

EpiCurator: an immunoinformatic workflow to predict and prioritize SARS-CoV-2 epitopes

Cristina S. Ferreira¹, Yasmin C. Martins¹, Rangel Celso Souza¹ and Ana Tereza R. Vasconcelos¹

Bioinformatics Laboratory, National Laboratory of Scientific Computation, Petrópolis, Rio de Janeiro, Brazil

ABSTRACT

The ongoing coronavirus 2019 (COVID-19) pandemic, triggered by the emerging SARS-CoV-2 virus, represents a global public health challenge. Therefore, the development of effective vaccines is an urgent need to prevent and control virus spread. One of the vaccine production strategies uses the *in silico* epitope prediction from the virus genome by immunoinformatic approaches, which assist in selecting candidate epitopes for *in vitro* and clinical trials research. This study introduces the EpiCurator workflow to predict and prioritize epitopes from SARS-CoV-2 genomes by combining a series of computational filtering tools. To validate the workflow effectiveness, SARS-CoV-2 genomes retrieved from the GISAID database were analyzed. We identified 11 epitopes in the receptor-binding domain (RBD) of Spike glycoprotein, an important antigenic determinant, not previously described in the literature or published on the Immune Epitope Database (IEDB). Interestingly, these epitopes have a combination of important properties: recognized in sequences of the current variants of concern, present high antigenicity, conservancy, and broad population coverage. The RBD epitopes were the source for a multi-epitope design to *in silico* validation of their immunogenic potential. The multi-epitope overall quality was computationally validated, endorsing its efficiency to trigger an effective immune response since it has stability, high antigenicity and strong interactions with Toll-Like Receptors (TLR). Taken together, the findings in the current study demonstrated the efficacy of the workflow for epitopes discovery, providing target candidates for immunogen development.

Submitted 3 July 2021
Accepted 4 November 2021
Published 30 November 2021

Corresponding author
Ana Tereza R. Vasconcelos,
atr@lncc.br

Academic editor
Elliot Lefkowitz

Additional Information and
Declarations can be found on
page 19

DOI 10.7717/peerj.12548

© Copyright
2021 Ferreira et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Genomics, Virology, Computational Science, COVID-19

Keywords Epitopes prediction, SARS-CoV-2, HLA alleles binding affinity, Immunoinformatic approaches, EpiCurator

INTRODUCTION

The emergence of the SARS-CoV-2 infection, causing COVID-19 disease, has spread rapidly worldwide and represents a global challenge for public health ([Cohen & Normile, 2020](#); [Chakraborty et al., 2020](#)). This virus was first reported in Wuhan in December 2019 ([Huang et al., 2020](#)) and quickly evolved, with the emergence of several variants ([Cella et al., 2021](#); [Brüssow, 2021](#)). According to the World Health Organization (WHO) classifications, there are four variants of concern (VOCs) currently spread worldwide designated as Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), and Delta (B.1.617.2) ([Faria et al., 2020](#); [Rambaut et al., 2020](#); [WHO, 2021](#); [Tegally et al., 2021](#)). In addition, the VOCs are

characterized by the possibility to impact the disease severity, such as the possible increased risk of hospitalization, mortality, and capability of evading vaccination-induced immune response ([Prévost & Finzi, 2021](#); [Geers et al., 2021](#)). The constant emergence of new variants keeps the contagiousness of SARS-CoV-2, which increases the uncertainty of virus spread ([Brüssow, 2021](#); [Naveca et al., 2021](#)).

Current research aims to develop effective interventions for controlling and preventing the COVID-19 pandemic; furthermore, vaccination is still the most economical and effective approach to prevent infection by the virus ([Shang et al., 2020](#)). At the time of writing, 22 vaccines against SARS-CoV-2 have been approved by at least one country ([McGill COVID19 Vaccine Tracker Team, 2021](#)). Five vaccines were approved for emergency use authorization and listed by WHO Emergency Use Listing (EUL) ([Bulla, 2021](#); [Mascellino et al., 2021](#)). The Pfizer/BioNTech vaccine is based on messenger RNA (mRNA), coding for viral spike (S) protein ([Badiani et al., 2020](#)). The Moderna vaccine is a lipid nano-particle-encapsulated mRNA-based vaccine that encodes a full-length spike ([Mahase, 2020](#)). The Johnson & Johnson (Janssen) is a recombinant, non-replicating adenovirus vector encoding a full-length S protein ([Livingston, Malani & Creech, 2021](#)). The Oxford-AstraZeneca is a chimpanzee adenovirus vectored DNA vaccine ([Knoll & Wonodi, 2021](#)), and the CoronaVac is an inactivated SARS-CoV-2 virus vaccine ([Gao et al., 2020](#); [Mallapaty, 2021](#)).

The current vaccines are generally based on B cell immunity with neutralizing antibody production ([Thanh Le et al., 2020](#); [Siracusano, Pastori & Lopalco, 2020](#); [Yoshida et al., 2021](#)). However, several studies report the capability of a new mutation interrupting the binding with some neutralizing antibodies ([Greaney et al., 2021](#); [Hoffmann et al., 2021](#); [Andreano et al., 2021](#)), which could clarify the spread of the current VOCs. The main reasons for this capability are the several changes (mutation and deletion) in the S protein, which is responsible by the increased affinity between the receptor-binding domain (RBD) and the human cellular receptor angiotensin-converting enzyme 2 (ACE2), promoting the antibodies escape ([Thomson et al., 2021](#); [Greaney et al., 2021](#); [Liu et al., 2021](#); [Rotondo et al., 2021](#)).

To improve the vaccine's effectiveness against SARS-CoV-2, global efforts in resources, cooperation, and innovation have contributed to the accelerated development of COVID-19 vaccines ([Bloom et al., 2021](#)). These efforts lead various researchers to carry out diversified methodologies of vaccine design such as a peptide-based vaccine, virus-like particle, replicating and non-replicating viral vectors, DNA or RNA, live attenuated virus, recombinant designed proteins, nanoparticles vaccine, and inactivated virus ([Medhi et al., 2020](#); [Krammer, 2020](#); [Di Natale et al., 2020](#); [Kyriakidis et al., 2021](#); [Shahcheraghi et al., 2021](#)).

This work highlights the peptide-based vaccine described in several studies about SARS-CoV-2 vaccines ([Malonis, Lai & Vergnolle, 2020](#); [Chakraborty et al., 2020](#); [Crooke et al., 2020](#); [Fatoba et al., 2021](#); [Naveed et al., 2021](#)). Currently, ten from 33 vaccines in phase 3 clinical trial, and four from 22 approved vaccines have a peptide-based design ([McGill COVID19 Vaccine Tracker Team, 2021](#)). The advantages of peptide-based vaccines include their capability to target very specific epitopes decreasing the risks associated with allergic

and autoimmune responses, besides involving minimal viral components to stimulate adaptive immunity (*Di Natale et al., 2020*). Additionally, their chemical or recombinant cloning synthesis allows large-scale production with low costs and high reproducibility (*Sun, 2013; Skwarczynski & Toth, 2016; Hudu, Shinkafi & Umar, 2016*).

The peptide-based vaccine design requires the immunoinformatic approach as part of the computational vaccinology strategy for epitopes prediction (*Ramana & Mehla, 2020; Oli et al., 2020; Lu et al., 2021*). This approach regards the wide availability of the SARS-CoV-2 NGS (Next-Generation Sequencing) information associated with human leukocyte antigen (HLA) profile (*Kazi et al., 2018; Oli et al., 2020; Sharma et al., 2020*) to identify T cell epitopes. Therefore, these epitopes have the capability to effectively bind to HLA molecules activating a long-lasting immune response mediated by CD8+ and CD4+ T cells (*Fast, Altman & Chen, 2020; Wang & Gui, 2020*).

T cell epitopes offer advantages for vaccine design since it does not depend on the recognition of structural proteins (*Bashir et al., 2021; Redd et al., 2021*) and are less affected by deletions and mutations of emergent variants (*Ribes, Chacour & Moncunill, 2021; Jin et al., 2021*). Concerning COVID-19 immune response, T cell epitopes have the potential to provide long-term protection from SARS-CoV-2. This characteristic allows the detection of memory T cell responses to multiple SARS-CoV-2 proteins, which might contribute to disease control (*Chen & John Wherry, 2020; Karlsson, Humbert & Buggert, 2020; Sette & Crotty, 2021*).

To provide effective T cell epitopes for *in vitro* peptide-based vaccine design a robust, refined and accurate *in silico* selection of epitopes is crucial. Despite the many immunoinformatic tools for epitope prediction, the curation for epitope selection is still limited and needs different web servers to complete the analysis. In this paper, we focus on a computational prediction, curation and validation of SARS-CoV-2 epitopes. We have as a central proposal a workflow (EpiCurator) that brings together different approaches for accurate selection of epitopes, providing the refined identification of promising SARS-CoV-2 epitopes. To validate the efficacy of this new tool, we use samples of circulating Brazilian lineages available in GISAID (<https://www.gisaid.org/>) (*Elbe & Buckland-Merrett, 2017*) from December 2020 to April 2021.

MATERIALS & METHODS

Genome retrieval and protein annotation

Genome sequences of 1,652 SARS-CoV-2 genome isolated in Brazil were retrieved from the GISAID database (*Elbe & Buckland-Merrett, 2017*) available from December 2020 to April 2021 (Table S1). The lineage distribution of the retrieved genomes includes P.1 (Gamma, $n = 770$), P.2 (Zeta, $n = 525$), B.1.1.28 ($n = 223$) and B.1.1.33 ($n = 136$). These genomes were analyzed by The Viral Annotation Pipeline and iIdentification (VAPiD) (*Shean et al., 2019*) to determine the amino acid sequence for all SARS-CoV-2 proteins using as reference NC_045512.2 from the NCBI database (<https://www.ncbi.nlm.nih.gov/>). The final genome processing includes clustering the amino acid sequences for each protein using the CD-HIT package (*Li & Godzik, 2006*) with a 100% sequence identity threshold.

Spike protein comparison

Clustal Omega and MView tool performed a homology analysis for the Spike protein (*Sievers & Higgins, 2018*); (*Brown, Leroy & Sander, 1998*) among 100 random samples of Brazilian lineages and the VOCs, retrieved from the GISAID database (*Elbe & Buckland-Merrett, 2017*), to measure the identity of the sequences to Gamma (P.1) samples.

Epitope prediction

The T and B cell epitopes prediction was performed using the softwares NetCTL, NetMHCpan, and NetMHCIIpan (<https://services.healthtech.dtu.dk/>) that allow the high-throughput computing analysis. These predictors support FASTA files containing amino acids marked with “X” in the sequence expanding the possibility of public genomes analysis, even with minor sequencing errors. This prediction features decreasing the pre-processing sequence steps and avoids the false-positive epitopes provided by joining sequences.

Prediction of SARS-CoV-2 epitopes

To predict CD8+ T cell epitopes, the FASTA sequences of SARS-CoV-2 proteins are processed using NetCTL v1.2 (*Larsen et al., 2007*). First, the sequences and the HLA class I supertypes, provided by the software, are submitted to select 9-mer peptides (*Chen et al., 1994; Sakaguchi et al., 1997; Gfeller et al., 2018*), then peptides with amino acids marked with “X” are removed (Fig. 1A). The predicted peptides in NetCTL are further processed using NetMHCpan v.4.0 software (*Hoof et al., 2009; Jurtz et al., 2017*) to identify epitopes with strong binding affinity to HLA class I alleles. The prediction parameter is based on the predicted percentile rank $\leq 0.5\%$ and half-maximal inhibitory concentration (IC50) $< 500 \text{ nM}$ (*Chen et al., 1994; Sakaguchi et al., 1997; Gfeller et al., 2018*) (Fig. 1A). To assess binding affinity, alleles of HLA-A, B, and C loci were selected from Allele Frequency Net Database (*Gonzalez-Galarza et al., 2020*) by their Brazilian population frequency $> 5\%$ (Table S2).

To predict CD4+ T cell epitopes and estimate binding affinity to HLA class II molecules, the FASTA sequences are processed using NetMHCIIpan v3.2 software (*Greenbaum et al., 2011*). This tool selects the epitopes with 15-mer peptide lengths based on the predicted percentile rank $\leq 2.0\%$ and IC50 $< 500 \text{ nM}$ (Fig. 1A). For the affinity prediction, the loci HLA-DRB1, HLA-DPA1-DPB1, and HLA-DQA1-DQB1 were selected by the phenotypic frequency $> 5\%$ in the Brazilian population (Table S2) from Allele Frequency Net Database (*Gonzalez-Galarza et al., 2020*). For both cells (CD8+ and CD4+ T cell), the epitopes with broad HLA affinity coverage (≥ 3 alleles) are filtered (Fig. 1A).

The prediction of linear B cell epitopes from SARS-CoV-2 structural proteins was performed by BepiPred v.2.0 software (*Jespersen et al., 2017*), with a threshold of 0.5 (corresponding specificity > 0.817 and sensitivity < 0.292) (Fig. 1A). Only the epitopes with more than seven and less than 50 amino acid residues in length and sequence without amino acids marked with “X” are considered for subsequent curation analysis.

Accurate selection of epitopes - EpiCurator

We developed a rigorous analysis workflow for accurate selection of epitopes, the EpiCurator, bringing together a set of filters according to pre-established criteria,

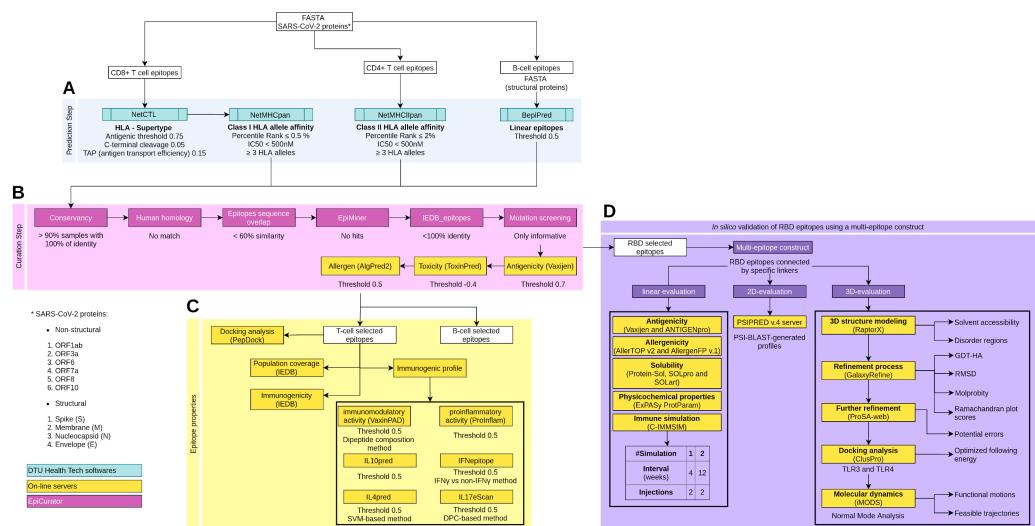


Figure 1 Schematic overview of the prediction, accurate selection and epitopes validation to identify new SARS-CoV-2 epitopes. The pipeline comprises four main analyses: prediction and curation epitopes selection, immune properties evaluation and multi-epitope validation. (A) The prediction encompasses four DTU Health Tech softwares (blue boxes) subdivided according to the type of identified epitopes: CD8+ T cell, CD4+ T cell and B cell epitopes. (B) The curation step uses the predicted epitopes for an accurate selection comprising our proposal EpiCurator workflow (pink boxes) and online server to evaluate the antigenicity, toxicity and allergen analysis. (C) The pipeline has a set of individual analyses that identify the population coverage, immunogenicity and other immunogenic properties using on-line servers (yellow boxes). (D) Additionally, a *in silico* validation of final RBD selected epitopes has been performed using a multi-epitope construct. Their linear, 2D and 3D sequences are evaluated using online servers (yellow boxes) to characterize the multi-epitope and the construct-TLR complexes as stable and immunogenic. Each analysis has its respective parameters presented below the boxes.

Full-size DOI: 10.7717/peerj.12548/fig-1

optimizing the analysis since it brings tools available to use by high-throughput computing architectures. They also group a series of analyses to guarantee the selection of unpublished and qualified epitopes. The workflow allows the identification of epitopes by the following analysis: conservancy, homology with the human genome, the overlap between epitopes of different classes, and the identification of epitopes previously published in PubMed Central® (PMC) or available in the IEDB database ([Vita et al., 2019](#)) beyond identifying their protein coordinates and mutations (Fig. 1B).

Prediction of epitope conservancy

The first module of EpiCurator calculates the conservancy of a predicted epitopes list in SARS-CoV-2 genomes using the BLAST command-line tools ([Madden, 2020](#)) (Fig. 1B). A customized BLAST database, optimized for shorter sequences (blastp-short task), was generated with proteins of some SARS-CoV-2 circulating lineages from Brazil (B.1.1.28, P.1, P.2) retrieved from the GISAID database (2,787 genome sequences) ([Elbe & Buckland-Merrett, 2017](#)). This database allows the comparison among the sequences of the predicted epitopes and the SARS-CoV-2 proteins. Thus, the analysis reports the percentage of identity (conservancy) per epitope using four criteria: 100% of identity; above 90%; between 70%

and 90%, and less than 70%. The epitopes conserved with 100% of identity in at least 90% of the genomes are used for further analysis (Fig. 1B).

Human homology

This module uses the human proteins dataset from Ensembl (GRCh38.p13) to identify the predicted epitopes sequence in the human genome. The workflow keeps the human protein sequences in memory to enhance the analysis readiness, searching strictly for the corresponding epitope sequences (exact match). Only the unmatched epitopes are selected, returning a filtered list of epitopes without human homology (Fig. 1B).

Epitope sequence overlap

A comparison is performed between the sequence of the epitopes derived from three groups of prediction (B Cell, HLA Class I, and HLA Class II T Cell). This analysis calculates the intersection and the identity among the lists of epitopes belonging to these groups returning four reports: (i) the intersection of all groups, (ii) B cell x Class I epitopes, B cell x Class II epitopes, and (iii) Class I x Class II epitopes. The result keeps the epitopes with less than 60% similarity in each report (Fig. 1B).

Search for epitopes from published articles - EpiMiner

The EpiMiner was developed to execute an automatic search of the predicted epitopes list on Pubmed and PMC published papers (Canese & Weis, 2013) (Fig. 1B). It is a pipeline performed in four steps. The first retrieves Pubmed and PMC articles that include the epitopes sequence. The second step extracts the sentences of the body, abstract or tables of the article, highlights the figure captions and breaks apart supplementary table data. The third step executes natural language processing techniques such as tokenization to divide the sentences into words and part-of-speech tagging to filter nouns and verbs (Chowdhury, 2005). These techniques contribute to reducing search time for epitope sequence recognition, eliminating uninformative sentences. The fourth step executes an entity recognition to identify epitopes sequence classified as nouns, saving the sentences and their respective publication into a report. Epitopes information is not searched if it is a part of the pictures or is in supplementary file in text format (*i.e.*, .doc, .docx, .txt, .pdf).

IEDB matching

To perform this analysis, the complete ensemble of epitopes from the Immune Epitope Database (IEDB) v3 release (https://www.iedb.org/database_export_v3.php) was retrieved (accessed on July 2021), and the structured files were parsed to filter the epitopes belonging to SARS-CoV-2 organisms ($n = 1,268$) (Vita et al., 2019). This filtered file is used to calculate the identity between the predicted list of epitopes and the ones retrieved, apprising the IEDB epitope ID and the respective similarity (Fig. 1B).

Mutation screening

Epitopes mutation is reported by comparing the epitope sequence with the SARS-CoV-2 Wuhan protein sequence and assigning the epitopes' coordinates in the respective whole protein. The alignment is conducted with the blastp function of the BLAST command-line tools (Madden, 2020). The SARS-CoV-2 reference protein sequences used for alignment are

published on the Uniprot Database (<https://covid-19.uniprot.org/>). This analysis describes the coordinates and presence of mutations in the predicted epitopes (Fig. 1B).

Epitope properties

Evaluation of antigenicity, toxicity, and immunogenic profile

The VaxiJen v2.0 server was applied to analyze the antigenicity of the predicted B cell and T cell epitopes with a conservative score threshold of 0.7 (Doytchinova & Flower, 2007a; Doytchinova & Flower, 2007b) (Fig. 1B). The toxicity is retrieved from the ToxinPred online server with support vector machine (SVM) based methods (threshold –0.4) and e-value cut-off 0.01 (Gupta et al., 2013) (Fig. 1B). The allergenic properties are retrieved from the AlgPred2 online server with a hybrid prediction model (threshold 0.5) (Sharma et al., 2020) (Fig. 1B). Despite being part of the accurate selection (Fig. 1B), they are not available for incorporation into the analytic workflow (EpiCurator) which prevents us from optimizing their assay. In addition, several online servers were used to evaluate the immunogenic profile of the T cell epitopes (Fig. 1C). The immunogenicity predictions were performed by the IEDB server (<https://www.iedb.org/>) (Paul et al., 2015; Calis et al., 2013; Dhanda et al., 2018; Vita et al., 2019). Likewise, their capability to induce interferon-gamma (IFN γ), interleukin-4 (IL-4), interleukin-10 (IL-10), and interleukin-17 (IL-17) (Dhanda, Vir & Raghava, 2013; Dhanda et al., 2013; Gupta et al., 2017; Nagpal et al., 2017) was evaluated. Additionally, their proinflammatory activity (Gupta et al., 2016) and immunomodulatory potential (Nagpal et al., 2018) was verified. The servers and parameters to evaluate the immunogenic profile are shown in Fig. 1C.

Estimation of population coverage

The IEDB's Population Coverage online tool (<http://tools.iedb.org/population/>) is used to analyze how T cell epitopes-HLA binding alleles diverge across ethnicities, regions, and countries around the world (Bui et al., 2006; Vita et al., 2019) (Fig. 1C). The predicted epitopes and their respective HLA binding alleles (Table S3) were inputted in the IEDB tool with separated allele class option and a list of countries/regions (Argentina, Brazil, England, France, Italy, Spain, United States, and World) was selected.

Docking analysis of the HLA-epitope complex

To validate the binding affinity of the predicted epitopes with HLAs structures, docking analysis was performed using the PepDock tool (Lee et al., 2015) of the GALAXY web server (Fig. 1C). To docking analysis, the Protein Data Bank archive (PDB) of eight HLA alleles (HLA-A*01:01, HLA-A*02:01, HLA-B*08:01, HLA-C*12:03, HLA-DRB1*03:01, HLA-DRB1*04:01, HLA-DRB1*12:02 and HLA-DRB1*15:01) was retrieved from the pHLA3D database (Menezes Teles et al., 2019) and RCSB PDB database (Berman et al., 2000; Burley et al., 2021).

The PepDock uses the HLA alleles PDBs and the epitopes sequence to perform a template-based model selection and then proceeds to the docking and refinement processes to optimize the energy score ranking ten complex models. The best model of each HLA-epitope complex was selected based on the major similarity score of protein structure and interaction beyond the highest estimated accuracy. Additionally, to identify the free energy

and the residue's contacts of the selected complexes, the Prodigy tool ([Xue et al., 2016](#)) and Chimera software ([Goddard, Huang & Ferrin, 2005](#)) were used respectively.

Genomes for EpiCurator pairwise comparison validation

Protein sequences for SARS-CoV-2 isolates reported by [Crooke et al. \(2020\)](#) were identified and retrieved from the Virus Pathogen Resource (ViPR) database ($n = 641,635$); additionally, six genome sequences reported by [Kiyotani et al. \(2020\)](#), two sequences for N and S protein reported by [Chen et al. \(2020\)](#), and five S sequences reported by [Chukwudzie et al. \(2021\)](#) were retrieved from NCBI GenBank. These samples were processed using prediction parameters of the HLA alleles reported by the authors with the approach reported in this Methods section (Epitope prediction and Accurate selection analysis). Regarding item Accurate selection of epitopes - EpiCurator of the Methods, we used only the three main analysis steps (conservancy, human homology and IEDB matching) to compare the selected epitopes ensemble by the papers, and ones chosen independently by our approach, leaving out the EpiMiner analysis since all the epitopes are published.

Multi-epitope construct and structural modelling for EpiCurator validation

The RBD epitopes ($n = 11$) were used to construct a multi-epitope sequence connected by specific linkers (Fig. S1). The linker aimed to separate the epitopes, so that it improved their expression, folding and stability beyond to prevent their fusion and facilitate the immune processing of antigen ([Arai et al., 2001](#); [Kar et al., 2020](#)). In the multi-epitope arrangement are also added an adjuvant (*Escherichia coli* 50S ribosomal protein L7/L12 (UniProt P0A7K2)) in the N-terminal sequence and a histidine hexamer in the C-terminal portion (Fig. 1D, Fig. S1).

Linear and secondary structure evaluation

To evaluate the properties and immune profile of multi-epitope, its linear sequence was submitted to several analyses (Fig. 1D) as follow: antigenicity ([Doytchinova & Flower, 2007a](#); [Magnan et al., 2010](#)), allergenicity ([Dimitrov et al., 2014b](#); [Dimitrov et al., 2014a](#)), and solubility for cell-free expression analysis ([Hebditch et al., 2017](#)), for overexpression analysis ([Magnan, Randall & Baldi, 2009](#)), and structurally solubility profile ([Hou et al., 2020](#)). In addition, we assess its physicochemical properties ([Gasteiger et al., 2005](#)). The servers and parameters to linear multi-epitope properties evaluation are shown in Fig. 1D.

The linear sequence was also analyzed by the C-IMMSIM server (<https://kraken.iac.rm.cnri.it/C-IMMSIM/>) to evaluate the *in silico* immune profile of the multi-epitope construct ([Rapin et al., 2010](#)). Two simulations were performed with intervals of 4 or 12 weeks ([Saad-Roy et al., 2021](#); [Cobey et al., 2021](#)) (Fig. 1D). Furthermore, the secondary structure of the multi-epitope was evaluated with PSIPRED v.4 with an accuracy of 84.2% ([Jones, 1999](#); [McGuffin, Bryson & Jones, 2000](#); [Buchan & Jones, 2019](#)) (Fig. 1D).

Multi-epitope 3D structure modelling, refinement, and evaluation

The tertiary structure multi-epitope construct is modeled by the RaptorX server (<http://raptordx.uchicago.edu/ContactMap>) that predicts structural properties such as

solvent accessibility (ACC) and disorder regions (DIS) (*Källberg et al., 2014*) (Fig. 1D). This tertiary structure was submitted to a refinement process using the GalaxyRefine (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) server which improves global and local model quality by rebuilding all side-chain conformations and applying structural relaxations (*Heo, Park & Seok, 2013*). The quality of the refined model is assessed by several parameters, as reported in Fig. 1D. The refined 3D structure is further evaluated by the ProSA-web server (<https://prosa.services.came.sbg.ac.at/prosa.php>) to validate the structural models checking for potential errors (*Berman et al., 2000; Wiederstein & Sippl, 2007*) (Fig. 1D). The final refined 3D structure was used to perform a docking by the ClusPro server (<https://cluspro.bu.edu>) (*Desta et al., 2020*) using the multi-epitope 3D structure and the Toll-Like Receptors TLR4 (PDB: 2Z63) and the TLR3 (PDB: 3CIG). The complex model with the lowest energy is chosen for each receptor (Fig. 1D). Their stability is assessed with molecular dynamics simulation on the iMODS server (<http://imods.chaconlab.org>) that performs Normal Mode Analysis (NMA) to describe functional motions between macromolecules in complexes and simulates feasible trajectories between two conformations (*López-Blanco et al., 2014*) (Fig. 1D).

RESULTS

Linear epitopes prediction from SARS-CoV-2 genome

Three main epitopes prediction comprise HLA Class I (CD8+ T cell), HLA Class II (CD4+ T cell), and B cell. To predict potential CD8+ T cell epitopes, the NetCTL and NetMHCpan, predictive algorithms were performed for all proteins annotated from 1,652 SARS-CoV-2 genomes (see Methods). Thus 9-mer epitopes were predicted with strong binding affinity assigned by percentile scores (rank) $\leq 0.5\%$ across HLA class I alleles frequent in the Brazilian population (Fig. 1A). Cumulatively reaching the prediction of 5,261 HLA class I epitopes with HLA promiscuity, binding to ≥ 3 HLA alleles, implying broad population coverage. The affinity to HLA-C alleles matches the highest epitope binding locus suggesting it could be the best grooves for these epitopes and would lead to the activation of the cell-mediated response (Table 1).

We also sought to predict potential 15-mer epitopes with binding affinity to HLA class II, using the NetMHCIIpan software (Fig. 1A), reaching 7,649 candidate HLA class II epitopes from the SARS-CoV-2 genomes. They have a strong binding affinity to ≥ 3 HLA alleles across a reference panel of HLA molecules (see Methods). These epitopes showed a preferential affinity for alleles from locus HLA-DRB1 ($n = 3,702$), suggesting that HLA-DRB1 alleles could lead these epitopes to activate the cellular immune response (Table 1).

To pairwise the cellular and humoral immune responses activation, the additional prediction of linear B cell epitopes was performed by BepiPred algorithm identifying 257 epitopes from SARS-CoV-2 structural proteins (Table 1).

Quality assessment analysis of the EpiCurator

A robust and refined accurate selection of epitopes is crucial to improve the development of peptide-derived vaccines. To this goal, the EpiCurator brings together six analyses

Table 1 Summary of SARS-CoV-2-derived epitopes predicted with high binding affinity to each HLA loci.

| Protein | Protein length (amino acids) | Number of epitopes | | | | | | | | | |
|--------------|---------------------------------|--|-------------|-------------|-------------|---|-------------|------------|-------------|-----------------|--|
| | | HLA class I-restricted T cell epitopes (rank ≤ 0.5) | | | | HLA class II-restricted T cell epitopes (rank ≤ 2) | | | | B cell epitopes | |
| | | All | HLA-A | HLA-B | HLA-C | All | HLA-DP | HLA-DQ | HLA-DR | All | |
| Envelope | 75 | 33 | 22 | 26 | 69 | 54 | — | — | 13 | 10 | |
| Membrane | 222 | 99 | 104 | 55 | 127 | 83 | 6 | 6 | 44 | 7 | |
| Nucleocapsid | 419 | 226 | 143 | 205 | 217 | 390 | 21 | 22 | 202 | 147 | |
| ORF10 | 38 | 23 | 12 | 12 | 29 | 10 | — | 1 | 8 | — | |
| ORF1ab | 7096 | 3621 | 2454 | 3517 | 4262 | 5225 | 1291 | 630 | 2608 | — | |
| ORF3a | 275 | 343 | 256 | 298 | 435 | 357 | 46 | 24 | 260 | — | |
| ORF6 | 61 | 32 | 15 | 20 | 26 | 47 | 13 | 4 | 24 | — | |
| ORF7a | 121 | 96 | 65 | 110 | 96 | 102 | 101 | 2 | 71 | — | |
| ORF7b | 43 | 15 | 9 | 7 | 15 | 3 | — | — | 1 | — | |
| ORF8 | 121 | 98 | 51 | 87 | 113 | 139 | 57 | 8 | 35 | — | |
| Spike | 1273 | 675 | 427 | 652 | 775 | 1239 | 531 | 120 | 436 | 93 | |
| TOTAL | - | 5261 | 3558 | 4989 | 6164 | 7649 | 2066 | 817 | 3702 | 257 | |

(Conservancy, Human homology, Epitopes overlap, EpiMiner, IEDB matching, and Mutation screening) to the accurately selected epitopes (Fig. 1B).

To the workflow's quality assessment analysis, we sought to characterize the number of epitopes taken for the main analysis. The epitopes conservancy selected 29.23% of the predicted epitopes with 100% identity across at least 90% of the SARS-CoV-2 samples. Nevertheless, 70.77% of the predicted epitopes do not correspond to the conservancy parameter (Fig. 2, Fig. S2). Human homology analysis identified only 0.1% of the predicted epitopes sequence in the human genome, keeping 99% accurately selected, unmatched with human sequences (Fig. 2, Fig. S2). In addition, the screening by the previously published epitopes (EpiMiner and IEDB matching analysis) allowed selected > 80% of new epitopes, since 17.57% of them have already been described in the literature and/or IEDB server (Fig. 2, Fig. S2). To confirm the effectiveness of the EpiMiner we provide all the articles IDs and DOI in which the epitopes were found in Table S4.

To further assess the quality of the EpiCurator analysis, a pairwise comparison validation was performed. The comparison used NGS data reported by four papers with substantial similarity with our methodologies (Kiyotani *et al.*, 2020; Chen *et al.*, 2020; Crooke *et al.*, 2020; Chukwudzie *et al.*, 2021). The prediction step allows identifying the same HLA class I and HLA class II-restricted T cell epitopes. At the same time, the accurate selection provided by the EpiCurator gathered 51.1% of the paper's epitopes (Table S5). The conservancy across SARS-CoV-2 genomes samples reached 95,6% of the paper's epitopes, with only 4.4% not conserved by our parameters. The human homology step certifies the absence of the article's selected epitopes in the Human genome since the analysis does not recognize the homology of any of them. Interestingly, the IEDB matching step identified 45.6% of the paper's epitopes as having already been published on the IEDB server showing

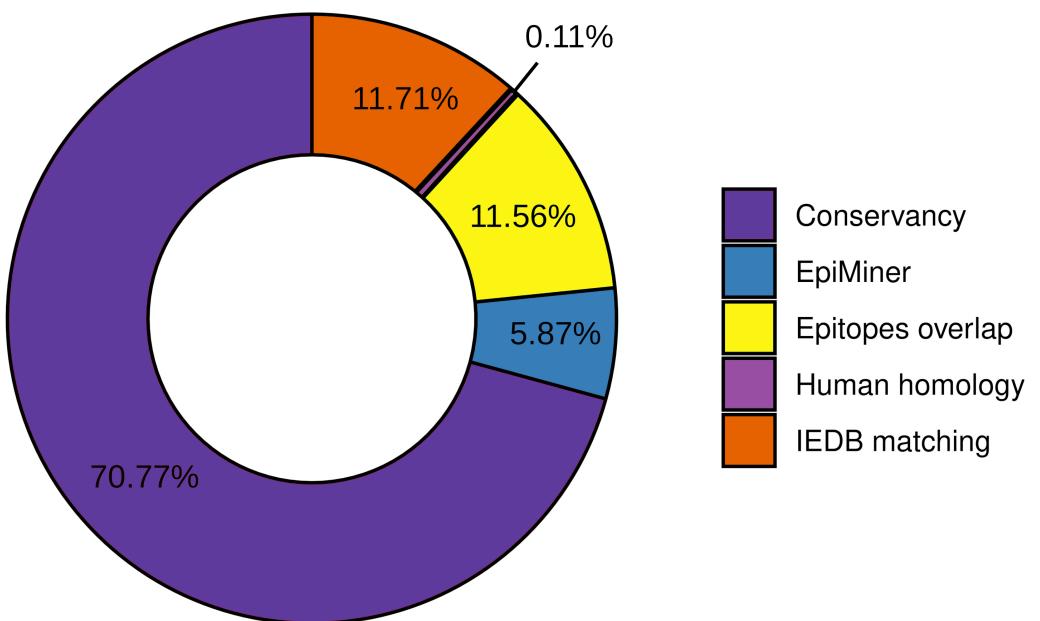


Figure 2 Efficiency of EpiCurator analysis for accurate selection of epitopes. The plot represents the percentage of epitopes removed in each different analysis.

[Full-size](#) DOI: 10.7717/peerj.12548/fig-2

the importance of this step analysis if we want to describe the epitopes by the first time ([Table S5](#)).

Properties of accurately selected epitopes

Assembling the workflow analysis results, 199 (3.78%) HLA class I-restricted T cell epitopes were selected from SARS-CoV-2 proteins, mainly identified in ORF1ab ($n = 154$ (77.4%)) and Spike glycoprotein ($n = 15$ (7.5%)) ([Table 2](#), [Fig. 3A](#)). The epitopes keep a high affinity for HLA-C alleles ([Fig. 3C](#)). The HLA class II epitopes' accurate selection reached 153 (2%) epitopes, also mainly identified in ORF1ab ($n = 111$ (72.5%)) and Spike glycoprotein ($n = 22$ (14.4%)). The HLA class II epitopes showed an affinity prevalence for the HLA-DPA1-DPB1 haplotypes ([Fig. 3D](#)). The comprehensive workflow analysis for B cell epitopes selected 14.6% of predicted epitopes with 60% of the epitopes from the Spike glycoprotein ([Table 2](#)). Interestingly, all epitopes have high antigenicity (0.92 ± 0.30), are non-toxic and non-allergenic ([Table S3](#)).

Furthermore, the T cell epitopes immunogenic potential was assessed to characterize the capability of inducing *in silico* protective immune responses. The analysis identified the profile of HLA class I and HLA class II epitopes respectively as follow: IL-4 inducer activity (26.6% and 23.8%), IFN γ production (22.1% and 39.7%), immunomodulatory activity (2% and 3.3%), and proinflammatory activity (78.9% and 81.4%) ([Table S3](#)).

In addition, to assess the capability to be an *in silico* Brazilian epitopes candidate, we sought to estimate their distribution among the Brazilian circulating lineages at the time of analysis (P.1, P.2 and B.1.1.28). Firstly, we identify the number of epitopes in common across P.1, P.2 and B.1.1.28 lineages identifying the highest proportion of HLA class II

Table 2 Summary of final SARS-CoV-2-derived epitopes accurately selected to each SARS-CoV-2 protein.

| Protein | HLA class I-restricted T cell epitopes | | HLA class II-restricted T cell epitopes | | B cell epitopes | |
|--------------|--|-------------------------------|---|-------------------------------|-----------------|-------------------------------|
| | Prediction (n) | Curation (n - %) [*] | Prediction (n) | Curation (n - %) [*] | Prediction (n) | Curation (n - %) [*] |
| Envelope | 33 | — | 54 | 1 (1.85) | 10 | 1 (10) |
| Membrane | 99 | 1 (1.01) | 83 | — | 7 | — |
| Nucleocapsid | 226 | 10 (4.25) | 390 | 5 (1.28) | 147 | 14 (9.52) |
| ORF10 | 23 | — | 10 | — | — | — |
| ORF1ab | 3621 | 154 (4.25) | 5225 | 111 (2.12) | — | — |
| ORF3a | 343 | 11 (3.21) | 357 | 4 (1.12) | — | — |
| ORF6 | 32 | — | 47 | — | — | — |
| ORF7a | 96 | 3 (3.12) | 102 | — | — | — |
| ORF7b | 15 | — | 3 | — | — | — |
| ORF8 | 98 | 5 (5.10) | 139 | 10 (7.19) | — | — |
| Spike | 675 | 15 (2.22) | 1239 | 22 (1.77) | 93 | 21 (22.58) |
| TOTAL | 5261 | 199 (3.78) | 7649 | 153 (2.00) | 257 | 36 (14.01) |

Notes.

*Percentage related to number of predicted epitope per protein.

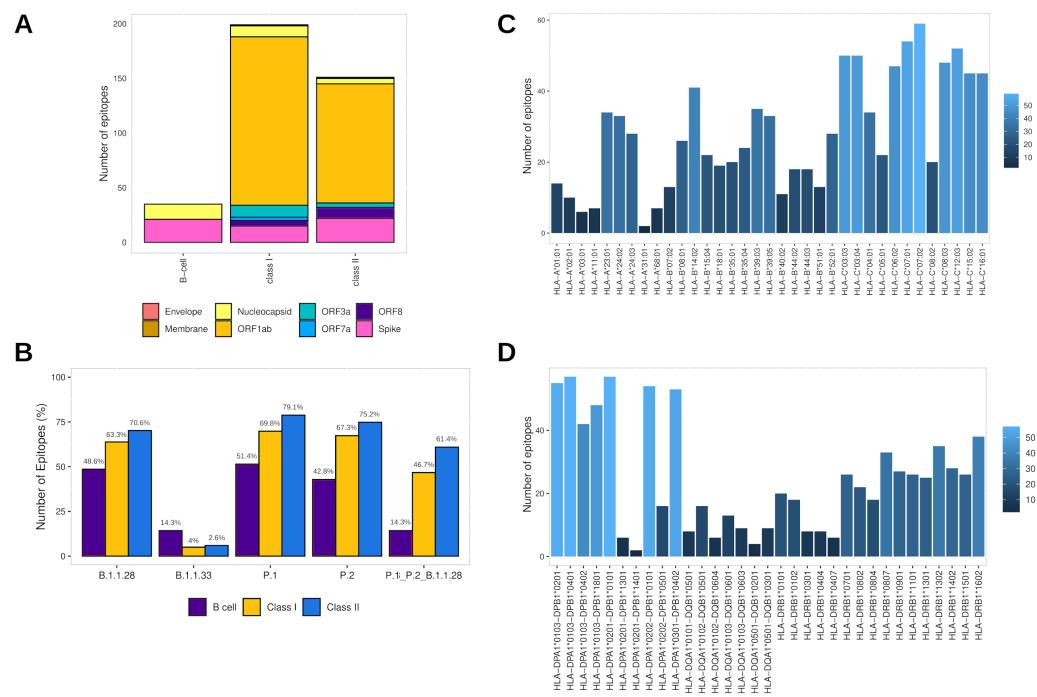


Figure 3 Identification of final SARS-CoV-2-derived HLA class I- and class II-binding epitopes. (A) Distribution of the accurately selected epitopes in the SARS-CoV-2 proteins. Each color represents a distinct protein. (B) Distribution of the accurately selected epitopes in each current Brazilian lineage. Each color represents a distinct epitope type. (C) Distribution of epitopes in HLA class I alleles. The color degree represents the variation of the number of epitopes. (D) Distribution of epitopes in HLA class II alleles. The color degree represents the variation of the number of epitopes.

Full-size DOI: 10.7717/peerj.12548/fig-3

epitopes (61.4%), following by HLA class I (46.7%) and B cell (14.3%) ([Fig. 3B](#)). Thereafter, considering the most representative lineage in Brazil, at the time data was retrieved (P.1), remarkably, proportions highest than 50% for all the epitopes were identified, with 70.1% of HLA class II epitopes, 69.8% of HLA class I, and 51.4% of B cell epitopes in P.1 ([Fig. 3B](#)).

In the last analysis, considering the HLA promiscuity epitopes selection, the population coverage for the T cell epitopes associated with their respective HLA allele binding ([Table S3](#)) was estimated by the IEDB server. Notably, the T cell epitopes have a wide population coverage, presenting 99.52% (HLA class I epitopes) and 100% (HLA class II epitopes) of cumulative Brazilian population coverage and 98.45% (HLA class I epitopes) and 99.87% (HLA class II epitopes) of worldwide population coverage ([Table S6](#)).

Epitope-specific RBD Spike as a baseline for validation of EpiCurator selection

The EpiCurator allows the selection of the majority of the epitopes in the ORF1ab and Spike glycoprotein. The SARS-CoV-2 Spike glycoprotein has greater prominence concerning the virus and host interaction and has been the main target in epitopes prediction ([Shang et al., 2020](#); [Walls et al., 2020](#)). Therefore, we assumed the spike epitopes as a baseline to validate the EpiCurator accurate selection. The Spike epitopes ($n = 58$) were mainly in the N-terminal domain (NTD) (53.4%) and receptor-binding domain (RBD) (18.9%) ([Fig. 4](#)).

The RBD epitopes ($n = 11$) comprise three of HLA class I (461-NYNYRYRLF-469, 506-QSYGFQPTY-514, 516-FGYQPYRVV-524), three of HLA class II (322-EKGIYQTSNFRVQPI-336, 322-EKGIYQTSNFRVQPR-336, 443-TGCVIAWNSKNLDSK-457), and five B cell epitopes (332-RVQPTES-338, 424-APGQTGK-430, 424-APGQTGT-430, 455-DSKVGGNYN-463, 474-LKPFERD-480) ([Fig. 4](#)).

To evaluate the robust and refined RBD epitopes selection, we perform a diversified analysis. They unveil the highest antigenicity (1.1 ± 0.27) ([Fig. 5A](#)) reflecting their ability to bind molecules of adaptive immunity. Notably, their affinity for several HLA alleles are high (0.49 ± 0.15) ([Fig. 5A](#)) and achieves promiscuity with ≥ 5 HLA allele binding per epitope ([Fig. 5B](#)). The RBD epitopes HLA data pairwise with population coverage that remains at around 80% in the Brazilian and worldwide population ([Fig. 5B](#)). To further characterize the RBD epitopes, they reach the conservancy of over 99% across SARS-CoV-2 genomes samples of lineages P.1, P.2 and B.1.1.28 published on GISAID (See Methods) ([Table S3](#)).

The high conservancy observed suggests that epitope sequences are shared in the circulating Brazilian lineages. Considering the natural homology among the SARS-CoV-2 lineages but regardless of the epitopes' conservancy with other lineages, we expand the identification of the RBD epitopes for diversified samples available in GISAID at the time of writing. Interestingly these epitopes were identified in more than 1 million samples for around 1 thousand lineages of GISAID ([Table S7](#)). These findings include some samples of current VOCs: Alpha (B.1.1.7), Beta (B.1.351), and Delta (B.1.617.2) ([Table S7](#)). Concern about the natural lineages' homology mentioned, we perform a comparison among spike glycoprotein of described Brazilian lineages and VOCs. The evaluation of random 100

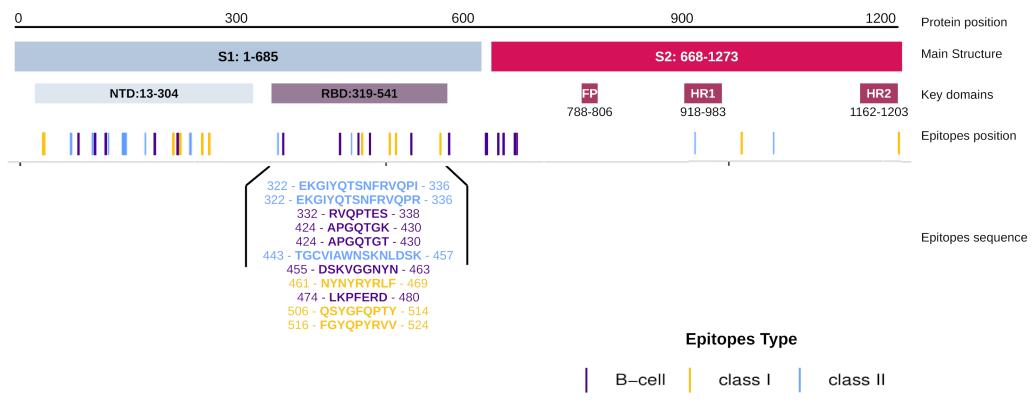


Figure 4 Distribution of the accurate selected epitopes in the structure of SARS-CoV-2 Spike glycoprotein. Representation of Spike protein structure and their main domains. The epitopes are distributed in the Spike structure by their coordinate in the protein sequence allowing the identification of the domains where the epitopes were found. The specific sequence and coordinate of the epitopes found in the RBD domain are shown. The epitopes are colored by their type (Class I, Class II and B cell epitopes).

Full-size DOI: 10.7717/peerj.12548/fig-4

samples available in GISAID of each lineage shows 99.1% of identity with 99.9% of coverage confirming the sequence lineages similarity (Table S8).

One of the mentioned properties of RBD epitopes is the high affinity for HLA alleles. To validate these findings the *in silico* structural binding performance was assessed by the PepDock docking tool. Thus, the RBD epitopes and the most major HLA alleles in studies with COVID-19 (see Methods) were structurally bonded in complex HLA-epitope (Fig. S3). All complexes were seen with high-performance scores demonstrating *in silico* epitope effectiveness in activating the cell-mediated immune response *via* MHC presentation. Their high estimated accuracy selected the two most significant HLA-epitope complexes (Table S9, Fig. 6). The HLA class I epitope 516-FGYQPYRVV-524 in complex with HLA-A*01:01, HLA-A*02:01, HLA-B*08:01, and HLA-C*12:03; and the HLA class II epitope 443-TGCVIAWNSKNLDSK-457 in complex with HLA-DRB1*03:01 and the HLA-DRB1*12:02 alleles (Table S9, Fig. 6). Additionally, the best measured free energy of the complexes was $\Delta G = -23$ kcal/mol for 461-NYNYRYRLF-469 HLA class I epitope in complex with HLA-A*01:01 and $\Delta G = -31.6$ kcal/mol for 322-EKGIVQTSNFRVQPR-336 HLA-class II epitope in complex with the HLA-DRB1*04:01 allele (Table S9, Fig. 6). An additional analysis was performed to assess the specific residues involved in the structural binding allowing to identify the specific amino acids in contact between the epitopes and the HLA alleles (Table S9).

Multi-epitope construct for *in silico* validation of RBD epitopes

To further validate the RBD epitopes and consequently confirm the effectiveness of EpiCurator in providing a curated selection, the epitopes were inserted in a multi-epitope construct. It consists of 287 amino residues, including 11 RBD selected epitopes joined by linkers (Fig. S1). The structural appraisal of the secondary structure predicted by the

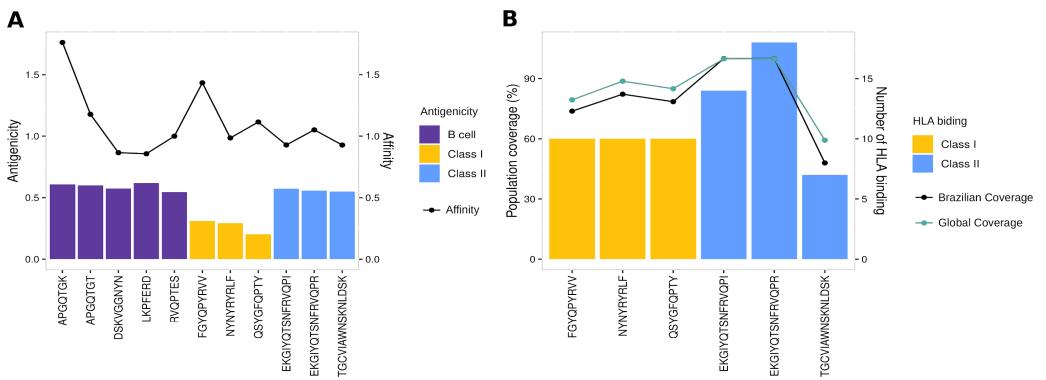


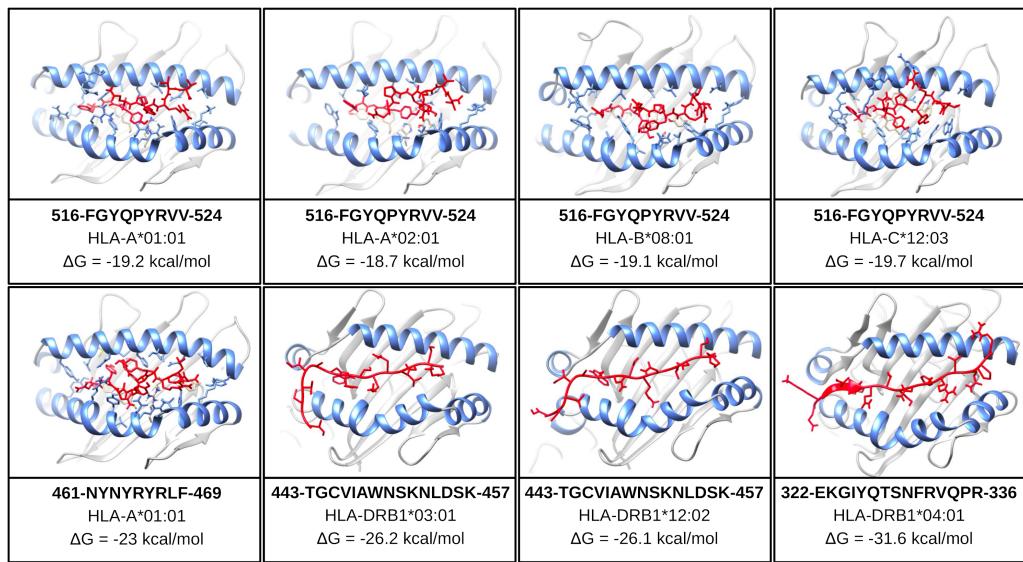
Figure 5 Properties of epitopes from Spike RBD domain. (A) The affinity values of the selected epitopes (x axis) are indicated as bars on the right y axis, and the antigenicity values are indicated as dots on the left y axis. (B) The number of HLA binding alleles for the selected epitopes (x axis) are indicated as bars on the left y axis, and the cumulative percentage of population coverage is depicted as dots on the right y axis. Each color of bars represents a distinct epitope type and each color of lines represents a distinct population.

Full-size DOI: 10.7717/peerj.12548/fig-5

PSIPRED server revealed 31.7% alpha-helix, 13.5% beta-strand and the disordered region was 12% (Fig. S4). Some evaluation of the linear sequence of multi-epitope were identified as follows: high antigenicity by the Vaxijen 2.0 (0.87) and the AntigenPro (0.93), good solubility by SolPro (0.91), Protein-sol (0.66), and SOLart (65.80%), non-allergenic feature by AllerTOP and AllergenFP, hydrophilic property (hydropathicity = -0.438) and stability (instability index = 20.56), with pI 7.03, and molecular weight calculated to be 29 kDa.

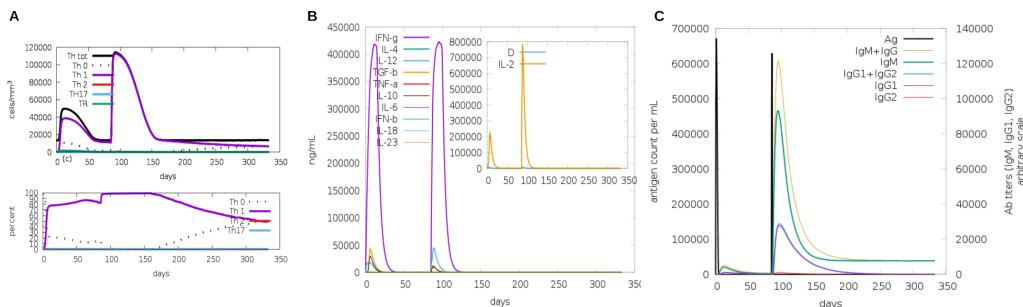
Another assessment of linear sequence includes the *in silico* immunogenic profile provided by the C-IMMSIM immune server. The analysis showed a gradual increase of IgM, IFN- γ (associated with both CD8+ T cell and CD4+ Th1 response as shown in Fig. 7A), and IL-2 level after each multi-epitope exposure indicated an elevated immune response (Fig. 7B). Besides, an adequate generation of both IgG1 and IgG2 was shown (Fig. 7C) with high levels of clonal proliferation of B cell and T cell population (Figs. S5A and S5B). In addition, the development of *in silico* immune memory was assessed by the abundance of different types of B cells and T cells (Figs. S5C and S5D).

The favorable *in silico* immunogenic profile of linear construct led to a multi-epitope 3D modelling to assess TLR binding capability adding validation for the curated selection provided by the EpiCurator. In this context, we modeled and evaluated the 3D multi-epitope to choose the best model for the TLR binding assay. The best predicted tertiary structure returned by RaptorX has an RMSD score of 11.218 (Fig. S6A), 62% of residues are predicted to be exposed and 30% are predicted to be disordered (Fig. S6B). The refinement by GalaxyRefine allows the choice of the best model based on its quality scores available in Table S10. This model highlights the epitopes according to their type (B cell, HLA class I T cell and HLA class II T cell) (Fig. S7). The overall model quality of the refined structure was further validated in the ProSA web (Fig. S8A), showing the energy results by position (Fig. S8B) with good local quality since all energy values of the residues are negative.

**Figure 6 Structure of the HLA-epitope complex for the main T cell epitopes in Spike RBD domain.**

Structure complexes provided by docking simulation shows the MHC binding grooves (blue ribbons), and the epitope (red structure). For each complex, the amino acids sequence of the epitope, HLA binding allele and the binding free energy are available.

Full-size DOI: [10.7717/peerj.12548/fig-6](https://doi.org/10.7717/peerj.12548/fig-6)

**Figure 7 In silico simulation of immune response using multi-epitope construct as an antigen.** The multi-epitope injection occurs on day 1 and within a 6-month interval. (A) Evolution of Th0, Th1, Th2 and Th17 response in cell/mm³ and percentage across the days. (B) Cytokine production across the days - specific subclasses are indicated as colored peaks. (C) multi-epitope injections (black vertical lines) promoting the immunoglobulin production - specific subclasses are indicated as colored peaks.

Full-size DOI: [10.7717/peerj.12548/fig-7](https://doi.org/10.7717/peerj.12548/fig-7)

Considering the accuracy of the refined model, a docking between our multi-epitope and the immune receptors (TLR3 and TLR4) was performed to check stability and binding affinity. The best complexes (Fig. 8) for each receptor had free energy values of $\Delta G = -1,048 \text{ kcal/mol}$ (TLR3) and $\Delta G = -1020 \text{ kcal/mol}$ (TLR4), indicating a high binding affinity. In addition, the stability and physical movements of the complexes were confirmed using molecular dynamics simulation in the iMODS server. The complex results are presented in Figs. S9 and S10.

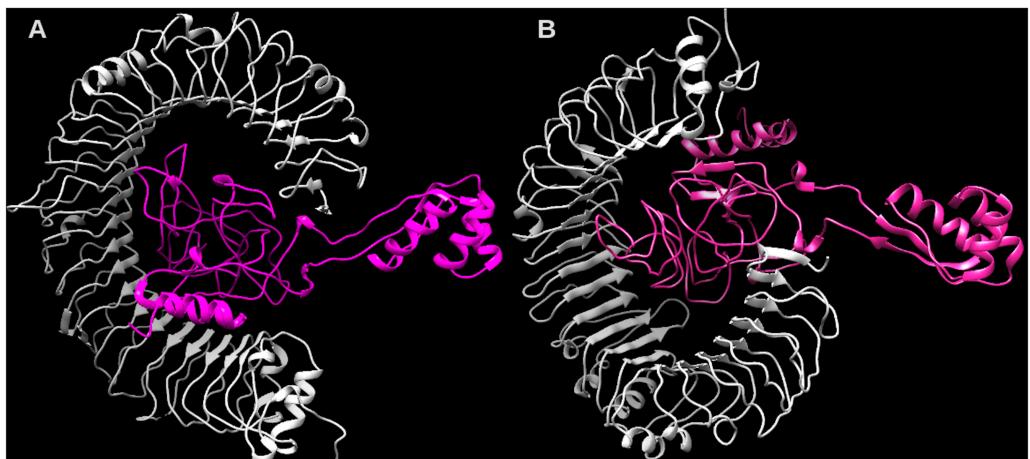


Figure 8 Structure of the multi-epitope-TLR complexes. The immune receptors TLR3 and TLR4 are demonstrated in gray color and the ligands (multi-epitope) are shown in pink color. (A) The multi-epitope-TLR3 complex (B) the multi-epitope-TLR4 complex.

Full-size DOI: 10.7717/peerj.12548/fig-8

DISCUSSION

Several studies have used immunoinformatic approaches to select B cell and T cell SARS-CoV-2 epitopes for epitopes-based vaccine formulation ([Ramana & Mehla, 2020](#); [Oli et al., 2020](#); [Siracusano, Pastori & Lopalco, 2020](#); [Yoshida et al., 2021](#); [Ribes, Chacour & Moncunill, 2021](#); [Jin et al., 2021](#)). However, current computational methods are limited since they identify a large number of epitopes and need different web servers for accurate selection ([Bui et al., 2006](#); [Doytchinova & Flower, 2007b](#); [Gupta et al., 2013](#)). This study developed an approach that gathered diversified analysis assembled in a workflow (EpiCurator). It accurately selects SARS-CoV-2 epitopes with useful *in silico* properties for an immunogen candidate target.

Its effectiveness was validated with a pairwise comparison analysis with four previously published papers ([Kiyotani et al., 2020](#); [Chen et al., 2020](#); [Crooke et al., 2020](#); [Chukwudiozie et al., 2021](#)). The main analysis of this validation was the IEDB matching that promotes a robust and refined selection of epitopes which confirms the importance of the IEDB database in the epitopes analysis studies ([Beaver, Bourne & Ponomarenko, 2007](#)). In addition, our approach can be used for extensive scale search and high-throughput computing analysis with the SARS-CoV-2 genomes available, an advantage to analyze a pandemic emergent virus ([Ojha et al., 2020](#); [Minervina et al., 2021](#); [Pham et al., 2021](#)).

Remarkably, our workflow conducts a thorough and facilitated accurate selection that enables us to prioritize the epitopes candidates. The EpiCurator recognizes patterns of cross-conservation with SARS-CoV-2 and human epitopes, eliminating the ones with significant human homology ([Meyers et al., 2021](#)). In addition, notably identifying epitopes in published articles, ensuring the selection *in silico* candidates with a plausible assumption to be described by the first time in this study. Consequently, this analysis might also validate the effectiveness of our workflow's accurate selection analysis, since our epitopes were

identified in different studies previously published ([Prachar et al., 2020](#)). The workflow also allows the selection of epitopes by high conservancy ($\geq 90\%$), across more than two thousand SARS-CoV-2 circulating Brazilian lineages in accordance with previous reports for SARS-CoV-2 epitopes selection ([Zaheer et al., 2020](#); [Mahapatra et al., 2020](#); [Mallajosyula et al., 2021](#)).

Three main phases comprise our approach to SARS-CoV-2 epitopes identification, the prediction, accurate selection and epitopes validation. Therefore, taking the spike glycoprotein epitopes as parameters, they represent 10% of the predicted epitopes and 2% of the accurately selected similarly to those presented in other reports ([Kiyotani et al., 2020](#); [Crooke et al., 2020](#)) with the resembling result for the other proteins. Indeed, the spike data give us an advantage since they were used to optimize the validation of the epitopes accurately selected by EpiCurator. Thus, they are taken as a baseline to conduct a thorough analysis. This baseline is reached since the Spike plays the most crucial role in the entry of viral particles into host cells, promoting an effective infection ([Ou et al., 2020](#)). These characteristics make the spike a chosen target for epitope screening leading to vaccine development ([Shang et al., 2020](#); [Walls et al., 2020](#); [Lin et al., 2020](#)). Pertaining to the Spike, the RBD region induces responses that block S protein binding with the human cell receptor, neutralizing SARS-CoV-2 infection ([Rakib et al., 2020](#); [Yang et al., 2020](#)), having substantial importance to peptide-based vaccine studies. Considering this importance, we prioritize 11 epitopes of this study identified on Spike RBD regions.

The RBD epitopes were identified in several SARS-CoV-2 lineages emphasizing samples of the current VOCs. They have favorable *in silico* structure interactions with the main HLA alleles related to the COVID-19 response ([Tavasolian et al., 2020](#); [Tomita et al., 2020](#); [Shkurnikov et al., 2021](#)). Thereby, it is plausible to assume that these epitopes could be responsible for the strong activation of the cell-mediated immune response ([Patronov et al., 2011](#); [Sarma, Olotu & Soliman, 2021](#)) in case of an experimental assay. In addition, the RBD epitopes have a broad populational coverage consistent with studies for SARS-CoV-2 epitopes identified in lineages isolated from India, England and the United States ([Mallavarpu Ambrose et al., 2021](#)). These findings combined with the high conservancy, antigenicity and immunogenicity suggest these epitopes' *in silico* immunogen profile ([Zheng & Song, 2020](#); [Mallavarpu Ambrose et al., 2021](#); [Jahangirian et al., 2021](#)), validating the accurate selection of EpiCurator.

Indeed, the RBD epitopes have substantial evidence and are seen as good epitope candidates. Despite sharing this evidence, they could together increase the immune responses and confer better protection against SARS-CoV-2 ([Jahangirian et al., 2021](#)). Accepting this assumption, they were used to design a multi-epitope construct to further validate their immunogenic properties ([Kalita et al., 2020](#); [Mohammad et al., 2020](#); [Singh et al., 2020](#); [Yang, Bogdan & Nazarian, 2021](#); [Lim et al., 2021](#); [Sharma et al., 2021](#)). The multi-epitope has the capability of *in silico* activation of memory B and T cell with Th1 response. Several experimental studies about immune response against COVID-19 endorse these findings ([Lipsitch et al., 2020](#); [Hartley et al., 2020](#); [Ghazavi et al., 2021](#); [Quast & Tarlinton, 2021](#); [Tarde et al., 2021](#)). On the other hand, no response Th17 was found in the *in silico* assay, a common response that characterizes the severe COVID-19

profile ([Wu & Yang, 2020](#)). Additionally, the multi-epitope had low binding energy and high stability of *in silico* structure interactions with TLR3 and TLR4 similar to other computational studies ([Kar et al., 2020](#); [Rahman et al., 2020](#); [Yang, Bogdan & Nazarian, 2021](#); [Nemati et al., 2021](#); [Saba et al., 2021](#)). This interaction is important since TLR is an innate immune receptor that recognizes viral proteins and triggers infection resistance ([O'Neill, Golenbock & Bowie, 2013](#)). Furthermore, TLR3 and TLR4 are specifically related by different studies as part of immune response in COVID-19 ([Patra, Chandra Das & Mukherjee, 2021](#); [Khanmohammadi & Rezaei, 2021](#); [Kaushik, Bhandari & Kuhad, 2021](#)). The findings in multi-epitope analysis confirm the importance of the RBD epitopes, selected by the EpiCurator, as good *in silico* immunogens candidates, reinforcing our approach's efficiency for a curated selection of epitopes.

CONCLUSIONS

It is important to reinforce that this work focuses on a computational prediction and selection of epitopes seeking to perform several *in silico* validation. Therefore, our approach is useful to researchers designing an experimental peptide-based vaccine to control the disease. Since, indeed, the development of an effective vaccine requires a detailed experimental investigation of the immunological correlations with SARS-CoV-2. Regarding the computational analysis, assembling the accurate selection of epitopes and their validation consolidate our approach as a helpful workflow analysis to provide epitopes for immunogens. These epitopes could significantly activate the *in silico* immune response against some circulating Brazilian variants. Furthermore, considering that the pandemic is still ongoing, our approach could contribute to continuous monitoring and identification of new SARS-CoV-2 epitopes according to the emergence of variants over time.

ACKNOWLEDGEMENTS

We would like to thank all the authors and the administrators of the GISAID and IEDB databases, allowing this study to be properly conducted. The authors acknowledge the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil) for providing HPC resources of the SDumont supercomputer, which has contributed to the results of the research reported within this paper. URL: <http://sdumont.lncc.br>.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was developed in the frameworks of Corona-ômica-RJ (FAPERJ = E-26/210.179/2020) and Rede Corona-ômica BR MCTI/FINEP (FINEP = 01.20.0029.000462/20, CNPq = 404096/2020-4) and BRICS/CNPq - (440931/2020-7). Ana Tereza R. Vasconcelos is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (303170/2017-4) and FAPERJ (E -26/202.826/2018). Yasmin C Martins is currently supported by FAPERJ (E-26/202.168/2020). The funders

had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq: 303170/2017-4.
FAPERJ: E -26/202.826/2018, E-26/202.168/2020.

Competing Interests

Ana Tereza R. Vasconcelos is an Academic Editor for PeerJ.

Author Contributions

- Cristina S. Ferreira conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Yasminn C. Martins conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Rangel Celso Souza analyzed the data, prepared figures and/or tables, and approved the final draft.
- Ana Tereza R. Vasconcelos conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data is available at GitHub: <https://github.com/YasCoMa/EpiCurator>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.12548#supplemental-information>.

REFERENCES

- Andreano E, Piccini G, Licastro D, Casalino L, Johnson NV, Paciello I, Dal Monego S, Pantano E, Manganaro N, Manenti A, Manna R, Casa E, Hyseni I, Benincasa L, Montomoli E, Amaro RE, McLellan JS, Rappuoli R. 2021. SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma. *Proceedings of the National Academy of Sciences of the United States of America* **118**(36):e2103154118 DOI [10.1073/pnas.2103154118](https://doi.org/10.1073/pnas.2103154118).
- Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. 2001. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering* **14**:529–532 DOI [10.1093/protein/14.8.529](https://doi.org/10.1093/protein/14.8.529).
- Badiani AA, Patel JA, Ziolkowski K, Nielsen FBH. 2020. Pfizer: the miracle vaccine for COVID-19? *Public Health in Practice (Oxford, England)* **1**:100061 DOI [10.1016/j.puhip.2020.100061](https://doi.org/10.1016/j.puhip.2020.100061).

- Bashir Z, Ahmad SU, Kiani BH, Jan Z, Khan N, Khan U, Haq I, Zahir F, Qadus A, Mahmood T.** 2021. Immunoinformatics approaches to explore B and T cell epitope-based vaccine designing for SARS-CoV-2 Virus. *Pakistan Journal of Pharmaceutical Sciences* 34:345–352 DOI [10.36721/pjps.2021.34.1.sup.345-352.1](https://doi.org/10.36721/pjps.2021.34.1.sup.345-352.1).
- Beaver JE, Bourne PE, Ponomarenko JV.** 2007. EpitopeViewer: a Java application for the visualization and analysis of immune epitopes in the Immune Epitope Database and Analysis Resource (IEDB). *Immunome Research* 3:3 DOI [10.1186/1745-7580-3-3](https://doi.org/10.1186/1745-7580-3-3).
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE.** 2000. The protein data bank. *Nucleic Acids Research* 28:235–242 DOI [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235).
- Bloom DE, Cadarette D, Ferranna M, Hyer RN, Tortorice DL.** 2021. How new models of vaccine development for COVID-19 have helped address an epic public health crisis. *Health Affairs* 40:410–418 DOI [10.1377/hlthaff.2020.02012](https://doi.org/10.1377/hlthaff.2020.02012).
- Brown NP, Leroy C, Sander C.** 1998. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* 14:380–381 DOI [10.1093/bioinformatics/14.4.380](https://doi.org/10.1093/bioinformatics/14.4.380).
- Brüssow H.** 2021. COVID-19: emergence and mutational diversification of SARS-CoV-2. *Microbial Biotechnology* 14:756–768 DOI [10.1111/1751-7915.13800](https://doi.org/10.1111/1751-7915.13800).
- Buchan DWA, Jones DT.** 2019. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Research* 47:W402–W407 DOI [10.1093/nar/gkz297](https://doi.org/10.1093/nar/gkz297).
- Bui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A.** 2006. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 7:153 DOI [10.1186/1471-2105-7-153](https://doi.org/10.1186/1471-2105-7-153).
- Bulla HAM.** 2021. COVID -19: efficacy and safety profile of main vaccines approved for emergency use authorization in 2021. *International Journal of Research - GRANTHAALAYAH* 9:271–283 DOI [10.29121/granthaalayah.v9.i7.2021.4062](https://doi.org/10.29121/granthaalayah.v9.i7.2021.4062).
- Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow GV, Christie CH, Dalenberg K, Di Costanzo L, Duarte JM, Dutta S, Feng Z, Ganesan S, Goodsell DS, Ghosh S, Green RK, Guranović V, Guzenko D, Hudson BP, Lawson CL, Liang Y, Lowe R, Namkoong H, Peisach E, Persikova I, Randle C, Rose A, Rose Y, Sali A, Segura J, Sekharan M, Shao C, Tao Y-P, Voigt M, Westbrook JD, Young JY, Zardecki C, Zhuravleva M.** 2021. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research* 49:D437–D451 DOI [10.1093/nar/gkaa1038](https://doi.org/10.1093/nar/gkaa1038).
- Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Keşmir C, Peters B.** 2013. Properties of MHC class I presented peptides that enhance immunogenicity. *PLOS Computational Biology* 9:e1003266 DOI [10.1371/journal.pcbi.1003266](https://doi.org/10.1371/journal.pcbi.1003266).
- Canese K, Weis S.** 2013. Pubmed: the bibliographic database. In: *The NCBI handbook*. Bethesda, MD: National Center for Biotechnology Information (US).
- Cella E, Benedetti F, Fabris S, Borsetti A, Pezzuto A, Ciotti M, Pascarella S, Cecarelli G, Zella D, Ciccozzi M, Giovanetti M.** 2021. SARS-CoV-2 lineages and

- sub-lineages circulating worldwide: a dynamic overview. *Cancer Therapy* 1–5
DOI 10.1159/000515340.
- Chakraborty C, Sharma AR, Bhattacharya M, Sharma G, Lee S-S. 2020.** The 2019 novel coronavirus disease (COVID-19) pandemic: a zoonotic prospective. *Asian Pacific Journal of Tropical Medicine* 13:242 DOI 10.4103/1995-7645.281613.
- Chen Z, John Wherry E. 2020.** T cell responses in patients with COVID-19. *Nature Reviews. Immunology* 20:529–536 DOI 10.1038/s41577-020-0402-6.
- Chen Y, Sidney J, Southwood S, Cox AL, Sakaguchi K, Henderson RA, Appella E, Hunt DF, Sette A, Engelhard VH. 1994.** Naturally processed peptides longer than nine amino acid residues bind to the class I MHC molecule HLA-A2.1 with high affinity and in different conformations. *Journal of Immunology* 152:2874–2881.
- Chen H-Z, Tang L-L, Yu X-L, Zhou J, Chang Y-F, Wu X. 2020.** Bioinformatics analysis of epitope-based vaccine design against the novel SARS-CoV-2. *Infectious Diseases of Poverty* 9:88 DOI 10.1186/s40249-020-00713-3.
- Chowdhury GG. 2005.** Natural language processing. *Annual Review of Information Science and Technology* 37:51–89 DOI 10.1002/aris.1440370103.
- Chukwudzie OS, Gray CM, Fagbayi TA, Chukwuanukwu RC, Oyebanji VO, Bankole TT, Adewole RA, Daniel EM. 2021.** Immuno-informatics design of a multimeric epitope peptide based vaccine targeting SARS-CoV-2 spike glycoprotein. *PLOS ONE* 16:e0248061 DOI 10.1371/journal.pone.0248061.
- Cobey S, Larremore DB, Grad YH, Lipsitch M. 2021.** Concerns about SARS-CoV-2 evolution should not hold back efforts to expand vaccination. *Nature Reviews. Immunology* 21:330–335 DOI 10.1038/s41577-021-00544-9.
- Cohen J, Normile D. 2020.** New SARS-like virus in China triggers alarm. *Science* 367:234–235 DOI 10.1126/science.367.6475.234.
- Crooke SN, Ovsyannikova IG, Kennedy RB, Poland GA. 2020.** Immunoinformatic identification of B cell and T cell epitopes in the SARS-CoV-2 proteome. *Scientific Reports* 10:14179 DOI 10.1038/s41598-020-70864-8.
- Desta IT, Porter KA, Xia B, Kozakov D, Vajda S. 2020.** Performance and its limits in rigid body protein-protein docking. *Structure* 28:1071–1081 DOI 10.1016/j.str.2020.06.006.
- Dhanda SK, Gupta S, Vir P, Raghava GPS. 2013.** Prediction of IL4 inducing peptides. *Clinical & Developmental Immunology* 2013:263952 DOI 10.1155/2013/263952.
- Dhanda SK, Karosiene E, Edwards L, Grifoni A, Paul S, Andreatta M, Weiskopf D, Sidney J, Nielsen M, Peters B, Sette A. 2018.** Predicting HLA CD4 immunogenicity in human populations. *Frontiers in Immunology* 9:1369 DOI 10.3389/fimmu.2018.01369.
- Dhanda SK, Vir P, Raghava GPS. 2013.** Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct* 8:30 DOI 10.1186/1745-6150-8-30.
- Dimitrov I, Bangov I, Flower DR, Doytchinova I. 2014a.** AllerTOP v.2—a server for in silico prediction of allergens. *Journal of Molecular Modeling* 20:2278 DOI 10.1007/s00894-014-2278-5.

- Dimitrov I, Naneva L, Doytchinova I, Bangov I. 2014b.** AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* **30**:846–851 DOI [10.1093/bioinformatics/btt619](https://doi.org/10.1093/bioinformatics/btt619).
- Di Natale C, La Manna S, De Benedictis I, Brandi P, Marasco D. 2020.** Perspectives in peptide-based vaccination strategies for syndrome coronavirus 2 pandemic. *Frontiers in Pharmacology* **11**:578382 DOI [10.3389/fphar.2020.578382](https://doi.org/10.3389/fphar.2020.578382).
- Doytchinova IA, Flower DR. 2007a.** VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* **8**:4 DOI [10.1186/1471-2105-8-4](https://doi.org/10.1186/1471-2105-8-4).
- Doytchinova IA, Flower DR. 2007b.** Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine* **25**:856–866 DOI [10.1016/j.vaccine.2006.09.032](https://doi.org/10.1016/j.vaccine.2006.09.032).
- Elbe S, Buckland-Merrett G. 2017.** Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* **1**:33–46 DOI [10.1002/gch2.1018](https://doi.org/10.1002/gch2.1018).
- Faria NR, Claro IM, Candido D, Franco LAM, Andrade PS, Coletti TM, Silva CAM, Sales FC, Manuli ER, Aguiar RS, Gaburo N, Camilo C da C, Fraiji NA, Crispim MAE, Carvalho M do PSS, Rambaut A, Loman N, Pybus OG, Sabino EC. 2020.** Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Available at <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586> (accessed on June 2021).
- Fast E, Altman RB, Chen B. 2020.** Potential T-cell and B-cell Epitopes of 2019-nCoV. ArXiv preprint. arXiv:2020.02.19.955484 DOI [10.1101/2020.02.19.955484](https://doi.org/10.1101/2020.02.19.955484).
- Fatoba AJ, Maharaj L, Adeleke VT, Okpeku M, Adeniyi AA, Adeleke MA. 2021.** Immunoinformatics prediction of overlapping CD8+ T-cell, IFN- γ and IL-4 inducer CD4+ T-cell and linear B-cell epitopes based vaccines against COVID-19 (SARS-CoV-2). *Vaccine* **39**:1111–1121 DOI [10.1016/j.vaccine.2021.01.003](https://doi.org/10.1016/j.vaccine.2021.01.003).
- Gao Q, Bao L, Mao H, Wang L, Xu K, Yang M, Li Y, Zhu L, Wang N, Lv Z, Gao H, Ge X, Kan B, Hu Y, Liu J, Cai F, Jiang D, Yin Y, Qin C, Li J, Gong X, Lou X, Shi W, Wu D, Zhang H, Zhu L, Deng W, Li Y, Lu J, Li C, Wang X, Yin W, Zhang Y, Qin C. 2020.** Development of an inactivated vaccine candidate for SARS-CoV-2. *Science* **369**:77–81 DOI [10.1126/science.abc1932](https://doi.org/10.1126/science.abc1932).
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Amos B. 2005.** Protein identification and analysis tools on the ExPASy server. In: Walker JM, ed. *The proteomics protocols handbook*. Totowa, NJ: Humana Press, 571–607.
- Geers D, Shamier MC, Bogers S, Hartog Gden, Gommers L, Nieuwkoop NN, Schmitz KS, Rijsbergen LC, Van Osch JAT, Dijkhuizen E, Smits G, Comvalius A, Van Mourik D, Caniels TG, Van Gils MJ, Sanders RW, Oude Munnink BB, Molenkamp R, De Jager HJ, Haagmans BL, De Swart RL, Koopmans MPG, Van Binnendijk RS, De Vries RD, GeurtsvanKessel CH. 2021.** SARS-CoV-2 variants of concern partially escape humoral but not T-cell responses in COVID-19 convalescent donors and vaccinees. *Science Immunology* **6**(59):eabj1750 DOI [10.1126/sciimmunol.abj1750](https://doi.org/10.1126/sciimmunol.abj1750).

- Gfeller D, Guillaume P, Michaux J, Pak H-S, Daniel RT, Racle J, Coukos G, Bassani-Sternberg M.** 2018. The length distribution and multiple specificity of naturally presented HLA-I ligands. *Journal of Immunology* **201**:3705–3716 DOI [10.4049/jimmunol.1800914](https://doi.org/10.4049/jimmunol.1800914).
- Ghazavi A, Ganji A, Keshavarzian N, Rabiemajd S, Mosayebi G.** 2021. Cytokine profile and disease severity in patients with COVID-19. *Cytokine* **137**:155323 DOI [10.1016/j.cyto.2020.155323](https://doi.org/10.1016/j.cyto.2020.155323).
- Goddard TD, Huang CC, Ferrin TE.** 2005. Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. *Structure* **13**:473–482 DOI [10.1016/j.str.2005.01.006](https://doi.org/10.1016/j.str.2005.01.006).
- Gonzalez-Galarza FF, McCabe A, Santos EJMD, Jones J, Takeshita L, Ortega-Rivera ND, Cid-Pavon GMD, Ramsbottom K, Ghattaoraya G, Alfirevic A, Middleton D, Jones AR.** 2020. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Research* **48**:D783–D788 DOI [10.1093/nar/gkz1029](https://doi.org/10.1093/nar/gkz1029).
- Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD.** 2021. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host & Microbe* **29**:463–476 DOI [10.1016/j.chom.2021.02.003](https://doi.org/10.1016/j.chom.2021.02.003).
- Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A.** 2011. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* **63**:325–335 DOI [10.1007/s00251-011-0513-0](https://doi.org/10.1007/s00251-011-0513-0).
- Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R.** 2013. In silico approach for predicting toxicity of peptides and proteins. *PLOS ONE* **8**:e73957 DOI [10.1371/journal.pone.0073957](https://doi.org/10.1371/journal.pone.0073957).
- Gupta S, Madhu MK, Sharma AK, Sharma VK.** 2016. ProInflam: a webserver for the prediction of proinflammatory antigenicity of peptides and proteins. *Journal of Translational Medicine* **14**:1–10 DOI [10.1186/s12967-016-0928-3](https://doi.org/10.1186/s12967-016-0928-3).
- Gupta S, Mittal P, Madhu MK, Sharma VK.** 2017. IL17eScan: a tool for the identification of peptides inducing IL-17 response. *Frontiers in Immunology* **8**:1430 DOI [10.3389/fimmu.2017.01430](https://doi.org/10.3389/fimmu.2017.01430).
- Hartley GE, Edwards ESJ, Aui PM, Varese N, Stojanovic S, McMahon J, Peleg AY, Boo I, Drummer HE, Hogarth PM, O'Hehir RE, Van Zelm MC.** 2020. Rapid generation of durable B cell memory to SARS-CoV-2 spike and nucleocapsid proteins in COVID-19 and convalescence. *Science Immunology* **5**(54):eabf8891 DOI [10.1126/sciimmunol.abf8891](https://doi.org/10.1126/sciimmunol.abf8891).
- Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J.** 2017. Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* **33**:3098–3100 DOI [10.1093/bioinformatics/btx345](https://doi.org/10.1093/bioinformatics/btx345).
- Heo L, Park H, Seok C.** 2013. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Research* **41**:W384–W388 DOI [10.1093/nar/gkt458](https://doi.org/10.1093/nar/gkt458).

- Hoffmann M, Arora P, Groß R, Seidel A, Hörnich BF, Hahn AS, Krüger N, Graichen L, Hofmann-Winkler H, Kempf A, Winkler MS, Schulz S, Jäck H-M, Jahrsdörfer B, Schrezenmeier H, Müller M, Kleger A, Münch J, Pöhlmann S.** 2021. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell* **184**:2384–2393 DOI [10.1016/j.cell.2021.03.036](https://doi.org/10.1016/j.cell.2021.03.036).
- Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M.** 2009. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* **61**:1–13 DOI [10.1007/s00251-008-0341-z](https://doi.org/10.1007/s00251-008-0341-z).
- Hou Q, Kwasigroch JM, Rooman M, Pucci F.** 2020. SOLart: a structure-based method to predict protein solubility and aggregation. *Bioinformatics* **36**:1445–1452 DOI [10.1093/bioinformatics/btz773](https://doi.org/10.1093/bioinformatics/btz773).
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B.** 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**:497–506 DOI [10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- Hudu SA, Shinkafi SH, Umar S.** 2016. An overview of recombinant vaccine technology, adjuvants and vaccine delivery methods. *International Journal of Pharmacy and Pharmaceutical Sciences* **8**:19–24 DOI [10.22159/ijpps.2016v8i11.14311](https://doi.org/10.22159/ijpps.2016v8i11.14311).
- Jahangirian E, Jamal GA, Nourooz M, Mohammadpour A.** 2021. A reverse vaccinology and immunoinformatics approach for designing a multiepitope vaccine against SARS-CoV-2. *Immunogenetics* **73**:1–19 DOI [10.1007/s00251-021-01228-3](https://doi.org/10.1007/s00251-021-01228-3).
- Jespersen MC, Peters B, Nielsen M, Marcatili P.** 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research* **45**:W24–W29 DOI [10.1093/nar/gkx346](https://doi.org/10.1093/nar/gkx346).
- Jin X, Ding Y, Sun S, Wang X, Zhou Z, Liu X, Li M, Chen X, Shen A, Wu Y, Liu B, Zhang J, Li J, Yang Y, Qiu H, Shen C, He Y, Zhao G.** 2021. Screening of HLA-A restricted T cell epitopes of SARS-CoV-2 and induction of CD8+ T cell responses in HLA-A transgenic mice. ArXiv preprint. arXiv:2021.04.01.438020.
- Jones DT.** 1999. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* **292**:195–202 DOI [10.1006/jmbi.1999.3091](https://doi.org/10.1006/jmbi.1999.3091).
- Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M.** 2017. NetMHCpan-4.0: improved Peptide-MHC Class I interaction predictions integrating eluted ligand and peptide binding affinity data. *Journal of Immunology* **199**:3360–3368 DOI [10.4049/jimmunol.1700893](https://doi.org/10.4049/jimmunol.1700893).
- Kalita P, Padhi AK, Zhang KYJ, Tripathi T.** 2020. Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. *Microbial Pathogenesis* **145**:104236 DOI [10.1016/j.micpath.2020.104236](https://doi.org/10.1016/j.micpath.2020.104236).
- Källberg M, Margaryan G, Wang S, Ma J, Xu J.** 2014. RaptorX server: a resource for template-based protein structure modeling. *Methods in Molecular Biology* **1137**:17–27 DOI [10.1007/978-1-4939-0366-5_2](https://doi.org/10.1007/978-1-4939-0366-5_2).

- Kar T, Narsaria U, Basak S, Deb D, Castiglione F, Mueller DM, Srivastava AP. 2020.** A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific Reports* **10**:10895 DOI [10.1038/s41598-020-67749-1](https://doi.org/10.1038/s41598-020-67749-1).
- Karlsson AC, Humbert M, Buggert M. 2020.** The known unknowns of T cell immunity to COVID-19. *Science Immunology* **5** DOI [10.1126/sciimmunol.abe8063](https://doi.org/10.1126/sciimmunol.abe8063).
- Kaushik D, Bhandari R, Kuhad A. 2021.** TLR4 as a therapeutic target for respiratory and neurological complications of SARS-CoV-2. *Expert Opinion on Therapeutic Targets* **25**:491–508 DOI [10.1080/14728222.2021.1918103](https://doi.org/10.1080/14728222.2021.1918103).
- Kazi A, Chuah C, Majeed ABA, Leow CH, Lim BH, Leow CY. 2018.** Current progress of immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design. *Pathogens and Global Health* **112**:123–131 DOI [10.1080/2047724.2018.1446773](https://doi.org/10.1080/2047724.2018.1446773).
- Khanmohammadi S, Rezaei N. 2021.** Role of Toll-like receptors in the pathogenesis of COVID-19. *Journal of Medical Virology* **93**:2735–2739 DOI [10.1002/jmv.26826](https://doi.org/10.1002/jmv.26826).
- Kiyotani K, Toyoshima Y, Nemoto K, Nakamura Y. 2020.** Bioinformatic prediction of potential T cell epitopes for SARS-CoV-2. *Journal of Human Genetics* **65**:569–575 DOI [10.1038/s10038-020-0771-5](https://doi.org/10.1038/s10038-020-0771-5).
- Knoll MD, Wonodi C. 2021.** Oxford-AstraZeneca COVID-19 vaccine efficacy. *The Lancet* **397**:72–74 DOI [10.1016/S0140-6736\(20\)32623-4](https://doi.org/10.1016/S0140-6736(20)32623-4).
- Krammer F. 2020.** SARS-CoV-2 vaccines in development. *Nature* **586**:516–527 DOI [10.1038/s41586-020-2798-3](https://doi.org/10.1038/s41586-020-2798-3).
- Kyriakidis NC, López-Cortés A, González EV, Grimaldos AB, Prado EO. 2021.** SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *NPJ Vaccines* **6**:28 DOI [10.1038/s41541-021-00292-w](https://doi.org/10.1038/s41541-021-00292-w).
- Larsen MV, Lundsgaard C, Lamberth K, Buus S, Lund O, Nielsen M. 2007.** Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* **8**:424 DOI [10.1186/1471-2105-8-424](https://doi.org/10.1186/1471-2105-8-424).
- Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, Mayhew S. 2020.** The COVID-19 vaccine development landscape. *Nature Reviews. Drug Discovery* **19**:305–306 DOI [10.1038/d41573-020-00073-5](https://doi.org/10.1038/d41573-020-00073-5).
- Lee H, Heo L, Lee MS, Seok C. 2015.** GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Research* **43**:W431–W435 DOI [10.1093/nar/gkv495](https://doi.org/10.1093/nar/gkv495).
- Li W, Godzik A. 2006.** Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**:1658–1659 DOI [10.1093/bioinformatics/btl158](https://doi.org/10.1093/bioinformatics/btl158).
- Lim HX, Lim J, Jazayeri SD, Poppema S, Poh CL. 2021.** Development of multi-epitope peptide-based vaccines against SARS-CoV-2. *Biomedical Journal* **44**:18–30 DOI [10.1016/j.bj.2020.09.005](https://doi.org/10.1016/j.bj.2020.09.005).
- Lin L, Ting S, Yufei H, Wendong L, Yubo F, Jing Z. 2020.** Epitope-based peptide vaccines predicted against novel coronavirus disease caused by SARS-CoV-2. *Virus Research* **288**:198082 DOI [10.1016/j.virusres.2020.198082](https://doi.org/10.1016/j.virusres.2020.198082).

- Lipsitch M, Grad YH, Sette A, Crotty S. 2020.** Cross-reactive memory T cells and herd immunity to SARS-CoV-2. *Nature Reviews. Immunology* **20**:709–713 DOI [10.1038/s41577-020-00460-4](https://doi.org/10.1038/s41577-020-00460-4).
- Liu Z, Van Blargan LA, Bloyet L-M, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico JM, Theel ES, Liebeskind MJ, Alford B, Buchser WJ, Ellebedy AH, Fremont DH, Diamond MS, Whelan SPJ. 2021.** Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host & Microbe* **29**:477–488 DOI [10.1016/j.chom.2021.01.014](https://doi.org/10.1016/j.chom.2021.01.014).
- Livingston EH, Malani PN, Creech CB. 2021.** The Johnson & Johnson Vaccine for COVID-19. *JAMA: The Journal of the American Medical Association* **325**:1575 DOI [10.1001/jama.2021.2927](https://doi.org/10.1001/jama.2021.2927).
- López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. 2014.** iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Research* **42**:W271–W276 DOI [10.1093/nar/gku339](https://doi.org/10.1093/nar/gku339).
- Lu G, Shan S, Zainab B, Ayaz Z, He J, Xie Z, Rashid U, Zhang D, Mehmood Abbasi A. 2021.** Novel vaccine design based on genomics data analysis: A review. *ScandInavian Journal of Immunology* **93**:e12986 DOI [10.1111/sji.12986](https://doi.org/10.1111/sji.12986).
- Madden T. 2020.** User manual. Bethesda, MD: National Center for Biotechnology Information (US).
- Magnan CN, Randall A, Baldi P. 2009.** SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics* **25**:2200–2207 DOI [10.1093/bioinformatics/btp386](https://doi.org/10.1093/bioinformatics/btp386).
- Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, Baldi P. 2010.** High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics* **26**:2936–2943 DOI [10.1093/bioinformatics/btq551](https://doi.org/10.1093/bioinformatics/btq551).
- Mahapatra SR, Sahoo S, Dehury B, Raina V, Patro S, Misra N, Suar M. 2020.** Designing an efficient multi-epitope vaccine displaying interactions with diverse HLA molecules for an efficient humoral and cellular immune response to prevent COVID-19 infection. *Expert Review of Vaccines* **19**:871–885 DOI [10.1080/14760584.2020.1811091](https://doi.org/10.1080/14760584.2020.1811091).
- Mahase E. 2020.** Covid-19: moderna vaccine is nearly 95% effective, trial involving high risk and elderly people shows. *BMJ* **371**:1 DOI [10.1136/bmj.m4471](https://doi.org/10.1136/bmj.m4471).
- Mallajosyula V, Ganjavi C, Chakraborty S, McSween AM, Pavlovitch-Bedzyk AJ, Wilhelmy J, Nau A, Manohar M, Nadeau KC, Davis MM. 2021.** CD8+ T cells specific for conserved coronavirus epitopes correlate with milder disease in COVID-19 patients. *Science Immunology* **6**(61):eabg5669 DOI [10.1126/sciimmunol.abg5669](https://doi.org/10.1126/sciimmunol.abg5669).
- Mallapaty S. 2021.** WHO approval of Chinese CoronaVac COVID vaccine will be crucial to curbing pandemic. *Nature* **594**:161–162 DOI [10.1038/d41586-021-01497-8](https://doi.org/10.1038/d41586-021-01497-8).
- Mallavarpu Ambrose J, Priya Veeraraghavan V, Kullappan M, Chellapandian P, Krishna Mohan S, Manivel VA. 2021.** Comparison of immunological profiles of SARS-CoV-2 variants in the COVID-19 pandemic trends: an immunoinformatics approach. *Antibiotics* **10**(5):535 DOI [10.3390/antibiotics10050535](https://doi.org/10.3390/antibiotics10050535).
- Malonis RJ, Lai JR, Vergnolle O. 2020.** Peptide-based vaccines: current progress and future challenges. *Chemical Reviews* **120**:3210–3229 DOI [10.1021/acs.chemrev.9b00472](https://doi.org/10.1021/acs.chemrev.9b00472).

- Mascellino MT, Di Timoteo F, De Angelis M, Oliva A.** 2021. Overview of the main Anti-SARS-CoV-2 vaccines: mechanism of action, efficacy and safety. *Individual Differences Research* 14:3459–3476 DOI [10.2147/IDR.S315727](https://doi.org/10.2147/IDR.S315727).
- McGill COVID19 Vaccine Tracker Team.** 2021. COVID-19 vaccine tracker. Available at <https://covid19.trackvaccines.org/vaccines/#approved> (accessed on September 2021).
- McGuffin LJ, Bryson K, Jones DT.** 2000. The PSIPRED protein structure prediction server. *Bioinformatics* 16:404–405 DOI [10.1093/bioinformatics/16.4.404](https://doi.org/10.1093/bioinformatics/16.4.404).
- Medhi R, Srinoi P, Ngo N, Tran H-V, Lee TR.** 2020. Nanoparticle-based strategies to combat COVID-19. *ACS Applied Nano Materials* 3:8557–8580 DOI [10.1021/acsanm.0c01978](https://doi.org/10.1021/acsanm.0c01978).
- Menezes Teles E, Oliveira D, Melo Santos De Serpa Brandão R, Claudio Demes da Mata Sousa L, das Chagas Alves Lima F, Jamil Hadad do Monte S, Sérgio Coelho Marroquim M, Vanildo De Sousa Lima A, Gilberto Borges Coelho A, Matheus Sousa Costa J, Martins Ramos R, Socorro da Silva A.** 2019. pHLA3D: an online database of predicted three-dimensional structures of HLA molecules. *Human Immunology* 80:834–841 DOI [10.1016/j.humimm.2019.06.009](https://doi.org/10.1016/j.humimm.2019.06.009).
- Meyers LM, Gutiérrez AH, Boyle CM, Terry F, McGonnigal BG, Salazar A, Princiotta MF, Martin WD, De Groot AS, Moise L.** 2021. Highly conserved, non-human-like, and cross-reactive SARS-CoV-2 T cell epitopes for COVID-19 vaccine design and validation. *NPJ Vaccines* 6:71 DOI [10.1038/s41541-021-00331-6](https://doi.org/10.1038/s41541-021-00331-6).
- Minervina AA, Komech EA, Titov A, Bensouda Koraichi M, Rosati E, Mamedov IZ, Franke A, Efimov GA, Chudakov DM, Mora T, Walczak AM, Lebedev YB, Pogorelyy MV.** 2021. Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T-cell memory formation after mild COVID-19 infection. *eLife* 10:e63502 DOI [10.7554/eLife.63502](https://doi.org/10.7554/eLife.63502).
- Mohammad MG, Ibrahim F, Navid N, Shirin M.** 2020. Multi-Epitope vaccines (MEVs), as a novel strategy against infectious diseases. *Current Proteomics* 17:354–364 DOI [10.2174/157016461766190919120140](https://doi.org/10.2174/157016461766190919120140).
- Nagpal G, Chaudhary K, Agrawal P, Raghava GPS.** 2018. Computer-aided prediction of antigen presenting cell modulators for designing peptide-based vaccine adjuvants. *Journal of Translational Medicine* 16:181 DOI [10.1186/s12967-018-1560-1](https://doi.org/10.1186/s12967-018-1560-1).
- Nagpal G, Usmani SS, Dhanda SK, Kaur H, Singh S, Sharma M, Raghava GPS.** 2017. Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports* 7:42851 DOI [10.1038/srep42851](https://doi.org/10.1038/srep42851).
- Naveca FG, Nascimento V, De Souza VC, Corado A De L, Nascimento F, Silva G, Costa Á, Duarte D, Pessoa K, Mejía M, Brandão MJ, Jesus M, Gonçalves L, Costa CFda, Sampaio V, Barros D, Silva M, Mattos T, Pontes G, Abdalla L, Santos JH, Arantes I, Dezordi FZ, Siqueira MM, Wallau GL, Resende PC, Delatorre E, Gräf T, Bello G.** 2021. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. *Nature Medicine* 27:1230–1238 DOI [10.1038/s41591-021-01378-7](https://doi.org/10.1038/s41591-021-01378-7).
- Naveed M, Tehreem S, Arshad S, Bukhari SA, Shabbir MA, Essa R, Ali N, Zaib S, Khan A, Al-Harrasi A, Khan I.** 2021. Design of a novel multiple epitope-based vaccine: an

immunoinformatics approach to combat SARS-CoV-2 strains. *Journal of Infection and Public Health* 14:938–946 DOI [10.1016/j.jiph.2021.04.010](https://doi.org/10.1016/j.jiph.2021.04.010).

Nemati AS, Tafrihi M, Sheikhi F, Tabari AR, Haditabar A. 2021. Designing a new multi epitope-based vaccine against COVID-19 disease: an immunoinformatic study based on reverse vaccinology approach. *Research Square* 1:1–21 DOI [10.21203/rs.3.rs-206270/v1](https://doi.org/10.21203/rs.3.rs-206270/v1).

Ojha R, Gupta N, Naik B, Singh S, Verma VK, Prusty D, Prajapati VK. 2020. High throughput and comprehensive approach to develop multiepitope vaccine against minacious COVID-19. *European Journal of Pharmaceutical Sciences: Official Journal of the European Federation for Pharmaceutical Sciences* 151:105375 DOI [10.1016/j.ejps.2020.105375](https://doi.org/10.1016/j.ejps.2020.105375).

Oli AN, Obialor WO, Ifeanyichukwu MO, Odimegwu DC, Okoyeh JN, Emechebe GO, Adejumo SA, Ibeanu GC. 2020. Immunoinformatics and vaccine development: an overview. *ImmunoTargets and Therapy* 9:13–30 DOI [10.2147/ITTT.S241064](https://doi.org/10.2147/ITTT.S241064).

O'Neill LAJ, Golenbock D, Bowie AG. 2013. The history of Toll-like receptors - redefining innate immunity. *Nature Reviews. Immunology* 13:453–460 DOI [10.1038/nri3446](https://doi.org/10.1038/nri3446).

Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J, Xiang Z, Mu Z, Chen X, Chen J, Hu K, Jin Q, Wang J, Qian Z. 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature Communications* 11:1620 DOI [10.1038/s41467-020-15562-9](https://doi.org/10.1038/s41467-020-15562-9).

Patra R, Chandra Das N, Mukherjee S. 2021. Targeting human TLRs to combat COVID-19: a solution? *Journal of Medical Virology* 93:615–617 DOI [10.1002/jmv.26387](https://doi.org/10.1002/jmv.26387).

Patronov A, Dimitrov I, Flower DR, Doytchinova I. 2011. Peptide binding prediction for the human class II MHC allele HLA-DP2: a molecular docking approach. *BMC Structural Biology* 11:32 DOI [10.1186/1472-6807-11-32](https://doi.org/10.1186/1472-6807-11-32).

Paul S, Lindestam Arlehamn CS, Scriba TJ, Dillon MBC, Oseroff C, Hinz D, McKinney DM, Carrasco Pro S, Sidney J, Peters B, Sette A. 2015. Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *Journal of Immunological Methods* 422:28–34 DOI [10.1016/j.jim.2015.03.022](https://doi.org/10.1016/j.jim.2015.03.022).

Pham T-H, Qiu Y, Zeng J, Xie L, Zhang P. 2021. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. *Nature Machine Intelligence* 3:247–257 DOI [10.1038/s42256-020-00285-9](https://doi.org/10.1038/s42256-020-00285-9).

Prachar M, Justesen S, Steen-Jensen DB, Thorgrimsmen S, Jurgons E, Winther O, Bagger FO. 2020. Identification and validation of 174 COVID-19 vaccine candidate epitopes reveals low performance of common epitope prediction tools. *Scientific Reports* 10:20465 DOI [10.1038/s41598-020-77466-4](https://doi.org/10.1038/s41598-020-77466-4).

Prévost J, Finzi A. 2021. The great escape? SARS-CoV-2 variants evading neutralizing responses. *Cell Host & Microbe* 29:322–324 DOI [10.1016/j.chom.2021.02.010](https://doi.org/10.1016/j.chom.2021.02.010).

Quast I, Tarlinton D. 2021. B cell memory: understanding COVID-19. *Immunity* 54:205–210 DOI [10.1016/j.immuni.2021.01.014](https://doi.org/10.1016/j.immuni.2021.01.014).

- Rahman MS, Hoque MN, Islam MR, Akter S, Rubay et Ul Alam ASM, Siddique MA, Saha O, Rahaman MM, Sultana M, Crandall KA, Hossain MA.** 2020. Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2, the etiologic agent of COVID-19 pandemic: an in silico approach. *PeerJ* 8:e9572 DOI 10.7717/peerj.9572.
- Rakib A, Sami SA, Mimi NJ, Chowdhury MM, Eva TA, Nainu F, Paul A, Shahriar A, Tareq AM, Emon NU, Chakraborty S, Shil S, Mily SJ, Ben Hadda T, Almalki FA, Emran TB.** 2020. Immunoinformatics-guided design of an epitope-based vaccine against severe acute respiratory syndrome coronavirus 2 spike glycoprotein. *Computers in Biology and Medicine* 124:103967 DOI 10.1016/j.combiomed.2020.103967.
- Ramana J, Mehla K.** 2020. Immunoinformatics and epitope prediction. In: Tomar N, ed. *Immunoinformatics*. New York: Springer US, 155–171 DOI 10.1007/978-1-0716-0389-5_6.
- Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, Connor T, Peacock T, Robertson DL, Volz E.** 2020. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Available at <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563> (accessed on June 2021).
- Rapin N, Lund O, Bernaschi M, Castiglione F.** 2010. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLOS ONE* 5:e9862 DOI 10.1371/journal.pone.0009862.
- Redd AD, Nardin A, Kared H, Bloch EM, Pekosz A, Laeyendecker O, Abel B, Fehlings M, Quinn TC, Tobian AA.** 2021. CD8+ T cell responses in COVID-19 convalescent individuals target conserved epitopes from multiple prominent SARS-CoV-2 circulating variants. *medRxiv: the preprint server for health sciences*. DOI 10.1101/2021.02.11.21251585.
- Ribes M, Chaccour C, Moncunill G.** 2021. Adapt or perish: sSARS-CoV-2 antibody escape variants defined by deletions in the Spike N-terminal Domain. *Signal Transduction and Targeted Therapy* 6:164 DOI 10.1038/s41392-021-00601-8.
- Rotondo JC, Martini F, Maritati M, Mazziotta C, Di Mauro G, Lanzillotti C, Barp N, Gallerani A, Tognon M, Contini C.** 2021. SARS-CoV-2 infection: new molecular, phylogenetic, and pathogenetic insights. Efficacy of current vaccines and the potential risk of variants. *Viruses* 13:1687 DOI 10.3390/v13091687.
- Saad-Roy CM, Morris SE, Metcalf CJ, Mina MJ, Baker RE, Farrar J, Holmes EC, Pybus OG, Graham AL, Levin SA, Grenfell BT, Wagner CE.** 2021. Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. *Science* 372:363–370 DOI 10.1126/science.abg8663.
- Saba AA, Adiba M, Saha P, Hosen MI, Chakraborty S, Nabi AHMN.** 2021. An in-depth in silico and immunoinformatics approach for designing a potential multi-epitope construct for the effective development of vaccine to combat against SARS-CoV-2 encompassing variants of concern and interest. *Computers in Biology and Medicine* 136:104703 DOI 10.1016/j.combiomed.2021.104703.

- Sakaguchi T, Ibe M, Miwa K, Kaneko Y, Yokota S, Tanaka K, Takiguchi M. 1997.** Binding of 8-mer to 11-mer peptides carrying the anchor residues to slow assembling HLA class I molecules (HLA-B*5101). *Immunogenetics* **45**:259–265 DOI [10.1007/s002510050201](https://doi.org/10.1007/s002510050201).
- Sarma VR, Olotu FA, Soliman MES. 2021.** Integrative immunoinformatics paradigm for predicting potential B-cell and T-cell epitopes as viable candidates for subunit vaccine design against COVID-19 virulence. *Biomedical Journal* DOI [10.1016/j.bj.2021.05.001](https://doi.org/10.1016/j.bj.2021.05.001).
- Sette A, Crotty S. 2021.** Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* **184**:861–880 DOI [10.1016/j.cell.2021.01.007](https://doi.org/10.1016/j.cell.2021.01.007).
- Shahcheraghi SH, Ayatollahi J, Aljabali AA, Shastri MD, Shukla SD, Chellappan DK, Jha NK, Anand K, Katari NK, Mehta M, Satija S, Dureja H, Mishra V, Almuttery AG, Alnuqaydan AM, Charbe N, Prasher P, Gupta G, Dua K, Lotfi M, Bakshi HA, Tambuwala MM. 2021.** An overview of vaccine development for COVID-19. *Therapeutic Delivery* **12**:235–244 DOI [10.4155/td-2020-0129](https://doi.org/10.4155/td-2020-0129).
- Shang W, Yang Y, Rao Y, Rao X. 2020.** The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines. *NPJ Vaccines* **5**:18 DOI [10.1038/s41541-020-0170-0](https://doi.org/10.1038/s41541-020-0170-0).
- Sharma A, Pal S, Panwar A, Kumar S, Kumar A. 2021.** In-silico immunoinformatic analysis of SARS-CoV-2 virus for the development of putative vaccine construct. *Immunobiology* **226**:152134 DOI [10.1016/j.imbio.2021.152134](https://doi.org/10.1016/j.imbio.2021.152134).
- Sharma N, Patiyal S, Dhall A, Pande A, Arora C, Raghava GPS. 2020.** AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. *Briefings in Bioinformatics* DOI [10.1093/bib/bbaa294](https://doi.org/10.1093/bib/bbaa294).
- Shean RC, Makhsous N, Stoddard GD, Lin MJ, Greninger AL. 2019.** VAPiD: a lightweight cross-platform viral annotation pipeline and identification tool to facilitate virus genome submissions to NCBI GenBank. *BMC Bioinformatics* **20**:1–8 DOI [10.1186/s12859-019-2606-y](https://doi.org/10.1186/s12859-019-2606-y).
- Shkurnikov M, Nersisyan S, Jankevici T, Galatenko A, Gordeev I, Vechorko V, Tonevitsky A. 2021.** Association of HLA Class I genotypes with severity of coronavirus disease-19. *Frontiers in Immunology* **12**:641900 DOI [10.3389/fimmu.2021.641900](https://doi.org/10.3389/fimmu.2021.641900).
- Sievers F, Higgins DG. 2018.** Clustal Omega for making accurate alignments of many protein sequences. *Protein Science: A Publication of the Protein Society* **27**:135–145 DOI [10.1002/pro.3290](https://doi.org/10.1002/pro.3290).
- Singh A, Thakur M, Sharma LK, Chandra K. 2020.** Designing a multi-epitope peptide based vaccine against SARS-CoV-2. *Scientific Reports* **10**:16219 DOI [10.1038/s41598-020-73371-y](https://doi.org/10.1038/s41598-020-73371-y).
- Siracusano G, Pastori C, Lopalco L. 2020.** Humoral immune responses in COVID-19 patients: a window on the state of the art. *Frontiers in Immunology* **11**:1049 DOI [10.3389/fimmu.2020.01049](https://doi.org/10.3389/fimmu.2020.01049).
- Skwarczynski M, Toth I. 2016.** Peptide-based synthetic vaccines. *Chemical Science* **7**:842–854 DOI [10.1039/c5sc03892h](https://doi.org/10.1039/c5sc03892h).
- Sun L. 2013.** Peptide-based drug development. *Modern Chemistry and Applications* **1**:1–2 DOI [10.4172/mca.1000e103](https://doi.org/10.4172/mca.1000e103).

- Tarke A, Sidney J, Methot N, Yu ED, Zhang Y, Dan JM, Goodwin B, Rubiro P, Sutherland A, Wang E, Frazier A, Ramirez SI, Rawlings SA, Smith DM, da Silva Antunes R, Peters B, Scheuermann RH, Weiskopf D, Crotty S, Grifoni A, Sette A. 2021. Impact of SARS-CoV-2 variants on the total CD4+ and CD8+ T cell reactivity in infected or vaccinated individuals. *Cell Reports. Medicine* 2:100355 DOI [10.1016/j.xcrm.2021.100355](https://doi.org/10.1016/j.xcrm.2021.100355).
- Tavasolian F, Rashidi M, Hatam GR, Jeddi M, Hosseini AZ, Mosawi SH, Abdollahi E, Inman RD. 2020. HLA, immune response, and susceptibility to COVID-19. *Frontiers in Immunology* 11:601886 DOI [10.3389/fimmu.2020.601886](https://doi.org/10.3389/fimmu.2020.601886).
- Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao N-Y, Korsman S, Davies M-A, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Kosakovsky Pond SL, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, De Oliveira T. 2021. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 592:438–443 DOI [10.1038/s41586-021-03402-9](https://doi.org/10.1038/s41586-021-03402-9).
- Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva Filipe A, Wojcechowskyj JA, Davis C, Piccoli L, Pascall DJ, Dillen J, Lytras S, Czudnochowski N, Shah R, Meury M, Jesudason N, De Marco A, Li K, Bassi J, O'Toole A, Pinto D, Colquhoun RM, Culap K, Jackson B, Zatta F, Rambaut A, Jaconi S, Sreenu VB, Nix J, Zhang I, Jarrett RF, Glass WG, Beltramello M, Nomikou K, Pizzuto M, Tong L, Cameroni E, Croll TI, Johnson N, Di Iulio J, Wickenhagen A, Ceschi A, Harbison AM, Mair D, Ferrari P, Smollett K, Sallusto F, Carmichael S, Garzoni C, Nichols J, Galli M, Hughes J, Riva A, Ho A, Schiuma M, Semple MG, Openshaw PJM, Fadda E, Baillie JK, Chodera JD, Investigators ISARIC4C, COVID-19 Genomics UK (COG-UK) Consortium, Rihm SJ, Lycett SJ, Virgin HW, Telenti A, Corti D, Robertson DL, Snell G. 2021. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* 184:1171–1187 DOI [10.1016/j.cell.2021.01.037](https://doi.org/10.1016/j.cell.2021.01.037).
- Tomita Y, Ikeda T, Sato R, Sakagami T. 2020. Association between HLA gene polymorphisms and mortality of COVID-19: an in silico analysis. *Immunity, Inflammation and Disease* 8:684–694 DOI [10.1002/iid.3.358](https://doi.org/10.1002/iid.3.358).
- Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. 2019. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Research* 47:D339–D343 DOI [10.1093/nar/gky1006](https://doi.org/10.1093/nar/gky1006).
- Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181:281–292 DOI [10.1016/j.cell.2020.02.058](https://doi.org/10.1016/j.cell.2020.02.058).
- Wang X, Gui J. 2020. Cell-mediated immunity to SARS-CoV-2. *Pediatric Investigation* 4:281–291 DOI [10.1002/ped4.12228](https://doi.org/10.1002/ped4.12228).

- WHO.** 2021. Weekly epidemiological update on COVID-19-25 2021. Available at <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---25-may-2021> (accessed on 27 May 2021).
- Wiederstein M, Sippl MJ.** 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* 35:W407–W410 DOI [10.1093/nar/gkm290](https://doi.org/10.1093/nar/gkm290).
- Wu D, Yang XO.** 2020. TH17 responses in cytokine storm of COVID-19: an emerging target of JAK2 inhibitor Fedratinib. *Journal of Microbiology, Immunology, and Infection = Wei Mian Yu Gan Ran Za Zhi* 53:368–370 DOI [10.1016/j.jmii.2020.03.005](https://doi.org/10.1016/j.jmii.2020.03.005).
- Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A.** 2016. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* 32:3676–3678 DOI [10.1093/bioinformatics/btw514](https://doi.org/10.1093/bioinformatics/btw514).
- Yang J, Wang W, Chen Z, Lu S, Yang F, Bi Z, Bao L, Mo F, Li X, Huang Y, Hong W, Yang Y, Zhao Y, Ye F, Lin S, Deng W, Chen H, Lei H, Zhang Z, Luo M, Gao H, Zheng Y, Gong Y, Jiang X, Xu Y, Lv Q, Li D, Wang M, Li F, Wang S, Wang G, Yu P, Qu Y, Yang L, Deng H, Tong A, Li J, Wang Z, Yang J, Shen G, Zhao Z, Li Y, Luo J, Liu H, Yu W, Yang M, Xu J, Wang J, Li H, Wang H, Kuang D, Lin P, Hu Z, Guo W, Cheng W, He Y, Song X, Chen C, Xue Z, Yao S, Chen L, Ma X, Chen S, Gou M, Huang W, Wang Y, Fan C, Tian Z, Shi M, Wang F-S, Dai L, Wu M, Li G, Wang G, Peng Y, Qian Z, Huang C, Lau JY-N, Yang Z, Wei Y, Cen X, Peng X, Qin C, Zhang K, Lu G, Wei X.** 2020. A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature* 586:572–577 DOI [10.1038/s41586-020-2599-8](https://doi.org/10.1038/s41586-020-2599-8).
- Yang Z, Bogdan P, Nazarian S.** 2021. An in silico deep learning approach to multi-epitope vaccine design: a SARS-CoV-2 case study. *Scientific Reports* 11:3238 DOI [10.1038/s41598-021-81749-9](https://doi.org/10.1038/s41598-021-81749-9).
- Yoshida S, Ono C, Hayashi H, Fukumoto S, Shiraishi S, Tomono K, Arase H, Matsuura Y, Nakagami H.** 2021. SARS-CoV-2-induced humoral immunity through B cell epitope analysis in COVID-19 infected individuals. *Scientific Reports* 11:5934 DOI [10.1038/s41598-021-85202-9](https://doi.org/10.1038/s41598-021-85202-9).
- Zaheer T, Waseem M, Waqar W, Dar HA, Shehroz M, Naz K, Ishaq Z, Ahmad T, Ullah N, Bakhtiar SM, Muhammad SA, Ali A.** 2020. Anti-COVID-19 multi-epitope vaccine designs employing global viral genome sequences. *PeerJ* 8:e9541 DOI [10.7717/peerj.9541](https://doi.org/10.7717/peerj.9541).
- Zheng M, Song L.** 2020. Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV. *Cellular & Molecular Immunology* 17:536–538 DOI [10.1038/s41423-020-0385-z](https://doi.org/10.1038/s41423-020-0385-z).