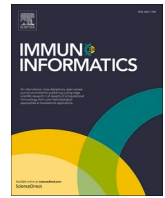




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

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.



Epitope identification of SARS-CoV-2 structural proteins using in silico approaches to obtain a conserved rational immunogenic peptide

Leonardo Pereira de Araújo^{a,b}, Maria Eduarda Carvalho Dias^a, Gislaine Cristina Scodeler^a, Ana de Souza Santos^a, Letícia Martins Soares^a, Patrícia Paiva Corsetti^a, Ana Carolina Barbosa Padovan^a, Nelson José de Freitas Silveira^b, Leonardo Augusto de Almeida^{a,*}

^a Laboratory of Molecular Biology of Microorganisms, Federal University of Alfenas, Alfenas, Minas Gerais, Brazil

^b Laboratory of Molecular Modeling and Computer Simulation, Federal University of Alfenas, Alfenas, Minas Gerais, Brazil

ARTICLE INFO

Keywords:
 COVID-19
 New coronavirus
 Reverse vaccinology
 SARS-CoV-2

ABSTRACT

The short time between the first cases of COVID-19 and the declaration of a pandemic initiated the search for ways to stop the spread of SARS-CoV-2. There are great expectations regarding the development of effective vaccines that protect against all variants, and in the search for it, we hypothesized the obtention of a predicted rational immunogenic peptide from structural components of SARS-CoV-2 might help the vaccine research direction. In the search for a candidate of an immunogenic peptide of the SARS-CoV-2 envelope (E), membrane (M), nucleocapsid (N), or spike (S) proteins, we access the predicted sequences of each protein after the genome sequenced worldwide. We obtained the consensus amino acid sequences of about 14,441 sequences of each protein of each continent and the worldwide consensus sequence. For epitope identification and characterization from each consensus structural protein related to MHC-I or MHC-II interaction and B-cell receptor recognition, we used the IEDB reaching 68 epitopes to E, 174 to M, 245 to N, and 833 to S proteins. To select an epitope with the highest probability of binding to the MHC or BCR, all epitopes of each consensus sequence were aligned. The curation indicated 1, 4, 8, and 21 selected epitopes for E, M, N, and S proteins, respectively. Those epitopes were tested in silico for antigenicity obtaining 16 antigenic epitopes. Physicochemical properties and allergenicity evaluation of the obtained epitopes were done. Ranking the results, we obtained one epitope of each protein except for the S protein that presented two epitopes after the selection. To check the 3D position of each selected epitope in the protein structure, we used molecular homology modeling. Afterward, each selected epitope was evaluated by molecular docking to reference MHC-I or MHC-II allelic protein sequences. Taken together, the results obtained in this study showed a rational search for a putative immunogenic peptide of SARS-CoV-2 structural proteins that can improve vaccine development using in silico approaches. The epitopes selected represent the most conserved sequence of new coronavirus and may be used in a variety of vaccine development strategies since they are also presented in the described variants of SARS-CoV-2.

1. Introduction

1.1. SARS-CoV-2 and COVID-19

Coming from a large family of single-stranded RNA viruses, SARS-CoV-2 is a polyadenylated, positive-sense virus that is responsible for the worsening of the severe acute respiratory syndrome, elevating the outbreak that began in December 2019 in Wuhan, China, to the pandemic status declared on March 11, 2020, by the World Health

Organization [1,2]. Seven coronaviruses capable of infecting humans are known, three of which are potential causes of serious diseases: MERS-CoV, SARS-CoV, and SARS-CoV-2; meanwhile, 229E, NL63, HKU1, and OC43 are linked to mild symptoms not related to complications [3]. Due to the higher transmissibility and faster spread of SARS-CoV-2, the WHO announced the pandemic status in February 2020 [1]. The coronavirus disease 2019 (COVID-19) causes a variety of symptoms from mild to severe cases characterized by pneumonia with a pro-inflammatory cytokine storm or extrapulmonary responses to virus

* Corresponding author.

E-mail address: leonardo.almeida@unifal-mg.edu.br (L.A. de Almeida).

<https://doi.org/10.1016/j.immuno.2022.100015>

Received 23 July 2021; Received in revised form 8 April 2022; Accepted 10 June 2022

Available online 11 June 2022

2667-1190/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

leading to systemic effects as vasculitis and subsequently inducing thrombosis [4].

1.2. Search for COVID-19 treatment and vaccines

Until now, about 8,000 clinical trials are in progress for COVID-19, including the development of new drugs or drug repositioning and vaccine development. However, despite the great efforts of scientists around the world, there is still no fully efficient treatment for COVID-19 [5]. An effective strategy for the control and prevent the new coronavirus spread is the use of vaccines as perceived with the decrease in the number of cases in countries with high rates of vaccinated people [6]. Although there are different approaches to obtain a vaccine to SARS-CoV-2, it is important to explore the structural proteins of the virus since they are fundamental to directly interacting with the host and may be the source of specific immunogenic components against it [7].

1.3. Immunoinformatics and COVID-19 vaccine development based on SARS-CoV-2 structural proteins

With the advent of immunoinformatics where the bioinformatics have helped a quick way to develop epitope-based vaccines *in silico* as a preliminary study for the *in vivo* validation study [8]. Sohail and colleagues [9] proposed this strategy to identify *in silico* T cell epitope identification for SARS-CoV-2. Since the ideal immune response against the new coronavirus should include an efficient antiviral innate immunity and a robust, specific cellular and humoral adaptive response against this virus [10] it is necessary to identify possible immunogenic epitopes based on the different SARS-CoV-2 genomes present in the DNA repositories. SARS-CoV-2 expresses 29 proteins, four of which are structural proteins: envelope protein (E), membrane glycoprotein (M), nucleocapsid phosphoprotein (N), and surface glycoprotein (S). They are closely associated with host interaction during the infection, which makes them important targets to be recognized by the immune system [11]. Among the structural proteins, the S protein is used as the most important to host recognition due to its presence around the virion, and it might be the best target for neutralizing antibodies. However, the SARS-CoV-2 variants described during the pandemic are mostly based on differences in amino acids in the primary sequences of the S protein, decreasing the capacity of neutralizing antibodies. In addition, researchers have observed interesting mechanisms of the new coronavirus to induce filopodia in the infected cells, which can augment the spread of the virus [12]. Therefore, a robust and efficient cellular immune response should be triggered by epitope-based vaccines that can induce elimination of the focus of infection using T cell activation-dependent mechanisms [13,14]. In order to obtain a rationally selected, specific immunogenic epitope from SARS-CoV-2 structural proteins that surpasses the possible issues associated with the virus variants or only one type of adaptive immune response, we identified a consensus sequence of the four structural proteins with a high possibility to elicit T and B cell responses against this virus. We used different established immunoinformatic approaches and structural analysis to identify the position of each selected epitope in the consensus 3D structure modeled by us for the structural proteins of SARS-CoV-2.

2. Methods

2.1. Obtaining SARS-CoV-2 putative structural protein consensus sequences

The methods followed the pipeline presented in Supplementary Fig. 1, showing the main immunoinformatic and molecular modeling and docking established tools used in this manuscript. The protein sequences for SARS-CoV-2 were obtained from the National Center of Biotechnology Information (NCBI) database. The specific database for SARS-CoV-2, present on the website in the “NCBI Virus” tab, was used by

means of the following identification: “taxid: 2697049” (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Protein&VirusLineage_ss=SARS-CoV-2,%20taxid:2697049). All sequences of the envelope protein (E), membrane glycoprotein (M), nucleocapsid phosphoprotein (N), and surface glycoprotein (S) from all continents present on the site were downloaded in FASTA format. The Jalview software (<https://www.jalview.org/>) was used to obtain the consensus sequences for each region and the global consensus sequence for each of the four target proteins [15].

2.2. Epitopes identification

For all epitope predictions of the target proteins, we used the algorithms of the Immune Epitope Database and Analysis Resource (IEDB; <https://www.iedb.org/>), of which only the methods were changed for the selection of each prediction.

2.3. MHC-I or MHC-II and B-cell receptor epitopes prediction

The NetMHCpan EL 4.1 server (<http://tools.iedb.org/mhci/>) was used and the option to obtain epitopes that interact with the 27 reference alleles of MHC-I [16] that simulate a greater coverage of the global population, among 97% and 99% and IC50 < 500nM [17] in sequence the cutoff of 1% of the total amount of each target structure was used; all options were recommended by the IEDB platform itself [18,19]. The “IEDB recommended 2.22” method (<http://tools.iedb.org/mhcii/>) was used, which consists of the consensus use of the NN-align [20], SMM-align [21], CombLib [22], and Sturniolo [23] methods, to obtain the best possible result for a given protein, as recommended by the IEDB. The combination of the 27 MHC-II alleles recommended for the worldwide coverage of the population was also used [24], for the cutoff, the epitopes that fit into two observations, including a consensus percentile threshold rank < 20.0 and interacting with more than 50% of the selected alleles [25], were selected. The BepiPred Linear Epitope Prediction 2.0 server (<http://tools.iedb.org/bcell/>) was used to obtain B cell interaction epitopes, based on a system of machine learning algorithms called Random Forest that classifies the amino acid sequences of the target proteins from simulations in crystallized proteins and amino acids not considered epitopes. A standard cutoff threshold of 0.5 was used to obtain promiscuous epitopes [26].

2.4. Alignment and determination of worldwide consensus epitopes

All epitopes found from each structural protein that interacts with B-cell receptors MHC-I and MHC-II were aligned using the MultiAlign server (<http://multalin.toulouse.inra.fr/multalin/>; [27] in addition to a manual alignment for comparison and greater precision. From the alignment, the epitopes with the highest repetition rate per region with an established size of 15 amino acids were selected. The selection of the most promiscuous 15 amino acids epitope was based on the peptide length able to fit in the MHC groves with best affinity for interaction, stability, and specificity to SARS-CoV-2 proteins.

2.5. Transmembrane epitopes identification

All regions of epitopes that were within the transmembrane regions of the SARS-CoV-2 structural proteins were removed in such a way the SOSUI server was used (<http://harrier.nagahama-i-bio.ac.jp/sosui/sosuisubmit.html>) [28], which identifies the transmembrane regions present in the proteins. It is worth mentioning that the nucleocapsid phosphoprotein is a completely soluble protein with no transmembrane regions.

2.6. Selected epitopes antigenicity, allergenicity, putative N-glycan sites and physical and chemical properties test

For the antigenicity test, the VaxiJen 2.0 server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>), with 89% accuracy, was used with the 0.5 and “probable antigen” cutoff [29]. For the allergenicity test, the server AlgPred 2.0, an update version of “AlgPred: Prediction of Allergenic Proteins and Mapping of IgE Epitopes” (<http://crrd.osdd.net/raghava/algpred/>), with sensitivity of 93.1%, of 95.36% specificity and 94.26% accuracy, was used with the hybrid method that consists of using the following five tools available on the server: SVMc, IgE, epitope, ARPs BLAST, and MAST [30]. The Putative N-glycan sites were determined using the NetNGlyc 1.0 Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>), 76% accuracy [31]. For the physical and chemical properties test, the ExPasy server's ProtParam tool (<http://web.expasy.org/protparam/>) was used with a cutoff of 40 for structure stability values [32].

2.7. SARS-CoV-2 structural proteins modeling

For the molecular modeling of SARS-CoV-2 envelope protein, membrane glycoprotein, nucleocapsid phosphoprotein, and surface glycoprotein membrane glycoprotein, the script AlphaFold v2.0 [33] (<https://github.com/deepmind/alphafold>) was used to predict the conformation of the proteins with an extremely high level of accuracy (>95%). Multiple databases as UniRef90 [34], MGnify [35], BFD [36], Uniclust30 [37], PDB70 and PDB [38] were used to homology modeling the structural proteins.

2.8. Refinement of human leukocyte antigen (HLA) and selected epitopes from structural proteins from SARS-CoV-2 for molecular docking

Receptor proteins for the docking process were selected from the 27 interaction alleles made available by the IEDB, and then a search was carried out on the PDB database for all these alleles to perform molecular docking. All proteins, both receptor (HLA) and ligands (Epitopes), were subjected to a preparation, in which hydrogens absent from the structures were added, the waters present in the .pdb files were deleted, the protein was scanned using the script complete_pdb. Py from the MODELLER software in order to find missing atoms, and the grid box in the active site of the structure was defined [39].

2.9. HLA and selected epitopes from structural proteins from SARS-CoV-2 docking

To carry out the dockings, the protein-peptide method was defined, which consists of determining the receptor protein as a rigid structure with some small conformations of the side chains as flexible and the ligand-protein as a flexible structure. For that, we used the MDockPep servers (<https://zougrouptoolkit.missouri.edu/mdockpep/index.html>; [40] XU X et al., 2018) and ClusPro (<https://cluspro.org/help.php>) [41].

In addition, the AutoDock Vina software was used with the intention of obtaining a greater range of results and thus having a greater precision of results [42]. LigPlot + [43] and Pymol [44] were used to determine the epitope-MHC interaction in 2D and 3D diagrams, respectively.

3. Results

3.1. SARS-CoV-2 structural proteins consensus sequence from different deposited genomes in a database.

From the NCBI database, we analyzed 53,765 putative SARS-CoV-2 structural proteins from translated, deposited genomes (table 1), representing sequences from six continents. North America is the continent with the highest number of obtained SARS-CoV-2 genome sequences with a total of 41,112 sequences (10,278 per protein). Asia presents 4,784 sequences (1,196 per protein), followed by Oceania with 5,672 sequences (1,418 per protein), Europe with 1,644 sequences (411 per protein), Africa with 428 sequences (107 per protein), and finally South America with 124 sequences (31 per proteins). After obtaining all sequences from all continents, the process of obtaining consensus sequences was carried out. In addition to the respective conservation values, in which a consensus sequence of each protein per continent was obtained (Supp. 1-4), we identified global consensus sequences from each new coronavirus structural protein to be used as targets in obtaining putative epitopes.

3.2. Putative immunogenic epitopes from the consensus sequence of SARS-CoV-2 structural proteins are described to interact with MHC-I, and the S protein contains a higher number of epitopes

According to Table 1, the higher number of epitopes from structural proteins from SARS-CoV-2 interact more with MHC-I and less with B-cell receptors. Since the surface glycoprotein is considered the highest immunogenic protein of SARS-CoV-2 and the bigger structural protein from this virus, the majority of putative epitopes were found in this protein (773 putative epitopes). On the other hand, the envelope protein with 75 amino acids presented only 68 putative epitopes in this sequence. All results are presented in Table 1, and they were obtained using the IEDB software performing a projection of 98,55% of population coverage according to “Population Coverage Calculation Result” tool.

3.3. Rational selection of putative immunogenic epitopes from SARS-CoV-2 structural proteins indicates at least one candidate epitope in each protein

To obtain the best candidates of putative immunogenic epitopes from each SARS-CoV-2 structural protein, the obtained epitopes were aligned, and the linear amino acids sequence performing about 15 amino acids that presenting repeating was selected. Furthermore, epitopes presented in the transmembrane anchoring site were excluded since they present

Table 1

Rational selection of consensus structural proteins epitopes of SARS-CoV-2.

	Africa	Asia	Europe	North America	Oceania	South America	Total
Genomes	428	4,784	1,644	41,112	5,672	124	53,764
Epitopes from consensus structural proteins of SARS-CoV-2							
	Envelope		Membrane		Nucleocapsid		Surface
B-cell receptor	2	6		11			33
MHC-I	36	116		222			625
MHC-II	30	52		12			115
Rational selection of the best epitopes							
	Envelope		Membrane		Nucleocapsid		Surface
Multalign	1	4		8			21
Vaxijen	1	3		3			9
ProtParam	1	1		2			6
Algpred	1	1		1			2

hydrophobic characteristics that cause difficult interactions with the amino acids present in the MHC-I or MHC-II grooves. Since the nucleocapsid phosphoprotein is a soluble protein, all repeated epitopes identified in this protein were kept for the next evaluations. After those selections, 34 epitopes were obtained, and they were in silico tested for antigenicity using the VaxiJen 2.0 software, obtaining 16 antigenic epitopes (Table 2). Physicochemical properties evaluation of the obtained epitopes was based on molecular weight, isoelectric point, structural formula, number of atoms, number of amino acids, half-life, structural stability, and n-glycan sites. At this point, 10 epitopes were selected. It is important to note that the epitope from envelope protein is the only epitope that did not reach the cutoff value for stability. To overcome it, it was necessary to augment the selected epitope to 17 amino acids to make it stable but with a n-glycan site. The epitope was subjected to the antigenicity evaluation again, yielding a score necessary to be considered a putative immunogenic epitope. To increase the confidence of using the selected epitopes in future in vitro and in vivo evaluation, the allergenicity was tested using the AlgPred software with the following five tools in the server: SVMc, IgE, epitope, ARPs BLAST, and MAST. At this point, only five epitopes were not allergenic. Ranking the results, we obtained one epitope of each protein except for the S protein that presented two epitopes after selection (Table 1).

3.4. Position identification of selected putative immunogenic epitopes in SARS-CoV-2 structural proteins after molecular modeling by homology

To identify the position of selected putative immunogenic epitopes in the structural proteins from SARS-CoV-2, we performed molecular modeling for each global consensus protein. Since there is no information about the 3D consensus protein in the PDB database, we used the opensource script AlphaFold v2.0 for the envelope protein (E; Fig. 1A), membrane glycoprotein (M; Fig. 1B), nucleocapsid phosphoprotein (N; Fig. 1C), and surface glycoprotein (S; Fig. 1D). The selected epitopes inserted in the 3D structure that were identified for each protein are highlighted in red in Fig. 1. The transmembrane helices are indicated in the models in green.

3.5. Molecular docking between the selected putative immunogenic epitopes in SARS-CoV-2 structural proteins and MHC-I and MHC-II shows high probabilities of interaction between them

Although not all structures of MHC-I and MHC-II were found in the database (IEDB), we used the structure of 14 proteins of MHC-I alleles and three of MHC-II alleles according to the underlined presented in Table 3. The results demonstrated great energy of interaction based on two different online servers and by AutoDock Vina software. To represent the molecular docking between the epitope and the MHC-I and MHC-II alleles, we chose the interaction with the highest power of interaction indicated by the servers and the software. The best interaction

between the envelope protein (E) and MHC-I was with the HLA-B*1501 allele (Figs. 2A and 2C) and MHC-II with the HLA-DRB1*04:01 allele (Figs. 2B and 2D). The points of contact are demonstrated in the 3D Figs. 2A and 2B, and twenty-two hydrophobic interactions between the epitope and MHC-I are 2D represented in Fig. 2C. For epitope and MHC-II interactions it was observed eighteen hydrogen bonds and twenty-eight hydrophobic interactions (Fig. 2D). For the membrane glycoprotein (M), we demonstrated the interaction between the HLA-B*57:01 allele for MHC-I (Figs. 3A and 3C) and the HLA-DRB1*15:01 allele for MHC-II (Figs. 3B and 3D). The points of contact are demonstrated in the 3D Figs. 3A and 3B, and eight hydrogen bonds and thirty-five hydrophobic interactions between the epitope and MHC-I are 2D represented in Fig. 3C. For epitope and MHC-II interactions it was observed six hydrogen bonds and thirty-two hydrophobic interactions (Fig. 3D). The epitope present in the nucleocapsid phosphoprotein (N) was docking with the HLA-B*35:01 allele for MHC-I (Figs. 4A and 4C) and HLA-DRB1*04:01 allele for MHC-II (Figs. 4B and 4D). The 3D diagram shows the points of contact in Figs. 4A and 4B. Seven hydrogen bonds and twenty-six hydrophobic interactions between the selected epitope and MHC-I are represented in the Fig. 4C, while ten hydrogen bonds and thirty-seven hydrophobic interactions between the selected epitope and MHC-II are represented in the Fig. 4D. The surface glycoprotein (S) presented two epitopes, and the results of the interaction are demonstrated in Figs. 5 and 6. For epitope 1, we demonstrated the interaction between the MHC-I allele HLA-A*02:03 (Figs. 5A and 5C), with twelve hydrogen bonds and twenty-four hydrophobic interactions between the epitope 1 and MHC-I (Fig. 5C). The interaction between the epitope 1 and MHC-II allele HLA-DRB1*04:01 showed nine hydrogen bonds and twenty-nine hydrophobic interactions (Fig. 5D). For epitope 2 from the S protein demonstrated the interaction between the allele HLA-B*35:01 (Figs. 6A and 6C) and allele HLA-DRB1*15:01 (Figs. 6B and 6D) with MHC-I and MHC-II, respectively. The points of contact are demonstrated in the 3D Figs. 6A and 6B, and nine hydrogen bonds and twenty-nine hydrophobic interactions between the epitope 2 and MHC-I (Fig. 6C). For epitope 2 and MHC-II interactions it was observed eight hydrogen bonds and twenty-seven hydrophobic interactions (Fig. 6D).

4. Discussion

4.1. Short time to find putative immunogenic epitopes to COVID-19 vaccine development by immunoinformatic

The urgency of an effective vaccine against emerging and rapidly transmitting pathogens, such as the new coronavirus, is a reality. The WHO report, after a year and a half of the COVID-19 pandemic status, that there are more than 284 vaccines in development, 100 of which are in clinical studies in different phases, nine are in phase III already published, two of messenger RNA technology (Pfizer, Moderna), four of viral vectors (AstraZeneca, Gamaleya, Jansen, CanSino), one of

Table 2
Antigenicity, allergenicity, and physical-chemical properties of selected epitopes from consensus structural proteins epitopes of SARS-CoV-2.

	Envelope	Membrane	Nucleocapsid	Surface 1	Surface 2
Sequence	LVKPSFYVYSRVKLNLS	TVATSRILSYKLGGA	MEVTPSGTWLTYTGA	IAIPTNFTISVITEI	ALQIPFAMQMAYRFN
Antigenicity	0,6582	0,7300	0,7886	0,7719	1,0112
MW	2014,36	1630,86	1613,80	1619,88	1801,15
PI	10,00	9,70	4,00	4,00	8,79
Stability	38,99	19,08	-7,92	7,39	33,91
Half-life	5,5 h	7,2 h	30 h	20 h	4,4 h
Reticulocytes					
Half-life	3 min	>20 h	>20	30 min	>20 h
Yeast					
Half-life	2 min	>10 h	>10 h	>10 h	>10 h
<i>E. coli</i>					
IgE epitope	No	No	No	No	No
Allergenicity	No	No	No	No	No
n-glycan site	Yes	No	No	Yes	No

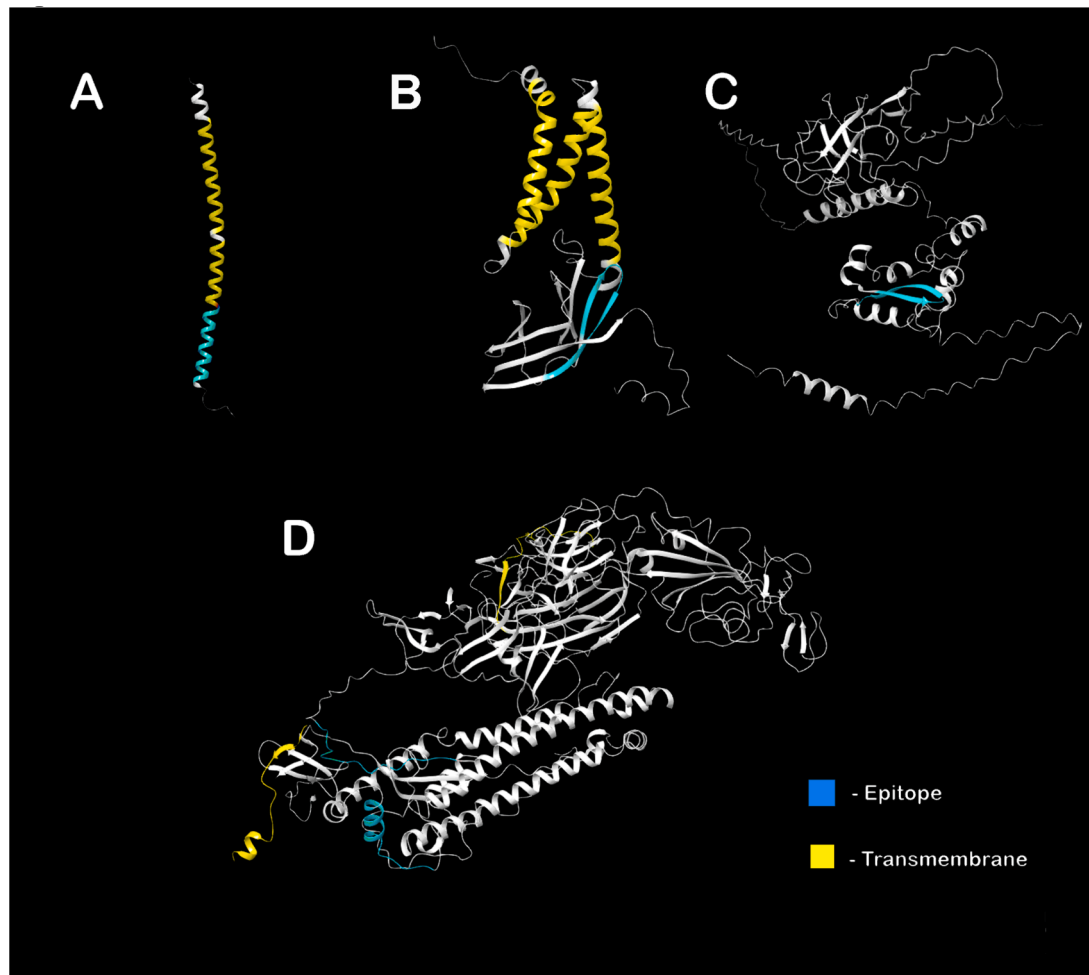


Fig. 1. Position identification of selected epitopes in consensus 3D structural proteins from SARS-CoV-2. Molecular modeling by homology for the envelope protein (A), membrane glycoprotein (B), nucleocapsid phosphoprotein (C), and surface glycoprotein (D). The molecular modeling of the proteins was based on homology using AlphaFold v2.0 scrip. The selected epitopes are highlighted in red, and the transmembrane helices are indicated in the models in green.

recombinant protein (Novavax), and two using an inactivated virus (Sinopharm). Here we used *in silico* approaches to obtain epitopes from structural proteins from SARS-CoV-2 to reach, in a short period of time, a specific immunogenic peptide able to be used as a safe peptide-based vaccine. Currently, numerous projects for the development of effective vaccines have been carried out, and some of these vaccines are already commercially available for humans [45]. However, traditional vaccine production techniques have some disadvantages, which can be overcome by using computational approaches [46]. In addition, recent advances in bioinformatics have provided a variety of tools and servers capable of reducing the cost and time of advancing the traditional vaccine [5]. Immunoinformatics approaches can be used to analyze pathogen antigens, predict their epitopes, and assess their immunogenicity [5]. Furthermore, reverse vaccinology, epitope prediction, structural vaccinology, rational approaches, and molecular docking are of great use in designing a potential vaccine against COVID-19 [46].

4.2. Global consensus amino acid sequence and modeling of SARS-CoV-2 structural proteins

Therefore, in this work, bioinformatics techniques were used to reach a consensus amino acid sequence from the four structural proteins of SARS-CoV-2 using the sequences of genomes deposited in the database from different regions of the world following obtaining a global consensus sequence. The strategy to obtain a consensus sequence of each protein may decrease the possibility of observing variants, but it

increases the possibility to define the conserved amino acid sequences from those proteins on available genomes around the world. Here, we were able to define the three-dimensional structure of the consensus proteins and identify their secondary structures using molecular modeling and artificial intelligence data. Furthermore, the immunoinformatics approaches may direct in a rational way the identification of putative immunogenic epitopes to design an epitope-based vaccine to control SARS-CoV-2 [47].

4.3. Rational selection of putative immunogenic epitopes from global consensus SARS-CoV-2 structural protein

We also demonstrated that sequentially using rational *in silico* techniques, based on well-established bioinformatic tools, made it possible to select the best epitope able to interact with the B-cell receptor, which may stimulate a specific adaptive humoral response, or MHC-I and MHC-II alleles, which may stimulate a cellular adaptive immune response based on TCD8+ or TCD4+ activation, respectively. The best example is the search for epitopes in the sequence of the surface glycoprotein (S). There are about 770 epitopes in the protein S, but with our sequence of analysis, only two potential epitopes are selected in this protein. S glycoprotein, also called spike, is one of the main proteins for therapeutic and vaccine targets, in addition to being one of the most important proteins to study, as it is responsible for the SARS-CoV-2 virus infecting host cells. It has a structure with covalently linked carbohydrate molecules that are N-glycans and is extremely glycosylated. This

Table 3
Molecular docking between selected epitopes of SARS-CoV-2 consensus structural proteins and HLAs.

	Epitope	Allele	ClusPro		MdockPep	AutoDock Vina
			Center	Lowest Energy		
MHC-I	Envelope	HLA-A*0203	-842,2	-842,2	-266,03	-7,1
		HLA-B*0801	-636,7	-963,7	-239,0	-7,3
		HLA-B*1501	-635,5	-791,4	-261,9	-7,4
		HLA-B*3501	-679,4	-923,0	-273,9	-6,4
		HLA-A*0206	-749,8	-749,8	-270,2	-5,7
		HLA-B*5701	-655,0	-850,3	-230,4	-8
	Membrane	HLA-B*3501	-693,9	-833,7	-221,1	-7,3
		HLA-B*5801	-578,8	-689,7	-220,4	-7,4
		HLA-A*0301	-731,9	-731,9	-223,6	-7,0
		HLA-A*0206	-809,0	-809,6	-247,7	-6,5
		HLA-B*3501	-711,2	-911,1	-203,6	-7,8
		HLA-B*4403	-681,4	-833,6	-204,3	-7,4
	Surface 1	HLA-A*0203	-792,0	-931,7	-187,7	-8,3
		HLA-B*5101	-796,5	-1112,8	-185,4	-7,9
		HLA-B*3501	-792,4	-843,8	-187,8	-8,3
		HLA-B*5701	-762	-914,2	-213,7	-7,7
		HLA-B*5801	-787,9	-905,3	-210,3	-7,2
		HLA-B*3501	-794,6	-1044,9	-234,9	-7,9
	Surface 2	HLA-A*2402	-664,7	-789,9	-229,3	-7,8
		HLA-B*5801	-730,5	-982,8	-264,4	-7,0
		HLA-A*0206	-754,5	-949,9	-255,2	-7,1
		HLA-B*1501	-752,5	-1016,6	-242,4	-6,6
		HLA-DRB1*04:01	-695,2	-806,9	-275,3	-7,2
		HLA-DRB1*15:01	-847,4	-911,3	-273,2	-6,9
MHC-II	Envelope	HLA-DRB1*01:01	-788,7	-917,0	-276,5	-5,8
		HLA-DRB1*04:01	-743,9	-747,3	-228,1	-6,6
	Membrane	HLA-DRB1*15:01	-814,2	-814,2	-219,9	-6,3
		HLA-DRB1*01:01	-762,1	-762,1	-230,2	-7,0
	Nucleocapsid	HLA-DRB1*04:01	-697,9	-809,8	-226,4	-8,2
		HLA-DRB1*15:01	-723,0	-858,8	-240,5	-6,5
		HLA-DRB1*01:01	-660,5	-751,0	-214,5	-6,3
	Surface 1	HLA-DRB1*04:01	-722,0	-844,1	-200,8	-7,2
		HLA-DRB1*15:01	-784,1	-857,2	-209,0	-6,7
		HLA-DRB1*01:01	-729,1	-827,8	-201,4	-7,1
	Surface 2	HLA-DRB1*04:01	-770,7	-806,3	-252,1	-7,7
		HLA-DRB1*15:01	-797,3	-996,6	-263,6	-7,7
		HLA-DRB1*01:01	-766,7	-888,1	-258,6	-7,4

glycosylation plays a very important role in the interaction of this protein with the ACE2 receptor because water molecules can disrupt the interaction of the S protein with this receptor. These glycans, like carbohydrates, are quite polar, and in addition to stabilizing this molecule, they work by capturing water molecules on the surface of S glycoprotein as well as capturing water molecules on the surface of ACE2, favoring its interaction with the receptor [48]. However, the S protein is where most variations are found between strains, favoring the evasion of the virus by the host's immune system [49,50]. Based on knowledge of how the infection occurs, which residues are responsible, and which atoms are involved, we can infer the impact of these mutations on the protein's activity.

4.4. The concerns about the SARS-CoV-2 new variants during vaccine development

Viruses mutate constantly, but just a few key mutations can affect in some way, making a variant more virulent or more lethal, and it depends a lot on where this mutation occurs [51]. One of the two selected epitopes in this work is not presented in the described variants, while the other, epitope 1, is related to a variant T716I and contains a n-glycan site (Fig. 7). If a mutation with changes in the amino acid sequence occurred in the same region of a selected epitope, there would be a loss of efficiency of the same, or even a loss of functionality of the immune system to efficiently recognize the pathogen. The n-glycan site may worsen the antibody interaction to neutralize the virus. However, it is not hampering the cellular response against the infection. Among the variants described the ones that cause the most concern is related to changes in the amino acid sequence of the S glycoprotein. The US Department of

Health and Human Services (HHS) created the Sars-CoV-2 Interagency Group (SIG), formed by the largest agencies for Disease Control and Prevention (CDC). The GIS is responsible for defining the classification of the new coronavirus mutations, observing the characteristics that a given mutation can generate; these three groups are the variant of interest (VOI), variant of concern (VOC), and variant of high consequence (VOHC). For a mutation to qualify for the VOI classification, it must have at least one of the characteristics, such as specific genetic markers, that lead the virus to increase its degree of transmission or change the escape mechanism to the immune system or any evidence to prove that it is responsible by an expansion of contamination. Currently, seven strains fall into this category, and they are Iota (United States), B.1.526.1 (United States), Eta (Nigeria/United Kingdom), B.1.617 (India), Kappa (India), B.1.617.3 (India), and Zeta (Brazil). To define a mutation as VOC, it must have, in addition to the attributes mentioned for VOI, any evidence that proves increased disease severity, impact on vaccines, treatments, diagnoses, increased transmissibility, and reduced vaccine effect among others. Currently, seven strains are present in this category: Alpha (United Kingdom), Gamma (Japan/Brazil), Delta (India), Beta (South Africa), Epsilon (B.1.427 and B.1.429; United States), and Omicron (South Africa). And finally, for a mutation to be characterized as VOHC, in addition to having all the characteristics mentioned for VOC, it must have an impact on medical countermeasures (MCM), evidence of a drastic reduction in the effect of vaccines, and an increase in hospitalizations with worsening clinical conditions; however, to date, no mutations fall into this category [52]. It is noteworthy that the variant VOC B.1.351 from South Africa was the object of a study for researchers at the Faculty of Medicine of São Paulo (FM-USP) in which the 417N mutation responsible for replacing the amino acid lysine with asparagine is

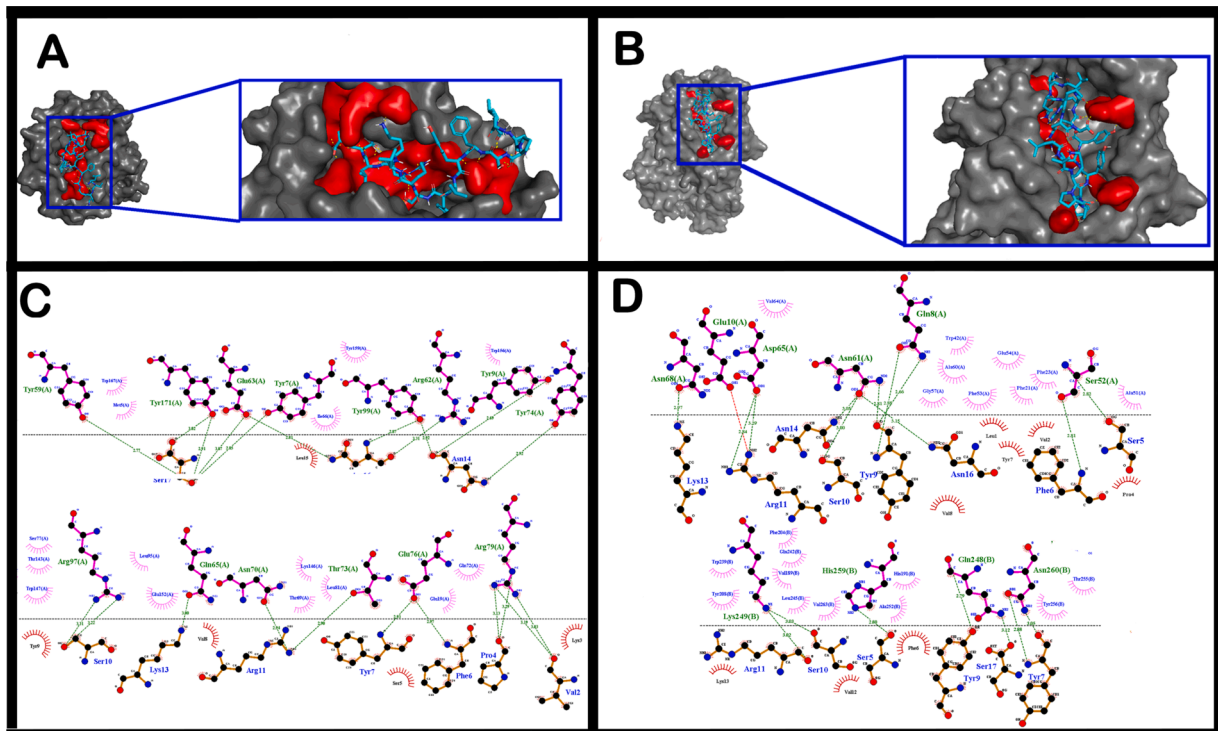


Fig. 2. Interaction between the selected epitope from the consensus sequence of the SARS-CoV-2 envelope protein and MHC-I and MHC-II alleles. The best interaction evaluated by molecular docking between the envelope protein epitope and the MHC-I allele HLA-B*1501 (A and C) and MHC-II allele HLA-DRB1*04:01 (B and D). The points of contact are demonstrated in A and B in red, and the points of interaction between the epitope and MHC molecules are 2D represented in C and D, respectively.

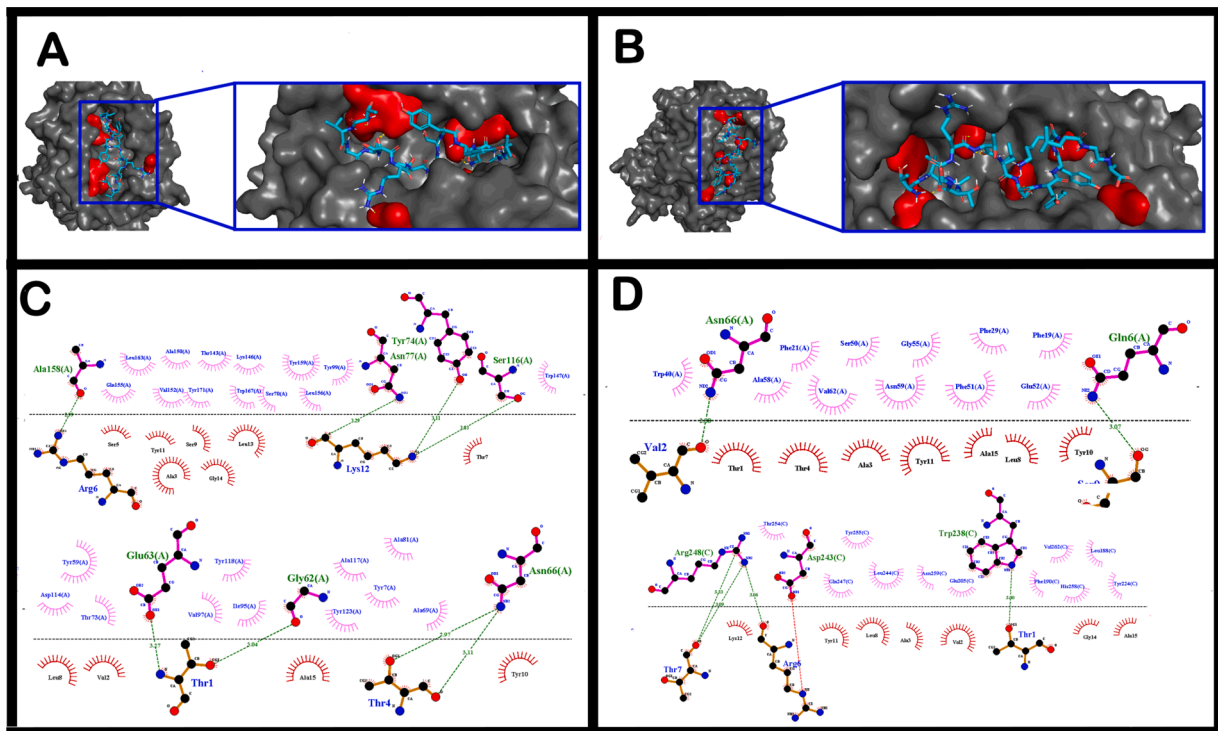


Fig. 3. Interaction between the selected epitope from the consensus sequence of the SARS-CoV-2 membrane glycoprotein and MHC-I and MHC-II alleles. The best interaction evaluated by molecular docking between the membrane glycoprotein epitope and the MHC-I allele HLA-B*57:01 (A and C), and MHC-II allele HLA-DRB1*15:01 (B and D). The points of contact are demonstrated in A and B in red, and the points of interaction between the epitope and MHC molecules are 2D represented in C and D, respectively.

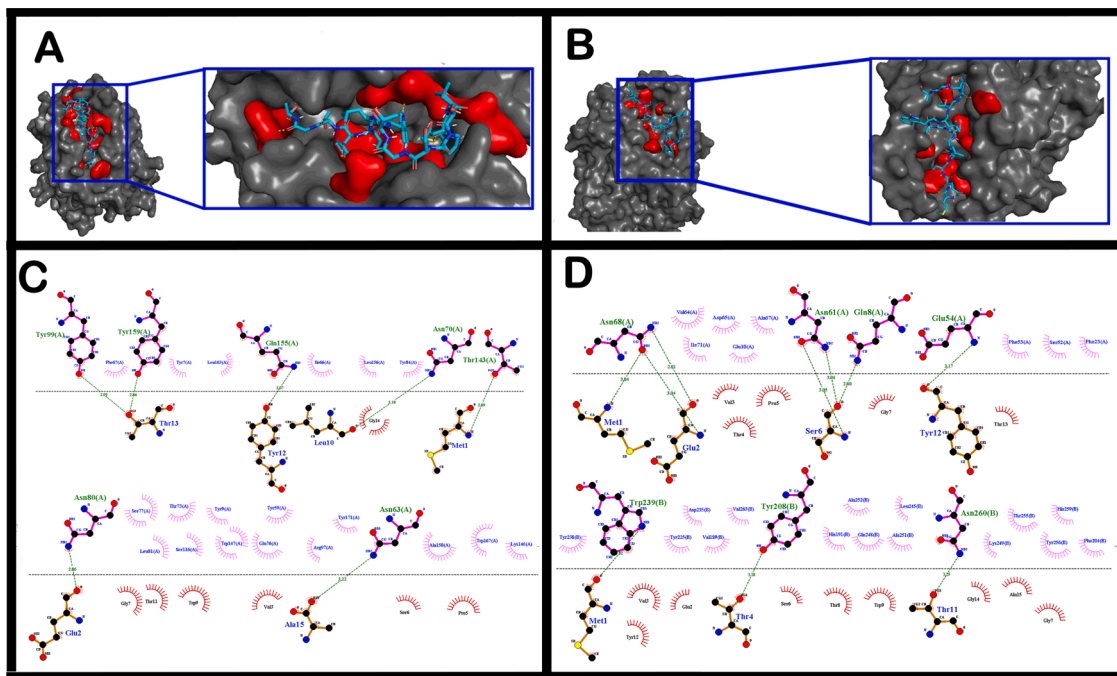


Fig. 4. Interaction between the selected epitope from the consensus sequence of the SARS-CoV-2 nucleocapsid phosphoprotein and MHC-I and MHC-II alleles. The best interaction evaluated by molecular docking between the nucleocapsid phosphoprotein epitope and the MHC-I allele HLA-B*35:01 (A and C), and MHC-II allele HLA-DRB1*04:01 (B and D). The points of contact are demonstrated in A and B in red, and the points of interaction between the epitope and MHC molecules are 2D represented in C and D, respectively.

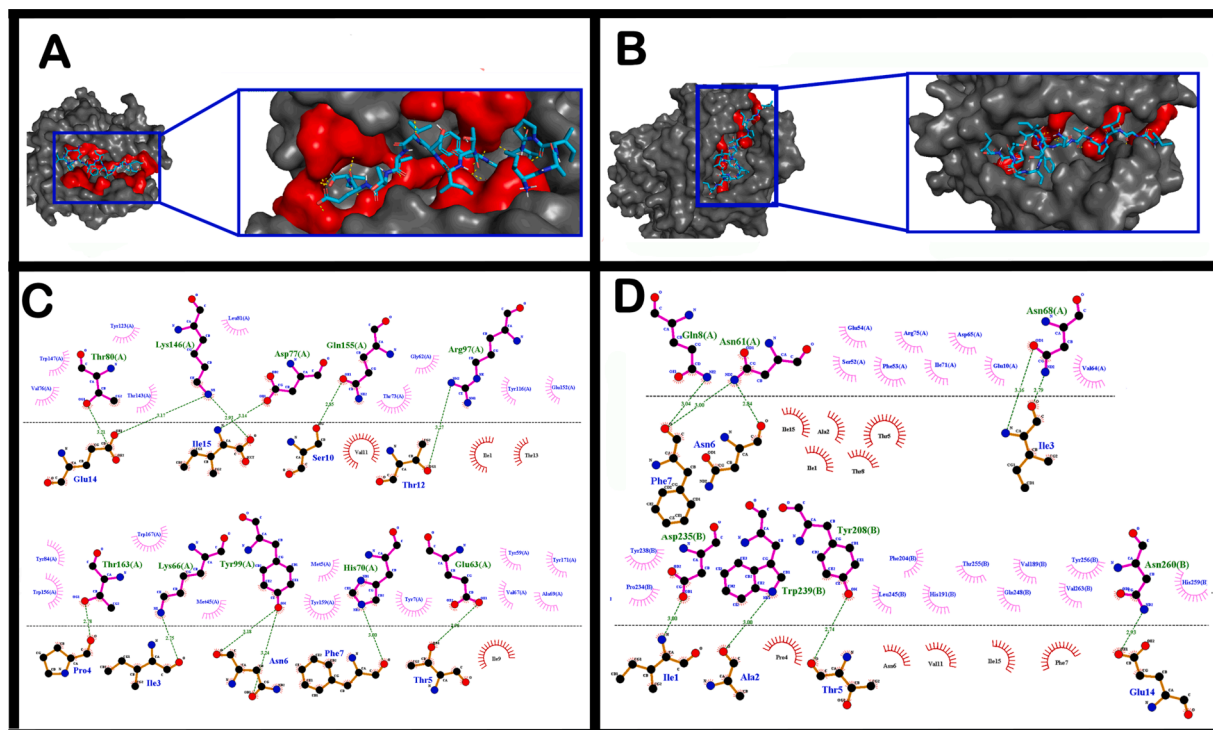


Fig. 5. Interaction between the selected epitope 1 from the SARS-CoV-2 surface glycoprotein consensus sequence and MHC-I and MHC-II alleles. The best interaction evaluated by molecular docking between the surface glycoprotein epitope 1 and the MHC-I allele HLA-A*02:03 (A and C) and MHC-II allele HLA-DRB1*04:01 (B and D). The points of contact are demonstrated in A and B in red, and the points of interaction between the epitope and MHC molecules are 2D represented in C and D, respectively.

responsible for glycolysis of proteins making an escape mechanism for the virus, changing its conformation [53]. Currently, the new Indian variant, B. 1617, of the new coronavirus raises the concern and warning

signal regarding transmission, lethality, and resistance to available vaccines [54,55]. Using the approaches indicated in this work, peptide-based vaccines facilitate the handling of variants.

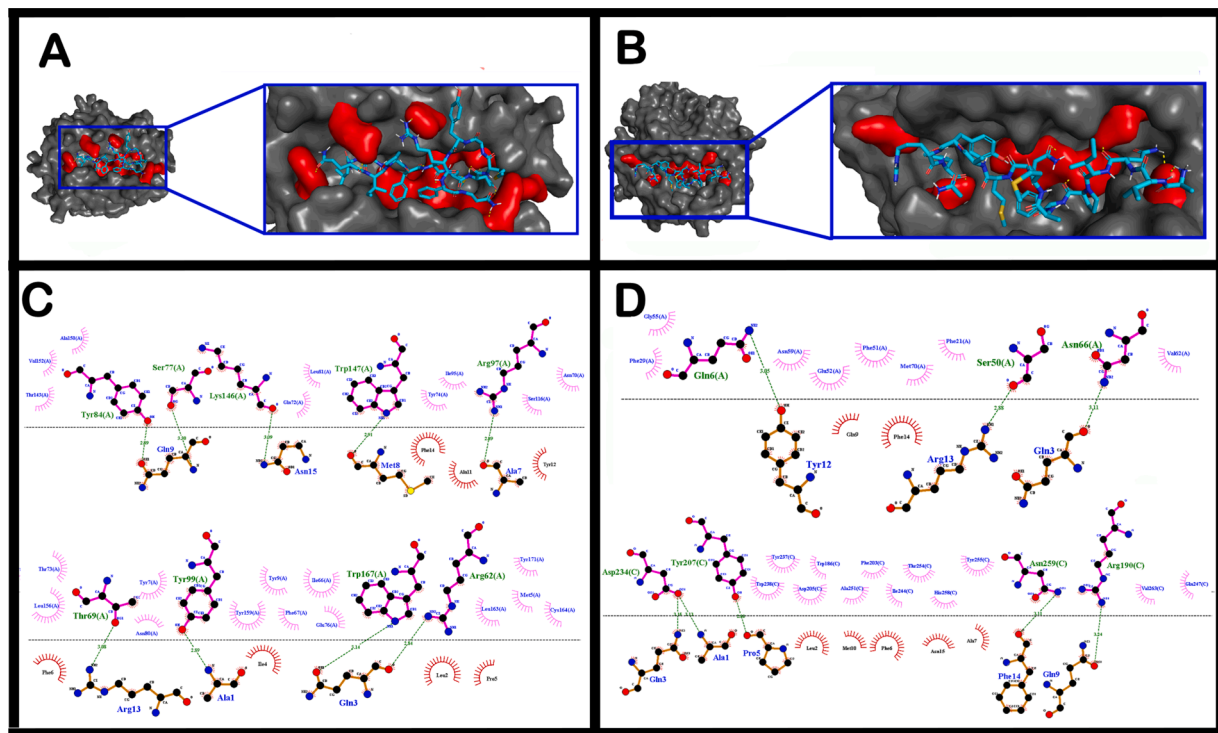


Fig. 6. Interaction between the selected epitope 2 from the SARS-CoV-2 surface glycoprotein consensus sequence and MHC-I and MHC-II alleles. The best interaction evaluated by molecular docking between the surface glycoprotein epitope 2 and the MHC-I allele HLA-B*35:01 (A and C), and MHC-II allele HLA-DRB1*15:01 (B and D). The points of contact are demonstrated in A and B in red, and the points of interaction between the epitope and MHC molecules are 2D represented in C and D, respectively.

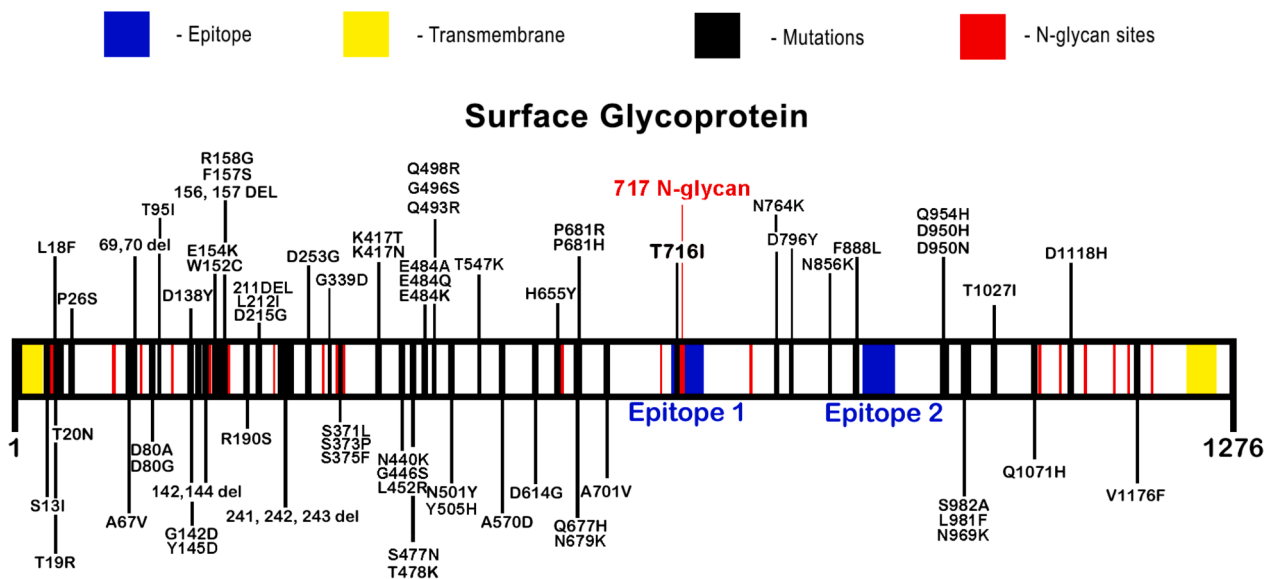


Fig. 7. SARS-CoV-2 surface glycoprotein variants. The described variants of SARS-CoV-2 based on amino acid changes in the surface glycoprotein indicate a variety of points across the protein indicated by a red mark. The selected epitopes are highlighted in blue. N-glycan sites are highlighted in green. The selected epitope 1 lies within a described variant T716I and a putative n-glycan site, while epitope 2 is not included within any yet-described variant.

4.5. Handling of a peptide-based putative vaccine against SARS-CoV-2 rationally selected by immunoinformatic

The only way to perform the management of a vaccine with the complete protein sequence is to change the sequence itself as a whole. On the other hand, peptide vaccines, which are the small regions of the protein that have characteristics of an immune response, can be managed in two ways: using a pool of peptides that escape and are not

within the most frequent areas that present mutation and replacing the protein or in the advent of the emergence of a new variant where that substitution is precisely in the peptide used and can perform only that substitution [56]. Besides, generating multiple target sites as other structural proteins of a specific pathogen may increase the possibilities for the host's immune system to recognize and eliminate the invader. The results presented in this work also identify the most likely epitope, using a variety of established bioinformatic tools with different accuracy

level, to be immunogenic in the envelope protein (E), membrane glycoprotein (M), or nucleocapsid phosphoprotein (N). It is important to note that variants in these proteins are not described as more transmissible, lethal, or resistant to vaccines. All epitopes were evaluated to their potential to interact with immune system molecules and the security to be used in humans since they are not allergenic but antigenic with a good half-life in mammals, yeast, or *E. coli*. Finally, the results obtained by our group are based on a relatively reduced cost and time compared to traditional techniques, but it is necessary to evaluate the action of selected peptides associated with adjuvants to functionally define them as immunogenic and possible candidates to be used as a peptide-based vaccine. Lee and colleagues (2021), remarks the importance of different approaches to correlate the immunogenicity prediction of MHC-bound peptides, including large scale *in vitro* or *in vivo* evaluation of selected epitopes to increase the confidence of *in silico* epitope selection [57]. Vaccination is the most effective and safest method of creating an immune barrier, capable of breaking the SARS-CoV-2 transmission and preventing the most serious pathology of the disease. When the effective vaccination campaign is combined with other preventive methods of social distancing and the use of a mask, over time, there is a significant reduction in the number of cases of the disease, even though the social life of its population returns [58].

CRedit authorship contribution statement

Leonardo Pereira de Araújo: Conceptualization, Data curation, Methodology, Writing – original draft. **Maria Eduarda Carvalho Dias:** Data curation, Methodology, Formal analysis, Writing – original draft. **Gislaine Cristina Scodeler:** Data curation, Methodology, Formal analysis, Writing – original draft. **Ana de Souza Santos:** Data curation, Methodology, Formal analysis, Writing – original draft. **Letícia Martins Soares:** Data curation, Methodology, Formal analysis. **Patrícia Paiva Corsetti:** Data curation, Methodology, Formal analysis, Supervision. **Ana Carolina Barbosa Padovan:** Data curation, Methodology, Formal analysis, Supervision. **Nelson José de Freitas Silveira:** Data curation, Methodology, Formal analysis, Supervision. **Leonardo Augusto de Almeida:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Project administration, Resources, Supervision.

Declaration of Competing Interests

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.

Acknowledgments

This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (Finance Code 001), Universidade Federal de Alenas and the Brazilian Ministry of Education (MEC) (EDITAL N° 002/2020 - PRPPG/REITORIA), and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, process 115462/2020-1).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.immuno.2022.100015.

References

- [1] WORLD HEALTH ORGANIZATION. WHO announces COVID-19 outbreak a pandemic. March 12, 2020. <https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic> (accessed 19 July 2021). Reference to a dataset:.
- [2] WU Di, et al. The SARS-CoV-2 outbreak: what we know. *Int J Infect Dis* 2020;94: 44–8. <https://doi.org/10.1016/j.ijid.2020.03.004>.
- [3] PETERSEN Eskild, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis* 2020. [https://doi.org/10.1016/S1473-3099\(20\)30484-9](https://doi.org/10.1016/S1473-3099(20)30484-9).
- [4] Yao XH, Luo T, et al. A cohort autopsy study defines COVID-19 systemic pathogenesis. *Cell Res* 2021 Jun 16. <https://doi.org/10.1038/s41422-021-00523-8>.
- [5] UL QAMAR, Tahir Muhammad, et al. Reverse vaccinology assisted designing of multiepitope-based subunit vaccine against SARS-CoV-2. *Infect Dis Poverty* 2020;9(1):1–14. <https://doi.org/10.1186/s40249-020-00752-w>.
- [6] MOORE Sam, et al. Modelling optimal vaccination strategy for SARS-CoV-2 in the UK. *PLoS Comput Biol* 2021;17(5):e1008849. <https://doi.org/10.1371/journal.pcbi.1008849>.
- [7] ABRAHAM PEELE K, et al. Design of multi-epitope vaccine candidate against SARS-CoV-2: a *in-silico* study. *J Biomol Struct Dyn* 2020:1–9. <https://doi.org/10.1080/07391102.2020.1770127>.
- [8] OLI Angus Nnamdi, et al. Immunoinformatics and vaccine development: an overview. *Immuno Targets Therapy* 2020;9(13). <https://doi.org/10.2147/ITT.S241064>. PMID: 32161726.
- [9] SOHAIL Muhammad Saqib, et al. In silico T cell epitope identification for SARS-CoV-2: progress and perspectives. *Adv Drug Deliv Rev* 2021. <https://doi.org/10.1016/j.addr.2021.01.007>.
- [10] SCHJNS V LAVELLE EC. Prevention and treatment of COVID-19 disease by controlled modulation of innate immunity. *Eur J Immunol* 2020;50(7):932–8. <https://doi.org/10.1002/eji.202048693>. Epub 2020 Jun 15.
- [11] ONG Edison, et al. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Front Immunol* 2020;11:1581. <https://doi.org/10.3389/fimmu.2020.01581>.
- [12] BARRETO-VIEIRA Debora Ferreira, et al. Morphology and morphogenesis of SARS-CoV-2 in Vero-E6 cells. *Memórias do Instituto Oswaldo Cruz* 2021;116. <https://doi.org/10.1590/0074-02760200443>.
- [13] SMITH Christof C, et al. Landscape and selection of vaccine epitopes in SARS-CoV-2. *Genome Med* 2021;13(1):1–23. <https://doi.org/10.1186/s13073-021-00910-1>.
- [14] LEHMANN Alexander A, et al. Deconvoluting the T cell response to SARS-CoV-2: specificity versus chance and cognate cross-reactivity. *Front Immunol* 2021;12. <https://doi.org/10.3389/fimmu.2021.635942>.
- [15] WATERHOUSE Andrew M, et al. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25(9):1189–91. <https://doi.org/10.1093/bioinformatics/btp033>.
- [16] WEISKOPF Daniela, et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc Natl Acad Sci* 2013;110(22):E2046–53. <https://doi.org/10.1073/pnas.1305227110>.
- [17] SETTE Alessandro, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol* 1994;152(12):5586–92.
- [18] MOUTAFTSI Magdalini, et al. A consensus epitope prediction approach identifies the breadth of murine T CD8+ cell responses to vaccinia virus. *Nat Biotechnol* 2006;24(7):817–9. <https://doi.org/10.1038/nbt1215>.
- [19] KOTTURI Maya F, et al. The CD8+ T-cell response to lymphocytic choriomeningitis virus involves the L antigen: uncovering new tricks for an old virus. *J Virol* 2007;81(10):4928–40. <https://doi.org/10.1128/JVI.02632-06>.
- [20] JENSEN Kamilla Kjaergaard, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* 2018;154(3):394–406. <https://doi.org/10.1111/imm.12889>.
- [21] NIELSEN Morten, LUNDEGAARD Claus, LUND Ole. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinf* 2007;8(1):1–12. <https://doi.org/10.1186/1471-2105-8-238>.
- [22] SIDNEY John, et al. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Res* 2008;4(1):1–14. <https://doi.org/10.1186/1745-7580-4-2>.
- [23] STURNIOLO Tiziana, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 1999;17(6):555–61. <https://doi.org/10.1038/9858>.
- [24] GREENBAUM Jason, et al. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* 2011;63(6):325–35. <https://doi.org/10.1007/s00251-011-0513-0>.
- [25] OSEROFF Carla, et al. Molecular determinants of T cell epitope recognition to the common Timothy grass allergen. *J Immunol* 2010;185(2):943–55. <https://doi.org/10.4049/jimmunol.1000405>.
- [26] JESPERSEN Martin Closter, et al. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 2017;45(W1):W24–9. <https://doi.org/10.1093/nar/gkx346>.
- [27] CORPET Florence. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 1988;16(22):10881–90. <https://doi.org/10.1093/nar/16.22.10881>.
- [28] MITAKU Shigeki, HIROKAWA Takatsugu, TSUJI Toshiyuki. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane–water interfaces. *Bioinformatics* 2002;18(4):608–16. <https://doi.org/10.1093/bioinformatics/18.4.608>.
- [29] DOYTCINOVA Irini A.; FLOWER Darren R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* 2007;8(1):1–7. <https://doi.org/10.1186/1471-2105-8-4>.
- [30] SHARMA N, PATTYAL S, DHALL A, PANDE A, ARORA C, RAGHAV A. GPs. AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE

- epitopes. *Brief Bioinform* 2021;22(4):bbaa294. <https://doi.org/10.1093/bib/bbaa294>. Jul 20.
- [31] SAHA Sudipto, RAGHAVA GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* 2006;34(2):W202–9. <https://doi.org/10.1093/nar/gkl343>. suppl.
- [32] GASTEIGER Elisabeth, et al. ExpASY: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003;31(13):3784–8. <https://doi.org/10.1093/nar/gkg563>.
- [33] Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* 2002:310–22.
- [34] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9. <https://doi.org/10.1038/s41586-021-03819-2>.
- [35] The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 2020;49:D480–9.
- [36] Mitchell AL, et al. MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Res* 2020;48:D570–8.
- [37] Steinegger M, Mirdita M, Söding J. Protein-level assembly increases protein sequence recovery from metagenomic samples manifold. *Nat Methods* 2019;16:603–6. <https://doi.org/10.1038/s41592-019-0437-4>.
- [38] Mirdita M, et al. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res* 2017;45:D170–6.
- [39] Consortium wwPDB. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res* 2018;47:D520–8.
- [40] WEBB B, Sali A. Protein structure modeling with MODELLER. *Methods Mol Biol* 2014;1137:1–15. https://doi.org/10.1007/978-1-4939-0366-5_1.
- [41] XU Xianjin, YAN Chengfei, ZOU Xiaoqin. MDockPeP: An ab-initio protein–peptide docking server. *J Comput Chem* 2018;39(28):2409–13. <https://doi.org/10.1002/jcc.25555>.
- [42] PORTER Kathryn A, et al. ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. *Bioinformatics* 2017;33(20):3299–301. <https://doi.org/10.1093/bioinformatics/btx216>.
- [43] TROTT Oleg, OLSON Arthur J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;31(2):455–61. <https://doi.org/10.1002/jcc.21334>.
- [44] Laskowski R A, Swindells M B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model* 2011;51:2778–86.
- [45] The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
- [46] HISHAM Yasmin, et al. Identification of highly conserved SARS-CoV-2 antigenic epitopes with wide coverage using reverse vaccinology approach. *Viruses* 2021;13(5):787. <https://doi.org/10.3390/v13050787>.
- [47] SINGH Nidhi, et al. Molecular characterization, pathogen-host interaction pathway and in silico approaches for vaccine design against COVID-19. *J Chem Neuroanat* 2020;101874. <https://doi.org/10.1016/j.jchemneu.2020.101874>.
- [48] NAVEED Muhammad, et al. Design of a novel multiple epitope-based vaccine: an immunoinformatics approach to combat SARS-CoV-2 strains. *J Infect Public Health* 2021. <https://doi.org/10.1016/j.jiph.2021.04.010>.
- [49] WALLS Alexandra C, et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020;181(2):281–92. <https://doi.org/10.1016/j.cell.2020.02.058>. e6.
- [50] HARVEY William T, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 2021;19(7):409–24. <https://doi.org/10.1038/s41579-021-00573-0>.
- [51] VERKHIVKER Gennady, et al. Dynamic profiling of binding and allosteric propensities of the SARS-CoV-2 spike protein with different classes of antibodies: mutational and perturbation-based scanning reveals the allosteric duality of functionally adaptable hotspots. *J Chem Theory Comput* 2021. <https://doi.org/10.1021/acs.jctc.1c00372>.
- [52] NELSON Gard, et al. Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y, V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. *bioRxiv* 2021.
- [53] CDC, Centers For Disease Control And Prevention. SARS-CoV-2 Variant Classifications and Definitions. 2021. <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html#print> (access: 20 april 2021) Reference to a dataset.
- [54] SANTOS Keity Souza, et al. Immunodominant B cell epitope in SARS-CoV-2 RBD comprises a B. 1.351 and P. 1 mutation hotspot: implications for viral spread and antiescape. *medRxiv* 2021.
- [55] QUINONEZ Elena, et al. Structural analysis of the novel variants of SARS-CoV-2 and forecasting in North America. *Viruses* 2021;13(5):930. <https://doi.org/10.3390/v13050930>.
- [56] WALL Emma C, et al. Neutralising antibody activity against SARS-CoV-2 VOCs B. 1.617. 2 and B. 1.351 by BNT162b2 vaccination. *Lancet North Am Ed* 2021;397(10292):2331–3. [https://doi.org/10.1016/S0140-6736\(21\)01290-3](https://doi.org/10.1016/S0140-6736(21)01290-3).
- [57] CHAUHAN Nidhi, et al. Interpretative immune targets and contemporary position for vaccine development against SARS-CoV-2: A systematic review. *J Med Virol* 2021;93(4):1967–82. <https://doi.org/10.1002/jmv.26709>.
- [58] LEE Chloe H, et al. To what extent does MHC binding translate to immunogenicity in humans? *Immuno Informatics* 2022. <https://doi.org/10.1016/j.immuno.2021.100006>.