interact with water easily (Ali et al., 2017). Since in vivo hydrophilic residues of a protein are typically located on its surface, whereas hydrophobic residues typically lie within the macromolecule, protein hydrophilic sites are closely related to protein antigen epitopes. The physicochemical properties of the 20 bioinformatics-predicted epitopes were listed in Table 2.

3.6. Worldwide population coverage

Evaluation of the population coverage was conducted in the selected CTL and HTL epitopes and their associated HLA alleles. The calculated CTL and HTL epitopes provide a 94.78 % population coverage in the Chinese population and an 87.02 % population coverage worldwide. The selected epitopes appear to be able to interact with several HLA alleles from different countries, including the United States (87.59 %), India (74.42 %), East Asia (94.99 %), North America (87.54 %) and Europe (86.06 %), which suggest that vaccines based on these epitopes could be effective for most people in the world.

3.7. Evaluation of humoral immune response

S₂₀₆₋₂₂₁ (KHTPINLVRDLPQGFS) is a peptide segment containing 9 helper T cell epitopes predicted by NetMHCIIpanv4.0 EL (Table 1). S₂₀₆₋₂₂₁ peptide segment belongs to the non-RBD domain of the S1 segment of the S protein. A notable D215G mutation and N211L, DEL212 mutations are present in B.1.351 and B.1.1.529 among the major mutant strains. S₄₀₃₋₄₂₅ (RGDVRQIAPGQTGKIADYNYKL) peptide segment in RBD contains both CTL epitopes S₄₀₈₋₄₂₅ and LBL epitopes S₄₀₇₋₄₂₀. A mutation of interest K417N was typical in B.1.351 and B.1.1.529 and mutation of interest K417T was typical in the P.1 virus. S₁₁₅₇₋₁₁₇₀ (NHTSPDVDLGDIS) is a B cell epitope located in the S2 segment and is conservative for different mutant strains. With good antigenicity and physicochemical properties, it is reckoned to be active in producing antibodies, thus has the potential to become candidate epitopes for vaccines.

The seropositivity of all the three epitope-based immunogens could be detected after the second immunization and reached a relatively high level after the third immunization. Elisa detection turned out that S₂₀₆₋₂₂₁, S₄₀₃₋₄₂₅, and S₁₁₅₇₋₁₁₇₀ reached the binding titers of serum antibody to the corresponding peptide, S protein, RBD protein were (415044, 2582, 209.3), (852819, 45238, 457767) and (357897, 10528, 13.79), respectively. The binding titers to Omicron S protein were 642, 12,878

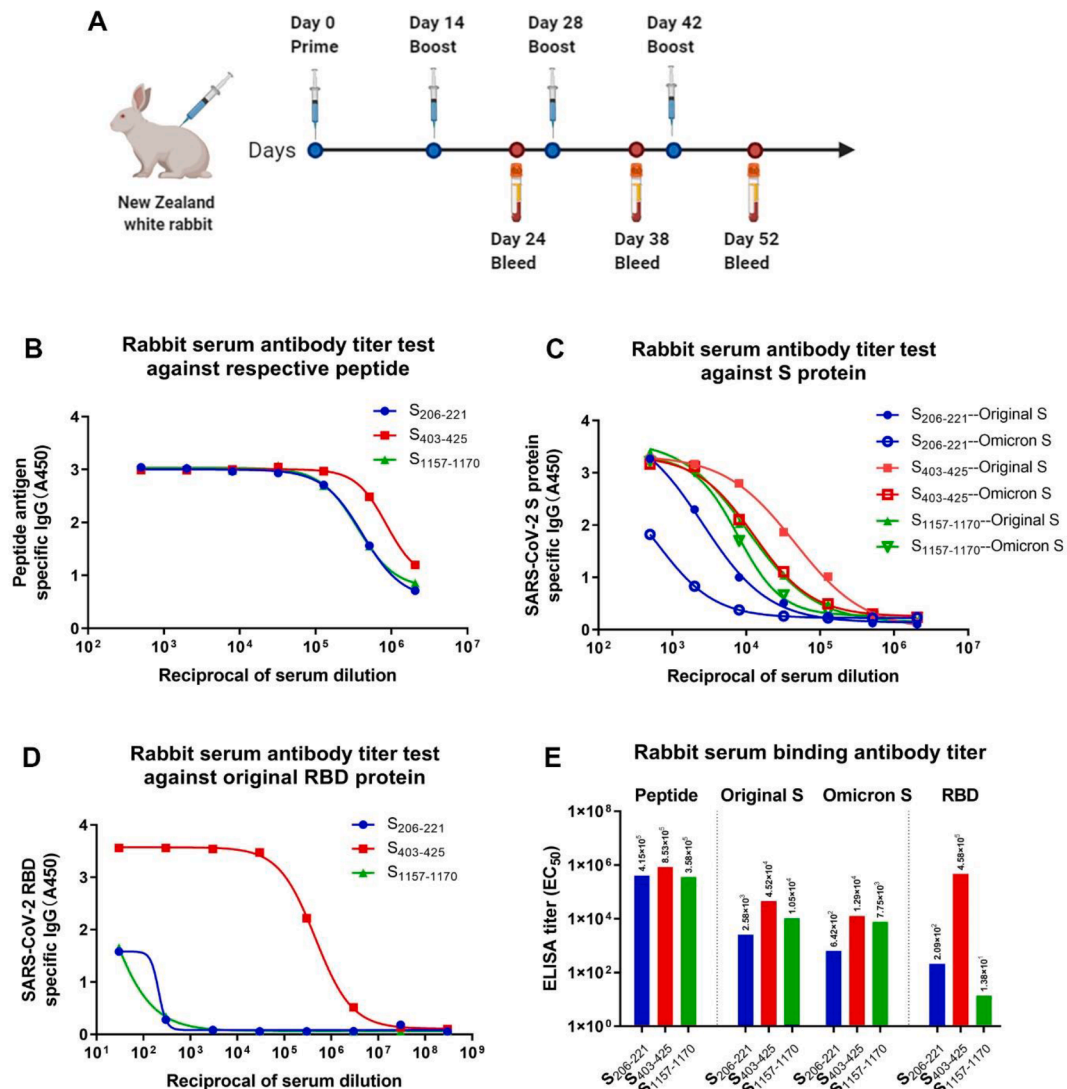


Fig. 5. Identification of serum antibody against peptide, S protein and RBD in rabbits.

and 7750, respectively, showing decreased affinity of S₂₀₆₋₂₂₁ and S₄₀₃₋₄₂₅ compared to the original S protein. The results were shown in Fig. 5.

4. Discussion

Epitopes are structures composed of special chemical groups in antigen molecules which can bind to B cell receptor (BCR) or T cell receptor (TCR), accordingly being called B cell epitopes or T cell epitopes. B cell epitopes identification depends on the structure analyzing of antigen-mono-clonal antibody (MAb) complex which based on the purification of MAb and antigen-MAb, a very complicated and difficult manipulation. T cell epitopes are linear epitopes and present great combinatorial structure diversity. It is laborious and time-consuming to find the right epitopes by experiments. Therefore, immunoinformatics methods has become an indispensable tool for epitope localization and is playing an increasingly important role in epitope finding. It is reported that immunoinformatics methods could improve epitopes finding by 10–20 times while reduce the experimental workload by 95 % (De Groot et al., 2002).

In this paper, we adopted multiple software and calculation tools to improve the accuracy of our prediction. Similar results from different methods were admitted so that the false positive rate were inhibited to relatively low level. For example, in CTL epitope prediction, MHC-I affinity prediction is first performed using the IEDB database that aggregates experimental data on antibody and T cell epitopes. At the same time, the NetMHCpan algorithm was used to evaluate the ability of antigen processing and transport process, and the comprehensive analysis obtained better prediction results than using single data training.

Bioinformatics method was applied to design HTL, CTL, and LBL epitopes against SARS-CoV-2. CTL represents one of several types of cells in the immune system that can kill infected cells directly (Xu et al., 2020). CTL execute cell-killing effect only after certain peptides on the major histocompatibility complex (MHC) molecules are presented to and recognized by them. HTL is also called CD4⁺T lymphocyte. After proteolytic cleavage of viral antigens, antigen-presenting cells such as B cells, macrophages, and dendritic cells present epitopes to HTLs in the epitope-MHC II complex (Couture et al., 2019). B cells are considered as the core component of the adaptive immune system and have the ability to secrete specific antibodies to neutralize invading viruses (Quast and Tarlinton, 2021). By differentiating long-lived plasma cells and memory B lymphocytes, B cells play a crucial role in long-term immunity. Investigation on B cell epitopes would be helpful to understand the pathogenic mechanism of the virus and develop vaccines against SARS-CoV-2.

In this study, we selected two peptides with high MHC-I binding rank (rank > 1%) in the NetMHCpan 4.1 EL and high MHC-I processing IC₅₀ (IC₅₀ > 500 nM) in the NetMHCpan 3.0 predicted process as negative controls: S₁₇₉₋₁₉₁ (LEGKQGNFKNLRE) in the S1 region and S₄₃₆₋₄₄₉ (WNSNNLDSKVGNGY) in the RBD protein. Their MHC-I binding rank were 100 % and 48 %, while their MHC-I processing IC₅₀ were 38245.2 nM and 29639.7 nM, respectively. The experimental results showed that the binding antibody levels to S protein (500 and 500) and RBD protein (0 and 700) were relatively low, which in turn proved the validity of our prediction. Meanwhile, immunization with S₂₀₆₋₂₂₁ (contains CTL epitope and HTL epitope) on the S1 segment, S₄₀₃₋₄₂₅ (contains CTL epitope and LBL epitope) on the RBD segment, and LBL epitope S₁₁₅₇₋₁₁₇₀ on the S2 segment were able to produce antibodies against the respective peptide antigen. Moreover, antibodies produced by S₄₀₃₋₄₂₅ in the RBD have a high affinity to RBD, while antibodies produced by the other two non-RBD domain epitopes (S₂₀₆₋₂₂₁, S₁₁₅₇₋₁₁₇₀) does not have affinity to RBD. These results verify the specificity of the immune response. The immunization results of S₄₀₃₋₄₂₅ and S₁₁₅₇₋₁₁₇₀ proved that the predictions are valid, and these two epitopes are worthy of further investigation. Since S₁₁₅₇₋₁₁₇₀ is the conservative sequence of different mutant strains of coronavirus, it is potential to be a component in vaccines against a variety of coronavirus variants. However, this work is

limited. Without a virus neutralization assay or a T cell activation assay, the neutralization capability or cellular immunity of the proposed vaccines was not further verified.

Omicron strain has over 50 mutations, of which over 30 mutations were in spike protein. There are mutations N211L and DEL212 in S₂₀₆₋₂₂₁, K417N in S₄₀₃₋₄₂₅, but no mutations in S₁₁₅₇₋₁₁₇₀. The results show obvious decrease in binding activity to Omicron S protein with S₂₀₆₋₂₂₁ and S₄₀₃₋₄₂₅, indicating that N211L, DEL212 and K417N mutations cause the reduction of the antibody binding activity. Although the binding antibody do not directly neutralize the virus, it is a prerequisite for neutralizing activity and can be employed to evaluate the immunogenicity of the peptide antigen.

5. Conclusion

Antigen epitopes can be predicted and screened effectively by bioinformatic methods. The predicted epitopes have good antigenicity, exhibit active binding with HLA-Alleles, and have broad population coverage for different geographical regions. Three candidate epitopes S₂₀₆₋₂₂₁, S₄₀₃₋₄₂₅, and S₁₁₅₇₋₁₁₇₀ predicted in this paper turned out to be good immunogen in vivo, and were competent for the development of SARS-CoV-2 peptide vaccines. N211L, DEL212 and K417N mutations at the Omicron S protein lead to the loss of antibody affinity to Omicron S protein and might help the variant strain evade from the original-vaccination-based immunity.

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.

Data availability

Data will be made available on request.

Acknowledgments

Feng-hua Xu contributed to the conception of the study; Qi Sun and Zhuan-qing Huang performed the experiment; Sen Yang, Yuan-yuan Li and Yue Ma, contributed significantly to analysis and manuscript preparation; Fei Yang and Ying Zhang helped perform the analysis with constructive discussions. All animal protocols were approved by the Institutional Animal Care and Use Committees at the Institute of PLA General Hospital. This study was performed in strict accordance with the Regulations for the Care and Use of Laboratory Animals and Guideline for Ethical Review of Animal (China, GB/T35892-2018).

Funding

This work is supported by the National Natural Science Foundations of China [82171814]. The funders have no roles in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imbio.2022.152287>.

References

- Al Kaabi, N., Zhang, Y., Xia, S., et al., 2021. Effect of 2 Inactivated SARS-CoV-2 Vaccines on Symptomatic COVID-19 Infection in Adults: A Randomized Clinical Trial. *JAMA* 326, 35–45. <https://doi.org/10.1001/jama.2021.8565>.
- Ali, M., Pandey, R.K., Khatoun, N., et al., 2017. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to

- battle against dengue infection. *Sci Rep* 7, 9232. <https://doi.org/10.1038/s41598-017-09199-w>.
- Bhattacharya, M., Sharma, A.R., Patra, P., et al., 2020. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-CoV-2): Immunoinformatics approach. *J Med Virol* 92, 618–631. <https://doi.org/10.1002/jmv.25736>.
- Boopathi, S., Poma, A.B., Kolandaivel, P., 2021. Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *J Biomol Struct Dyn* 39, 3409–3418. <https://doi.org/10.1080/07391102.2020.1758788>.
- Huynh-Hoa Bui, John Sidney, Kenny Dinh, et al. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *Bmc Bioinformatics* 2006;7,doi:10.1186/1471-2105-7-153.
- Calis, J.J.A., Maybeno, M., Greenbaum, J.A., et al., 2013. Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *Plos Comput. Biol.* 9, e1003266.
- Corbett, K.S., Edwards, D., Leist, S.R., et al., 2020. SARS-CoV-2 mRNA Vaccine Development Enabled by Prototype Pathogen Preparedness. *Nature* 586, 567–571. <https://doi.org/10.1101/2020.06.11.145920>.
- Couture, A., Garnier, A., Docagne, F., et al., 2019. HLA-Class II Artificial Antigen Presenting Cells in CD4(+) T Cell-Based Immunotherapy. *Front. Immunol.* 10, 1081. <https://doi.org/10.3389/fimmu.2019.01081>.
- De Groot, A.S., Sbai, H., Aubin, C.S., et al., 2002. Immuno-informatics: Mining genomes for vaccine components [J/OL]. *Immunol. Cell Biol.* 80 (3), 255–269. <https://doi.org/10.1046/j.1440-1711.2002.01092.x>.
- Deng, X., Garcia-Knight, M. A., Khalid, M. M., et al., 2021, Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R spike protein mutation. medRxiv, doi:10.1101/2021.03.07.21252647.
- Dhanda, S.K., Gupta, S., Vir, P., et al., 2013a. Prediction of IL4 Inducing Peptides. *Clin. Dev. Immunol.* 2013 <https://doi.org/10.1155/2013/263952>.
- Dhanda, S.K., Vir, P., Raghava, G.P.S., 2013b. Designing of interferon-gamma inducing MHCclass-II binders. *Biol. Direct* 8. <https://doi.org/10.1186/1745-6150-8-30>.
- Dimitrov, I., Bangov, I., Flower, D.R., et al., 2014. AllerTOP vol 2-a server for in silico prediction of allergens. *J. Mol. Model.* 20, 2278–2283. <https://doi.org/10.1007/s00894-014-2278-5>.
- Doytchinova, I. A., Flower, D. R., 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *Bmc Bioinformatics* 8, doi:10.1186/1471-2105-8-4.
- Doytchinova, I.A., Flower, D.R., 2007. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine* 25, 856–866. <https://doi.org/10.1016/j.vaccine.2006.09.032>.
- El-Manzalawy, Y., Dobbs, D., Honavar, V., 2008. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recog.* 21, 243–255. <https://doi.org/10.1002/jmr.893>.
- Jonathan M. Gershoni, A. R.-B., Dror D. Siman-Tov, Natalia Tarnovitski Freund and Weiss, Y. Epitope mapping: the first step in developing epitope-based vaccines. *Biodrugs : Clinical Immunotherapeutics, Biopharmaceuticals and Gene Therapy* 2007;21:145-146, doi:10.2165/00063030-200721030-00002.
- Grifoni, A., Sidney, J., Zhang, Y., et al., 2020. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* 27, 671–680. <https://doi.org/10.1016/j.chom.2020.03.002>.
- Gupta, S., Kapoor, P., Chaudhary, K., et al., 2013. In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS ONE* 8, e73957. <https://doi.org/10.1371/journal.pone.0073957>.
- He, Y., Zhou, Y., Wu, H., et al., 2004. Identification of immunodominant sites on the spike protein of severe acute respiratory syndrome (SARS) coronavirus: implication for developing SARS diagnostics and vaccines. *J. Immunol.* 173, 4050–4057. <https://doi.org/10.4049/jimmunol.173.6.4050>.
- He, Y., Li, J., Mao, W., et al., 2018. HLA common and well-documented alleles in China. *HLA* 92, 199–205. <https://doi.org/10.1111/tan.13358>.
- Hongjing, G.u., Chen, Q.i., Yang, G., et al., 2020. Adaptation of SARS-CoV-2 in BALB c mice fortesting vaccine efficacy. *Science* 369, 1603–1607. <https://doi.org/10.1126/science.abc4730>.
- Ikai, A., 1980. Thermostability and Aliphatic Index of Globular Proteins. *J. Biochem.* 88, 1895–1898. <https://doi.org/10.1093/oxfordjournals.jbchem.a133168>.
- Irini A. Doytchinova*, a. D. R. F. Bioinformatic Approach for Identifying Parasite and Fungal Candidate Subunit Vaccines. *The Open Vaccine Journal* 2008;1:22-26, doi: 10.2174/1875035400801010022.
- Jain, R., Jain, A., Verma, S.K., 2021. Prediction of Epitope based Peptides for Vaccine Development from Complete Proteome of Novel Corona Virus (SARS-COV-2) Using Immunoinformatics. *Int. J. Pept. Res. Ther.* 1–12. <https://doi.org/10.1007/s10989-021-10205-z>.
- Jespersen, M. C., Peters, B., Nielsen, M., et al., 2017, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*; 45: W24-W29, doi:10.1093/nar/gkx346.
- Juraszek, J., Rutten, L., Blokland, S., et al., 2020. Stabilizing the Closed SARS-CoV-2 Spike Trimer. *Nat. Commun.* 12, 244. <https://doi.org/10.1101/2020.07.10.197814>.
- Kalita, P., Padhi, A.K., Zhang, K.Y.J., et al., 2020. Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. *Microb. Pathog.* 145, 104236 <https://doi.org/10.1016/j.micpath.2020.104236>.
- Keech, C., Albert, G., Cho, I., et al., 2020. Phase 1–2 Trial of a SARS-CoV-2 Recombinant Spike Protein Nanoparticle Vaccine. *N. Engl. J. Med.* 383, 2320–2332. <https://doi.org/10.1056/NEJMoa2026920>.
- Kiyotani, K., Toyoshima, Y., Nemoto, K., et al., 2020. Bioinformatic prediction of potential T cell epitopes for SARS-Cov-2. *J. Hum. Genet.* 65, 569–575. <https://doi.org/10.1038/s10038-020-0771-5>.
- Korber, B., Fischer, W. M., Gnanakaran, S., et al., 2020, Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*; 182: 812-827 e819, doi:10.1016/j.cell.2020.06.043.
- Lopez Bernal, J., Andrews, N., Gower, C., et al., 2021, Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *N Engl J Med*; 385:585-594, doi:10.1056/NEJMoa2108891.
- Manavalan, B., Govindaraj, R. G., Shin, T. H., et al., 2018, iBCE-EL: A New Ensemble Learning Framework for Improved Linear B-Cell Epitope Prediction. *Frontiers in Immunology*; 9, doi:10.3389/fimmu.2018.01695.
- Mcbride, R., Van Zyl, M., Fielding, B.C., 2014. The coronavirus nucleocapsid is a multifunctional protein [J/OL]. *Viruses* 6 (8), 2991–3018. <https://doi.org/10.3390/v6082991>.
- Navagi, R.J., Toth I, M, S. Peptide Applications in Biomedicine, Biotechnology and Bioengineering. 2018, doi:10.2165/00063030-200721030-00002.
- Nielsen, M., Andreatta, M. NetMHCpan-3.0: improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* 2016;8:33, doi:10.1186/s13073-016-0288-x.
- Quast, I., Tarlinton, D., 2021. B cell memory: understanding COVID-19. *Immunity* 54, 205–210. <https://doi.org/10.1016/j.immuni.2021.01.014>.
- Reynisson, B., Alvarez, B., Paul, S., et al., 2020, NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*; 48: W449-W454, doi:10.1093/nar/gkaa379.
- Safavi, A., Kefayat, A., Mahdevar, E., et al., 2020. Exploring the out of sight antigens of SARS-CoV-2 to design a candidate multi-epitope vaccine by utilizing immunoinformatics approaches. *Vaccine* 38, 7612–7628. <https://doi.org/10.1016/j.vaccine.2020.10.016>.
- Saha, S., Raghava, G.P., 2006. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 65, 40–48. <https://doi.org/10.1002/prot.21078>.
- Sahin, U., Muik, A., Derhovanessian, E., et al., 2020, Concurrent human antibody and TH1 type T-cell responses elicited by a COVID-19 RNA vaccine., doi:10.1101/2020.07.17.20140533.
- Starr, T. N., Greaney, A. J., Hilton, S. K., et al., 2020. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*; 182:1295-1310 e1220, doi:10.1016/j.cell.2020.08.012.
- Wilkins, M.R., Gasteiger, E., Bairoch, A., et al., 1999. Protein identification and analysis tools in the ExpASY server. *Methods Mol. Biol.* 112, 531–552. <https://doi.org/10.1385/1-59259-584-7:531>.
- Xu, Z., Shi, L., Wang, Y., et al., 2020. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respiratory Med.* 8, 420–422. [https://doi.org/10.1016/s2213-2600\(20\)30076-x](https://doi.org/10.1016/s2213-2600(20)30076-x).
- Xu, W., Wang, M., Yu, D., et al., 2020. Variations in SARS-CoV-2 Spike Protein Cell Epitopes and Glycosylation Profiles During Global Transmission Course of COVID-19. *Front. Immunol.* 11, 565278 <https://doi.org/10.3389/fimmu.2020.565278>.
- Zhang, J., Zeng, H., Gu, J., et al., 2020. Progress and Prospects on Vaccine Development against SARS-CoV-2 [J/OL]. *Vaccines* 8 (2), 153.
- Zhang, Y., Zeng, G., Pan, H., et al., 2021. Safety, tolerability, and immunogenicity of an inactivated SARS-CoV-2 vaccine in healthy adults aged 18–59 years: a randomised, double-blind, placebo-controlled, phase 1/2 clinical trial. *Lancet Infect. Dis.* 21, 181–192. [https://doi.org/10.1016/s1473-3099\(20\)30843-4](https://doi.org/10.1016/s1473-3099(20)30843-4).
- Zhu, F.-C., Li, Y.-H., Guan, X.-H., et al., 2020. Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. *Lancet* 395, 1845–1854. [https://doi.org/10.1016/s0140-6736\(20\)31208-3](https://doi.org/10.1016/s0140-6736(20)31208-3).