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

Cristina dos Santos Ferreira[1], Yasmmin Côrtes Martins[1], Rangel Celso Souza[1], Ana Tereza Ribeiro de Vasconcelos[#1]

[1]Laboratório de Bioinformática, Laboratório Nacional de Computação Científica, Petrópolis, Brazil.

Corresponding Author:
Ana Tereza Ribeiro de Vasconcelos[#1]
Getúlio Vargas Av., 333, Quitandinha
Petrópolis - Rio de Janeiro
CEP 25651-075 - Brazil
Email address: atrv@lncc.br

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 with immunogen development.

1. 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;

42   Chakraborty et al., 2020). This virus was first reported in Wuhan in December 2019 (Huang et
43   al., 2020) and quickly evolved, with the emergence of several variants (Cella et al., 2021;
44   Brüssow, 2021). According to the World Health Organization (WHO) classifications, there are
45   four variants of concern (VOCs) currently spread worldwide designated as Alpha (B.1.1.7), Beta
46   (B.1.351), Gamma (P.1), and Delta (B.1.617.2) (Faria et al., 2020; Rambaut et al., 2020; WHO,
47   2021; Tegally et al., 2021). In addition, the VOCs are characterized by the possibility to impact
48   the disease severity, such as the possible increased risk of hospitalization, mortality, and
49   capability of evading vaccination-induced immune response (Prévost & Finzi, 2021; Geers et al.,
50   2021). The constant emergence of new variants keeps the contagiousness of SARS-CoV-2, which
51   increases the uncertainty of virus spread (Brüssow, 2021; Naveca et al., 2021).
52   Current research aims to develop effective interventions for controlling and preventing the
53   COVID-19 pandemic, furthermore, vaccination is still the most economical and effective
54   approach to prevent infection by the virus (Shang et al., 2020). At the time of writing, 22
55   vaccines against SARS-CoV-2 have been approved by at least one country (McGill COVID19
56   Vaccine Tracker Team, 2021). Five vaccines were approved for emergency use authorization and
57   listed by WHO Emergency Use Listing (EUL) (Bulla, 2021; Mascellino et al., 2021). The Pfizer/
58   BioNTech vaccine is based on messenger RNA (mRNA), coding for viral spike (S) proteins
59   (Badiani et al., 2020). The Moderna vaccine is a lipid nano-particle-encapsulated mRNA-based
60   vaccine that encodes a full-length spike (Mahase, 2020). The Johnson & Johnson (Janssen) is a
61   recombinant, non-replicating adenovirus vector encoding a full-length S protein (Livingston,
62   Malani & Creech, 2021). The Oxford-AstraZeneca is a chimpanzee adenovirus vectored DNA
63   vaccine (Knoll & Wonodi, 2021), and the CoronaVac is an inactivated virus COVID-19 (Gao et
64   al., 2020; Mallapaty, 2021).
65   The current vaccines are generally based on B cell immunity with neutralizing antibody
66   production (Thanh Le et al., 2020; Siracusano, Pastori & Lopalco, 2020; Yoshida et al., 2021).
67   However, several studies report the capability of a new mutation interrupting the binding with
68   some neutralizing antibodies (Greaney et al., 2021; Hoffmann et al., 2021; Andreano et al.,
69   2021), which could clarify the spread of the current VOCs. The main reasons for this capability
70   are the several changes (mutation and deletion) in the S protein, which is responsible by the
71   increased affinity between the receptor-binding domain (RBD) and the human cellular receptor
72   angiotensin-converting enzyme 2 (ACE2), promoting the antibodies escape (Thomson et al.,
73   2021; Greaney et al., 2021; Liu et al., 2021; Rotondo et al., 2021).
74   To improve the vaccine's effectiveness against SARS-CoV-2, global efforts in resources,
75   cooperation, and innovation have contributed to the accelerated development of COVID-19
76   vaccines (Bloom et al., 2021). These efforts lead various researchers to carry out diversified
77   methodologies of vaccine design such as a peptide-based vaccine, virus-like particle, replicating
78   and non-replicating viral vectors, DNA or RNA, live attenuated virus, recombinant designed
79   proteins, nanoparticles vaccine, and inactivated virus (Medhi et al., 2020; Krammer, 2020; Di
80   Natale et al., 2020; Kyriakidis et al., 2021; Shahcheraghi et al., 2021).
81   This work highlights the peptide-based vaccine described in several studies about SARS-CoV-2
82   vaccines (Malonis, Lai & Vergnolle, 2020; Chakraborty et al., 2020; Crooke et al., 2020; Fatoba
83   et al., 2021; Naveed et al., 2021). Currently, ten from 33 vaccines in phase 3 clinical trial, and

84   four from 22 approved vaccines have a peptide-based design (McGill COVID19 Vaccine Tracker
85   Team, 2021). The advantages of peptide-based vaccines include their capability to target very
86   specific epitopes decreasing the risks associated with allergic and autoimmune responses, besides
87   involving minimal viral components to stimulate adaptive immunity (Di Natale et al., 2020).
88   Additionally, their chemical or recombinant cloning synthesis allows large-scale production with
89   low costs and high reproducibility (Sun, 2013; Skwarczynski & Toth, 2016; Hudu, Shinkafi &
90   Umar, 2016).
91   The peptide-based vaccine design requires the immunoinformatic approach as part of the
92   computational vaccinology strategy for epitopes prediction (Ramana & Mehla, 2020; Oli et al.,
93   2020; Lu et al., 2021). This approach regards the wide availability of the SARS-CoV-2 NGS
94   (Next-Generation Sequencing) information associated with human leukocyte antigen (HLA)
95   profile (Kazi et al., 2018; Oli et al., 2020; Sharma et al., 2020) to identify T cell epitopes.
96   Therefore, these epitopes have the capability to effectively bind to HLA molecules activating a
97   long-lasting immune response mediated by CD8+ and CD4+ T cells (Fast, Altman & Chen, 2020;
98   Wang & Gui, 2020).
99   T cell epitopes offer advantages for vaccine design since it does not depend on the recognition of
100   structural proteins (Bashir et al., 2021; Redd et al., 2021) and are less affected by deletions and
101   mutations of emergent variants (Ribes, Chaccour & Moncunill, 2021; Jin et al., 2021).
102   Concerning COVID-19 immune response, T cell epitopes have the potential to provide long-term
103   protection from SARS-CoV-2. This characteristic allows the detection of memory T cell
104   responses to multiple SARS-CoV-2 proteins, which might contribute to disease control (Chen &
105   John Wherry, 2020; Karlsson, Humbert & Buggert, 2020; Sette & Crotty, 2021).
106   To provide effective T cell epitopes for *in vitro* peptide-based vaccine design a robust, refined
107   and accurate *in silico* selection of epitopes is crucial. Despite the many immunoinformatic tools
108   for epitope prediction, the curation for epitope selection is still limited and needs different web
109   servers to complete the analysis. In this paper, we focus on a computational prediction, curation
110   and validation of SARS-CoV-2 epitopes. We have as a central proposal a workflow (EpiCurator)
111   that brings together different approaches for accurate selection epitopes, providing the refined
112   identification of promising SARS-CoV-2 epitopes. To validate the efficacy of this new tool, we
113   use samples of circulating Brazilian lineages available in GISAID (https://www.gisaid.org/) (Elbe
114   & Buckland-Merrett, 2017) from December 2020 to April 2021.

115   2. Materials & Methods

116   2.1. Genome retrieval and protein annotation

117   Genome sequences of 1,652 SARS-CoV-2 genome isolated in Brazil were retrieved from the
118   GISAID database (Elbe & Buckland-Merrett, 2017) available from December 2020 to April 2021
119   (Table S1). The lineage distribution of the retrieved genomes includes P.1 (Gamma, n=770), P.2
120   (Zeta, n=525), B.1.1.28 (n=223) and B.1.1.33 (n=136). These genomes were analyzed by The
121   Viral Annotation Pipeline and iDentification (VAPiD) (Shean et al., 2019) to determine the
122   amino acid sequence for all SARS-CoV-2 proteins using as reference NC_045512.2 from the
123   NCBI database (https://www.ncbi.nlm.nih.gov/). The final genome processing includes clustering
124   the amino acid sequences for each protein using the CD-HIT package (Li & Godzik, 2006) with a
125   100% sequence identity threshold.

126 **2.2. Spike protein comparison**
127 Clustal Omega performed a homology analysis for the Spike protein (Sievers & Higgins, 2018)
128 and MView tool (Brown, Leroy & Sander, 1998) among 100 random samples of Brazilian
129 lineages and the VOCs, retrieved from the GISAID database (Elbe & Buckland-Merrett, 2017),
130 to measure the identity of the sequences to Gamma (P.1) samples.

131 **2.3. Epitope Prediction**
132 The T and B cell epitopes prediction was performed using the predictors NetCTL, NetMHCpan,
133 and NetMHCIIpan (https://services.healthtech.dtu.dk/) that allow the high-throughput computing
134 analysis. These predictors support FASTA files containing amino acids marked with "X" in the
135 sequence expanding the possibility of public genomes analysis, even with minor sequencing
136 errors. This prediction features decreasing the pre-processing sequence steps and avoids the false-
137 positive epitopes provided by joining sequences.

138 **2.3.1. Prediction of SARS-CoV-2 epitopes.**
139 To predict CD8+ T cell epitopes, the FASTA sequences of SARS-CoV-2 proteins are processed
140 using NetCTL v1.2 (Larsen et al., 2007). First, the sequences and the HLA class I supertypes,
141 provided by the software, are submitted to select 9-mer peptides (Chen et al., 1994; Sakaguchi et
142 al., 1997; Gfeller et al., 2018), then peptides with amino acids marked with "X" are removed
143 (Figure 1A). The predicted peptides in NetCTL are further processed using NetMHCpan v.4.0
144 software (Hoof et al., 2009; Jurtz et al., 2017) to identify epitopes with strong binding affinity to
145 HLA class I alleles. The prediction parameter is based on the predicted percentile rank ≤ 0.5%
146 and half-maximal inhibitory concentration (IC50) < 500 nM (Chen et al., 1994; Sakaguchi et al.,
147 1997; Gfeller et al., 2018) (Figure 1A). To assess binding affinity, alleles of HLA-A, B, and C
148 loci were selected from Allele Frequency Net Database (Gonzalez-Galarza et al., 2020) by their
149 Brazilian population frequency > 5% (Table S2).

150 To predict CD4+ T cell epitopes and estimate binding affinity to HLA class II molecules, the
151 FASTA sequences are processed using NetMHCIIpan v3.2 software (Greenbaum et al., 2011).
152 This tool selects the epitopes with 15-mer peptide lengths based on the predicted percentile rank
153 ≤ 2.0% and IC50 < 500 nM (Figure 1A). For the affinity prediction, the loci HLA-DRB1, HLA-
154 DPA1-DPB1, and HLA-DQA1-DQB1 were selected by the phenotypic frequency > 5% in the
155 Brazilian population (Table S2) from Allele Frequency Net Database (Gonzalez-Galarza et al.,
156 2020). For both cells (CD8+ and CD4+ T cell), the epitopes with broad HLA affinity coverage
157 (≥3 alleles) are filtered (Figure 1A).

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

163 **2.4. Accurate selection of epitopes - EpiCurator**
164 We developed a rigorous analysis workflow for accurate selection of epitopes, the EpiCurator,
165 bringing together a set of filters according to pre-established criteria, optimizing the analysis
166 since it brings tools available to use by high-throughput computing architectures. They also group
167 a series of analyses to guarantee the selection of unpublished and qualified epitopes. The

168   workflow allows the identification of epitopes by the following analysis: conservancy, homology

169   with the human genome, the overlap between epitopes of different classes, and the identification

170   of epitopes previously published in PubMed Central® (PMC) or available in the IEDB database

171   (Vita et al. 2019) beyond identifying their protein coordinates and mutations (Figure 1B).

172   **2.4.1. Prediction of Epitope Conservancy**

173   The first module of EpiCurator calculates the conservancy of a predicted epitopes list in SARS-

174   CoV-2 genomes using the BLAST command-line tools (Madden, 2020) (Figure 1B). A

175   customized BLAST database, optimized for shorter sequences (blastp-short task), was generated

176   with proteins of some SARS-CoV-2 circulating lineages from Brazil (B.1.1.28, P.1, P.2)

177   retrieved from the GISAID database (2,787 genome sequences) (Elbe & Buckland-Merrett,

178   2017). This database allows the comparison among the sequences of the predicted epitopes and

179   the SARS-CoV-2 proteins. Thus, the analysis reports the percentage of identity (conservancy) per

180   epitope using four criteria: 100% of identity; above 90%; between 70% and 90%, and less than

181   70%. The epitopes conserved with 100% of identity in at least 90% of the genomes are used for

182   further analysis (Figure 1B).

183   **2.4.2. Human homology**

184   This module uses the human proteins dataset from Ensembl (GRCh38.p13) to identify the

185   predicted epitopes sequence in the human genome. The workflow keeps the human protein

186   sequences in memory to enhance the analysis readiness, searching strictly for the corresponding

187   epitope sequences (exact match). Only the unmatched epitopes are selected, returning a filtered

188   list of epitopes without human homology (Figure 1B).

189   **2.4.3. Epitope sequence overlap**

190   A comparison is performed between the sequence of the epitopes derived from three groups of

191   prediction (B Cell, HLA Class I, and HLA Class II T Cell). This analysis calculates the

192   intersection and the identity among the lists of epitopes belonging to these groups returning four

193   reports: (i) the intersection of all groups, (ii) B-cell x Class I epitopes, B-Cell x Class II epitopes,

194   and (iii) Class I x Class II epitopes. The result keeps the epitopes with less than 60% similarity in

195   each report (Figure 1B).

196   **2.4.4. Search for epitopes from published articles - EpiMiner**

197   The EpiMiner was developed to execute an automatic search of the predicted epitopes list on

198   Pubmed and PMC published papers (Canese & Weis, 2013) (Figure 1B). It is a pipeline

199   performed in four steps. The first retrieves Pubmed and PMC articles that include the epitopes

200   sequence. The second step extracts the sentences of the body, abstract or tables of the article,

201   highlights the figure captions and breaks apart supplementary table data. The third step executes

202   natural language processing techniques such as tokenization to divide the sentences into words

203   and part-of-speech tagging to filter nouns and verbs (Chowdhury, 2005). These techniques

204   contribute to reducing search time for epitope sequence recognition, eliminating uninformative

205   sentences. The fourth step executes an entity recognition to identify epitopes sequence classified

206   as nouns, saving the sentences and their respective publication into a report. Epitopes information

207   is not searched if it is a part of the pictures or is in supplementary file in text format

208   (i.e. .doc, .docx, .txt, .pdf).

209   **2.4.5. IEDB matching**

210 To perform this analysis, the complete ensemble of epitopes from the Immune Epitope Database
211 (IEDB) v3 release (https://www.iedb.org/database_export_v3.php) was retrieved (accessed on
212 July 2021), and the structured files were parsed to filter the epitopes belonging to SARS-CoV-2
213 organisms (n= 1268) (Vita et al. 2019). This filtered file is used to calculate the identity between
214 the predicted list of epitopes and the ones retrieved, apprising the IEDB epitope ID and the
215 respective similarity (Figure 1B).

216 **2.4.6. Mutation screening**
217 Epitopes mutation is reported by comparing the epitope sequence with the SARS-CoV-2 Wuhan
218 protein sequence and assigning the epitopes' coordinates in the respective whole protein. The
219 alignment is conducted with the blastp function of the BLAST command-line tools (Madden,
220 2020). The SARS-CoV-2 reference protein sequences used for alignment are published on the
221 Uniprot Database (https://covid-19.uniprot.org/). This analysis describes the coordinates and
222 presence of mutations in the predicted epitopes (Figure 1B).

223 **2.5. Epitope properties**
224 **2.5.1. Evaluation of antigenicity, toxicity, and immunogenic profile**
225 The VaxiJen v2.0 server was applied to analyze the antigenicity of the predicted B cell and T cell
226 epitopes with a conservative score threshold of 0.7 (Doytchinova & Flower, 2007a,b) (Figure
227 1B). The toxicity is retrieved from the ToxinPred online server with support vector machine
228 (SVM) based methods (threshold -0.4) and e-value cut-off 0.01 (Gupta et al., 2013) (Figure 1B).
229 The allergenic properties are retrieved from the AlgPred2 online server with a hybrid prediction
230 model (threshold 0.5) (Sharma et al., 2020) (Figure 1B). Despite being part of the accurate
231 selection (Figure 1B), they are not available for incorporation into the analytic workflow (item
232 2.4) which prevents us from optimizing their assay. In addition, several on-line servers were used
233 to evaluate the immunogenic profile of the T cell epitopes (Figure 1C). The immunogenicity
234 predictions were performed by the IEDB server (https://www.iedb.org/) (Paul et al. 2015; Calis et
235 al. 2013; Dhanda et al. 2018; Vita et al. 2019). Likewise, their capability to induce interferon-
236 gamma (IFNγ), interleukin-4 (IL-4), interleukin-10 (IL-10), and interleukin-17 (IL-17) (Dhanda,
237 Vir & Raghava, 2013; Dhanda et al., 2013; Gupta et al., 2017; Nagpal et al., 2017) was evaluated.
238 Additionally, their proinflammatory activity (Gupta et al., 2016) and immunomodulatory
239 potential (Nagpal et al., 2018) was verified. The servers and parameters to evaluate the
240 immunogenic profile are shown in figure 1C.

241 **2.5.2. Estimation of population coverage**
242 The IEDB's Population Coverage on-line tool (http://tools.iedb.org/population/) is used to
243 analyze how T cell epitopes-HLA binding alleles diverge across ethnicities, regions, and
244 countries around the world (Bui et al. 2006; Vita et al. 2019) (Figure 1C). The predicted epitopes
245 and their respective HLA binding alleles (Table S3) were inputted in the IEDB tool with
246 separated allele class option and a list of countries/regions (Argentina, Brazil, England, France,
247 Italy, Spain, United States, and World) was selected.

248 **2.5.3. Docking analysis of the HLA-epitope complex**
249 To validate the binding affinity of the predicted epitopes with HLAs structures, docking analysis
250 was performed using the PepDock tool (Lee et al., 2015) of the GALAXY web server (Figure
251 1C). To docking analysis, the Protein Data Bank archive (PDB) of eighth HLA alleles (HLA-

252   A*01:01, HLA-A*02:01, HLA-B*08:01, HLA-C*12:03, HLA-DRB1*03:01, HLA-
253   DRB1*04:01, HLA-DRB1*12:02 and HLA-DRB1*15:01) was retrieved from the pHLA3D
254   database (Menezes Teles E Oliveira et al., 2019) and RCSB PDB database (Berman et al., 2000;
255   Burley et al., 2021).
256   The PepDock uses the HLA alleles PDBs and the epitopes sequence to perform a template-based
257   model selection and then proceeds to the docking and refinement processes to optimize the
258   energy score ranking ten complex models. The best model of each HLA-epitope complex was
259   selected based on the major similarity score of protein structure and interaction beyond the
260   highest estimated accuracy. Additionally, to identify the free energy and the residue's contacts of
261   the selected complexes, the Prodigy tool (Xue et al., 2016) and Chimera software (Goddard,
262   Huang & Ferrin, 2005) were used respectively.
263   **2.6. Genomes for EpiCurator pairwise comparison validation**
264   Protein sequences for SARS-CoV-2 isolates reported by (Crooke et al., 2020)(2020(Crooke et al.,
265   2020) were identified and retrieved from the Virus Pathogen Resource (ViPR) database (n=
266   641,635); additionally, six genome sequences reported by (Kiyotani et al., 2020)(2020(Kiyotani
267   et al., 2020), two sequences for N and S protein reported by Chen et al., (2020), and five S
268   sequences reported by Chukwudozie et al., (2021) were retrieved from NCBI GenBank. These
269   samples were processed using prediction parameters of the HLA alleles reported by the authors
270   with the approach reported in this Methods section (2.3. and 2.4 analysis). Regarding item 2.4 of
271   Methods, we used only the three main analysis steps (conservancy, human homology and IEDB
272   matching) to compare the selected epitopes ensemble by the papers, and ones chosen
273   independently by our approach, leaving out the EpiMiner analysis since all the epitopes are
274   published.
275   **2.7. Multi-epitope construct and structural modelling for EpiCurator validation**
276   The RBD epitopes (n = 11) were used to construct a multi-epitope sequence connected by
277   specific linkers. The linker aimed to separate the epitopes, so that it improved their expression,
278   folding and stability beyond to prevent their fusion and facilitate the immune processing of
279   antigen (Arai et al., 2001; Kar et al., 2020). In the multi-epitope arrangement are also added an
280   adjuvant (*Escherichia coli* 50S ribosomal protein L7/L12 (UniProt P0A7K2)) in the N-terminal
281   sequence and a histidine hexamer in the C-terminal portion (Figure 1D, Figure S1).
282   **2.7.1. Linear and secondary structure evaluation**
283   To evaluate the properties and immune profile of multi-epitope, its linear sequence was submitted
284   to several analyses (Figure 1D) as follow: antigenicity (Doytchinova & Flower, 2007a; Magnan
285   et al., 2010), allergenicity (Dimitrov et al., 2014b,a), and solubility for cell-free expression
286   analysis (Hebditch et al., 2017), for overexpression analysis (Magnan, Randall & Baldi, 2009),
287   and structurally solubility profile (Hou et al., 2020). In addition, we assess its physicochemical
288   properties (Gasteiger et al., 2005). The servers and parameters to linear multi-epitope properties
289   evaluation are shown in figure 1D.
290   The linear sequence was also analyzed by the C-IMMSIM server (https://kraken.iac.rm.cnr.it/C-
291   IMMSIM/) to evaluate the *in silico* immune profile of the multi-epitope construct (Rapin et al.,
292   2010). Two simulations were performed with intervals of 4 or 12 weeks (Saad-Roy et al., 2021;
293   Cobey et al., 2021) (Figure 1D). Furthermore, the secondary structure of the multi-epitope was

294   evaluated with PSIPRED v.4 with an accuracy of 84.2% (Jones, 1999; McGuffin, Bryson &

295   Jones, 2000; Buchan & Jones, 2019) (Figure 1D)

296   **2.7.2. Multi-epitope 3D structure modelling, refinement, and evaluation**

297   The tertiary structure multi-epitope construct is modeled by the RaptorX server

298   (http://raptorx.uchicago.edu/ContactMap) that predicts structural properties such as solvent

299   accessibility (ACC) and disorder regions (DIS) (Källberg et al., 2014) (Figure 1D). This tertiary

300   structure was submitted to a refinement process using the GalaxyRefine

301   (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) server which improves global and

302   local model quality by rebuilding all side-chain conformations and applying structural relaxations

303   (Heo, Park & Seok, 2013). The quality of the refined model is assessed by several parameters, as

304   reported in figure 1D. The refined 3D structure is further evaluated by the ProSA-web server

305   (https://prosa.services.came.sbg.ac.at/prosa.php) to validate structural models checking for

306   potential errors (Berman et al., 2000; Wiederstein & Sippl, 2007) (Figure 1D). The final refined

307   3D structure was used to perform a docking by the ClusPro server (https://cluspro.bu.edu) (Desta

308   et al., 2020) using the multi-epitope 3D structure and the Toll-Like Receptors TLR4 (PDB: 2Z63)

309   and the TLR3 (PDB: 3CIG). The complex model with the lowest energy is chosen for each

310   receptor (Figure 1D). Their stability is assessed with molecular dynamics simulation on the

311   iMODS server (http://imods.chaconlab.org) that performs Normal Mode Analysis (NMA) to

312   describe functional motions between macromolecules in complexes and simulates feasible

313   trajectories between two conformations (López-Blanco et al., 2014) (Figure 1D).

314   # 3. Results

315   **3.1. Linear epitopes prediction from SARS-CoV-2 genome**

316   Three main epitopes prediction comprise HLA Class I (CD8+ T cell), HLA Class II (CD4+ T

317   cell), and B cell. To predict potential CD8+ T cell epitopes, the NetCTL and NetMHCpan,

318   predictive algorithms were performed for all proteins annotated from 1,652 SARS-CoV-2

319   genomes (see Methods). Thus 9-mer epitopes were predicted with strong binding affinity

320   assigned by percentile scores (rank) ≤0.5% across HLA class I alleles frequent in the Brazilian

321   population (Figure 1A). Cumulatively reaching the prediction of 5,261 HLA class I epitopes with

322   HLA promiscuity, binding to ≥3 HLA alleles, implying broad population coverage. The affinity

323   to HLA-C alleles matches the highest epitope binding locus suggesting it could be the best

324   grooves for these epitopes and would lead to the activation of the cell-mediated response (Table

325   1).

326   We also sought to predict potential 15-mer epitopes with binding affinity to HLA class II, using

327   the NetMHCIIpan software (Figure 1A), reaching 7,649 candidate HLA class II epitopes from the

328   SARS-CoV-2 genomes. They have a strong binding affinity to ≥3 HLA alleles across a reference

329   panel of HLA molecules (see Methods). These epitopes showed a preferential affinity for alleles

330   from locus HLA-DRB1 (n= 3,702), suggesting that HLA-DRB1 alleles could lead these epitopes

331   to activate the cellular immune response (Table 1).

332   To pairwise the cellular and humoral immune responses activation, the additional prediction of

333   linear B cell epitopes was performed by BepiPred algorithm identifying 257 epitopes from

334   SARS-CoV-2 structural proteins (Table 1).

335   **3.2. Quality assessment analysis of the EpiCurator**

336 A robust and refined accurate selection of epitopes is crucial to improve the development of
337 peptide-derived vaccines. To this goal, the EpiCurator brings together six analyses (Conservancy,
338 Human homology, Epitopes overlap, EpiMiner, IEDB matching, and Mutation screening) to the
339 accurately selected epitopes (Figure 1B).
340 To the workflow's quality assessment analysis, we sought to characterize the number of epitopes
341 taken for the main analysis. The epitopes conservancy selected 29.23% of the predicted epitopes
342 with 100% identity across at least 90% of the SARS-CoV-2 samples. Nevertheless, 70.77% of the
343 predicted epitopes do not correspond to the conservancy parameter (Figure 2, Figure S2). Human
344 homology analysis identified only 0.1% of the predicted epitopes sequence in the human genome,
345 keeping 99% accurately selected, unmatched with human sequences (Figure 2, Figure S2). In
346 addition, the screening by the previously published epitopes (EpiMiner and IEDB matching
347 analysis) allowed selected > 80% of new epitopes, since 17.57% of them have already been
348 described in the literature and/or IEDB server (Figure 2, Figure S2). To confirm the effectiveness
349 of the EpiMiner we provide all the articles IDs and DOI in which the epitopes were found in table
350 S4.
351 To further assess the quality of the EpiCurator analysis, a pairwise comparison validation was
352 performed. The comparison used NGS data reported by four papers with substantial similarity
353 with our methodologies (Kiyotani et al., 2020; Chen et al., 2020; Crooke et al., 2020;
354 Chukwudozie et al., 2021). The prediction step allows identifying the same HLA class I and HLA
355 class II-restricted T cell epitopes. At the same time, the accurate selection provided by the
356 EpiCurator gathered 51.1% of the paper's epitopes (Table S5). The conservancy across SARS-
357 CoV-2 genomes samples reached 95,6% of the paper's epitopes, with only 4.4% not conserved by
358 our parameters. The human homology step certifies the absence of the article's selected epitopes
359 in the Human genome since the analysis does not recognize the homology of any of them.
360 Interestingly, the IEDB matching step identified 45.6% of the paper's epitopes as having already
361 been published on the IEDB server showing the importance of this step analysis if we want to
362 describe the epitopes by the first time (Table S5).
363 **3.3. Properties of accurately selected epitopes**
364 Assembling the workflow analysis results, 199 (3.78%) HLA class I-restricted T cell epitopes
365 were selected from SARS-CoV-2 proteins, mainly identified in ORF1ab (n = 154 (77.4%)) and
366 Spike glycoprotein (n = 15 (7.5%)) (Table 2, Figure 3A). The epitopes keep a high affinity for
367 HLA-C alleles (Figure 3C). The HLA class II epitopes' accurate selection reached 153 (2%)
368 epitopes, also mainly identified in ORF1ab (n = 111 (72.5%)) and Spike glycoprotein (n = 22
369 (14.4%)). The HLA class II epitopes showed an affinity prevalence for the HLA-DPA1-DPB1
370 haplotypes (Figure 3D). The comprehensive workflow analysis for B cell epitopes selected
371 14.6% of predicted epitopes with 60% of the epitopes from the Spike glycoprotein (Table 2).
372 Interestingly, all epitopes have high antigenicity (0.92±0.30), are non-toxic and non-allergenic
373 (Table S3).
374 Furthermore, the T cell epitopes immunogenic potential was assessed to characterize the
375 capability of inducing *in silico* protective immune responses. The analysis identified the profile
376 of HLA class I and HLA class II epitopes respectively as follow: IL-4 inducer activity (26.6%

377 and 23.8%), IFNγ production (22.1% and 39.7%), immunomodulatory activity (2% and 3.3%),
378 and proinflammatory activity (78.9% and 81.4%) (Table S3).
379 In addition, to assess the capability to be an *in silico* Brazilian epitopes candidate, we sought to
380 estimate their distribution among the Brazilian circulating lineages at the time of analysis (P.1,
381 P.2 and B1.1.28). Firstly, we identify the number of epitopes in common across P.1, P.2 and
382 B.1.1.28 lineages identifying the highest proportion of HLA class II epitopes (61.4%), following
383 by HLA class I (46.7%) and B cell (14.3%) (Figure 3B). Thereafter, considering the most
384 representative lineage in Brazil, at the time data was retrieved (P.1). Remarkably, proportions
385 highest than 50% for all the epitopes were identified, with 70.1% of HLA class II epitopes,
386 69.8% of HLA class I, and 51.4% of B cell epitopes (Figure 3B).
387 In the last analysis, considering the HLA promiscuity epitopes selection, the population coverage
388 for the T cell epitopes associated with their respective HLA allele binding (Table S3) was
389 estimated by the IEDB server. Notably, the T cell epitopes have a wide population coverage,
390 presenting 99.52% (HLA class I epitopes) and 100% (HLA class II epitopes) of cumulative
391 Brazilian population coverage and 98.45% (HLA class I epitopes) and 99.87% (HLA class II
392 epitopes) of worldwide population coverage (Table S6).
393 **3.4. Epitope-specific RBD Spike as a baseline for validation of EpiCurator selection**
394 The EpiCurator allows the selection of the majority of the epitopes in the ORF1ab and Spike
395 glycoprotein. The SARS-CoV-2 Spike glycoprotein has greater prominence concerning the virus
396 and host interaction and has been the main target in epitopes prediction (Shang et al., 2020; Walls
397 et al., 2020). Therefore, we assumed the spike epitopes as a baseline to validate the EpiCurator
398 accurate selection. The Spike epitopes (n = 58) were mainly in the N-terminal domain (NTD)
399 (53.4%) and receptor-binding domain (RBD) (18.9%) (Figure 4).
400 The RBD epitopes (n = 11) comprise three of HLA class I (461-NYNYRYRLF-469, 506-
401 QSYGFQPTY-514, 516-FGYQPYRVV-524), three of HLA class II (322-
402 EKGIYQTSNFRVQPI-336, 322-EKGIYQTSNFRVQPR-336, 443-TGCVIAWNSKNLDSK-
403 457), and five B cell epitopes (332-RVQPTES-338, 424-APGQTGK-430, 424-APGQTGT-430,
404 455-DSKVGGNYN-463, 474-LKPFERD-480) (Figure 4).
405 To evaluate the robust and refined RBD epitopes selection, we perform a diversified analysis.
406 They unveil the highest antigenicity (1.1 ± 0.27) (Figure 5A) reflecting their ability to bind
407 molecules of adaptive immunity. Notably, their affinity for several HLA alleles are high (0.49 ±
408 0.15) (Figure 5A) and achieves promiscuity with ≥ 5 HLA allele binding per epitope (Figure 5B).
409 The RBD epitopes HLA data pairwise with population coverage that remains at around 80% in
410 the Brazilian and worldwide population (Figure 5B). To further characterize the RBD epitopes,
411 they reach the conservancy of over 99% across SARS-CoV-2 genomes samples of lineages P.1,
412 P.2 and B.1.1.28 published on GISAID (See Methods) (Table S3).
413 The high conservancy observed suggests that epitope sequences are shared in the circulating
414 Brazilian lineages. Considering the natural homology among the SARS-CoV-2 lineages but
415 regardless of the epitopes' conservancy with other lineages, we expand the identification of the
416 RBD epitopes for diversified samples available in GISAID at the time of writing. Interestingly
417 these epitopes were identified in more than 1 million samples for around 1 thousand lineages of
418 GISAID (Table S7). These findings include some samples of current VOCs: Alpha (B.1.1.7),

419   Beta (B.1.351), and Delta (B.1.617.2) (Table S7). Concern about the natural lineages' homology
420   mentioned, we perform a comparison among spike glycoprotein of described Brazilian lineages
421   and VOCs. The evaluation of random 100 samples available in GISAID of each lineage shows
422   99.1% of identity with 99.9% of coverage confirming the sequence lineages similarity (Table
423   S8).
424   One of the mentioned properties of RBD epitopes is the high affinity for HLA alleles. To validate
425   these findings the *in silico* structural binding performance was assessed by the PepDock docking
426   tool. Thus, the RBD epitopes and the most major HLA alleles in studies with COVID-19 (see
427   Methods) were structurally bonded in complex HLA-epitope (Figure S3). All complexes were
428   seen with high-performance scores demonstrating *in silico* epitope effectiveness in activating the
429   cell-mediated immune response via MHC presentation. Their high estimated accuracy selected
430   the two most significant HLA-epitope complexes (Table S9, Figure 6). The HLA class I epitope
431   516-FGYQPYRVV-524 in complex with HLA-A*01:01, HLA-A*02:01, HLA-B*08:01, and
432   HLA-C*12:03; and the HLA class II epitope 443-TGCVIAWNSKNLDSK-457 in complex with
433   HLA-DRB1*03:01 and the HLA-DRB1*12:02 alleles (Table S9, Figure 6). Additionally, the best
434   measured free energy of the complexes was ΔG = -23 kcal/mol for 461-NYNYRYRLF-469 HLA
435   class I epitope in complex with HLA-A*01:01 and ΔG = -31.6 kcal/mol for 322-
436   EKGIYQTSNFRVQPR-336 HLA-class II epitope in complex with the HLA-DRB1*04:01 allele
437   (Table S9, Figure 6). An additional analysis was performed to assess the specific residues
438   involved in the structural binding allowing to identify the specific amino acids in contact between
439   the epitopes and the HLA alleles (Table S9).
440   **3.5. Multi-epitope construct for *in silico* validation of RBD epitopes**
441   To further validate the RBD epitopes and consequently confirm the effectiveness of EpiCurator
442   in providing a curated selection, the epitopes were inserted in a multi-epitope construct. It
443   consists of 287 amino residues, including 11 RBD selected epitopes joined by linkers (Figure
444   S1). The structural appraisal of the secondary structure predicted by the PSIPRED server
445   revealed 31.7% alpha-helix, 13.5% beta-strand and the disordered region was 12% (Figure S4).
446   Some evaluation of the linear sequence of multi-epitope were identified as follows: high
447   antigenicity by the Vaxijen 2.0 (0.87) and the AntigenPro (0.93), good solubility by SolPro
448   (0.91), Protein-sol (0.66), and SOLart (65.80%), non-allergenic feature by AllerTOP and
449   AllergenFP, hydrophilic property (hydropathicity = -0.438) and stability (instability index =
450   20.56), with pI 7.03, and molecular weight calculated to be 29 kDa.
451   Another assessment of linear sequence includes the *in silico* immunogenic profile provided by the
452   C-IMMSIM immune server. The analysis showed a gradual increase of IgM, IFN-γ (associated
453   with both CD8+ T-cell and CD4+ Th1 response as shown in Figure 7A), and IL-2 level after each
454   multi-epitope exposure indicated an elevated immune response (Figure 7B). Besides, an adequate
455   generation of both IgG1 and IgG2 was shown (Figure 7C) with high levels of clonal proliferation
456   of B-cell and T-cell population (Figure S5A and B). In addition, the development of *in silico*
457   immune memory was assessed by the abundance of different types of B-cells and T-cells (Figure
458   S5C and D).
459   The favorable *in silico* immunogenic profile of linear construct led to a multi-epitope 3D
460   modelling to assess TLR binding capability adding validation for the curated selection provided

461 by the EpiCurator. In this context, we modeled and evaluated the 3D multi-epitope to choose the
462 best model for the TLR binding assay. The best predicted tertiary structure returned by RaptorX
463 has an RMSD score of 11.218 (Figure S6A), 62% of residues are predicted to be exposed and
464 30% are predicted to be disordered (Figure S6B). The refinement by GalaxyRefine allows the
465 choice of the best model based on its quality scores available in table S10. This model highlights
466 the epitopes according to their type (B cell, HLA class I T cell and HLA class II T cell) (Figure
467 S7). The overall model quality of the refined structure was further validated in the ProSA web
468 (Figure S8A), showing the energy results by position (Figure S8B) with good local quality since
469 all energy values of the residues are negative.
470 Considering the accuracy of the refined model, a docking between our multi-epitope and the
471 immune receptors (TLR3 and TLR4) was performed to check stability and binding affinity. The
472 best complexes (Figure 8) for each receptor had free energy values of ΔG = -1048 kcal/mol
473 (TLR3) and ΔG = -1020 kcal/mol (TLR4), indicating a high binding affinity. In addition, the
474 stability and physical movements of the complexes were confirmed using molecular dynamics
475 simulation in the iMODS server. The complex results are presented in Figures S9 and S10.
476

4. Discussion

478 Several studies have used immunoinformatic approaches to select B cell and T cell SARS-CoV-2
479 epitopes for epitopes-based vaccine formulation (Ramana & Mehla, 2020; Oli et al., 2020;
480 Siracusano, Pastori & Lopalco, 2020; Yoshida et al., 2021; Ribes, Chaccour & Moncunill, 2021;
481 Jin et al., 2021). However, current computational methods are limited since they identify a large
482 number of epitopes and need different web servers for accurate selection (Bui et al., 2006;
483 Doytchinova & Flower, 2007b; Gupta et al., 2013). This study developed an approach that
484 gathered diversified analysis assembled in a workflow (EpiCurator). It accurately selects SARS-
485 CoV-2 epitopes with useful *in silico* properties for an immunogen candidate target.
486 Its effectiveness was validated with a pairwise comparison analysis with four previously
487 published papers (Kiyotani et al., 2020; Chen et al., 2020; Crooke et al., 2020; Chukwudozie et
488 al., 2021). The main analysis of this validation was the IEDB matching that promotes a robust
489 and refined selection of epitopes which confirms the importance of the IEDB database in the
490 epitopes analysis studies (Beaver, Bourne & Ponomarenko, 2007). In addition, our approach can
491 be used for extensive scale search and high-throughput computing analysis with the SARS-CoV-
492 2 genomes available, an advantage to analyze a pandemic emergent virus (Ojha et al., 2020;
493 Minervina et al., 2021; Pham et al., 2021).
494 Remarkably, our workflow conducts a thorough and facilitated accurate selection that enables us
495 to prioritize the epitopes candidates. The EpiCurator recognizes patterns of cross-conservation
496 with SARS-CoV-2 and human epitopes, eliminating the ones with significant human homology
497 (Meyers et al., 2021). In addition, notably identifying epitopes in published articles, ensuring the
498 selection *in silico* candidates with a plausible assumption to be described by the first time in this
499 study. Consequently, this analysis might also validate the effectiveness of our workflow's
500 accurate selection analysis, since our epitopes were identified in different studies previously
501 published (Prachar et al., 2020). The workflow also allows the selection of epitopes by high
502 conservancy (≥90%), across more than two thousand SARS-CoV-2 circulating Brazilian lineages

503   in accordance with previous reports for SARS-CoV-2 epitopes selection (Zaheer et al., 2020;
504   Mahapatra et al., 2020; Mallajosyula et al., 2021).
505   Three main phases comprise our approach to SARS-CoV-2 epitopes identification, the prediction,
506   accurate selection and epitopes validation. Therefore, taking the spike glycoprotein epitopes as
507   parameters, they represent 10% of the predicted epitopes and 2% of the accurately selected
508   similarly to those presented in other reports (Kiyotani et al., 2020; Crooke et al., 2020) with the
509   resembling result for the other proteins. Indeed, the spike data give us an advantage since they
510   were used to optimize the validation of the epitopes accurately selected by EpiCurator. Thus, they
511   are taken as a baseline to conduct a thorough analysis. This baseline is reached since the Spike
512   plays the most crucial role in the entry of viral particles into host cells, promoting an effective
513   infection (Ou et al., 2020). These characteristics make the spike a chosen target for epitope
514   screening leading to vaccine development (Shang et al., 2020; Walls et al., 2020; Lin et al.,
515   2020). Pertaining to the Spike, the RBD region induces responses that block S protein binding
516   with the human cell receptor, neutralizing SARS-CoV-2 infection (Rakib et al., 2020; Yang et al.,
517   2020), having substantial importance to peptide-based vaccine studies. Considering this
518   importance, we prioritize 11 epitopes of this study identified on Spike RBD regions.
519   The RBD epitopes were identified in several SARS-CoV-2 lineages emphasizing samples of the
520   current VOCs. They have favorable *in silico* structure interactions with the main HLA alleles
521   related to the COVID-19 response (Tavasolian et al., 2020; Tomita et al., 2020; Shkurnikov et al.,
522   2021). Thereby, it is plausible to assume that these epitopes could be responsible for the strong
523   activation of the cell-mediated immune response (Patronov et al., 2011; Sarma, Olotu & Soliman,
524   2021) in case of an experimental assay. In addition, the RBD epitopes have a broad populational
525   coverage consistent with studies for SARS-CoV-2 epitopes identified in lineages isolated from
526   India, England and the United States (Mallavarpu Ambrose et al., 2021). These findings
527   combined with the high conservancy, antigenicity and immunogenicity suggest these epitopes' *in*
528   *silico* immunogen profile (Zheng & Song, 2020; Mallavarpu Ambrose et al., 2021; Jahangirian et
529   al., 2021), validating the accurate selection of EpiCurator.
530   Indeed, the RBD epitopes have substantial evidence and are seen as good epitope candidates.
531   Despite sharing this evidence, they could together increase the immune responses and confer
532   better protection against SARS-CoV-2 (Jahangirian et al., 2021). Accepting this assumption, they
533   were used to design a multi-epitope construct to further validate their immunogenic properties
534   (Kalita et al., 2020; Mohammad et al., 2020; Singh et al., 2020; Yang, Bogdan & Nazarian, 2021;
535   Lim et al., 2021; Sharma et al., 2021). The multi-epitope has the capability of *in silico* activation
536   of memory B and T-cell with Th1 response. Several experimental studies about immune response
537   against COVID-19 endorse these findings (Lipsitch et al., 2020; Hartley et al., 2020; Ghazavi et
538   al., 2021; Quast & Tarlinton, 2021; Tarke et al., 2021). On the other hand, no response Th17 was
539   found in the *in silico* assay, a common response that characterizes the severe COVID-19 profile
540   (Wu & Yang, 2020). Additionally, the multi-epitope had low binding energy and high stability of
541   *in silico* structure interactions with TLR3 and TLR4 similar to other computational studies (Kar
542   et al., 2020; Rahman et al., 2020; Yang, Bogdan & Nazarian, 2021; Nemati et al., 2021; Saba et
543   al., 2021). This interaction is important since TLR is an innate immune receptor that recognizes
544   viral proteins and triggers infection resistance (O'Neill, Golenbock & Bowie, 2013).

545 Furthermore, TLR3 and TLR4 are specifically related by different studies as part of immune
546 response in COVID-19 (Patra, Chandra Das & Mukherjee, 2021; Khanmohammadi & Rezaei,
547 2021; Kaushik, Bhandari & Kuhad, 2021). The findings in multi-epitope analysis confirm the
548 importance of the RBD epitopes, selected by the EpiCurator, as good *in silico* immunogens
549 candidates, reinforcing our approach's efficiency for a curated selection of epitopes.
550

551 5. Conclusions

552 It is important to reinforce that this work focuses on a computational prediction and selection of
553 epitopes seeking to perform several *in silico* validation. Therefore, our approach is useful to
554 researchers designing an experimental peptide-based vaccine to control the disease. Since,
555 indeed, the development of an effective vaccine requires a detailed experimental investigation of
556 the immunological correlations with SARS-CoV-2. Regarding the computational analysis,
557 assembling the accurate selection of epitopes and their validation consolidate our approach as a
558 helpful workflow analysis to provide epitopes for immunogens. These epitopes could
559 significantly activate the *in silico* immune response against some circulating Brazilian variants.
560 Furthermore, considering that the pandemic is still ongoing, our approach could contribute to
561 continuous monitoring and identification of new SARS-CoV-2 epitopes according to the
562 emergence of variants over time.
563

564 6. Acknowledgements

565

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

571
572
573

574 7. References

575

576 Andreano E, Piccini G, Licastro D, Casalino L, Johnson NV, Paciello I, Dal Monego S, Pantano
577   E, Manganaro N, Manenti A, Manna R, Casa E, Hyseni I, Benincasa L, Montomoli E,
578   Amaro RE, McLellan JS, Rappuoli R. 2021. SARS-CoV-2 escape from a highly neutralizing
579   COVID-19 convalescent plasma. *Proceedings of the National Academy of Sciences of the
580   United States of America* 118. DOI: 10.1073/pnas.2103154118.
581 Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. 2001. Design of the linkers which
582   effectively separate domains of a bifunctional fusion protein. *Protein engineering* 14:529–
583   532. DOI: 10.1093/protein/14.8.529.
584 Badiani AA, Patel JA, Ziolkowski K, Nielsen FBH. 2020. Pfizer: The miracle vaccine for
585   COVID-19? *Public health in practice (Oxford, England)* 1:100061. DOI:
586   10.1016/j.puhip.2020.100061.
587 Bashir Z, Ahmad SU, Kiani BH, Jan Z, Khan N, Khan U, Haq I, Zahir F, Qadus A, Mahmood T.

588   2021. Immunoinformatics approaches to explore B and T cell epitope-based vaccine
589   designing for SARS-CoV-2 Virus. *Pakistan journal of pharmaceutical sciences* 34:345–352.
590   DOI: 10.36721/pjps.2021.34.1.sup.345-352.1.
591   Beaver JE, Bourne PE, Ponomarenko JV. 2007. EpitopeViewer: a Java application for the
592   visualization and analysis of immune epitopes in the Immune Epitope Database and Analysis
593   Resource (IEDB). *Immunome research* 3:3. DOI: 10.1186/1745-7580-3-3.
594   Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE.
595   2000. The Protein Data Bank. *Nucleic acids research* 28:235–242. DOI:
596   10.1093/nar/28.1.235.
597   Bloom DE, Cadarette D, Ferranna M, Hyer RN, Tortorice DL. 2021. How New Models Of
598   Vaccine Development For COVID-19 Have Helped Address An Epic Public Health Crisis.
599   *Health affairs* 40:410–418. DOI: 10.1377/hlthaff.2020.02012.
600   Brown NP, Leroy C, Sander C. 1998. MView: a web-compatible database search or multiple
601   alignment viewer. *Bioinformatics* 14:380–381. DOI: 10.1093/bioinformatics/14.4.380.
602   Brüssow H. 2021. COVID-19: emergence and mutational diversification of SARS-CoV-2.
603   *Microbial biotechnology* 14:756–768. DOI: 10.1111/1751-7915.13800.
604   Buchan DWA, Jones DT. 2019. The PSIPRED Protein Analysis Workbench: 20 years on.
605   *Nucleic acids research* 47:W402–W407. DOI: 10.1093/nar/gkz297.
606   Bui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. 2006. Predicting population
607   coverage of T-cell epitope-based diagnostics and vaccines. *BMC bioinformatics* 7:153. DOI:
608   10.1186/1471-2105-7-153.
609   Bulla HAM. 2021. COVID -19: EFFICACY AND SAFETY PROFILE OF MAIN VACCINES
610   APPROVED FOR EMERGENCY USE AUTHORIZATION IN 2021. *International Journal*
611   *of Research -GRANTHAALAYAH* 9:271–283. DOI:
612   10.29121/granthaalayah.v9.i7.2021.4062.
613   Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow GV, Christie CH, Dalenberg K, Di
614   Costanzo L, Duarte JM, Dutta S, Feng Z, Ganesan S, Goodsell DS, Ghosh S, Green RK,
615   Guranović V, Guzenko D, Hudson BP, Lawson CL, Liang Y, Lowe R, Namkoong H,
616   Peisach E, Persikova I, Randle C, Rose A, Rose Y, Sali A, Segura J, Sekharan M, Shao C,
617   Tao Y-P, Voigt M, Westbrook JD, Young JY, Zardecki C, Zhuravleva M. 2021. RCSB
618   Protein Data Bank: powerful new tools for exploring 3D structures of biological
619   macromolecules for basic and applied research and education in fundamental biology,
620   biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic acids research*
621   49:D437–D451. DOI: 10.1093/nar/gkaa1038.
622   Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Keşmir C, Peters B.
623   2013. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS*
624   *computational biology* 9:e1003266. DOI: 10.1371/journal.pcbi.1003266.
625   Canese K, Weis S. 2013. Pubmed: the bibliographic database. In: *The NCBI Handbook*
626   *[Internet]*. National Center for Biotechnology Information (US),.
627   Cella E, Benedetti F, Fabris S, Borsetti A, Pezzuto A, Ciotti M, Pascarella S, Ceccarelli G, Zella
628   D, Ciccozzi M, Giovanetti M. 2021. SARS-CoV-2 Lineages and Sub-Lineages Circulating
629   Worldwide: A Dynamic Overview. *Chemotherapy*:1–5. DOI: 10.1159/000515340.
630   Chakraborty C, Sharma AR, Bhattacharya M, Sharma G, Lee S-S. 2020. The 2019 novel
631   coronavirus disease (COVID-19) pandemic: A zoonotic prospective. *Asian Pacific journal*
632   *of tropical medicine* 13:242. DOI: 10.4103/1995-7645.281613.
633   Chen Z, John Wherry E. 2020. T cell responses in patients with COVID-19. *Nature reviews.*
634   *Immunology* 20:529–536. DOI: 10.1038/s41577-020-0402-6.
635   Chen Y, Sidney J, Southwood S, Cox AL, Sakaguchi K, Henderson RA, Appella E, Hunt DF,
636   Sette A, Engelhard VH. 1994. Naturally processed peptides longer than nine amino acid

637   residues bind to the class I MHC molecule HLA-A2.1 with high affinity and in different
638   conformations. *Journal of immunology* 152:2874–2881.
639 Chen H-Z, Tang L-L, Yu X-L, Zhou J, Chang Y-F, Wu X. 2020. Bioinformatics analysis of
640   epitope-based vaccine design against the novel SARS-CoV-2. *Infectious diseases of poverty*
641   9:88. DOI: 10.1186/s40249-020-00713-3.
642 Chowdhury GG. 2005. Natural language processing. *Annual review of information science and*
643   *technology* 37:51–89. DOI: 10.1002/aris.1440370103.
644 Chukwudozie OS, Gray CM, Fagbayi TA, Chukwuanukwu RC, Oyebanji VO, Bankole TT,
645   Adewole RA, Daniel EM. 2021. Immuno-informatics design of a multimeric epitope peptide
646   based vaccine targeting SARS-CoV-2 spike glycoprotein. *PloS one* 16:e0248061. DOI:
647   10.1371/journal.pone.0248061.
648 Cobey S, Larremore DB, Grad YH, Lipsitch M. 2021. Concerns about SARS-CoV-2 evolution
649   should not hold back efforts to expand vaccination. *Nature reviews. Immunology* 21:330–
650   335. DOI: 10.1038/s41577-021-00544-9.
651 Cohen J, Normile D. 2020. New SARS-like virus in China triggers alarm. *Science* 367:234–235.
652   DOI: 10.1126/science.367.6475.234.
653 Crooke SN, Ovsyannikova IG, Kennedy RB, Poland GA. 2020. Immunoinformatic identification
654   of B cell and T cell epitopes in the SARS-CoV-2 proteome. *Scientific reports* 10:14179.
655   DOI: 10.1038/s41598-020-70864-8.
656 Desta IT, Porter KA, Xia B, Kozakov D, Vajda S. 2020. Performance and Its Limits in Rigid
657   Body Protein-Protein Docking. *Structure* 28:1071–1081.e3. DOI: 10.1016/j.str.2020.06.006.
658 Dhanda SK, Gupta S, Vir P, Raghava GPS. 2013. Prediction of IL4 inducing peptides. *Clinical &*
659   *developmental immunology* 2013:263952. DOI: 10.1155/2013/263952.
660 Dhanda SK, Karosiene E, Edwards L, Grifoni A, Paul S, Andreatta M, Weiskopf D, Sidney J,
661   Nielsen M, Peters B, Sette A. 2018. Predicting HLA CD4 Immunogenicity in Human
662   Populations. *Frontiers in immunology* 9:1369. DOI: 10.3389/fimmu.2018.01369.
663 Dhanda SK, Vir P, Raghava GPS. 2013. Designing of interferon-gamma inducing MHC class-II
664   binders. *Biology direct* 8:30. DOI: 10.1186/1745-6150-8-30.
665 Dimitrov I, Bangov I, Flower DR, Doytchinova I. 2014a. AllerTOP v.2--a server for in silico
666   prediction of allergens. *Journal of molecular modeling* 20:2278. DOI: 10.1007/s00894-014-
667   2278-5.
668 Dimitrov I, Naneva L, Doytchinova I, Bangov I. 2014b. AllergenFP: allergenicity prediction by
669   descriptor fingerprints. *Bioinformatics* 30:846–851. DOI: 10.1093/bioinformatics/btt619.
670 Di Natale C, La Manna S, De Benedictis I, Brandi P, Marasco D. 2020. Perspectives in Peptide-
671   Based Vaccination Strategies for Syndrome Coronavirus 2 Pandemic. *Frontiers in*
672   *pharmacology* 11:578382. DOI: 10.3389/fphar.2020.578382.
673 Doytchinova IA, Flower DR. 2007a. VaxiJen: a server for prediction of protective antigens,
674   tumour antigens and subunit vaccines. *BMC bioinformatics* 8:4. DOI: 10.1186/1471-2105-8-
675   4.
676 Doytchinova IA, Flower DR. 2007b. Identifying candidate subunit vaccines using an alignment-
677   independent method based on principal amino acid properties. *Vaccine* 25:856–866. DOI:
678   10.1016/j.vaccine.2006.09.032.
679 Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative
680   contribution to global health. *Global challenges (Hoboken, NJ)* 1:33–46. DOI:
681   10.1002/gch2.1018.
682 Faria NR, Claro IM, Candido D, Franco LAM, Andrade PS, Coletti TM, Silva CAM, Sales FC,
683   Manuli ER, Aguiar RS, Gaburo N, Camilo C da C, Fraiji NA, Crispim MAE, Carvalho M do
684   PSS, Rambaut A, Loman N, Pybus OG, Sabino EC. 2020.Genomic characterisation of an
685   emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *Available at*

686   *https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-*
687   *manaus-preliminary-findings/586* (accessed June 2021).
688   Fast E, Altman RB, Chen B. 2020. Potential T-cell and B-cell Epitopes of 2019-nCoV.
689   *bioRxiv*:2020.02.19.955484. DOI: 10.1101/2020.02.19.955484.
690   Fatoba AJ, Maharaj L, Adeleke VT, Okpeku M, Adeniyi AA, Adeleke MA. 2021.
691   Immunoinformatics prediction of overlapping CD8+ T-cell, IFN-γ and IL-4 inducer CD4+
692   T-cell and linear B-cell epitopes based vaccines against COVID-19 (SARS-CoV-2). *Vaccine*
693   39:1111–1121. DOI: 10.1016/j.vaccine.2021.01.003.
694   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
695   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,
696   Zhu L, Deng W, Li Y, Lu J, Li C, Wang X, Yin W, Zhang Y, Qin C. 2020. Development of
697   an inactivated vaccine candidate for SARS-CoV-2. *Science* 369:77–81. DOI:
698   10.1126/science.abc1932.
699   Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Amos B. 2005. Protein
700   Identification and Analysis Tools on the ExPASy Server. In: Walker JM ed. *The Proteomics*
701   *Protocols Handbook*. Humana Press, 571–607.
702   Geers D, Shamier MC, Bogers S, den Hartog G, Gommers L, Nieuwkoop NN, Schmitz KS,
703   Rijsbergen LC, van Osch JAT, Dijkhuizen E, Smits G, Comvalius A, van Mourik D, Caniels
704   TG, van Gils MJ, Sanders RW, Oude Munnink BB, Molenkamp R, de Jager HJ, Haagmans
705   BL, de Swart RL, Koopmans MPG, van Binnendijk RS, de Vries RD, GeurtsvanKessel CH.
706   2021. SARS-CoV-2 variants of concern partially escape humoral but not T-cell responses in
707   COVID-19 convalescent donors and vaccinees. *Science immunology* 6. DOI:
708   10.1126/sciimmunol.abj1750.
709   Gfeller D, Guillaume P, Michaux J, Pak H-S, Daniel RT, Racle J, Coukos G, Bassani-Sternberg
710   M. 2018. The Length Distribution and Multiple Specificity of Naturally Presented HLA-I
711   Ligands. *Journal of immunology* 201:3705–3716. DOI: 10.4049/jimmunol.1800914.
712   Ghazavi A, Ganji A, Keshavarzian N, Rabiemajd S, Mosayebi G. 2021. Cytokine profile and
713   disease severity in patients with COVID-19. *Cytokine* 137:155323. DOI:
714   10.1016/j.cyto.2020.155323.
715   Goddard TD, Huang CC, Ferrin TE. 2005. Software extensions to UCSF chimera for interactive
716   visualization of large molecular assemblies. *Structure* 13:473–482. DOI:
717   10.1016/j.str.2005.01.006.
718   Gonzalez-Galarza FF, McCabe A, Santos EJMD, Jones J, Takeshita L, Ortega-Rivera ND, Cid-
719   Pavon GMD, Ramsbottom K, Ghattaoraya G, Alfirevic A, Middleton D, Jones AR. 2020.
720   Allele frequency net database (AFND) 2020 update: gold-standard data classification, open
721   access genotype data and new query tools. *Nucleic acids research* 48:D783–D788. DOI:
722   10.1093/nar/gkz1029.
723   Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD. 2021.
724   Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that
725   affect recognition by polyclonal human plasma antibodies. *Cell host & microbe* 29:463–
726   476.e6. DOI: 10.1016/j.chom.2021.02.003.
727   Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. 2011. Functional classification of
728   class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a
729   surprising degree of repertoire sharing across supertypes. *Immunogenetics* 63:325–335. DOI:
730   10.1007/s00251-011-0513-0.
731   Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery
732   Consortium, Raghava GPS. 2013. In silico approach for predicting toxicity of peptides and
733   proteins. *PloS one* 8:e73957. DOI: 10.1371/journal.pone.0073957.
734   Gupta S, Madhu MK, Sharma AK, Sharma VK. 2016. ProInflam: a webserver for the prediction

735   of proinflammatory antigenicity of peptides and proteins. *Journal of translational medicine*
736   14:1–10. DOI: 10.1186/s12967-016-0928-3.
737   Gupta S, Mittal P, Madhu MK, Sharma VK. 2017. IL17eScan: A Tool for the Identification of
738   Peptides Inducing IL-17 Response. *Frontiers in immunology* 8:1430. DOI:
739   10.3389/fimmu.2017.01430.
740   Hartley GE, Edwards ESJ, Aui PM, Varese N, Stojanovic S, McMahon J, Peleg AY, Boo I,
741   Drummer HE, Hogarth PM, O'Hehir RE, van Zelm MC. 2020. Rapid generation of durable
742   B cell memory to SARS-CoV-2 spike and nucleocapsid proteins in COVID-19 and
743   convalescence. *Science immunology* 5. DOI: 10.1126/sciimmunol.abf8891.
744   Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. 2017. Protein-Sol: a web
745   tool for predicting protein solubility from sequence. *Bioinformatics* 33:3098–3100. DOI:
746   10.1093/bioinformatics/btx345.
747   Heo L, Park H, Seok C. 2013. GalaxyRefine: Protein structure refinement driven by side-chain
748   repacking. *Nucleic acids research* 41:W384–8. DOI: 10.1093/nar/gkt458.
749   Hoffmann M, Arora P, Groß R, Seidel A, Hörnich BF, Hahn AS, Krüger N, Graichen L,
750   Hofmann-Winkler H, Kempf A, Winkler MS, Schulz S, Jäck H-M, Jahrsdörfer B,
751   Schrezenmeier H, Müller M, Kleger A, Münch J, Pöhlmann S. 2021. SARS-CoV-2 variants
752   B.1.351 and P.1 escape from neutralizing antibodies. *Cell* 184:2384–2393.e12. DOI:
753   10.1016/j.cell.2021.03.036.
754   Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M. 2009.
755   NetMHCpan, a method for MHC class I binding prediction beyond humans.
756   *Immunogenetics* 61:1–13. DOI: 10.1007/s00251-008-0341-z.
757   Hou Q, Kwasigroch JM, Rooman M, Pucci F. 2020. SOLart: a structure-based method to predict
758   protein solubility and aggregation. *Bioinformatics* 36:1445–1452. DOI:
759   10.1093/bioinformatics/btz773.
760   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
761   J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang
762   R, Gao Z, Jin Q, Wang J, Cao B. 2020. Clinical features of patients infected with 2019 novel
763   coronavirus in Wuhan, China. *The Lancet* 395:497–506. DOI: 10.1016/S0140-
764   6736(20)30183-5.
765   Hudu SA, Shinkafi SH, Umar S. 2016. AN OVERVIEW OF RECOMBINANT VACCINE
766   TECHNOLOGY, ADJUVANTS AND VACCINE DELIVERY METHODS. *International*
767   *journal of pharmacy and pharmaceutical sciences*:19–24. DOI:
768   10.22159/ijpps.2016v8i11.14311.
769   Jahangirian E, Jamal GA, Nouroozi M, Mohammadpour A. 2021. A reverse vaccinology and
770   immunoinformatics approach for designing a multiepitope vaccine against SARS-CoV-2.
771   *Immunogenetics*. DOI: 10.1007/s00251-021-01228-3.
772   Jespersen MC, Peters B, Nielsen M, Marcatili P. 2017. BepiPred-2.0: improving sequence-based
773   B-cell epitope prediction using conformational epitopes. *Nucleic acids research* 45:W24–
774   W29. DOI: 10.1093/nar/gkx346.
775   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
776   J, Yang Y, Qiu H, Shen C, He Y, Zhao G. 2021. Screening of HLA-A restricted T cell
777   epitopes of SARS-CoV-2 and induction of CD8+ T cell responses in HLA-A transgenic
778   mice. *bioRxiv*:2021.04.01.438020. DOI: 10.1101/2021.04.01.438020.
779   Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring
780   matrices. *Journal of molecular biology* 292:195–202. DOI: 10.1006/jmbi.1999.3091.
781   Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. 2017. NetMHCpan-4.0:
782   Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and
783   Peptide Binding Affinity Data. *Journal of immunology* 199:3360–3368. DOI:

784   10.4049/jimmunol.1700893.
785   Kalita P, Padhi AK, Zhang KYJ, Tripathi T. 2020. Design of a peptide-based subunit vaccine
786   against novel coronavirus SARS-CoV-2. *Microbial pathogenesis* 145:104236. DOI:
787   10.1016/j.micpath.2020.104236.
788   Källberg M, Margaryan G, Wang S, Ma J, Xu J. 2014. RaptorX server: a resource for template-
789   based protein structure modeling. *Methods in molecular biology* 1137:17–27. DOI:
790   10.1007/978-1-4939-0366-5_2.
791   Karlsson AC, Humbert M, Buggert M. 2020. The known unknowns of T cell immunity to
792   COVID-19. *Science immunology* 5. DOI: 10.1126/sciimmunol.abe8063.
793   Kar T, Narsaria U, Basak S, Deb D, Castiglione F, Mueller DM, Srivastava AP. 2020. A
794   candidate multi-epitope vaccine against SARS-CoV-2. *Scientific reports* 10:10895. DOI:
795   10.1038/s41598-020-67749-1.
796   Kaushik D, Bhandari R, Kuhad A. 2021. TLR4 as a therapeutic target for respiratory and
797   neurological complications of SARS-CoV-2. *Expert opinion on therapeutic targets* 25:491–
798   508. DOI: 10.1080/14728222.2021.1918103.
799   Kazi A, Chuah C, Majeed ABA, Leow CH, Lim BH, Leow CY. 2018. Current progress of
800   immunoinformatics approach harnessed for cellular- and antibody-dependent vaccine design.
801   *Pathogens and global health* 112:123–131. DOI: 10.1080/20477724.2018.1446773.
802   Khanmohammadi S, Rezaei N. 2021. Role of Toll-like receptors in the pathogenesis of COVID-
803   19. *Journal of medical virology* 93:2735–2739. DOI: 10.1002/jmv.26826.
804   Kiyotani K, Toyoshima Y, Nemoto K, Nakamura Y. 2020. Bioinformatic prediction of potential
805   T cell epitopes for SARS-Cov-2. *Journal of human genetics* 65:569–575. DOI:
806   10.1038/s10038-020-0771-5.
807   Knoll MD, Wonodi C. 2021. Oxford-AstraZeneca COVID-19 vaccine efficacy. *The Lancet*
808   397:72–74. DOI: 10.1016/S0140-6736(20)32623-4.
809   Krammer F. 2020. SARS-CoV-2 vaccines in development. *Nature* 586:516–527. DOI:
810   10.1038/s41586-020-2798-3.
811   Kyriakidis NC, López-Cortés A, González EV, Grimaldos AB, Prado EO. 2021. SARS-CoV-2
812   vaccines strategies: a comprehensive review of phase 3 candidates. *NPJ vaccines* 6:28. DOI:
813   10.1038/s41541-021-00292-w.
814   Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. 2007. Large-scale
815   validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*
816   8:424. DOI: 10.1186/1471-2105-8-424.
817   Lee H, Heo L, Lee MS, Seok C. 2015. GalaxyPepDock: a protein-peptide docking tool based on
818   interaction similarity and energy optimization. *Nucleic acids research* 43:W431–5. DOI:
819   10.1093/nar/gkv495.
820   Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein
821   or nucleotide sequences. *Bioinformatics* 22:1658–1659. DOI:
822   10.1093/bioinformatics/btl158.
823   Lim HX, Lim J, Jazayeri SD, Poppema S, Poh CL. 2021. Development of multi-epitope peptide-
824   based vaccines against SARS-CoV-2. *Biomedical journal* 44:18–30. DOI:
825   10.1016/j.bj.2020.09.005.
826   Lin L, Ting S, Yufei H, Wendong L, Yubo F, Jing Z. 2020. Epitope-based peptide vaccines
827   predicted against novel coronavirus disease caused by SARS-CoV-2. *Virus research*
828   288:198082. DOI: 10.1016/j.virusres.2020.198082.
829   Lipsitch M, Grad YH, Sette A, Crotty S. 2020. Cross-reactive memory T cells and herd immunity
830   to SARS-CoV-2. *Nature reviews. Immunology* 20:709–713. DOI: 10.1038/s41577-020-
831   00460-4.
832   Liu Z, VanBlargan LA, Bloyet L-M, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico JM,

833   Theel ES, Liebeskind MJ, Alford B, Buchser WJ, Ellebedy AH, Fremont DH, Diamond MS,
834   Whelan SPJ. 2021. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal
835   and serum antibody neutralization. *Cell host & microbe* 29:477–488.e4. DOI:
836   10.1016/j.chom.2021.01.014.
837   Livingston EH, Malani PN, Creech CB. 2021. The Johnson & Johnson Vaccine for COVID-19.
838   *JAMA: the journal of the American Medical Association* 325:1575. DOI:
839   10.1001/jama.2021.2927.
840   López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. 2014. iMODS: internal coordinates
841   normal mode analysis server. *Nucleic acids research* 42:W271–6. DOI: 10.1093/nar/gku339.
842   Lu G, Shan S, Zainab B, Ayaz Z, He J, Xie Z, Rashid U, Zhang D, Mehmood Abbasi A. 2021.
843   Novel vaccine design based on genomics data analysis: A review. *Scandinavian journal of*
844   *immunology* 93:e12986. DOI: 10.1111/sji.12986.
845   Madden T. 2020. *User Manual*. National Center for Biotechnology Information (US).
846   Magnan CN, Randall A, Baldi P. 2009. SOLpro: accurate sequence-based prediction of protein
847   solubility. *Bioinformatics* 25:2200–2207. DOI: 10.1093/bioinformatics/btp386.
848   Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, Baldi P. 2010. High-
849   throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics*
850   26:2936–2943. DOI: 10.1093/bioinformatics/btq551.
851   Mahapatra SR, Sahoo S, Dehury B, Raina V, Patro S, Misra N, Suar M. 2020. Designing an
852   efficient multi-epitope vaccine displaying interactions with diverse HLA molecules for an
853   efficient humoral and cellular immune response to prevent COVID-19 infection. *Expert*
854   *review of vaccines* 19:871–885. DOI: 10.1080/14760584.2020.1811091.
855   Mahase E. 2020. Covid-19: Moderna vaccine is nearly 95% effective, trial involving high risk
856   and elderly people shows. *BMJ* 371. DOI: 10.1136/bmj.m4471.
857   Mallajosyula V, Ganjavi C, Chakraborty S, McSween AM, Pavlovitch-Bedzyk AJ, Wilhelmy J,
858   Nau A, Manohar M, Nadeau KC, Davis MM. 2021. CD8+ T cells specific for conserved
859   coronavirus epitopes correlate with milder disease in COVID-19 patients. *Science*
860   *immunology* 6. DOI: 10.1126/sciimmunol.abg5669.
861   Mallapaty S. 2021. WHO approval of Chinese CoronaVac COVID vaccine will be crucial to
862   curbing pandemic. *Nature* 594:161–162. DOI: 10.1038/d41586-021-01497-8.
863   Mallavarpu Ambrose J, Priya Veeraraghavan V, Kullappan M, Chellapandiyan P, Krishna Mohan
864   S, Manivel VA. 2021. Comparison of Immunological Profiles of SARS-CoV-2 Variants in
865   the COVID-19 Pandemic Trends: An Immunoinformatics Approach. *Antibiotics (Basel,*
866   *Switzerland)* 10. DOI: 10.3390/antibiotics10050535.
867   Malonis RJ, Lai JR, Vergnolle O. 2020. Peptide-Based Vaccines: Current Progress and Future
868   Challenges. *Chemical reviews* 120:3210–3229. DOI: 10.1021/acs.chemrev.9b00472.
869   Mascellino MT, Di Timoteo F, De Angelis M, Oliva A. 2021. Overview of the Main Anti-SARS-
870   CoV-2 Vaccines: Mechanism of Action, Efficacy and Safety. *Individual differences*
871   *research: IDR* 14:3459–3476. DOI: 10.2147/IDR.S315727.
872   McGill COVID19 Vaccine Tracker Team. 2021.COVID-19 Vaccine Tracker. *Available at*
873   *https://covid19.trackvaccines.org/vaccines/#approved* (accessed September 2021).
874   McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server.
875   *Bioinformatics* 16:404–405. DOI: 10.1093/bioinformatics/16.4.404.
876   Medhi R, Srinoi P, Ngo N, Tran H-V, Lee TR. 2020. Nanoparticle-Based Strategies to Combat
877   COVID-19. *ACS Applied Nano Materials* 3:8557–8580. DOI: 10.1021/acsanm.0c01978.
878   Menezes Teles E Oliveira D, Melo Santos de Serpa Brandão R, Claudio Demes da Mata Sousa L,
879   das Chagas Alves Lima F, Jamil Hadad do Monte S, Sérgio Coelho Marroquim M, Vanildo
880   de Sousa Lima A, Gilberto Borges Coelho A, Matheus Sousa Costa J, Martins Ramos R,
881   Socorro da Silva A. 2019. pHLA3D: An online database of predicted three-dimensional

882   structures of HLA molecules. *Human immunology* 80:834–841. DOI:
883   10.1016/j.humimm.2019.06.009.
884   Meyers LM, Gutiérrez AH, Boyle CM, Terry F, McGonnigal BG, Salazar A, Princiotta MF,
885   Martin WD, De Groot AS, Moise L. 2021. Highly conserved, non-human-like, and cross-
886   reactive SARS-CoV-2 T cell epitopes for COVID-19 vaccine design and validation. *NPJ*
887   *vaccines* 6:71. DOI: 10.1038/s41541-021-00331-6.
888   Minervina AA, Komech EA, Titov A, Bensouda Koraichi M, Rosati E, Mamedov IZ, Franke A,
889   Efimov GA, Chudakov DM, Mora T, Walczak AM, Lebedev YB, Pogorelyy MV. 2021.
890   Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T-cell
891   memory formation after mild COVID-19 infection. *eLife* 10. DOI: 10.7554/eLife.63502.
892   Mohammad MG, Ibrahim F, Navid N, Shirin M. 2020. Multi-Epitope vaccines (MEVs), as a
893   novel strategy against infectious diseases. *Current proteomics* 17:354–364.
894   Nagpal G, Chaudhary K, Agrawal P, Raghava GPS. 2018. Computer-aided prediction of antigen
895   presenting cell modulators for designing peptide-based vaccine adjuvants. *Journal of*
896   *translational medicine* 16:181. DOI: 10.1186/s12967-018-1560-1.
897   Nagpal G, Usmani SS, Dhanda SK, Kaur H, Singh S, Sharma M, Raghava GPS. 2017.
898   Computer-aided designing of immunosuppressive peptides based on IL-10 inducing
899   potential. *Scientific reports* 7:42851. DOI: 10.1038/srep42851.
900   Naveca FG, Nascimento V, de Souza VC, Corado A de L, Nascimento F, Silva G, Costa Á,
901   Duarte D, Pessoa K, Mejía M, Brandão MJ, Jesus M, Gonçalves L, da Costa CF, Sampaio V,
902   Barros D, Silva M, Mattos T, Pontes G, Abdalla L, Santos JH, Arantes I, Dezordi FZ,
903   Siqueira MM, Wallau GL, Resende PC, Delatorre E, Gräf T, Bello G. 2021. COVID-19 in
904   Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence.
905   *Nature medicine*. DOI: 10.1038/s41591-021-01378-7.
906   Naveed M, Tehreem S, Arshad S, Bukhari SA, Shabbir MA, Essa R, Ali N, Zaib S, Khan A, Al-
907   Harrasi A, Khan I. 2021. Design of a novel multiple epitope-based vaccine: An
908   immunoinformatics approach to combat SARS-CoV-2 strains. *Journal of infection and*
909   *public health*. DOI: 10.1016/j.jiph.2021.04.010.
910   Nemati AS, Tafrihi M, Sheikhi F, Tabari AR, Haditabar A. 2021. Designing a new multi epitope-
911   based vaccine against COVID-19 disease: an immunoinformatic study based on reverse
912   vaccinology approach. *Research Square*. DOI: 10.21203/rs.3.rs-206270/v1.
913   Ojha R, Gupta N, Naik B, Singh S, Verma VK, Prusty D, Prajapati VK. 2020. High throughput
914   and comprehensive approach to develop multiepitope vaccine against minacious COVID-19.
915   *European journal of pharmaceutical sciences: official journal of the European Federation*
916   *for Pharmaceutical Sciences* 151:105375. DOI: 10.1016/j.ejps.2020.105375.
917   Oli AN, Obialor WO, Ifeanyichukwu MO, Odimegwu DC, Okoyeh JN, Emechebe GO, Adejumo
918   SA, Ibeanu GC. 2020. Immunoinformatics and Vaccine Development: An Overview.
919   *ImmunoTargets and therapy* 9:13–30. DOI: 10.2147/ITT.S241064.
920   O'Neill LAJ, Golenbock D, Bowie AG. 2013. The history of Toll-like receptors - redefining
921   innate immunity. *Nature reviews. Immunology* 13:453–460. DOI: 10.1038/nri3446.
922   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,
923   Chen J, Hu K, Jin Q, Wang J, Qian Z. 2020. Characterization of spike glycoprotein of
924   SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature*
925   *communications* 11:1620. DOI: 10.1038/s41467-020-15562-9.
926   Patra R, Chandra Das N, Mukherjee S. 2021. Targeting human TLRs to combat COVID-19: A
927   solution? *Journal of medical virology* 93:615–617. DOI: 10.1002/jmv.26387.
928   Patronov A, Dimitrov I, Flower DR, Doytchinova I. 2011. Peptide binding prediction for the
929   human class II MHC allele HLA-DP2: a molecular docking approach. *BMC structural*
930   *biology* 11:32. DOI: 10.1186/1472-6807-11-32.

931 Paul S, Lindestam Arlehamn CS, Scriba TJ, Dillon MBC, Oseroff C, Hinz D, McKinney DM,
932  Carrasco Pro S, Sidney J, Peters B, Sette A. 2015. Development and validation of a broad
933  scheme for prediction of HLA class II restricted T cell epitopes. *Journal of immunological*
934  *methods* 422:28–34. DOI: 10.1016/j.jim.2015.03.022.
935 Pham T-H, Qiu Y, Zeng J, Xie L, Zhang P. 2021. A deep learning framework for high-
936  throughput mechanism-driven phenotype compound screening and its application to
937  COVID-19 drug repurposing. *Nature machine intelligence* 3:247–257. DOI:
938  10.1038/s42256-020-00285-9.
939 Prachar M, Justesen S, Steen-Jensen DB, Thorgrimsen S, Jurgons E, Winther O, Bagger FO.
940  2020. Identification and validation of 174 COVID-19 vaccine candidate epitopes reveals low
941  performance of common epitope prediction tools. *Scientific reports* 10:20465. DOI: 10.1038/
942  s41598-020-77466-4.
943 Prévost J, Finzi A. 2021. The great escape? SARS-CoV-2 variants evading neutralizing
944  responses. *Cell host & microbe* 29:322–324. DOI: 10.1016/j.chom.2021.02.010.
945 Quast I, Tarlinton D. 2021. B cell memory: understanding COVID-19. *Immunity* 54:205–210.
946  DOI: 10.1016/j.immuni.2021.01.014.
947 Rahman MS, Hoque MN, Islam MR, Akter S, Rubayet Ul Alam ASM, Siddique MA, Saha O,
948  Rahaman MM, Sultana M, Crandall KA, Hossain MA. 2020. Epitope-based chimeric peptide
949  vaccine design against S, M and E proteins of SARS-CoV-2, the etiologic agent of COVID-
950  19 pandemic: an in silico approach. *PeerJ* 8:e9572. DOI: 10.7717/peerj.9572.
951 Rakib A, Sami SA, Mimi NJ, Chowdhury MM, Eva TA, Nainu F, Paul A, Shahriar A, Tareq
952  AM, Emon NU, Chakraborty S, Shil S, Mily SJ, Ben Hadda T, Almalki FA, Emran TB.
953  2020. Immunoinformatics-guided design of an epitope-based vaccine against severe acute
954  respiratory syndrome coronavirus 2 spike glycoprotein. *Computers in biology and medicine*
955  124:103967. DOI: 10.1016/j.compbiomed.2020.103967.
956 Ramana J, Mehla K. 2020. Immunoinformatics and Epitope Prediction. In: Tomar N ed.
957  *Immunoinformatics*. New York, NY: Springer US, 155–171. DOI: 10.1007/978-1-0716-
958  0389-5_6.
959 Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, Connor T, Peacock T,
960  Robertson DL, Volz E. 2020.Preliminary genomic characterisation of an emergent SARS-
961  CoV-2 lineage in the UK defined by a novel set of spike mutations. *Available at*
962  *https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-*
963  *lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563* (accessed June 2021).
964 Rapin N, Lund O, Bernaschi M, Castiglione F. 2010. Computational immunology meets
965  bioinformatics: the use of prediction tools for molecular binding in the simulation of the
966  immune system. *PloS one* 5:e9862. DOI: 10.1371/journal.pone.0009862.
967 Redd AD, Nardin A, Kared H, Bloch EM, Pekosz A, Laeyendecker O, Abel B, Fehlings M,
968  Quinn TC, Tobian AA. 2021. CD8+ T cell responses in COVID-19 convalescent individuals
969  target conserved epitopes from multiple prominent SARS-CoV-2 circulating variants.
970  *medRxiv : the preprint server for health sciences*. DOI: 10.1101/2021.02.11.21251585.
971 Ribes M, Chaccour C, Moncunill G. 2021. Adapt or perish: SARS-CoV-2 antibody escape
972  variants defined by deletions in the Spike N-terminal Domain. *Signal transduction and*
973  *targeted therapy* 6:164. DOI: 10.1038/s41392-021-00601-8.
974 Rotondo JC, Martini F, Maritati M, Mazziotta C, Di Mauro G, Lanzillotti C, Barp N, Gallerani A,
975  Tognon M, Contini C. 2021. SARS-CoV-2 Infection: New Molecular, Phylogenetic, and
976  Pathogenetic Insights. Efficacy of Current Vaccines and the Potential Risk of Variants.
977  *Viruses* 13:1687. DOI: 10.3390/v13091687.
978 Saad-Roy CM, Morris SE, Metcalf CJE, Mina MJ, Baker RE, Farrar J, Holmes EC, Pybus OG,
979  Graham AL, Levin SA, Grenfell BT, Wagner CE. 2021. Epidemiological and evolutionary

980   considerations of SARS-CoV-2 vaccine dosing regimes. *Science* 372:363–370. DOI:
981   10.1126/science.abg8663.
982 Saba AA, Adiba M, Saha P, Hosen MI, Chakraborty S, Nabi AHMN. 2021. An in-depth in silico
983   and immunoinformatics approach for designing a potential multi-epitope construct for the
984   effective development of vaccine to combat against SARS-CoV-2 encompassing variants of
985   concern and interest. *Computers in biology and medicine* 136:104703. DOI:
986   10.1016/j.compbiomed.2021.104703.
987 Sakaguchi T, Ibe M, Miwa K, Kaneko Y, Yokota S, Tanaka K, Takiguchi M. 1997. Binding of 8-
988   mer to 11-mer peptides carrying the anchor residues to slow assembling HLA class I
989   molecules (HLA-B*5101). *Immunogenetics* 45:259–265. DOI: 10.1007/s002510050201.
990 Sarma VR, Olotu FA, Soliman MES. 2021. Integrative immunoinformatics paradigm for
991   predicting potential B-cell and T-cell epitopes as viable candidates for subunit vaccine
992   design against COVID-19 virulence. *Biomedical journal*. DOI: 10.1016/j.bj.2021.05.001.
993 Sette A, Crotty S. 2021. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* 184:861–880.
994   DOI: 10.1016/j.cell.2021.01.007.
995 Shahcheraghi SH, Ayatollahi J, Aljabali AA, Shastri MD, Shukla SD, Chellappan DK, Jha NK,
996   Anand K, Katari NK, Mehta M, Satija S, Dureja H, Mishra V, Almutary AG, Alnuqaydan
997   AM, Charbe N, Prasher P, Gupta G, Dua K, Lotfi M, Bakshi HA, Tambuwala MM. 2021.
998   An overview of vaccine development for COVID-19. *Therapeutic delivery* 12:235–244.
999   DOI: 10.4155/tde-2020-0129.
1000 Shang W, Yang Y, Rao Y, Rao X. 2020. The outbreak of SARS-CoV-2 pneumonia calls for viral
1001   vaccines. *NPJ vaccines* 5:18. DOI: 10.1038/s41541-020-0170-0.
1002 Sharma A, Pal S, Panwar A, Kumar S, Kumar A. 2021. In-silico immunoinformatic analysis of
1003   SARS-CoV-2 virus for the development of putative vaccine construct. *Immunobiology*
1004   226:152134. DOI: 10.1016/j.imbio.2021.152134.
1005 Sharma N, Patiyal S, Dhall A, Pande A, Arora C, Raghava GPS. 2020. AlgPred 2.0: an improved
1006   method for predicting allergenic proteins and mapping of IgE epitopes. *Briefings in
1007   bioinformatics*. DOI: 10.1093/bib/bbaa294.
1008 Shean RC, Makhsous N, Stoddard GD, Lin MJ, Greninger AL. 2019. VAPiD: a lightweight
1009   cross-platform viral annotation pipeline and identification tool to facilitate virus genome
1010   submissions to NCBI GenBank. *BMC bioinformatics* 20:1–8. DOI: 10.1186/s12859-019-
1011   2606-y.
1012 Shkurnikov M, Nersisyan S, Jankevic T, Galatenko A, Gordeev I, Vechorko V, Tonevitsky A.
1013   2021. Association of HLA Class I Genotypes With Severity of Coronavirus Disease-19.
1014   *Frontiers in immunology* 12:641900. DOI: 10.3389/fimmu.2021.641900.
1015 Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many protein
1016   sequences. *Protein science: a publication of the Protein Society* 27:135–145. DOI: 10.1002/
1017   pro.3290.
1018 Singh A, Thakur M, Sharma LK, Chandra K. 2020. Designing a multi-epitope peptide based
1019   vaccine against SARS-CoV-2. *Scientific reports* 10:16219. DOI: 10.1038/s41598-020-
1020   73371-y.
1021 Siracusano G, Pastori C, Lopalco L. 2020. Humoral Immune Responses in COVID-19 Patients:
1022   A Window on the State of the Art. *Frontiers in immunology* 11:1049. DOI:
1023   10.3389/fimmu.2020.01049.
1024 Skwarczynski M, Toth I. 2016. Peptide-based synthetic vaccines. *Chemical science* 7:842–854.
1025   DOI: 10.1039/c5sc03892h.
1026 Sun L. 2013. Peptide-Based Drug Development. *Modern Chemistry and Applications* 1. DOI:
1027   10.4172/mca.1000e103.
1028 Tarke A, Sidney J, Methot N, Yu ED, Zhang Y, Dan JM, Goodwin B, Rubiro P, Sutherland A,

1029   Wang E, Frazier A, Ramirez SI, Rawlings SA, Smith DM, da Silva Antunes R, Peters B,
1030   Scheuermann RH, Weiskopf D, Crotty S, Grifoni A, Sette A. 2021. Impact of SARS-CoV-2
1031   variants on the total CD4+ and CD8+ T cell reactivity in infected or vaccinated individuals.
1032   *Cell reports. Medicine* 2:100355. DOI: 10.1016/j.xcrm.2021.100355.
1033 Tavasolian F, Rashidi M, Hatam GR, Jeddi M, Hosseini AZ, Mosawi SH, Abdollahi E, Inman
1034   RD. 2020. HLA, Immune Response, and Susceptibility to COVID-19. *Frontiers in*
1035   *immunology* 11:601886. DOI: 10.3389/fimmu.2020.601886.
1036 Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay
1037   S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T,
1038   Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G,
1039   Hsiao N-Y, Korsman S, Davies M-A, Tyers L, Mudau I, York D, Maslo C, Goedhals D,
1040   Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell
1041   BT, Lourenço J, Alcantara LCJ, Kosakovsky Pond SL, Weaver S, Martin D, Lessells RJ,
1042   Bhiman JN, Williamson C, de Oliveira T. 2021. Detection of a SARS-CoV-2 variant of
1043   concern in South Africa. *Nature* 592:438–443. DOI: 10.1038/s41586-021-03402-9.
1044 Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, Mayhew S.
1045   2020. The COVID-19 vaccine development landscape. *Nature reviews. Drug discovery*
1046   19:305–306. DOI: 10.1038/d41573-020-00073-5.
1047 Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva Filipe A, Wojcechowskyj JA, Davis
1048   C, Piccoli L, Pascall DJ, Dillen J, Lytras S, Czudnochowski N, Shah R, Meury M, Jesudason
1049   N, De Marco A, Li K, Bassi J, O'Toole A, Pinto D, Colquhoun RM, Culap K, Jackson B,
1050   Zatta F, Rambaut A, Jaconi S, Sreenu VB, Nix J, Zhang I, Jarrett RF, Glass WG, Beltramello
1051   M, Nomikou K, Pizzuto M, Tong L, Cameroni E, Croll TI, Johnson N, Di Iulio J,
1052   Wickenhagen A, Ceschi A, Harbison AM, Mair D, Ferrari P, Smollett K, Sallusto F,
1053   Carmichael S, Garzoni C, Nichols J, Galli M, Hughes J, Riva A, Ho A, Schiuma M, Semple
1054   MG, Openshaw PJM, Fadda E, Baillie JK, Chodera JD, ISARIC4C Investigators, COVID-
1055   19 Genomics UK (COG-UK) Consortium, Rihn SJ, Lycett SJ, Virgin HW, Telenti A, Corti
1056   D, Robertson DL, Snell G. 2021. Circulating SARS-CoV-2 spike N439K variants maintain
1057   fitness while evading antibody-mediated immunity. *Cell* 184:1171–1187.e20. DOI: 10.1016/
1058   j.cell.2021.01.037.
1059 Tomita Y, Ikeda T, Sato R, Sakagami T. 2020. Association between HLA gene polymorphisms
1060   and mortality of COVID-19: An in silico analysis. *Immunity, inflammation and disease*
1061   8:684–694. DOI: 10.1002/iid3.358.
1062 Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters
1063   B. 2019. The Immune Epitope Database (IEDB): 2018 update. *Nucleic acids research*
1064   47:D339–D343. DOI: 10.1093/nar/gky1006.
1065 Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. 2020. Structure, Function,
1066   and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181:281–292.e6. DOI:
1067   10.1016/j.cell.2020.02.058.
1068 Wang X, Gui J. 2020. Cell-mediated immunity to SARS-CoV-2. *Pediatric investigation* 4:281–
1069   291. DOI: 10.1002/ped4.12228.
1070 WHO. 2021.Weekly epidemiological update on COVID-19 - 25 May 2021. *Available at https://*
1071   *www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---25-may-*
1072   *2021* (accessed May 27, 2021).
1073 Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors
1074   in three-dimensional structures of proteins. *Nucleic acids research* 35:W407–10. DOI:
1075   10.1093/nar/gkm290.
1076 Wu D, Yang XO. 2020. TH17 responses in cytokine storm of COVID-19: An emerging target of
1077   JAK2 inhibitor Fedratinib. *Journal of microbiology, immunology, and infection = Wei mian*

1078   *yu gan ran za zhi* 53:368–370. DOI: 10.1016/j.jmii.2020.03.005.
1079   Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. 2016. PRODIGY: a web server for
1080   predicting the binding affinity of protein-protein complexes. *Bioinformatics* 32:3676–3678.
1081   DOI: 10.1093/bioinformatics/btw514.
1082   Yang Z, Bogdan P, Nazarian S. 2021. An in silico deep learning approach to multi-epitope
1083   vaccine design: a SARS-CoV-2 case study. *Scientific reports* 11:3238. DOI:
1084   10.1038/s41598-021-81749-9.
1085   Yang J, Wang W, Chen Z, Lu S, Yang F, Bi Z, Bao L, Mo F, Li X, Huang Y, Hong W, Yang Y,
1086   Zhao Y, Ye F, Lin S, Deng W, Chen H, Lei H, Zhang Z, Luo M, Gao H, Zheng Y, Gong Y,
1087   Jiang X, Xu Y, Lv Q, Li D, Wang M, Li F, Wang S, Wang G, Yu P, Qu Y, Yang L, Deng H,
1088   Tong A, Li J, Wang Z, Yang J, Shen G, Zhao Z, Li Y, Luo J, Liu H, Yu W, Yang M, Xu J,
1089   Wang J, Li H, Wang H, Kuang D, Lin P, Hu Z, Guo W, Cheng W, He Y, Song X, Chen C,
1090   Xue Z, Yao S, Chen L, Ma X, Chen S, Gou M, Huang W, Wang Y, Fan C, Tian Z, Shi M,
1091   Wang F-S, Dai L, Wu M, Li G, Wang G, Peng Y, Qian Z, Huang C, Lau JY-N, Yang Z, Wei
1092   Y, Cen X, Peng X, Qin C, Zhang K, Lu G, Wei X. 2020. A vaccine targeting the RBD of the
1093   S protein of SARS-CoV-2 induces protective immunity. *Nature* 586:572–577. DOI:
1094   10.1038/s41586-020-2599-8.
1095   Yoshida S, Ono C, Hayashi H, Fukumoto S, Shiraishi S, Tomono K, Arase H, Matsuura Y,
1096   Nakagami H. 2021. SARS-CoV-2-induced humoral immunity through B cell epitope
1097   analysis in COVID-19 infected individuals. *Scientific reports* 11:5934. DOI:
1098   10.1038/s41598-021-85202-9.
1099   Zaheer T, Waseem M, Waqar W, Dar HA, Shehroz M, Naz K, Ishaq Z, Ahmad T, Ullah N,
1100   Bakhtiar SM, Muhammad SA, Ali A. 2020. Anti-COVID-19 multi-epitope vaccine designs
1101   employing global viral genome sequences. *PeerJ* 8:e9541. DOI: 10.7717/peerj.9541.
1102   Zheng M, Song L. 2020. Novel antibody epitopes dominate the antigenicity of spike glycoprotein
1103   in SARS-CoV-2 compared to SARS-CoV. *Cellular & molecular immunology* 17:536–538.
1104   DOI: 10.1038/s41423-020-0385-z.

1105

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 Selected epitopes validation. (A) The prediction encompasses three DTU Health Tech software (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 on-line 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.

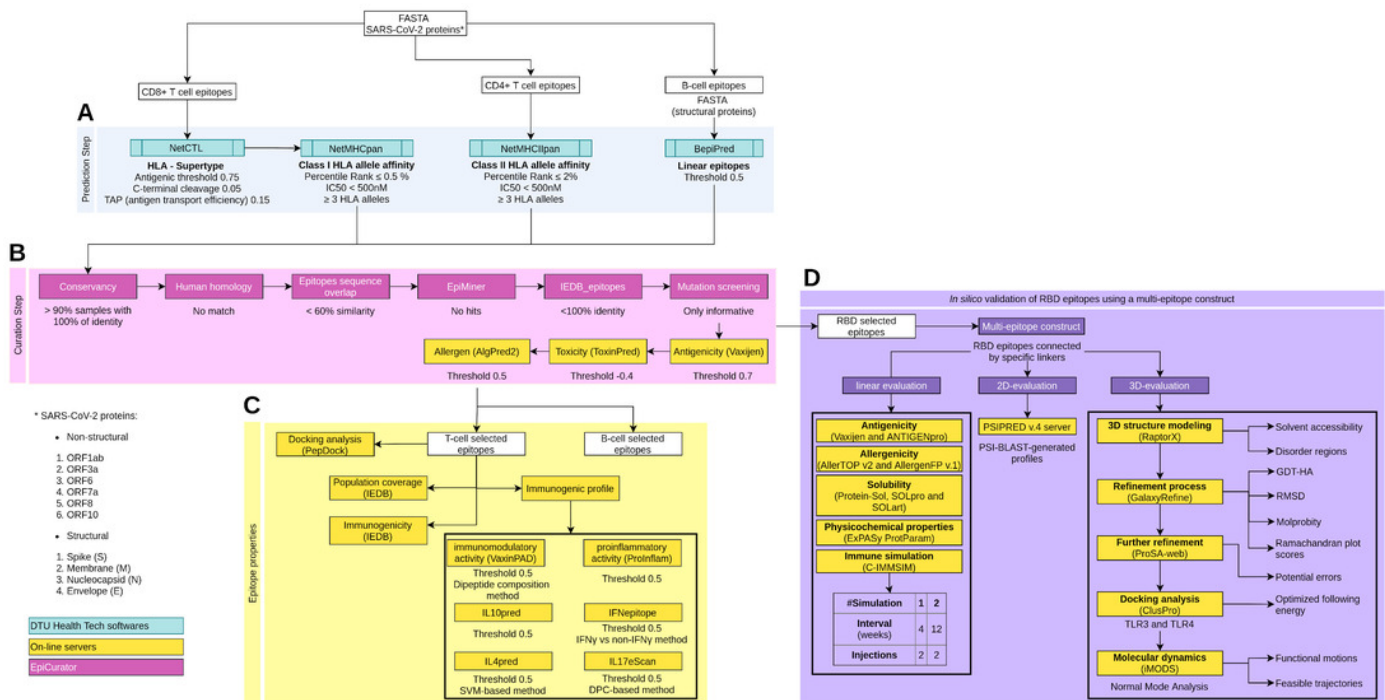



Figure 2

Efficiency of EpiCurator analysis for accurate selection of epitopes.

The plot represents the percentage of epitopes removed in each different analysis.

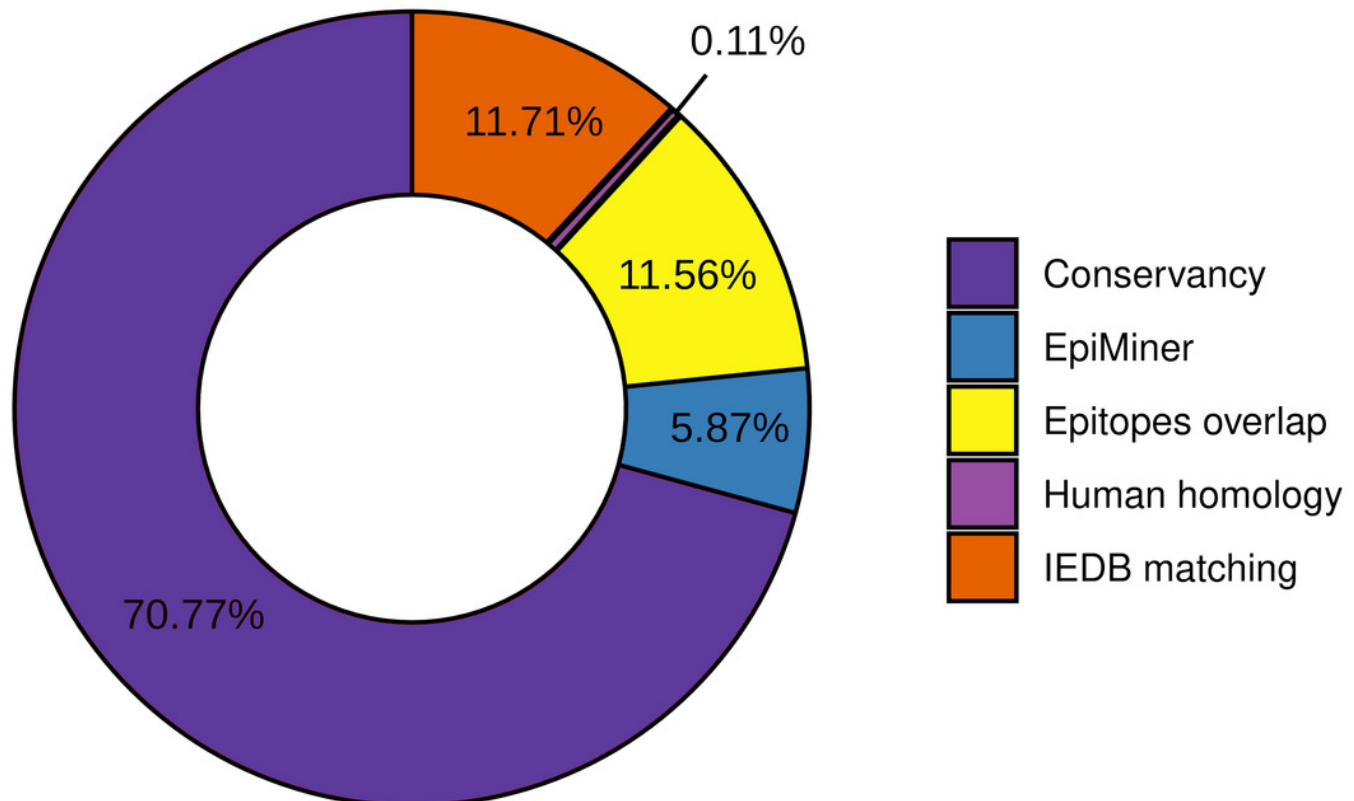



Figure 3

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.

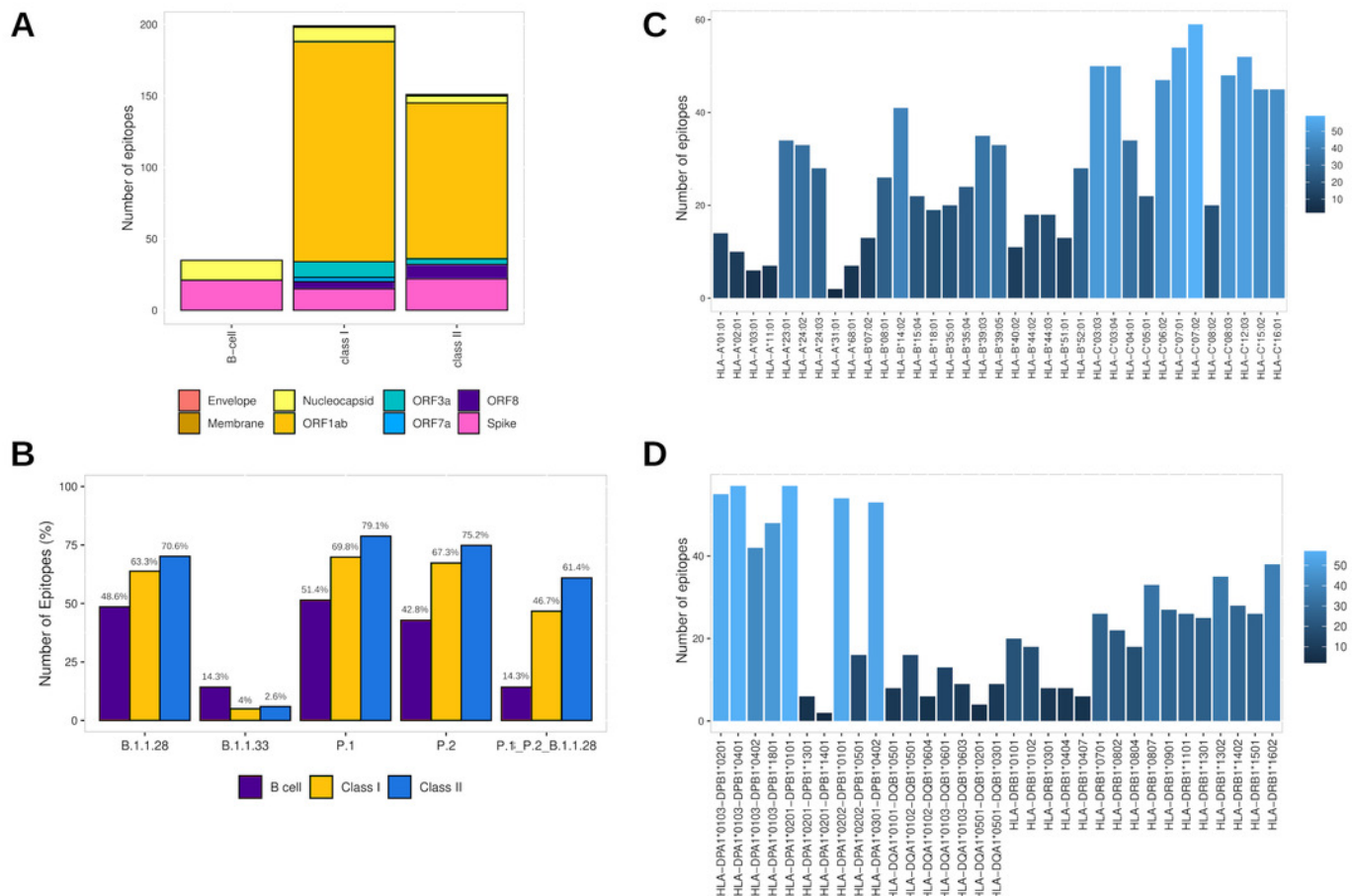



Figure 4

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).

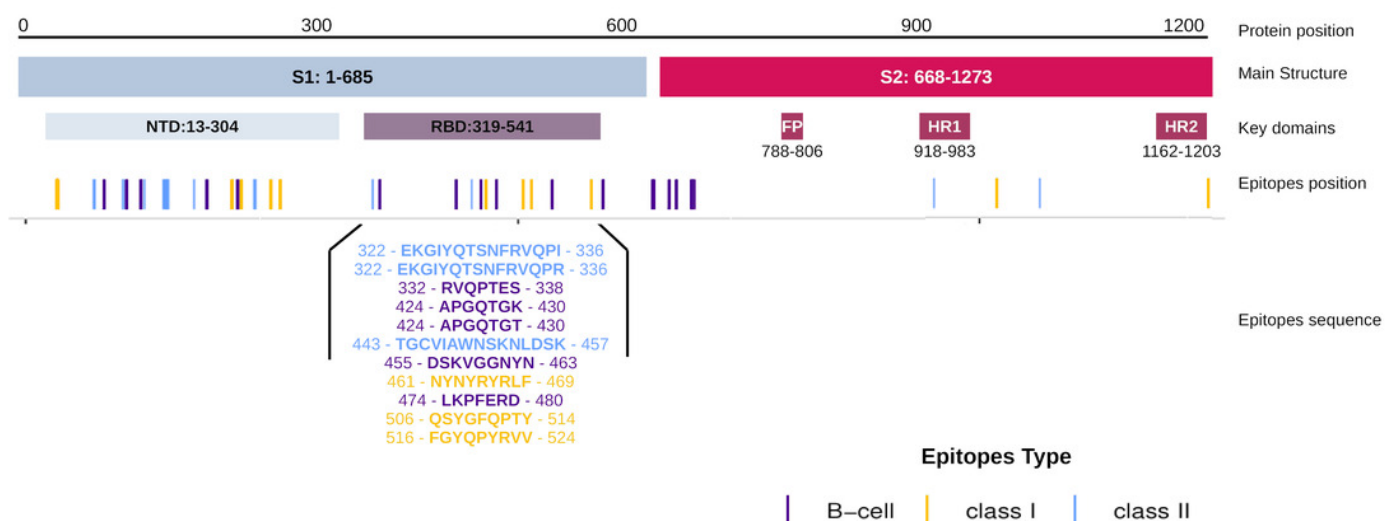



Figure 5

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.

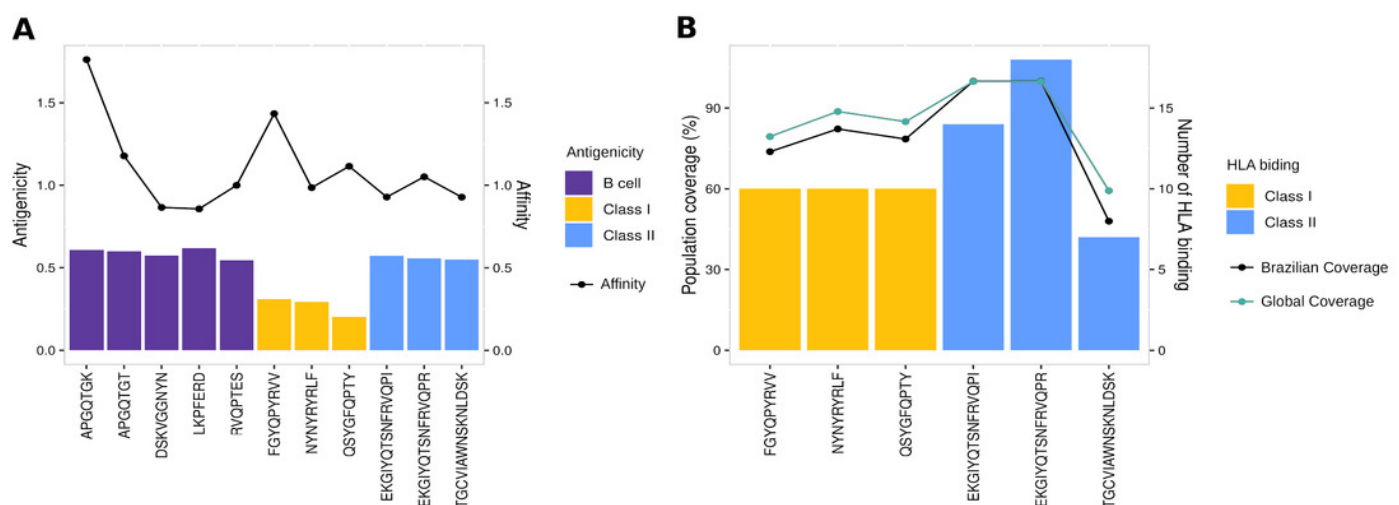
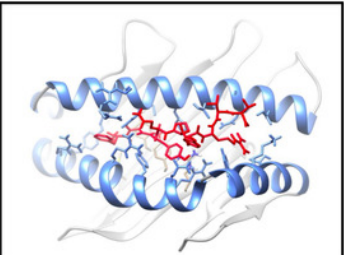
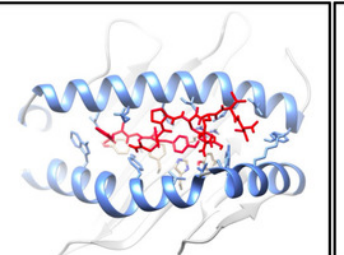
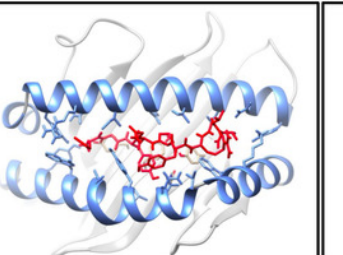
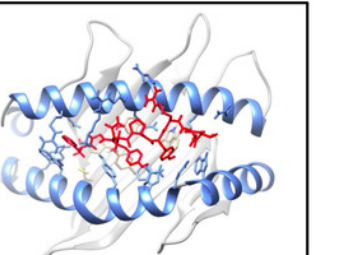
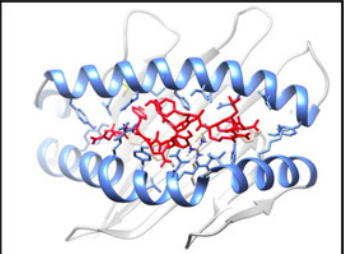
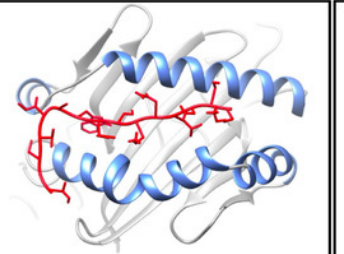
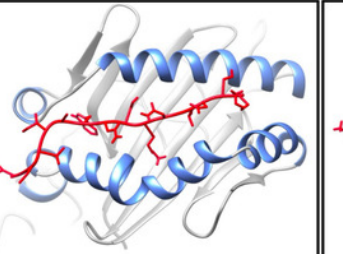
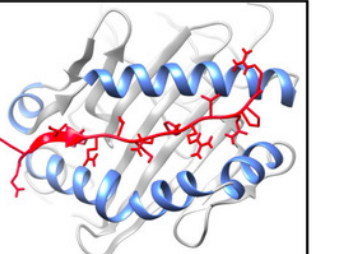



Figure 6

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.

516-FGYQPYRVV-524 HLA-A*01:01 ΔG = -19.2 kcal/mol	**516-FGYQPYRVV-524** HLA-A*02:01 ΔG = -18.7 kcal/mol	**516-FGYQPYRVV-524** HLA-B*08:01 ΔG = -19.1 kcal/mol	**516-FGYQPYRVV-524** HLA-C*12:03 ΔG = -19.7 kcal/mol
461-NYNYRYRLF-469 HLA-A*01:01 ΔG = -23 kcal/mol	**443-TGCVIAWNSKNLDSK-457** HLA-DRB1*03:01 ΔG = -26.2 kcal/mol	**443-TGCVIAWNSKNLDSK-457** HLA-DRB1*12:02 ΔG = -26.1 kcal/mol	**322-EKGIYQTSNFRVQPR-336** HLA-DRB1*04:01 ΔG = -31.6 kcal/mol

Figure 7

In silico simulation of immune response using multi-epitope construct as an antigen.

The multi-epitopes 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.

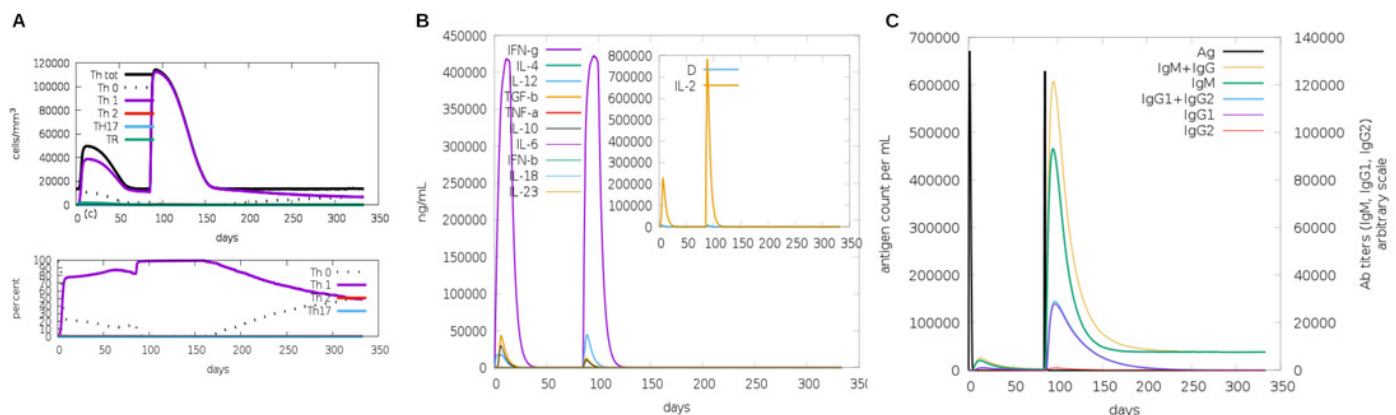
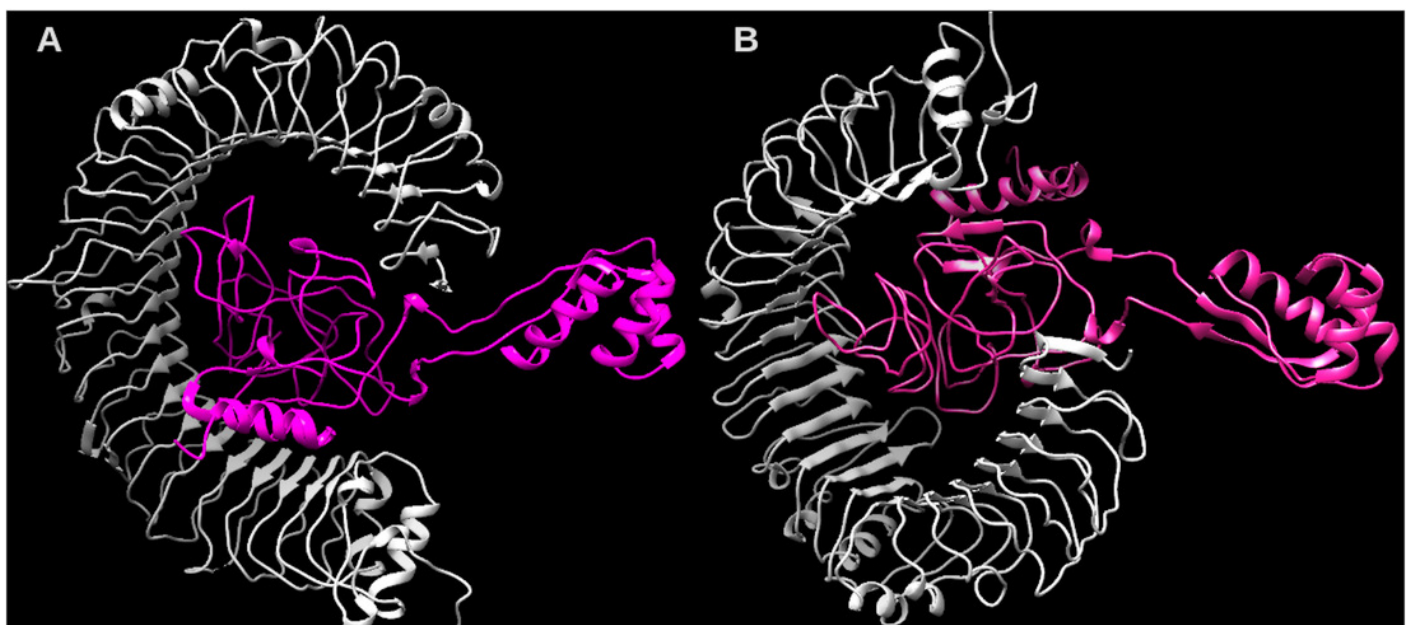



Figure 8

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.

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**

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)**

* Percentage related to number of predicted epitope per protein