

1 Immuno-informatics Design of a Multimeric Epitope Peptide 2 Based Vaccine Targeting SARS-CoV-2 Spike Glycoprotein

3
4 **Author:** ¹Onyeka S. Chukwudozie, ²Clive M. Gray, ¹Tawakalt A. Fagbayi,
5 ³Rebecca C. Chukwuanukwu, ⁴Victor O. Oyebanji, ¹Taiwo T. Bankole, ¹A.
6 Richard Adewole, ⁵Eze M. Daniel

7
8 ¹Department of Cell Biology and Genetics, University of Lagos, Lagos, Nigeria.
9 ²Division of Immunology, Institute of Infectious Disease and Molecular Medicine
10 and Department of Pathology, University of Cape town, South Africa.
11 ³Immunology Unit, Medical Laboratory Science Department, Nnamdi Azikiwe
12 University, Nnewi Campus, Nigeria.
13 ⁴Department of Veterinary Pathology, University of Ibadan, Ibadan, Nigeria.
14 ⁵Public Health Biotechnology Unit, Institute of Child Health, University College
15 Hospital, University of Ibadan, Nigeria.

16
17 **Correspondence:** Solomononyeka84@gmail.com
18 **Tel:** +2348188293195
19
20
21
22
23
24
25
26
27

28

ABSTRACT

29 Developing an efficacious vaccine to SARS-CoV-2 infection is critical to stem COVID-19
30 fatalities and providing the global community with immune protection. We have used a
31 bioinformatic approach to aid in the design of an epitope peptide-based vaccine against the spike
32 protein of the virus. Five antigenic B cell epitopes with viable antigenicity and a total of 27
33 discontinuous B cell epitopes were mapped out structurally in the spike protein for antibody
34 recognition. We identified eight CD8⁺ T cell 9-mers along with 12 CD4⁺ T cell 14-15-mer as
35 promising candidate epitopes putatively restricted by a large number of MHC-I and II alleles
36 respectively. We used this information to construct an *in silico* chimeric peptide vaccine whose
37 translational rate was highly expressed when cloned in pET28a (+) vector. The vaccine construct
38 was predicted to elicit high antigenicity and cell-mediated immunity when given as a
39 homologous prime-boost, with triggering of toll-like receptor 5 by the adjuvant linker. The
40 vaccine was characterized by an increase in IgM and IgG and an array of Th1 and Th2 cytokines.
41 Upon *in silico* challenge with SARS-CoV-2, there was a decrease in antigen levels using our
42 immune simulations. We therefore propose that potential vaccine designs consider this approach.

43

44 Keyword: Covid-19, SARS-CoV-2, Epitope Vaccine, Reverse Vaccinology, Molecular dynamics
45 simulation

46

47

48

49

50

51

52 **Introduction**

53 An unprecedeted pneumonia disease outbreak was reported in late December 2019, after
54 several deaths were recorded in Wuhan, China [1]. There was a rapid spread of the disease from
55 the city of Wuhan to many countries including the United States with thousands infected and
56 many dying within months of initial spread [1, 2]. On the 31st of December 2019, the disease
57 outbreak was traced to a novel strain of coronavirus [1], which was later termed SARS-CoV-2
58 and the disease COVID-19 by the WHO [3, 4]. Reports showed that as of 16th June 2020, there
59 have been at least 440,421 confirmed deaths and more than 8,144,359 confirmed cases [5].
60 SARS-CoV-2 been identified as a new strain from group 2B Coronaviruses, with approximately
61 70% genetic similarity to SARS-CoV, from the 2003 outbreak [4]. The virus has a 96%
62 similarity to a bat coronavirus, so it is widely suspected to originate from bats [6, 7]. The
63 pandemic has resulted in travel restrictions and nationwide lockdowns in several countries and
64 resulted in economic mayhem [1]. In the search for solutions, genome sequences have been
65 deposited in several online repositories, and developing a universally available vaccine is critical.

66 Coronaviruses are group of viruses that generally causes disease in mammals and Aves. In the
67 case of humans, coronaviruses are responsible for respiratory tract infections that can range from
68 mild, such as cases of the common cold and low fever, and others that can be lethal, such
69 as SARS, MERS, and COVID-19 [4]. Cases have shown that symptoms can vary depending on
70 the host, where for example, upper respiratory tract disease has been found in chickens and
71 diarrhea in cows and pigs [6]. The Coronavirus is grouped under the subfamily
72 *Orthocoronavirinae*, in the family Coronaviridae, order Nidovirales [5, 6]. They are positive-
73 sense single stranded RNA virus with nucleocapsid of helical symmetry. They are also

74 enveloped with spike proteins and the genome size of many of the coronaviruses range from 27-
75 34 kilobases, the largest among known retroviruses [8].

76 The current zoonotic jump to humans is of concern as the upper respiratory tract infection can
77 lead to fatal disease in some individuals. There is yet to be an effective vaccine to prevent or
78 treat any human coronavirus infections.

79 Structural analysis has revealed that the Spike protein S1 attaches the virion to the cell
80 membrane by interacting with ACE2 and CLEC4M/DC-SIGNR receptors. The Internalization of
81 the virus into the endosomes of the host cell induces conformational changes in the S
82 glycoprotein [9]. Proteolysis by cathepsin may unmask the fusion peptide of S2 and activate
83 membranes fusion within host endosomes. Spike protein S2 mediates fusion of the virion and
84 cellular membranes by acting as a class I viral fusion protein. Under the current model, the
85 protein has at least three conformational states: pre-fusion native state, pre-hairpin intermediate
86 state, and post-fusion hairpin state [9]. During viral and target cell membrane fusion, the coiled
87 regions (heptad repeats) assume a trimer-of- hairpin structures, positioning the fusion peptide in
88 close proximity to the C-terminal region of the ectodomain [10]. The formation of this structure
89 appears to drive apposition and subsequent fusion of viral and target cell membranes. Spike
90 protein S2' acts as a viral fusion peptide which is unmasked following S2 cleavage occurring
91 upon virus endocytosis [10]. The cellular receptor of SARS-CoV-2 and the receptor-binding
92 domain (RBD) on the S protein has been identified [11, 12]. Previous studies have shown that
93 the RBD in the S1 region plays a critical role in neutralizing antibody induction, angiotensin-
94 converting enzyme 2 (ACE2) binding and virus entry [13]. Depletion of RBD-specific antibodies
95 from sera significantly reduced serum-neutralizing capability, indicating that this domain is
96 dominant in neutralizing antibody induction [14]. The pivotal role of the S protein in viral

97 infection has made it a top candidate for vaccine production. There are currently over 90 SARS-
98 CoV-2 vaccine candidates [1] and an epitope-based vaccine may provide a useful complimentary
99 approach that would steer immunity to immunogenic epitopes on the S protein. With the
100 immuno-informatics, it is now possible to evaluate the immunogenic properties of proteins via
101 computational methods (*in silico*) with high efficiency and confidence [15-17]. We used such an
102 approach to design an epitope peptide-based vaccine against SARS-CoV-2 spike glycoprotein
103 and then *in silico* mimic the range of responses in a prime-boost scenario.

104 **Materials and Methods**

105 **Data retrieval, Structural and Physiochemical Analysis of SARS-CoV-2 Spike Protein**

106 The protein sequence from different geographical regions was retrieved from the NCBI
107 repository with their corresponding accession numbers: Wuhan, China (Genbank ID:
108 QHD43416.1), Japan (Genbank ID: BCA87361.1), California, USA (Genbank ID:
109 QHQ71963.1), Washington, USA (Genbank ID: QHO60594.1), and Valencia, Spain (Genbank
110 ID: QIQ08790.1). The protein structure of the SARS-CoV-2 spike (PDB: 6VSB) was
111 downloaded from the protein data bank. The physiochemical properties of the protein sequence
112 such as the GRAVY (Grand average of hydropathicity), half-life, molecular weight, instability
113 index, aliphatic index, and amino acid atomic composition was bio-computed via an online tool
114 Protparam (<http://web.expasy.org/protparam/>) [18].

115 **Prediction of B Cell Linear and Discontinuous Epitopes**

116 The Bepipred server from the Immune-Epitope-Database and Analysis-Resource (IEDB)
117 database was used for this prediction was used to identify B cell linear epitopes [19]. Bepipred-
118 2.0 is based on a random forest algorithm trained on epitopes annotated from antibody-antigen
119 protein structures [19]. This method is superior to other available tools for sequence-based

120 epitope prediction with regards to both epitope data derived from solved 3D structures and a
121 large collection of linear epitopes downloaded from the IEDB database [19]. The following
122 criteria such as the specificity at 75% and 14-15 mers (residues) was assumed to bind to MHC.
123 Several conditions such as antigenicity, accessibility of surface, flexibility, hydrophilicity is
124 imperative for the prediction of the B cell linear epitope. These conditions are taken into
125 consideration when making predictions with the Bepipred linear epitope prediction and Parker
126 hydrophilicity prediction algorithms.

127 SVMTriP (<http://sysbio.unl.edu/SVMTriP/>) was also used in the prediction of the B cell linear
128 epitopes. The SVMTriP is a Support Vector Machine method used to predict linear antigenic
129 epitopes which combine the Tri-peptide similarity and Propensity scores (SVMTriP).
130 Application of SVMTriP to non-redundant linear B-cell epitopes extracted from IEDB achieved
131 a sensitivity of 80.1% and a precision of 55.2% with five-fold cross-validation. For antigenicity
132 testing, these epitopes were subjected Vaxijen 2.0 [20]. We further predicted the discontinuous
133 epitopes which possesses greater attributes than the linear epitopes and discovery of
134 discontinuous B-cell epitopes is a major challenge in vaccine design. Previous epitope prediction
135 methods have mostly been based on protein sequences and are not very effective. Therefore, the
136 DiscoTope server was used to predict the surface accessibility and amino acids that form
137 discontinuous B cell epitopes found from X-ray crystallography of antigen/antibody protein
138 buildings. Pymol was utilized to examine the positions of forecast epitopes on the 3D structure of
139 SARS-CoV-2 protein [21].

140 **Prediction of epitopes restricted by class I Human Leukocyte Antigen (HLA) CD8+ (CTL)**
141 **and class II HLA CD4+ T cells (HTL)**

142 For *de novo* prediction of Covid-19 spike glycoprotein CD8⁺ T cell epitopes (peptides), we used
143 IEDB MHC I binding prediction algorithms (<http://tools.iedb.org/mhci>). This method integrates
144 the prediction of epitopes restricted to a large number of MHC class I alleles and proteasomal C-
145 terminal cleavage, using artificial neural network application. For better predictive accuracy,
146 other software such as artificial neural network (ANN), stabilized matrix method (SMM), MHC
147 binding energy covariance matrix (SMMPMBEC), NetMHCpan, pickpocket, and
148 NetMHCstapan, were adopted for this purpose. All of these predictive tools are archived on the
149 IEDB (Immune-Epitope-Database and Analysis-Resource) database with mathematical threshold
150 before best-fit epitopes are selected from each online server. To predict the CD4⁺ T cell epitopes
151 (peptides), we used the MHC II binding predictions tool (<http://tools.iedb.org/mhcii>) found in
152 the IEDB database. First, we selected the epitopes whose binding diversities with the different
153 HLA serotypes were higher, and we further subjected these epitopes to Vaxijen 2.0 server to test
154 for their antigenicity at a recommended threshold of 0.7. We also considered for further analysis,
155 by subjecting the top-scoring predicted epitopes from each tool that had been predicted by five or
156 more different methods and submitted them to IEDB T cell Class I Immunogenicity predictor
157 (<http://tools.iedb.org/immunogenicity>). Results were given in descending score values.
158 However, the table can also be sorted by clicking on individual column headers. The higher
159 score indicates a greater probability of eliciting an immune response.

160 **Profiling of the selected T cells Epitopes**

161 Following the selection of HLA-restricted CD8⁺ and CD4⁺ T cell epitopes, critical features such
162 as peptide toxicity predicted from the ToxinPred server (<http://crdd.osdd.net/raghava/toxinpred>),
163 allergenicity, predicted from AllergenFP 1.0 and digestion predicted from protein digest server

164 were made. All of these criteria were considered before the final selection of the T cell epitopes.
165 Epitopes with no toxicity were selected. Antigenicity testing was conducted through the Vaxijen
166 v2.0 server (<http://www.ddgpharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) [20], which operate
167 based on auto- and cross-covariance transformation of the input protein sequence into uniform
168 vectors of principal amino acid properties. The antigenicity index is generated at a threshold of
169 0.7.

170 **Epitope Conservancy in related SARS-CoV-2 Spike protein from different geographical
171 locations**

172 Conservation analysis of selected epitopes is the fraction of a protein sequence that contains the
173 epitope, while the identity is the degree of correspondence (similarity) between the sequences.
174 The degree of epitope conservancy was computed within the SARS-CoV-2 spike glycoprotein
175 sequence and set at a given identity level of 100 using the IEDB conservation-analysis-tool.

176 **The HLA-A 02*01 Allelic Affinity of the CD8⁺ T cell epitopes**

177 The Molecular docking of the antigenic epitopic peptides was conducted with the alleles they
178 were mostly restricted to, of which the HLA-A 02*01 allele was included. The X ray crystal
179 structure of the molecule was retrieved from the protein data bank (PDB: 4U6Y) and dock with
180 the epitopes. The refined binding free and dissociation energies were determined from the
181 docked complex.

182 **Population Coverage Analysis of CD8+ and CD4+ T cell epitopes.**

183 The selected epitopes from the HLA class I and class II families together with their respective
184 binding leukocyte antigens were subjected to IEDB Population Coverage tool
185 (<http://tools.iedb.org/population/>). This calculated the distribution or fraction of individuals
186 predicted to show a response to the selected epitopes with known HLA background. The tool

187 also computes the average number of epitope hits/HLA allele combinations recognized by the
188 entire population and the maximum and minimum number of epitope hits recognized by 90% of
189 the selected population. The HLA genotypic frequencies are calculated and T cell epitopes
190 queried based on the area, ethnicity and country. The entire world population was selected,
191 followed by subcontinents and countries. For countries like Nigeria and Ghana with no deposited
192 information on the IEDB database, they were included as part of the West African population.

193 **Designing of Multi-Epitope Vaccine Construct**

194 Selected antigenic epitopes were scrutinized to determine which could potentially induce
195 different Th1 and Th2 cytokines. Those with this attribute were selected for the vaccine
196 construct. To construct a multi-epitope vaccine, we finally selected CTL, HTL, and B cell linear
197 epitopes that were linked together with the help of AAY, GPGPG, and KK linkers, respectively.
198 To boost the immunogenic profile of the selected profile epitopes, an adjuvant would be
199 required. The outer membrane protein A (OmpA) (GenBank: AFS89615.1) was retrieved for this
200 purpose, because its serves as agonist to the human immune receptor by interacting with antigen
201 presenting cells [22]. The adjuvant was putatively added through the EAAAK linker, with the B
202 and HTL epitopes which were linked together through the GPGPG linkers. These complexes
203 were subsequently added to the CTL epitopes through the AAY linkers. The tag (6xHis-tag) was
204 added at the C terminal end of the vaccine construct. The 6xHis-tag is one of the simplest and
205 most widely used purification tags, with six or more consecutive histidine residues. These
206 residues readily coordinate with transition metal ions such as Ni^{2+} or Co^{2+} immobilized on beads
207 or resin for purification [23]. The intrinsic solubility properties of the vaccine peptide were
208 conducted using the CamSol tool, which yields a solubility profile where regions (residues) with
209 scores greater than 1 signifies soluble regions, while scores lesser than -1 represents poorly

210 soluble regions [24]. An overall score is generated for the entire sequence, as these amino
211 residue scores are ranked based on their level of solubility.

212

213

214 **Structural Modelling, Assessment, and Validation**

215 All the predicted peptides 3D structures were modelled via PEPFOLD server at RPBS MOBYL
216 portal [25]. PEP-FOLD is a *de novo* approach aimed at predicting peptide structures from amino
217 acid sequences [25, 26, 27]. This method, based on structural alphabet SA letters to describe the
218 conformations of four consecutive residues, couples the predicted series of SA letters to a greedy
219 algorithm and a coarse-grained force field [25, 26]. The predicted models are in cluster ranks
220 which are defined according to their scores. The cluster representatives correspond to the models
221 of the clusters having the best scores, i.e. with the lowest sOPEP energy (representing the highest
222 *tm* value) [25]. The PSIPRED v4.0 server was adopted for the prediction of the vaccine's
223 secondary structure [28], while the Swiss dock online tool was used for the tertiary structure
224 prediction of both the vaccine construct and the human HLA class II histocompatibility antigen,
225 DR alpha chain. To validate the generated protein structure, Procheck online tool together with
226 Ramachandran plot analysis was generated [29]. The plot analysis was able to show the allowed
227 and disallowed dihedral angles psi (ψ) and phi (ϕ) of an amino acid which is calculated based on
228 van der Waal radius of the side chain. The corresponding percentage value of both the allowed
229 and disallowed region of the separate plots of glycine and proline residues of the modeled
230 structure was generated.

231 **Molecular Docking Studies**

232 One of the best ways to access the immune response mounted by the epitopes is by studying their
233 binding affinity characterizing their molecular interaction with the human HLA class I and I
234 molecules. The binding pockets on the HLA-class I and II molecules and the human immune
235 receptor (TRL5) was predicted using the CASTp server (<http://sts.bioe.uic.edu/castp/>). The
236 CASTp server provides comprehensive and detailed quantitative characterization of topographic
237 features of a protein. The geometric modeling principle involves the calculation strategy of
238 alpha-shape and discrete-flow methods that are applied to the protein binding site, also the
239 measurement of pocket-size by the program [30]. The protein pocket atom is identified and then
240 the volume and area are calculated [31, 32]. The program also identifies the atoms forming the
241 rims of pocket mouth, computes how many mouth openings for each pocket, predict the area and
242 circumference of mouth openings, finally locates cavities and calculates their size. The
243 secondary structures were calculated by DSSP [30, 32]. The predicted structure of the HLA class
244 I and class II allele protein were utilized for molecular docking analysis with the selected
245 epitopes (peptides) to evaluate their binding affinities. The protein structure was chemically
246 manipulated by expulsion of water and ligand molecules. For the peptide-protein interaction,
247 HPEPDOCK (<http://huanglab.phys.hust.edu.cn/hpepdock/>) was utilized for this purpose [33]. It
248 uses a hierarchical algorithm and instead of running lengthy simulations to refine peptide
249 conformations, HPEPDOCK also considers peptide flexibility. UCSF Chimera and Pymol tools
250 were utilized to produce figures of docked complexes. ZDOCK server [34] was adopted for the
251 molecular docking between the multiple epitope vaccine peptides and the human immune
252 receptor (PDB: 3J0A). ZDOCK is based on the rigid-body docking program that predicts
253 protein-protein complexes and symmetric multimers. ZDOCK achieves high predictive accuracy

254 on protein-protein docking benchmarks, with >70% success in the top 1000 predictions for rigid-
255 body [35].

256 **Molecular Dynamics Simulation Studies**

257 The biological molecules in a solution of the peptide vaccine construct was studied, using the
258 small -and wide-angle X-ray scattering (SWAXS) [36]. The generated curves require accurate
259 prediction from the structural model. The predictions are complicated by scattering contributions
260 from the hydration layer and by effects from thermal fluctuations. The MD simulations provide a
261 realistic model for both the hydration layer and the excluded solvent, thereby avoiding any
262 solvent-related fitting parameters, while naturally accounting for thermal fluctuations [36]. To
263 determine the protein compactness, the radius of gyration of the biomolecule through the Guinier
264 analysis was also conducted. The interacting complex between the vaccine and the toll-like
265 receptor (PDB: 3J0A), was thoroughly accessed based on the existing coordinates between the
266 docked protein complex. Parameters considered were the deformability, B factor, eigenvalues
267 associated with the normal mode which represents the motion stiffness. Its value is directly
268 related to the energy required to deform the structure. The lower the eigenvalue, the easier the
269 deformation. The covariance matrix was also considered for the simulation. It indicates the
270 coupling between the pairs of residues. The correlation matrix is computed using the Ca
271 Cartesian coordinates. The elastic network of the docked complex was also computed [36].

272 ***In Silico* Codon Adaptation and Cloning**

273 For the maximum expression of the vaccine in the host, a codon optimization was conducted.
274 This was done using the Java Codon Adaptation Tool (JCat), with the aim of boosting the
275 vaccine translational rate in *E. coli* K12. Restriction enzymes cleavage sites, prokaryote
276 ribosomal binding site, and finally rho-independent transcription termination, were all avoided

277 during the option selection. Codon adaptation index (CAI) value and GC content of the adapted
278 sequence was obtained and compared with the ideal range. The obtained refined nucleotide was
279 cloned into the pET28a (+) vector, utilizing the SnapGene 4.2 tool.

280 **Immune Simulation of the Chimeric Peptide Vaccine**

281 The entire predicted conjugate vaccine peptide was accessed for their immunogenicity and
282 immune response attributes using the C-ImmSim online server ([http://150.146.2.1/C-
283 IMMSIM/index.php](http://150.146.2.1/C-IMMSIM/index.php)) [37]. The server uses a machine-learning basis in predicting the epitopes
284 and the associated immune interactions. It automatically simulates three anatomical
285 compartments which include: (i) bone, where the hematopoietic stem cells are stimulated and
286 myeloid cells are produced, (ii) the lymphatic organ and (iii) the thymus where naive T cells are
287 selected to avoid autoimmunity. Three injections containing the designed peptide vaccine was
288 administered at an interval of four weeks. From the default parameters, each time step were
289 positioned at 1, 84, and 168 meaning that each time step is 8hours and time step 1 is the injection
290 administered at time zero. So, three injections were administered at four weeks interval.
291 However, eight injections were administered four weeks apart to stimulate repeated exposure to
292 the antigen. In this scenario, the T cell memory will undergo continuous assessment. The
293 Simpson index was graphically interpreted from the plot analysis [37].

294 **Results**

295 **Linear and Discontinuous B cell Epitopes**

296 There were 5 promising linear B cell epitopes with non-allergenic attributes. The peptide
297 “VRQIAPGQTGKIAD” comparatively had the highest antigenic index than the other predicted
298 B cell epitope candidates. A characteristic non-toxic peptide attribute makes the selected
299 antigenic epitopes safe for vaccine design. The antigen conservancy of the epitopes across the
300 retrieved spike protein from different geographical locations was 100% [**Table 1**].

301 **Table 1:** B cells linear epitopes of SARS-CoV-2 Spike glycoprotein and their immunogenic
302 properties.

No.	Start	End	Peptide	Length	Antigenicity	Toxicity	Allergenicity	Conservancy
1	407	420	VRQIAPGQTGKIAD	14	1.261	Non-toxic	Non-allergen	100
2	18	32	LTTRTQLPPAYTNSF	15	0.79	Non-toxic	Non-allergen	100
3	4	18	FLVLLPLVSSQCVNL	15	0.8302	Non-toxic	Non-allergen	100
4	749	762	SNLLLQYGSFCTQL	14	0.800	Non-toxic	Non-allergen	100
5	1056	1069	PHGVVFLHVTVYVPA	14	0.806	Non-toxic	Non-allergen	100

303 **NB:** All of the selected B cell are tagged B1-B5.

304

305 Graphically, using the Kolaskar and Tongaonkar antigenicity scale, Emini surface accessibility,
306 and the Chou and Fasman beta-turn predictions, regions with viable immunogenic properties
307 were determined. The scale was able to show the favorable regions across the protein that are
308 potentially antigenic [supplementary data: Figure s1a-s1c]. The resulting B cell linear epitopes
309 were mapped out from the spike protein [Figure 1a]. The predicted discontinuous epitopes were
310 selected from the entire protein chain component (A, B, and C) of the virus spike protein (PDB:
311 6VSB), and ranked based on their propensity scores. A total of 27 discontinuous epitopes were
312 mapped out from the protein structure as shown in Figure 1b. Across the discontinuous epitopes
313 of the protein chain components, the maximum contact number was 10, and the least was 7. The
314 chain C component of the spike protein had a higher number of contact residues [supplementary
315 data: S-Table 2].

316 **The CTL and HTL Epitopes**

317 The spike protein sequence was scanned across multiple HLA class 1 alleles. The peptides were
318 selected based on their percentage rankings and the number of alleles they potentially bind to.

319 Additionally, the peptides were subjected to antigenicity test using Vaxijen 2.0. Based on the
320 antigenicity scores, 16 epitopes were selected for the next stage of screening. The most important
321 peptides are those with the capacity of binding with a higher number of HLA class I molecules
322 and showing a non-allergenic attribute. Before vaccine design can be considered, the
323 allergenicity prediction is crucial, as there is a possibility of vaccine candidates eliciting a Type
324 II hypersensitivity reaction. Allergen 1.0 online was adopted for this analysis and the
325 allergenicity scores show that these epitopes were non-allergenic. The non-toxicity attribute of
326 the peptides also makes them suitable for vaccine production. Eight peptides were allergenic and
327 eight were also non-allergenic.

328 The non-allergenic peptides were: GAEHVNNSY which is putatively restricted to HLA-
329 A*01:01, HLA-B*15:01, and HLA-C*02:02, KTSVDCTMY attaches to 5 alleles: HLA-
330 A*01:01, HLA-B*15:01, HLA-C*02:02, HLA-A*03:01, HLA-A*30:01, TTEILPVSM would be
331 able to bind to 3 alleles: HLA-A*01:01, HLA-C*02:02, HLA-C*01:02. Other selected epitopes
332 had similar putative restricted attachments such as “ILDITPCSF” with an antigenicity score of
333 1.184 and potentially attaches to 7 alleles: HLA-A*01:01, HLA-C*04:01, HLA-C*02:02, HLA-
334 C*01:02, HLA-B*15:01, HLA-A*02:01, HLA-B*13:01. The peptide GVYFASTEK would be
335 able to bind to 3 alleles: HLA-A*03:01, HLA-A*30:01, HLA-C*02:02, GVYYHKNNK,
336 ASANLAATK, VLKGVKLHY had the same similar attribute of binding to two, three and five
337 alleles respectively [Table 2].

338 **Table 2:** MHC-I T cell epitopes of SARS-CoV-2 spike glycoprotein.

Alleles	Start	End	Length	Peptide	Allergenicity	Antigenicity
HLA-A*01:01, HLA-B*15:01, HLA-C*02:02	652	660	9	GAEHVNNSY*	Non-allergen	0.9

HLA-A*01:01, HLA-B*15:01, HLA-C*02:02, HLA-A*03:01, HLA-A*30:01	733	741	9	KTSVDCTMY*	Non-allergen	1.1824
HLA-A*01:01, HLA-C*02:02, HLA-C*01:02	723	731	9	TTEILPVSM*	Non-allergen	1.2262
HLA-A*01:01, HLA-C*04:01, HLA-C*02:02, HLA-C*01:02, HLA-B*15:01, HLA-A*02:01, HLA-B*13:01	584	592	9	ILDITPCSF*	Non-allergen	1.184
HLA-A*01:01, HLA-B*15:01	441	449	9	LDSKVGGNY	Allergen	0.7814
HLA-A*02:01, HLA-C*02:02, HLA-C*01:02, HLA-B*13:01, HLA-C*04:01, HLA-A*30:01, HLA-B*15:01, HLA-B*08:01, HLA-B*07:02	417	425	9	KIADYNYKL	Allergen	1.664
HLA-A*02:01, HLA-C*02:02, HLA-C*01:02, HLA-A*30:01, HLA-B*08:01, HLA-C*04:01, HLA-B*13:01	1060	1068	9	VVFLHVTYV	Allergen	1.512
HLA-A*02:01, HLA-B*08:01, HLA-C*04:01, HLA-C*01:02, HLA-C*02:02, HLA-A*01:01, HLA-B*13:01, HLA-B*07:02	109	117	9	TLDSKTQSL	Allergen	1.0685
HLA-A*03:01, HLA-A*30:01, HLA-C*02:02	89	97	9	GVYFASTEK*	Non-allergen	0.7112
HLA-A*03:01, HLA-A*30:01	142	150	9	GVYYHKNNK*	Non-allergen	0.8264
HLA-A*03:01, HLA-A*30:01, HLA-C*02:02	1020	1028	9	ASANLAATK*	Non-allergen	0.7014
HLA-A*03:01, HLA-A*30:01	409	417	9	QIAPGQTGK	Allergen	1.8297
HLA-A*03:01, HLA-A*30:01, HLA-B*15:01, HLA-C*02:02, HLA-A*01:01	1264	1272	9	VLKGVKLHY*	Non-allergen	1.2378
HLA-A*03:01, HLA-A*30:01, HLA-C*02:02	349	357	9	SVYAWNRKR	Allergen	0.765
HLA-A*03:01, HLA-A*30:01	725	733	9	EILPVSMTK	Allergen	1.6842
HLA-A*03:01, HLA-A*30:01	378	386	9	KCYGVSPK	Allergen	1.4199

339 * The selected epitopes

340

341 For the HLA class II T cells epitopes, the spike protein sequence was also scanned through a
342 large number of the MHC-II alleles. Twelve epitopes were selected based on their antigenic
343 properties. All of these selected non-allergenic epitopes are capable of eliciting an immune
344 response by inducing either or all of IFN- γ , IL-4 and IL-10 cytokines. The peptide
345 “GYFKIYSKHTPINLV” was the only candidate to induce all of the three cytokines, which was
346 intriguing. Selection of these epitopes was also centered on their putative bindings to a large
347 number of MHC-II alleles. The peptide “FAMQMAYRF”, with an antigenicity score of 1.0278,
348 attaches to 8 HLA alleles: -DRB1*01:02, -DRB1*01:04, -DRB1*01:03, -DRB1*01:01, -
349 DRB1*01:05, -DRB1*07:01, -DRB1*04:01, and -DRB1*03:01. The epitope “FRVQPTESI”,
350 with an antigenicity score of 0.9396, is also restricted to 8 HLA alleles: -DRB1*04:01, -
351 DRB1*01:01, -DRB1*01:05, -DRB1*07:01, -DRB1*01:03, -DRB1*01:02, -DRB1*03:01, and -
352 DRB1*01:04. The HLA-DRB1 is the most common and versatile MHC-II molecule. The entire
353 putative attachments of the selected epitopes are summarized [Table 3]. The conservancy level
354 of the epitopes across the retrieved spike protein sequences from different geographical location
355 was 100%.

356 **Table 3:** MHC class II T-cell Epitopes of SARS-CoV-2 spike glycoprotein

S/ N	Core peptides	Peptides	Star t	End	Length	Allergenic ity	IFN- γ	IL -4	IL -10	MHC II Alleles	Antigenicit y
1.	FAMQMAYRF	IPFAMQMAYRFNGIG	896	910	15	Non-allergenic	-	+	-	DRB1*01:02, DRB1*01:04, DRB1*01:03, DRB1*01:01, DRB1*01:05, DRB1*07:01, DRB1*04:01, DRB1*03:01	1.0278
2.	FRVQPTESI	IYQTSNFRVQPTESI	312	326	15	Non-allergenic	+	+	-	DRB1*04:01, DRB1*01:01, DRB1*01:05, DRB1*07:01, DRB1*01:03,	0.9396

									DRB1*01:02, DRB1*03:01, DRB1*01:04		
3.	FQTRAGCLI	STGSNVFQTRAGCLI	637	651	15	Non-allergenic	+	-	+	DRB1*01:03	1.7332
4.	CVLGQSKRV	KMSECVLGQSKRVDF	1028	1042	15	Non-allergenic	-	+	-	DRB1*01:03	0.9083
5.	FLHVTYVPA	VVFLHVTYVPAQEKN	1060	1074	15	Non-allergenic	+	-	+	DRB1*04:01, DRB1*07:01, DRB1*01:01, DRB1*01:05	1.3346
6.	LQIPFAMQM	ALQIPFAMQMAYRFN	893	907	15	Non-allergenic	-	+	-	DRB1*01:02, DRB1*01:04, DRB1*01:03, DRB1*01:01, DRB1*01:05	1.0680
7.	IGINITRFQ	IGINITRFQTLLALH	231	245	15	Non-allergenic	-	-	+	DRB1*03:01	1.3386
8.	VFQTRAGCL	NVFQTRAGCLIGAEH	641	655	15	Non-allergenic	-	-	+	DRB1*07:01, DRB1*01:01, DRB1*01:05	1.7094
9.	FTISVTTEI	TNFTISVTTEILPVS	716	730	15	Non-allergenic	-	+	+	DRB1*07:01, DRB1*04:01, DRB1*01:05, DRB1*01:01	0.8535
10.	YFKIYSKHT	GYFKIYSKHTPINLV	199	212	15	Non-allergenic	+	+	+	DRB1*01:03, DRB1*01:02, DRB1*04:01, DRB1*01:01, DRB1*01:05, DRB1*07:01, DRB1*01:04	0.9056
12.	AALQIPFAM	AALQIPFAMQMAYRF	892	906	15	Non-allergenic	-	+	-	DRB1*01:02	0.7747
13.	VLSFELLHA	RVVVLSFELLHAPAT	509	523	15	Non-allergenic	+	+	-	DRB1*01:02	1.0776

357 + induced, - non-induced. Antigenicity threshold at 0.7

358

359 **Population Coverage of the CTL and HTL Epitopes**

360 HLA allelic distribution differs among diverse geographical regions and ethnic groups around
 361 the globe. It is therefore imperative to consider the population coverage in designing a viable
 362 epitope-based vaccine relevant for global populations. The selected CD8⁺ T cell epitopes

363 exhibited a higher individual percentage cover when queried with the entire world population.
 364 The HLA hits across the entire population revealed that approximately 81% of the world
 365 individuals are capable of responding to a median of 3 CTL epitopes [Table 4].

366
 367
 368
 369
 370

371 **Table 4:** Coverage of individual epitope (MHC class I) in the world

Epitope	Coverage	HLA allele (genotypic frequency (%))									Total 1 HLA hits
		HLA-A*01:0 1 (10.09)	HLA-A*02:0 1 (24.39)	HLA-A*03:0 1 (9.77)	HLA-A*30:0 1 (2.18)	HLA-B*13:0 1 (1.57)	HLA-B*15:0 1 (5.65)	HLA-C*01:0 2 (6.1)	HLA-C*02:0 2 (5.52)	HLA-C*04:0 1 (11.93)	
ILDITPCSF	73.40%	+	+	-	-	+	+	+	+	+	7
KTSVDCTMY	46.77%	+	-	+	+	-	+	-	+	-	5
VLKGVKLHY	46.77%	+	-	+	+	-	+	-	+	-	5
TTEILPVSM	33.48%	+	-	-	-	-	-	+	+	-	3
GAEHVNNSY	31.54%	+	-	-	-	-	+	-	+	-	3
GVYFASTEK	27.95%	-	-	+	+	-	-	-	+	-	3
ASANLAATK	27.95%	-	-	+	+	-	-	-	+	-	3
GVYYHKNNK	20.35%	-	-	+	+	-	-	-	-	-	2
Epitope set	81.05%	5	1	5	5	1	4	2	7	1	31

372 +: restricted

373 -: not restricted

374 shaded column: genotypic frequency of this allele is 0 (zero)

375

376 However, the population coverage for the CD4⁺ T cell epitopes was comparatively lower
377 compared to the CD8⁺ T cell epitopes, with average population coverage of 55.23% and
378 recognition of a median of 2 epitopes [Table 5].

379
380
381
382
383

384 **Table 5:** Coverage of individual epitope (MHC class II) in the world

Epitope	Coverage	HLA allele (genotypic frequency (%))								Total H L A hits	
		Class II	HLA-DRB1 *01:01 (6.65)	HLA-DRB1 *01:02 (1.8)	HLA-DRB1 *01:03 (0.92)	HLA-DRB1 *01:04 (0.01)	HLA-DRB1 *01:05 (0)	HLA-DRB1 *03:01 (10.47)	HLA-DRB1 *04:01 (6.46)	HLA-DRB1 *07:01 (10.71)	
IPFAMQMA YRFNGIG	55.23 %	+	+	+	+	+	+	+	+	+	8
IYQTSNFRV QPTESI	55.23 %	+	+	+	+	+	+	+	+	+	8
GYFKIYSKH TPINLV	41.83 %	+	+	+	+	+	-	+	+	+	7
VVFLHVTY VPAQEKN	38.05 %	+	-	-	-	+	-	+	+	+	4
TNFTISVTT EILPVS	38.05 %	+	-	-	-	+	-	+	+	+	4
NVFQTRAG CLIGAEH	28.63 %	+	-	-	-	+	-	-	+	+	3
IGINITRFQT LLALH	17.84 %	-	-	-	-	-	+	-	-	-	1
ALQIPFAM QMAYRFN	16.07 %	+	+	+	+	+	-	-	-	-	5
AALQIPFA MQMAYRF	3.19 %	-	+	-	-	-	-	-	-	-	1
RVVVLSFEL LHAPAT	3.19 %	-	+	-	-	-	-	-	-	-	1
KMSECVLG QSKRVDF	1.64 %	-	-	+	-	-	-	-	-	-	1
STGSNVFQT RAGCLI	1.64 %	-	-	+	-	-	-	-	-	-	1

Epitope set	55.23 %	8	6	6	4	8	3	6	7	48
-------------	---------	---	---	---	---	---	---	---	---	----

385 +: restricted

386 -: not restricted

387 The genotypic frequency of this allele is 0 (zero)

388

389 Based on the selection of the continents and countries, the European populace would likely show
390 a significant response to the selection of putative HLA class I restricted epitopes. England,
391 France, United States, Italy, and Oceania had the highest population coverage of 92.31%,
392 85.75%, 82.22%, 80.39%, and 75.07% respectively, while the Pakistan population had the lowest
393 population coverage at 35.8%. The population cover for the MHC class II epitopes in contrast to
394 the MHC class I epitopes was considerably lower. The striking observation was 0% coverage
395 exhibited by the Pakistan population [**Figure 2**].

396 **Binding Orientations of the CTL and HTL Epitopes in HLA-A*02:01 and HLA-
397 DRB1*01:01 Groove.**

398 The selected CTL and HTL antigenic epitopes were docked individually with the alleles they
399 were highly restricted to, which was HLA-A*02:01 for the CTL epitopes and HLA-DRB1*01
400 for the HTL epitopes. The differential binding patterns of the CTL epitopes were examined
401 [**Figure 3ai-3aix**]. Major class histocompatibility class II amino acid sequences are highly
402 polymorphic within a population, and correlate with individual differences in response to
403 infectious agents and vaccines. It is therefore expedient to structurally examine how the CD4+
404 epitopes recognize peptide fragments of antigens that lie in the antigen groove of the MHC-II
405 protein. The protein structure of the retrieved human HLA class II histocompatibility antigen,
406 DRB1 beta chain (human leukocyte antigen DRB1, HLA- DRB1*01:01) (PDB: 1AQD) was
407 retrieved for the molecular docking of the HTL epitopes because it was the most occurring allele

408 that the peptides were restricted to. The HTL epitopes with good population cover were chosen
409 for molecular docking with HLA-DRB1*01:01.
410 The peptides: IPFAMQMAYRFNGIG, IYQTSNFRVQPTESI, VVFLHVTYVPAQEKN,
411 TNTTISVTTEILPVS, and GYFKIYSKHTPINLV, were the selected epitopes. The binding free
412 energy characterizing the HLA-DRB1 antigenic binding groove and the interacting epitopes
413 alongside the corresponding amino residues was evaluated. The epitopes exhibited different
414 binding pattern with the MHC class II groove. Few of the binding peptides had a flanking region
415 outside the groove. Amino acids that are outside of the “core” peptide region extends out of the
416 open MHC-II binding groove forming the peptide flanking regions at both the N- and C-
417 terminus [Figure 3bi-3biv]. The epitope “IYQTSNFRVQPTESI” had the most extensive
418 flanking non-binding region with some of part of the peptide protruding completely out of the
419 groove. IPFAMQMAYRFNGIG had the best binding free energy score, with
420 TNTTISVTTEILPVS with the least binding energy [Table 6].

421 **Table 6:** Docking properties of HLA-DRB1*01:01 restricted epitopes

	Epitopes	Binding free energy (kcal/mol)	Interacting residues of the MHC II protein
a.	IPFAMQMAYRFNGIG	-18.92	PHE 15, GLN 146, PHE 152, PHE 14, VAL 88
b.	IYQTSNFRVQPTESI	-26.59	TRP 58, TRP 6, LEU 8, LEU 64, TYP 57
c.	VVFLHVTYVPAQEKN	-28.34	ARG 45, LEU 24, ARG 36, VAL 41, VAL 47
d.	TNTTISVTTEILPVS	-30.89	SER 123, LYS 9, VAL 126, HIS 13, LEU 144
e.	GYFKIYSKHTPINLV	-14.48	ILE 124, LEU 144, VAL 139, ALA 137, SER 141

422
423 **Construction of the Peptide Vaccine**
424 The multiple epitope peptide vaccine consists of 553 amino residues from 25 selected antigenic
425 B and T cells epitopes, covalently linked with an immuno-adjuvant [Figure 4a]. The tertiary
426 structure of the multiple epitope vaccine was also obtained [Figure 4b], and the structural
427 validation was assessed using ProSA-web which predicts the overall quality of the model

428 indicated in the form of z-score. If the z-scores of the predicted model are outside the range of
429 the characteristic for native proteins, it indicates the erroneous structure. The Z-score was -2.32
430 for the vaccine predicted model indicating a relatively good model [Figure 4c]. Before the
431 addition of the OmpA protein adjuvant, the conjugated vaccine was highly antigenic with a score
432 of 0.8 after subjecting it to Vaxijen server, signifying that the vaccine is viable at inducing
433 cellular and humoral immune response without the aid of an adjuvant. However, an adjuvant was
434 added to further boost the immunogenic properties to 0.85. A structural appraisal of the
435 secondary structure of the vaccine revealed 14% alpha helix, 41% beta strand and the disordered
436 region was 17%.

437

438

439 **Physiochemical, Solubility and Solvation properties of the Vaccine**

440 The physicochemical parameters and solubility properties of a vaccine candidates help to define
441 the efficacy and effectiveness of the vaccine. The molecular weight of the vaccine was 60728.51
442 Da and the bio-computed theoretical pI was 9.30, with an estimated half-life of 30 hours. The
443 instability index was 27.84, signifying that the vaccine is stable in a solvent environment (>40
444 signifies instability). The aliphatic index is computed to be 88.21, with a GRAVY score of -
445 0.056. The intrinsic vaccine solubility at a neutral pH 7 revealed the hydrophilic and
446 hydrophobic core of the vaccine construct [Figure 4d]. The overall solubility value of the
447 vaccine was -2.632908 signifying hydrophilic property. The stability of the vaccine construct
448 was assessed considering the radius of gyration and solvent density. The solvent density of the
449 vaccine is 334 e/nm^{-3} , the envelope distance is 7 Å, number of q values is 101 and the heavy
450 atoms is a total of 1589. The protein contains 3143 solutes atoms and 14365 water molecules.

451 The solute has zero charges with the distance of envelope from the solute at 0.7nm while the
452 maximum diameter of the solute is 7.3673 nm [**Figure 4e**].

453 **Molecular Docking between the Adjuvant Linker of the vaccine and the Toll-like receptor**
454 **(TLR5)**

455 TLR5 was selected due to its immunomodulatory ability to trigger IFN- β as well as activation of
456 type I IFN responses. This was clearly attested as our selected CD4+ epitopes elicited the Th1
457 and Th2 cytokines. The molecular interaction between the vaccine and the TLR5 (PDB: 3J0A),
458 was evaluated considering their refined binding energies and various interacting residues. The
459 conformational triggering of the TLR5 receptor was influenced by the adjuvant linker and not the
460 conjugated epitopes. The adjuvant linker binds to the A chain monomer of the toll-like receptor.
461 The interface amino residues of the receptor were: PRO 20, GLN 21, VAL 22, LEU 23, ASN 24,
462 THR 25, PRO 45, and PHE 46 respectively from the A chain, forming a hydrogen bond with the
463 interacting adjuvant residues: ASN 91, GLN 143, HIS 89, ALA 117, LEU 118, VAL 119, ARG
464 120, THR 142 and SER 141 [**Figure 5**]. The binding energy (ΔG) and dissociation constant (K_d)
465 predicted values of the protein-protein complex were -12.2 kcal mol⁻¹ and 1.0E-09 at 25.0°C
466 respectively.

467 **Molecular dynamics simulations**

468 The rigidity of the peptide vaccine system was examined by evaluating the radius of gyration
469 (R_g) values. The analyzed data shows that the average R_g value was 21.0067 nm, indicating that
470 the protein system retained its stability throughout the 85.5 ns time span of the MD simulation
471 [**Figure 6a**], as this signifies that the peptide vaccine is relatively stable. The molecular
472 interaction between the vaccine peptide and the TLR5 was screened for their protein stability, B
473 factor mobility, and deformity. This analysis relies on the associated coordinates of the docked
474 protein complex [**Figure 6b-6f**]. The eigenvalue found for the complex was 5.812952e-06. The

475 low eigenvalue for the complex signifies easier deformation of the complex, indicating that the
476 docking analysis between the vaccine and the TLR5 will activate immune cascades for
477 destroying the antigens.

478 **Codon Optimization and *In Silico* Cloning**

479 The length of the optimized vaccine codon sequence was 1659 nucleotides. The GC content of
480 the cDNA sequence and codon adaptive index was calculated as 50.8%, which still falls within
481 the recommended range of 30-70%, for effective translational efficiency. The codon adaptive
482 index was calculated as 0.93, falling within the range of 0.8-1.0, signifying the effective
483 expression of the vaccine constructs in the *E. coli*. EagI-NotI and SAI^I sites were subsequently
484 cloned into the pET28a (+) vector. The estimated length of the clone was 7.028 kbp [**Figure 7**].

485

486

487 **Immune Simulation of the Chimeric Peptide Vaccine**

488 At every administration and repetitive exposure to the attenuated peptide vaccine, there was a
489 significant increase in the antibody response with a simultaneous decrease in the antigen level.
490 This is characterized by an increase in IgM concentration. The observed increase in
491 immunoglobulin activities involving the combination of IgG1 + IgG2, IgG + IgM and IgM
492 antibodies components, was another evidence that the vaccine stimulated a good immune
493 response [**Figure 8a**]. This same immune response was observed for the B cell population, with
494 increase in the population of B cell memory formation. The subsequent increase in cytokine
495 levels was also induced by the vaccine [**Figure 8b**]. Similar elevated responses were observed
496 for the CD4⁺ and CD8⁺ cells population with a significant increase in memory formation [**Figure**
497 **8c-8f**]. The concomitant increase in dendritic cells and natural killer cells were another good
498 immune response attribute exhibited by the vaccine [**Figure 8g-8j**].

499

500 **DISCUSSION**

501 The emergence of new coronavirus strain SARS-CoV-2 viral diseases is a global threat,
502 responsible for the death of many across the globe, including health care workers [1]. Therefore,
503 there is an urgent need for therapeutics and preventive measures that could confer protection
504 against this enigma. Our study was therefore centered on using an epitope peptide-based vaccine
505 design against the SARS-CoV-2 spike protein complex. We successfully developed a peptide
506 vaccine after a rigorous round of screenings and conditions in selecting the epitopes, using an
507 array of immune-informatics tools. Criteria such as the elicitation of immune response with
508 lesser or no potential infectious abilities are taken into consideration before each epitope
509 selection, using stipulated thresholds. Both the B cell and T cell epitopes were predicted in this
510 study. B cells recognize solvent-exposed antigens through antigen receptors, named as B cell
511 receptors (BCR), consisting of membrane-bound immunoglobulins [38]. The B cell epitopes
512 were selected based on surface accessibility, and Kolaskar and Tongaonkar antigenicity scale
513 methods. Five antigenic B cell epitopes were predicted in the study. The peptide also has a non-
514 toxicity attribute, making it a safe vaccine candidate.

515 The T cell antigenic epitopes capable of binding a large number of MHC I and MHC II alleles
516 were predicted using various tools, and selections were made based on the recommended
517 thresholds. T cell epitopes presented by MHC class I molecules are typically peptides between 8
518 and 11 amino acids in length, whereas MHC class II molecules present longer peptides, 13-17
519 amino acids in length [39]. The CD8⁺ T cell recognizes the antigen of a pathogen after its
520 attachment with the MHC I molecules, therefore triggering a cytotoxic response against the
521 pathogen [40]. Eight promising CD8⁺ T cell epitopes were predicted. These peptides are capable
522 of eliciting a cytotoxic response with their respective antigenic properties.

523 Twelve helper T cell epitopes were predicted based on their virtual attachments with the beta
524 chain of antigen-presenting major histocompatibility complex class II (MHC II) molecule. In
525 complex with the beta chain HLA-DR, the T cell epitopes display antigenic peptides on
526 professional antigen-presenting cells (APCs) for recognition by alpha-beta T cell receptor (TCR)
527 on HLA-DRB1-restricted CD4-positive T cells [40, 41]. This guides antigen-specific T helper
528 effector functions, both antibody-mediated immune response and macrophage activation, to
529 ultimately eliminate the infectious agents and transformed cells [40]. The 12 selected epitopes
530 were highly antigenic, with each capable of inducing any of the Th1 and Th2 cytokines, which
531 corroborates the findings that T cell CD4+ is capable of inducing an adaptive immune response
532 in the human cells [41], as adaptive immunity is articulated by lymphocytes, more specifically by
533 B- and T-cells, which are responsible for the humoral and cell-mediated immunity [41].
534 Prediction of peptide binding to MHC II molecules readily discriminate CD4 T-cell epitopes, but
535 cannot tell their ability to activate the response of specific CD4 T-cell subsets (e.g., Th1, Th2,
536 and Treg). However, there is evidence that some CD4 T-cell epitopes appear to stimulate specific
537 subsets of Th cells [40, 41]. The ability to distinguish the epitopes capable of inducing distinct
538 responses is highly imperative in vaccine development. This study was able to differentiate the
539 abilities of each MHC class II epitopes of eliciting either the Th1 and Th2 cytokines or both.

540 The docking scores involving the predicted epitopes and the MHC II molecules were
541 comparatively evaluated. Considering the docking attributes of the MHC II epitopes, they
542 displayed a varied binding putative attribute. Some of the epitopes core peptides had a flanking
543 region away from the MHC class II binding groove. Generally, MHC-II peptides contain a
544 central binding motif of nine core amino residues that specifically attach to the MHC II binding

545 groove. These core peptides interact with the allelic specific pockets of the MHCII binding
546 groove.

547 Considering the population cover for the MHC-I and II epitopes, the European population
548 significantly show a potential response to the selected epitopes, however, we suggest that the
549 IEDB population coverage tool have less MHC class I and II deposition of alleles from
550 continents like Africa and Asia, compared to the European and Americans. Immunization of the
551 MHCII T cell epitopes will confer protection to 80.88% of the world population, while MHCII T
552 cell epitopes will confer protection to 55.23%. It is imperative to know that specific interactions
553 with high binding affinity epitope / HLA allele class II molecule unleash protective and specific
554 adaptive immune response [39-42].

555 The designed vaccine construct was predicted to be stable, soluble (i.e. hydrophilic) and with
556 increased thermostability, as depicted in its physicochemical characteristics. The molecular
557 weight of the vaccine and its high pI value signifies the efficacy as well as the stability of the
558 vaccine construct since proteins having <110kD molecular weight are considered good vaccine
559 candidates [43, 44]. Apart from size, surface properties like surface charge and hydrophobicity
560 can affect a designed vaccine candidate. Neutral or negatively charged molecules are preferred
561 and a balance between its hydrophobicity and hydrophilicity is crucial in designing vaccine
562 candidates [45].

563 **Conclusion**

564 This is a novel approach to predicting SARS-CoV-2 epitope peptide-based vaccine targeting the
565 spike protein, utilizing immune-informatics tools and immune simulation measures. These
566 predicted antigenic epitopes would hasten the production of protective vaccine for patients

567 around the world whose immune system has been compromised. Our selected epitopes (B and T
568 cells) will constitute a good vaccine candidate against the spike protein. In future studies, other
569 effective stimulants that could aid the rapid response of cells to antigens will be considered and
570 assessed.

571 **Acknowledgement**

572 We sincerely appreciate the effort of Sblend Digital overseen by Mr. Agboola Olamide, for
573 providing some of the sophisticated computing accessories and work space to conduct this
574 research.

575 **Data Availability**

576 Data are available upon request and may be obtained by contacting the corresponding author

577 **Conflict of Interest**

578 None

579 **Funding statement**

580 This research did not receive any specific grant from funding agencies in the public, commercial,
581 or not-for-profit sectors

582

583 **REFERENCES**

- 584 1. World Health Organization (WHO). Statement Regarding Cluster of Pneumonia Cases in
585 Wuhan, China. www.who.int. (2020)
- 586 2. Cohen J. Wuhan seafood market may not be source of novel virus spreading globally. Science
587 Mag American Association for the Advancement of Science. AAAS; 2020.
- 588 3. Center for Disease Control (CDC). Archived from the original on 2020-01-20.
589 Retrieved 2020-01-23. www.cdc.gov (2020).
- 590 4. Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV
591 epidemic threat of novel coronaviruses to global health - The latest 2019 novel
592 coronavirus outbreak in Wuhan, China. Int J Infect Dis. 2020; 91:
593 2646. <https://doi.org/10.1016/j.ijid.2020.01.009> PMID:31953166

594 5. Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering
595 (CSSE) at Johns Hopkins University. (JHU). ArcGIS. Johns Hopkins CSSE; 2020.

596 6. Cohen J. Wuhan seafood market may not be source of novel virus spreading
597 globally. ScienceMag American Association for the Advancement of Science. AAAS;
598 2020.

599 7. Eschner K. We're still not sure where the COVID-19 really came from. Popular Science;
600 2020.

601 8. Sexton NR, Smith EC, Blanc H, Vignuzzi M, Peersen OB, Denison MR. Homology-based
602 identification of a mutation in the coronavirus RNA-dependent RNA polymerase that
603 confers resistance to multiple mutagens. *J Virol.* 2016;90(16):7415–28.
604 <https://doi.org/10.1128/JVI.00080-16> PMID:27279608

605 9. Simmons G, Zmora P, Gierer S, Heurich A, Pöhlmann S. Proteolytic activation of the SARS-
606 coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research.
607 *Antiviral Res.* 2013;100(3):605–
608 14. <https://doi.org/10.1016/j.antiviral.2013.09.028> PMID:24121034

609 10. Fehr AR, Perlman S. Maier, H. J., Bickerton, E., Britton P (eds.). *Coronaviruses: an*
610 *overview of their replication and pathogenesis. Methods in Molecular Biology.*
611 Springer. **1282**, 1–23 (2015).

612 11. McIntosh K, Arber W, Haas R, Henle W, Hofschneider PH, Jerne NK, Koldovský P,
613 Koprowski H, Maaløe O, Rott R (eds.). *Coronaviruses: A Comparative Review. Current*
614 *Topics in Microbiology and Immunology / Ergebnisse der Mikrobiologie und*
615 *Immunitätsforschung. Current Topics in Microbiology and Immunology / Ergebnisse der*
616 *Mikrobiologie und Immunitätsforschung.* Berlin, Heidelberg: Springer: 87. (1974).

617 12. Kahn JS, McIntosh K. History and recent advances in coronavirus discovery. *Pediatr Infect*
618 *Dis J.* 2005;24(11 Suppl):S223
619 <https://doi.org/10.1097/01.inf.0000188166.17324.60> PMID:16378050

620 13. Geller C, Varbanov M, Duval RE. Human coronaviruses: insights into environmental
621 resistance and its influence on the development of new antiseptic strategies. *Viruses.*
622 2012;4(11):3044–68. <https://doi.org/10.3390/v4113044> PMID:23202515

623 14. Almeida JD, Berry DM, Cunningham CH, Hamre D, Hofstad MS, Mallucci L, et al.
624 *Virology: coronaviruses.* *Nature.* 1968;220(5168):650. <https://doi.org/10.1038/220650b0>.

625 15. Khan MK, Zaman S, Chakraborty S, Chakravorty R, Alam MM, Bhuiyan TR, et al. *In*
626 *silico* predicted mycobacterial epitope elicits in vitro T-cell responses. *Mol Immunol.*
627 2014;61(1):16–22. <https://doi.org/10.1016/j.molimm.2014.04.009> PMID:24853589

628 16. Naz RK, Dabir P. Peptide vaccines against cancer, infectious diseases, and conception. *Front*
629 *Biosci.* 2007;12(12):1833–44. <https://doi.org/10.2741/2191> PMID:17127424

630 17. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. *Open Biol.*
631 2013;3(1):120139. <https://doi.org/10.1098/rsob.120139> PMID:23303307

632 18. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch
633 A. Protein Identification and Analysis Tools on the ExPASy
634 Server; <https://doi.org/10.1385/1-59259-890-0:571>.

635 19. Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, et al. The immune epitope database
636 and analysis resource: from vision to blueprint. *PLoS Biol.* 2005;3(3):e91.
637 <https://doi.org/10.1371/journal.pbio.0030091> PMID:15760272

638 20. Irini A. Doytchinova, and Darren, R Flower. Bioinformatic Approach for Identifying Parasite
639 and Fungal Candidate Subunit Vaccines. *Open Vaccine J.* 2008;1(1):22–6.
640 <https://doi.org/10.2174/1875035400801010022>.

641 21. DeLano, W. L. Pymol: an open-source molecular graphics tool. *CCP4 Newslett Prot*
642 *Crystallogr.* 40, 82–92 (2002).

643 22. Jeannin P, Magistrelli G, Goetsch L, Haeuw JF, Thieblemont N, Bonnefoy JY, et al. Outer
644 membrane protein A (OmpA): a new pathogen-associated molecular pattern that interacts
645 with antigen presenting cells-impact on vaccine strategies. *Vaccine.* 2002;20(4 Suppl 4):
646 A23–7. [https://doi.org/10.1016/S0264-410X\(02\)00383-3](https://doi.org/10.1016/S0264-410X(02)00383-3) PMID:12477424

647 23. Malhotra A. Tagging for protein expression. *Methods Enzymol.* 2009; 463:239–58.
648 [https://doi.org/10.1016/S0076-6879\(09\)63016-0](https://doi.org/10.1016/S0076-6879(09)63016-0) PMID:19892176

649 24. Sormanni P, Aprile FA, Vendruscolo M. The CamSol method of rational design of protein
650 mutants with enhanced solubility. *J Mol Biol.* 2015;427(2):478–90.
651 <https://doi.org/10.1016/j.jmb.2014.09.026> PMID:25451785

652 25. Maupetit J, Derreumaux P, Tuffery P. PEP-FOLD: an online resource for de novo peptide
653 structure prediction. *Nucleic Acids Res.* 2009;37(Web Server issue):W498–503.
654 <https://doi.org/10.1093/nar/gkp323> PMID:19433514

655 26. Shen Y, Maupetit J, Derreumaux P, Tufféry P. Improved PEP-FOLD approach for peptide
656 and miniprotein structure prediction. *J. J Chem Theory Comput.* 2014;10(10):4745–58.
657 <https://doi.org/10.1021/ct500592m> PMID:26588162

658 27. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: an updated
659 de novo structure prediction server for both linear and disulfide bonded cyclic peptides.
660 *Nucleic Acids Res.* 2012;40(Web Server issue):W288–93.
661 <https://doi.org/10.1093/nar/gks419> PMID:22581768

662 28. Daniel WA, Buchan, David T Jones. The PSIPRED Protein Analysis Workbench: 20 years
663 on. *Nucleic Acids Res.* 2019;47(1):402–7.

664 29. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK - a program to check
665 the stereochemical quality of protein structures. *J Appl Cryst.* 1993;26(2):283–91.
666 <https://doi.org/10.1107/S0021889892009944>.

667 30. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography
668 of proteins. *Nucleic Acids Res.* 2018;46 W1:W363–7.
669 <https://doi.org/10.1093/nar/gky473> PMID:29860391

670 31. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.*
671 1998;7(9):1884–97. <https://doi.org/10.1002/pro.5560070905> PMID:9761470

673 32. Velankar S, Dana JM, Jacobsen J, van Ginkel G, Gane PJ, Luo J, et al. SIFTS: Structure
674 Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Res.*
675 2013;41(Database issue):D483–9. PMID:23203869

676 33. Zhou P, Jin B, Li H, Huang SY. HPEPDOCK: a web server for blind peptide-protein docking
677 based on a hierarchical algorithm. *Nucleic Acids Res.* 2018;46 W1:W443–50.
678 <https://doi.org/10.1093/nar/gky357> PMID:29746661

679 34. Pierce BG, Hourai Y, Weng Z. Accelerating protein docking in ZDOCK using an advanced
680 3D convolution library. *PLoS One.* 2011;6(9): e24657.
681 <https://doi.org/10.1371/journal.pone.0024657> PMID:21949741

682 35. Christopher JK, Jochen S. Hub WAXSiS: a web server for the calculation of SAXS/WAXS
683 curves based on explicit-solvent molecular dynamics. *Nucleic Acids Res.* 2015;43
684 W1:225–30. <https://doi.org/10.1093/nar/gkv309>.

685 36. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic
686 fluctuations in molecular dynamics and normal mode simulations. *Proteins.*
687 1991;11(3):205–17. <https://doi.org/10.1002/prot.340110305> PMID:1749773

688 37. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets
689 bioinformatics: the use of prediction tools for molecular binding in the simulation of the
690 immune system [PLoS]. *PLoS One.* 2010;5(4):e9862.
691 <https://doi.org/10.1371/journal.pone.0009862> PMID:20419125

692 38. Sun P, Ju H, Liu Z, Ning Q, Zhang J, Zhao X, et al. Bioinformatics resources and tools for
693 conformational B-cell epitope prediction. *Comput Math Methods Med.*
694 2013;2013:943636. <https://doi.org/10.1155/2013/943636> PMID:23970944

695 39. Yuri O. Poluektov, AeRyon Kim, and Scheherazade Sadegh-Nasseri. HLA-DO and Its Role
696 in MHC Class II Antigen Presentation. *Front Immunol*

697 40. Koretzky GA. Multiple roles of CD4 and CD8 in T cell activation. *J Immunol.* 2010
698 Sep;185(5):2643–4. <https://doi.org/10.4049/jimmunol.1090076> PMID:20724729

699 41. Isra K., Ibtihal, O., Islam, Z., Abdalgadir, F., Hanaa, A., Hajr, A. et al. Design of an epitope-
700 based peptide vaccine against *Cryptococcus neoformans*. *bioRxiv*. doi:
701 <https://doi.org/10.1101/434779>

702 42. Jose L. Sanchez-Trincado, Marta Gomez-Perosanz, and Pedro A. Reche. Peptide-Based
703 Immunotherapeutics and Vaccines. *J Immunol Res.* 2017 ().
704 <https://doi.org/10.1155/2017/2680160>

705 43. Naz A, et al. (2015) "Identification of putative vaccine candidates against *Helicobacter pylori*
706 exploiting exoproteome and secretome: A reverse vaccinology-based
707 approach," *Infection, Genetics and Evolution*. Elsevier, 32, pp. 280–291. doi:
708 [10.1016/j.meegid.2015.03.027](https://doi.org/10.1016/j.meegid.2015.03.027)

709 44. Dar HA, Zaheer T, Shehroz M, Ullah N, Naz K, Muhammad SA, et al. "Immuno-
710 informatics-aided design and evaluation of a potential multi-epitope vaccine against
711 *klebsiella pneumoniae," Vaccines. Vaccines (Basel).* 2019;7(3):E88.
712 <https://doi.org/10.3390/vaccines7030088> PMID:31409021

713 45. Liu H, Irvine DJ. Guiding principles in the design of molecular bioconjugates for vaccine
714 applications. *Bioconjug Chem.* 2015;26(5):791–801.
715 <https://doi.org/10.1021/acs.bioconjchem.5b00103> PMID:25822926

716 46. López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. iMODS: internal coordinates
717 normal mode analysis server. *Nucleic Acids Res.* 2014;42(Web Server issue):W271-6.
718 <https://doi.org/10.1093/nar/gku339> PMID:24771341

719

720

721

722

723

724

725 **Figure Legends**

726 **Figure 1a:** SARS-CoV-2 crystal spike glycoprotein showing the mapped-out B cell linear
727 epitopes. The linear epitopes are highlighted in red.

728 **Figure 1b:** Site of the predicted discontinuous B cell epitopes overlaid on the crystal structure of
729 SARS-CoV-2 envelope S protein. The discontinuous epitopes are shown in clusters of red
730 spheres.

731 **Figure 2:** Selected continents and countries' coverage of the MHC class I and II T cell epitopes.

732 **Figure 3a-e:** Differential binding of the HTL epitopes with HLA-DRB1*01:01. The MHC
733 protein is displayed in surface brown and the epitopes are highlighted in yellow licorice.

734 **Figure 4a.** Schematic presentation of the vaccine containing an adjuvant (green) linked with the
735 multi-epitope sequence through an EAAAK linker. The B and HTL epitopes are linked together
736 via the GPGPG linkers while the CTL epitopes are linked with the help of AAY linkers. The 6x-
737 His tag at the carboxyl end. **4b:** Tertiary structure of the vaccine. **4c:** Validation of the structure
738 with a Z score of -2.32. **4d:** Intrinsic solubility profile. Residues lesser than -1 depicts the
739 hydrophobic core of the vaccine peptide.

740 **Figure 4e:** The solute and solvation layer of the vaccine

741

742 **Figure 5:** The molecular interaction of the vaccine and TLR5 receptor. The vaccine chain is
743 highlighted in red and the toll like receptor in blue.

744 **Figure 6a:** The radius of gyration of the peptide vaccine. **6b-6f:** Molecular dynamics simulation
745 of the vaccine-TLR5 complex, showing (a) eigenvalue; (b) deformability; (c) B-factor; (d)
746 Covariance matrix indicates coupling between pairs of residues (red), uncorrelated (white) or
747 anti-correlated (blue) motions. and (e) elastic network analysis which defines which pairs of
748 atoms are connected by springs.

749

750 **Figure 7:** *In silico* cloning of the final vaccine construct into pET28a (+) expression vector
751 where the red part indicates the coding gene for the vaccine surrounded between EagI-NotI (166)
752 and SAI (1838) while the vector backbone has shown in a black circle. MCS represents the
753 multiple cloning site.

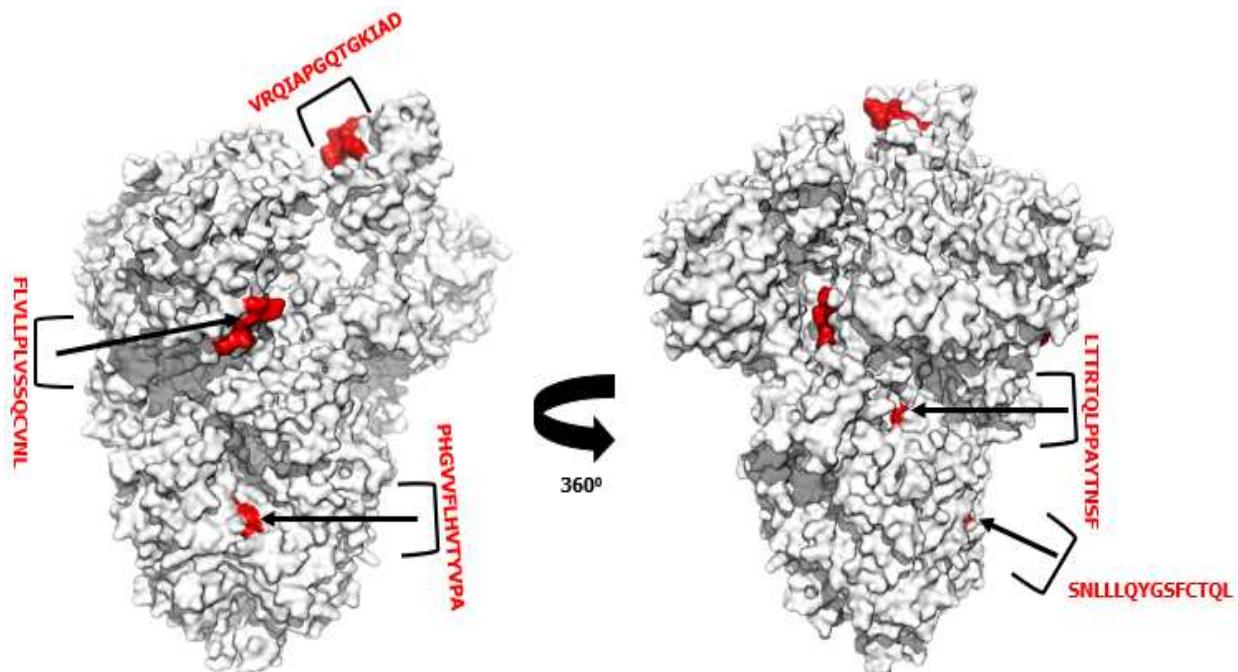
754 **Figure 8a:** Increase in immunoglobulin activities and concomitant decrease in antigen level in 365
755 days of vaccination. Antibodies are sub-divided per isotype.

756 **Figure 8b:** Concentration of cytokines and interleukins. D in the inset plot is danger signal

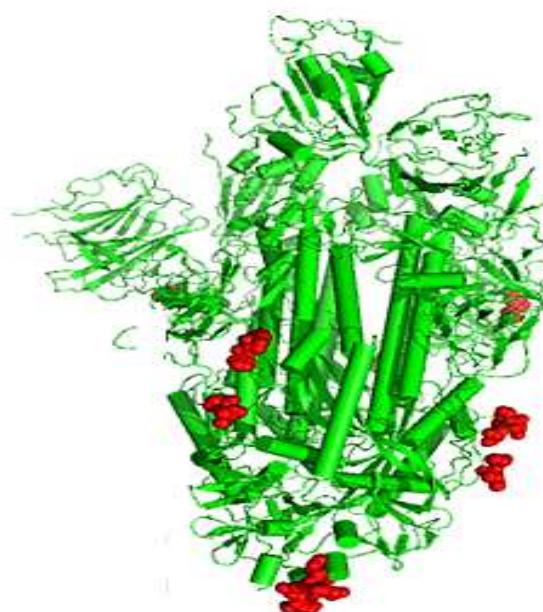
757 **Figure 8c-8i:** **c.** B lymphocytes: total count, memory cells, and sub-divided in isotypes IgM,
758 IgG1 and IgG2. **d.** CD4 T-helper lymphocytes count. **e.** CD4 T-helper lymphocytes count. **f.**
759 CD8 T-cytotoxic lymphocytes count. **g.** CD8 T-cytotoxic lymphocytes count per entity-state. **h.**
760 Natural Killer cells (total count). **i.** Dendritic cells.

761

762 **Figures**



763
764
765
Figure 1a



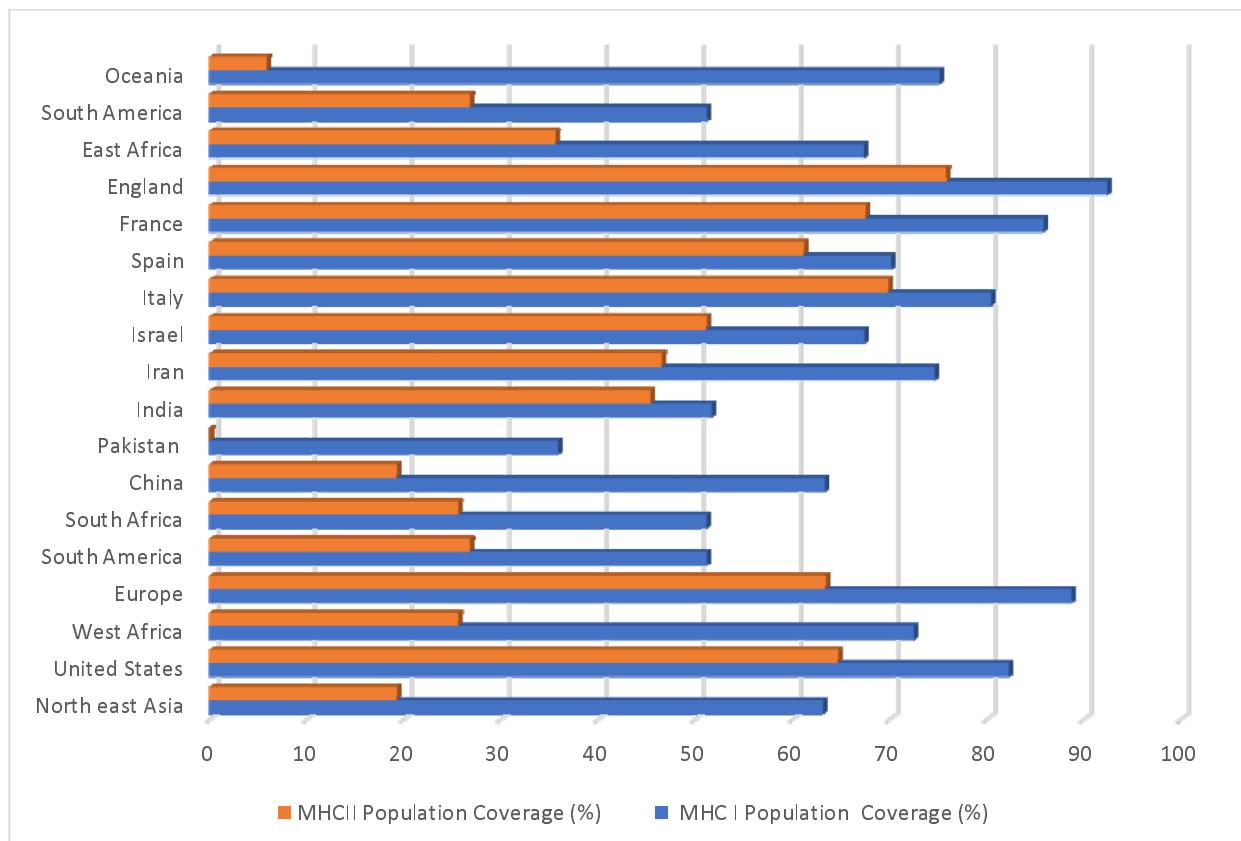
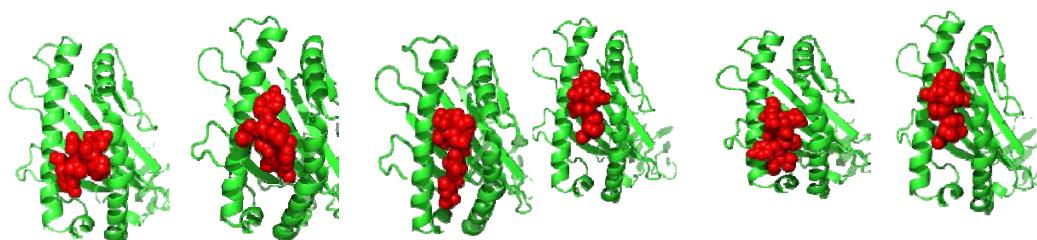
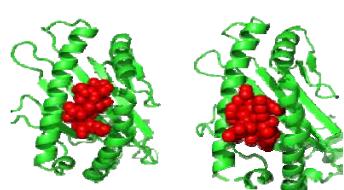


Figure 2

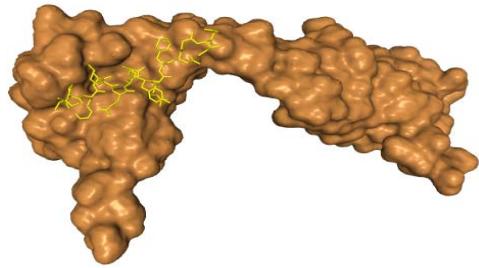


3ai 3aii 3aiii 3aiv 3avi 3av

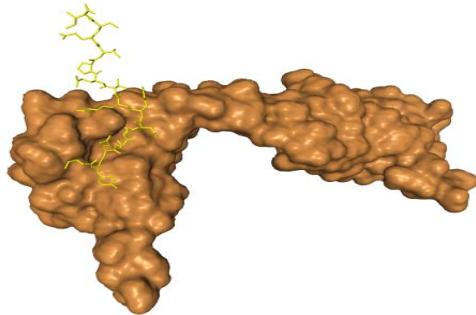


3aviii 3aix

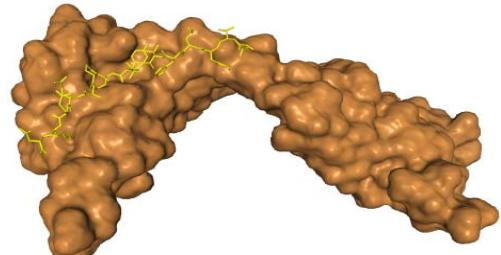
772



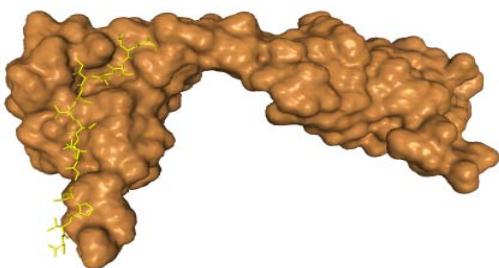
3bii



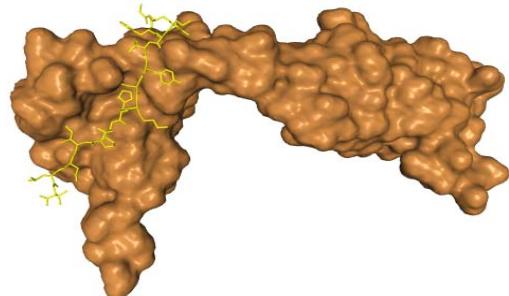
3bii



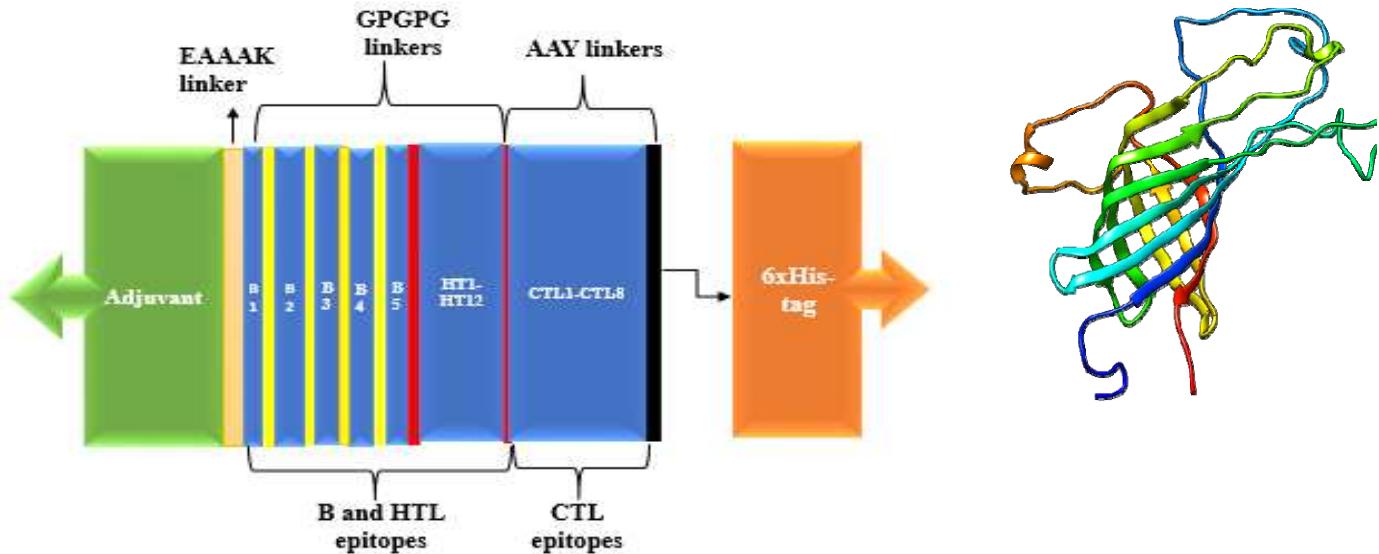
3biii



3biv

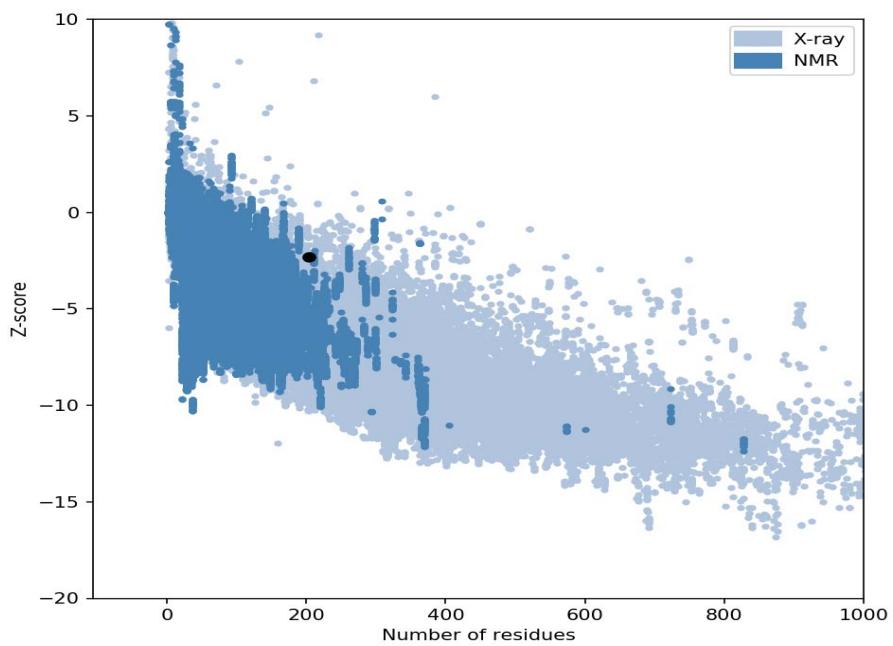


3bvi



4b

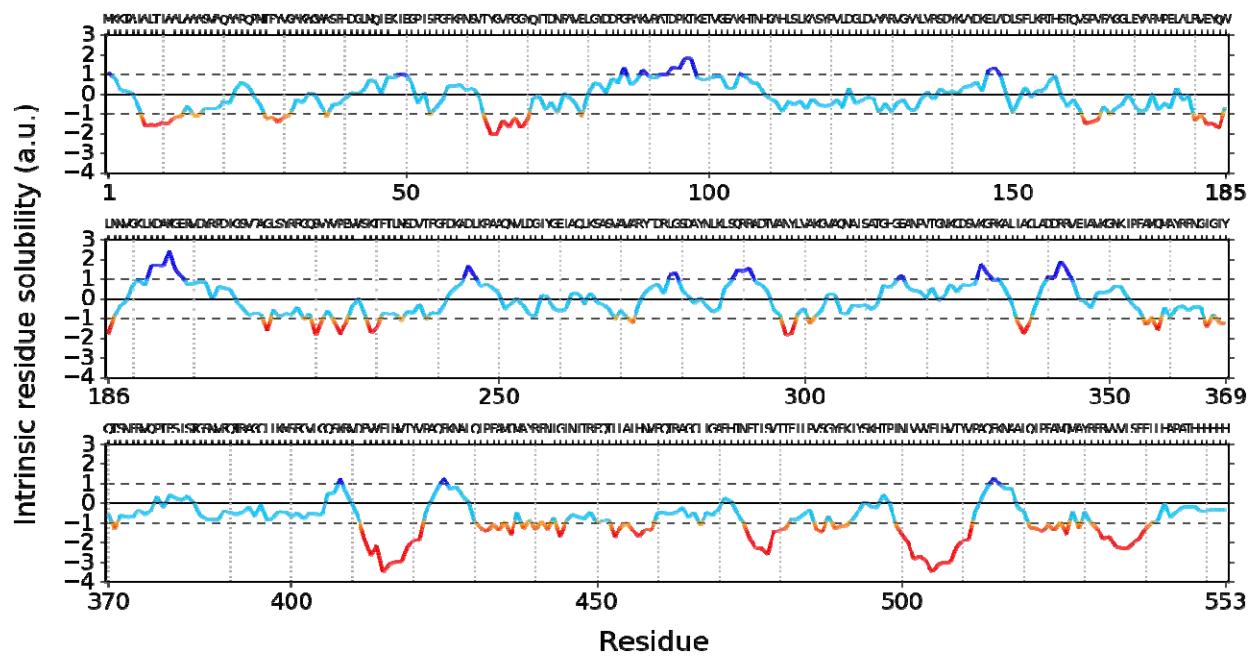
4a



774

4c

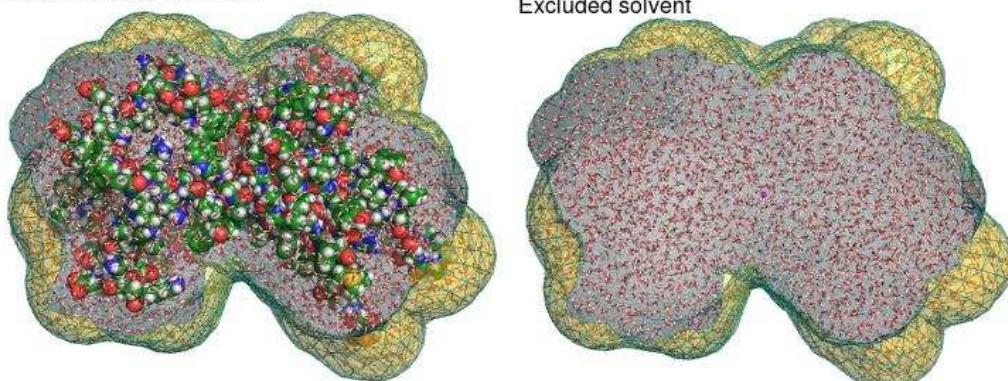
775



776

4d

Solute and solvation layer



778

4e

779

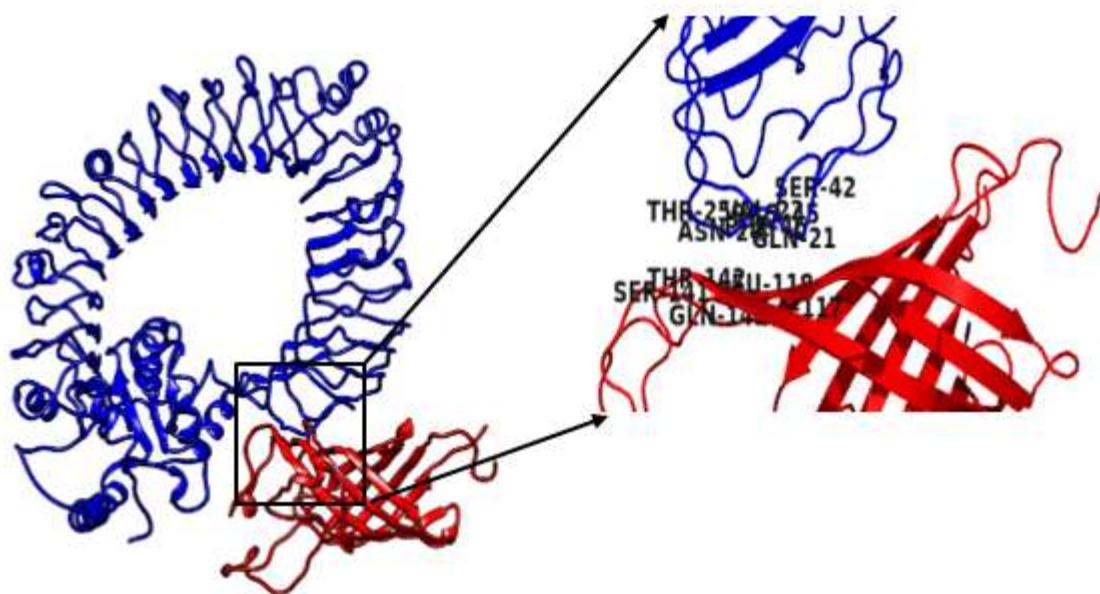
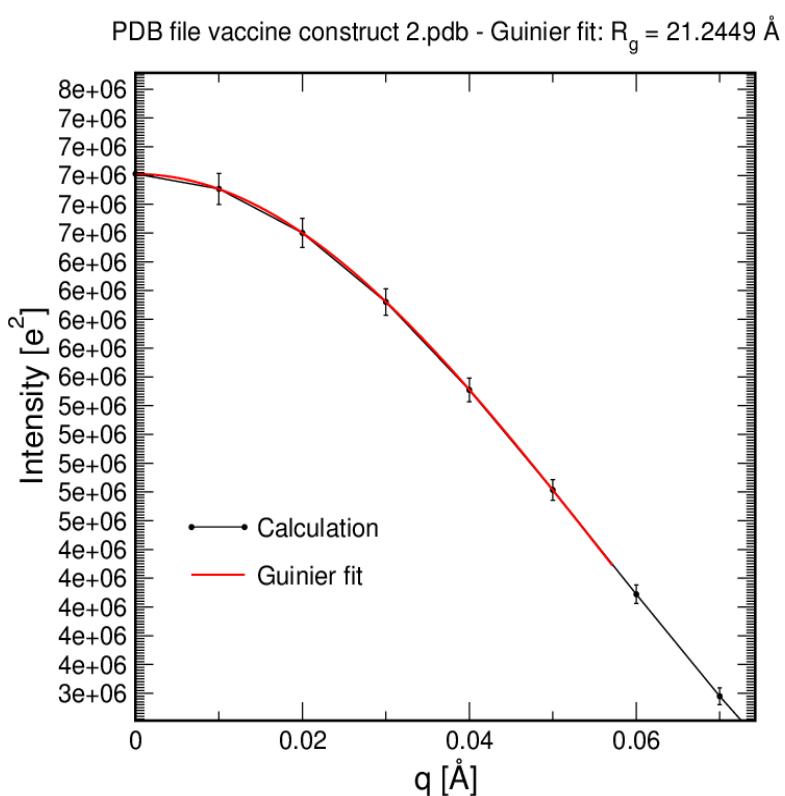


Figure 5

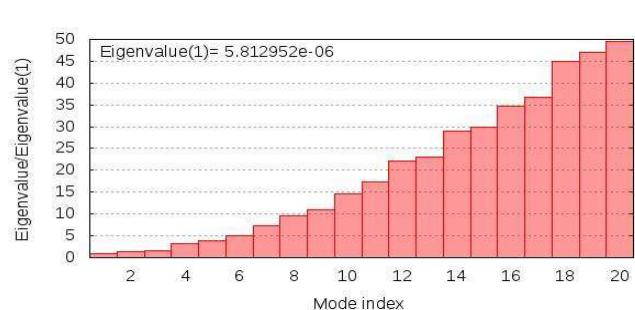
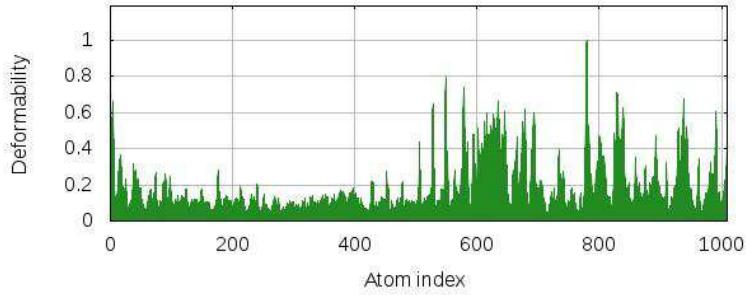
780
781
782



783

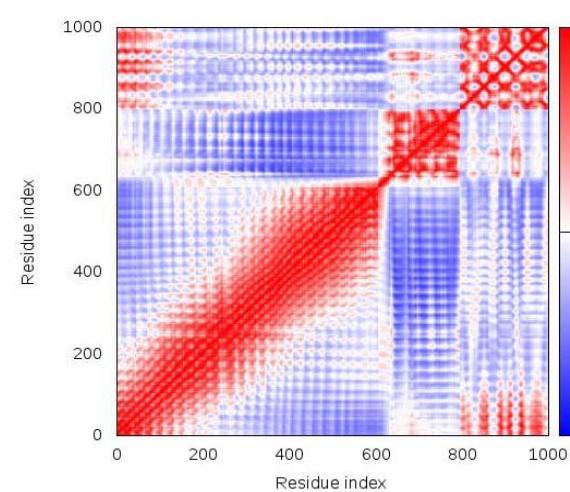
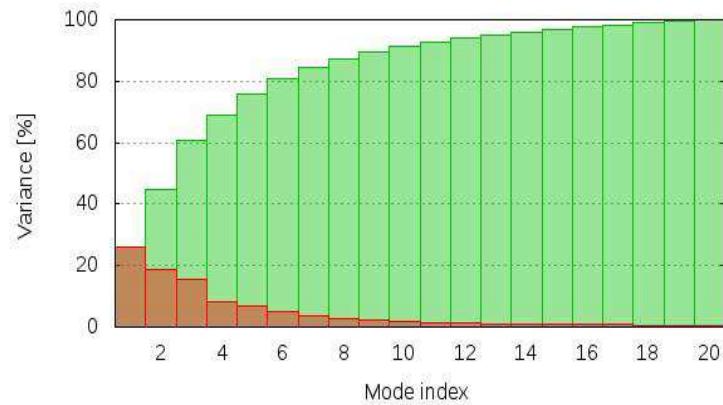
784

6a



6b

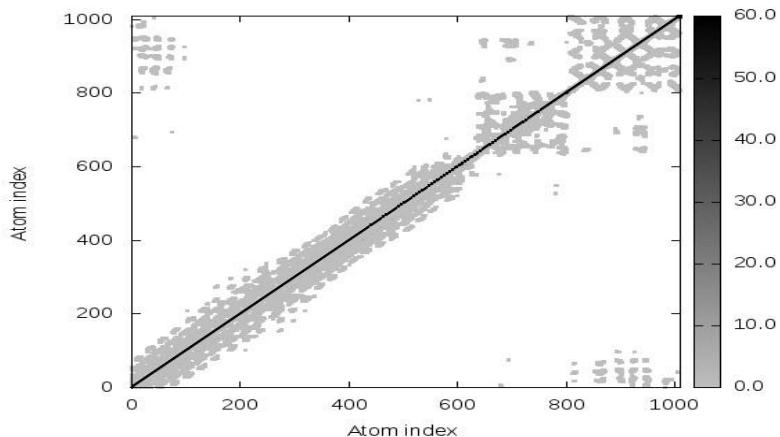
6c



6d

6e

785



786

787

6f.

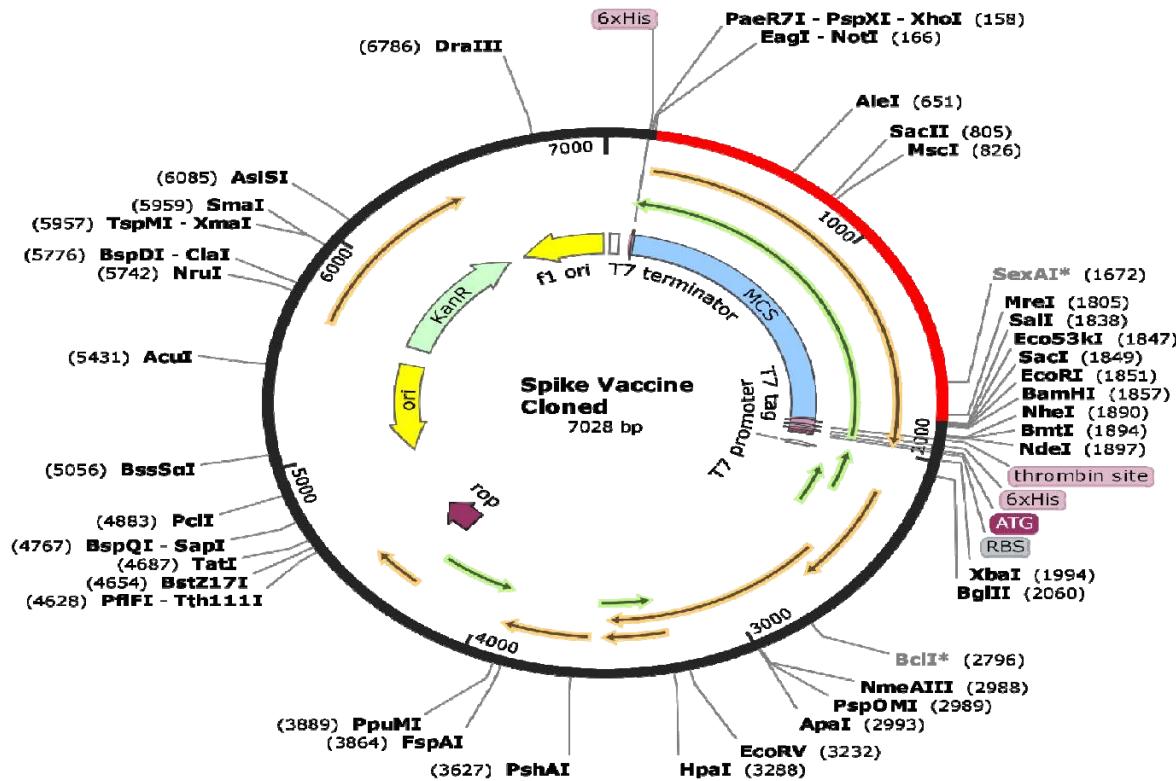
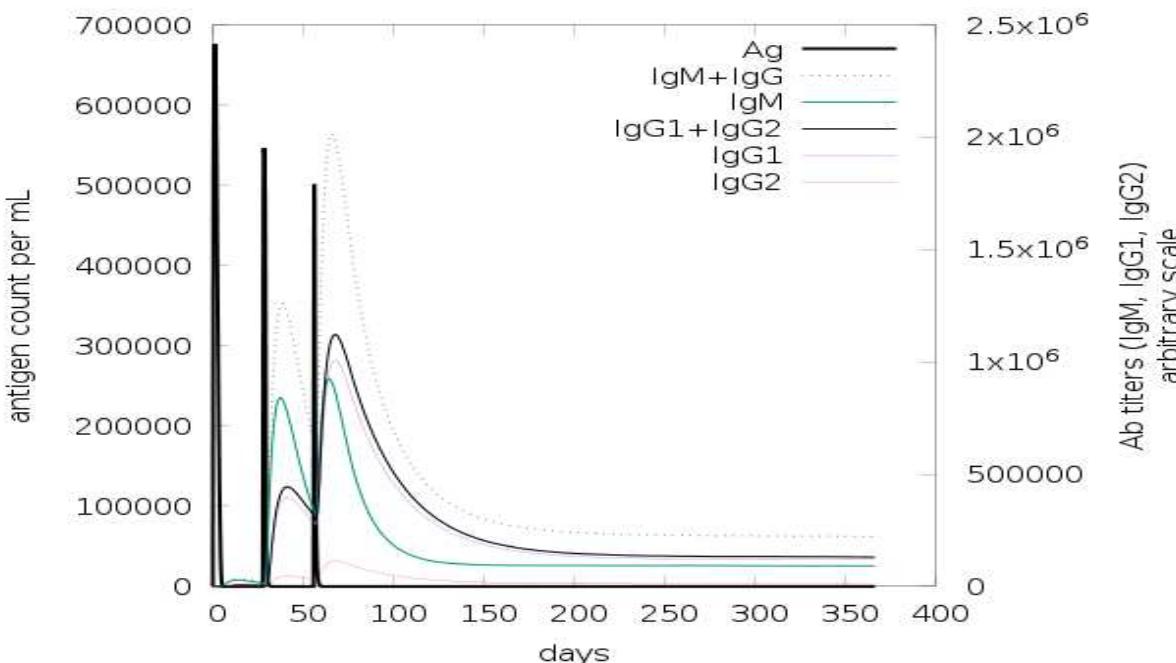
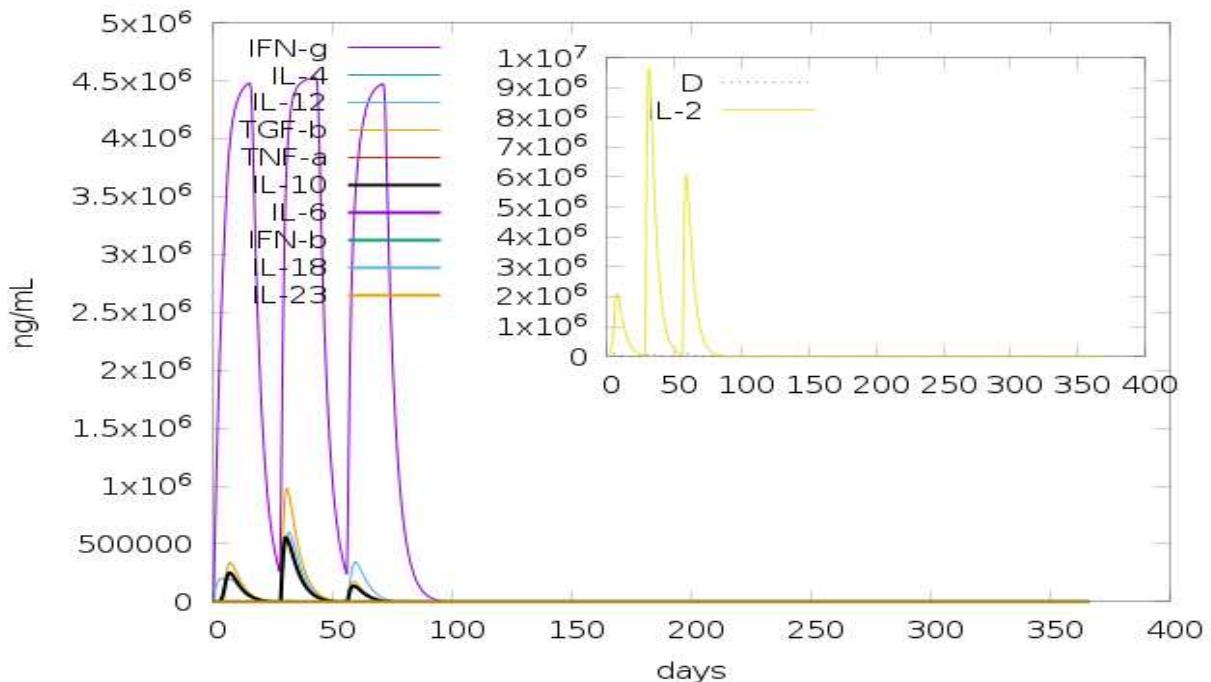


Figure 7



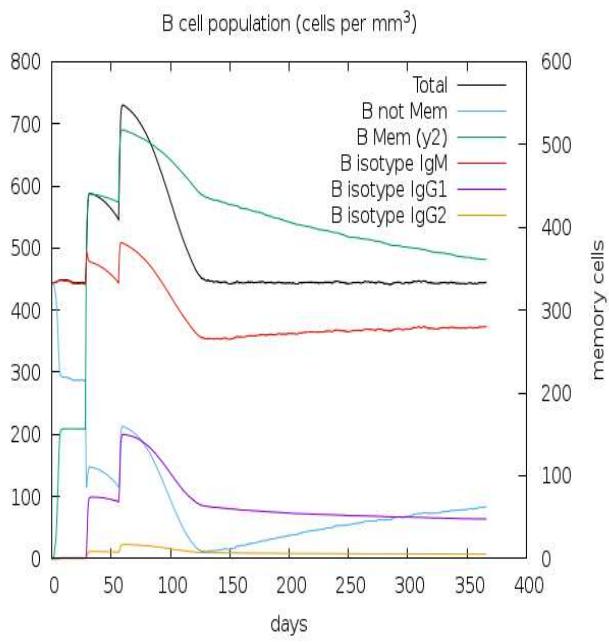
8a



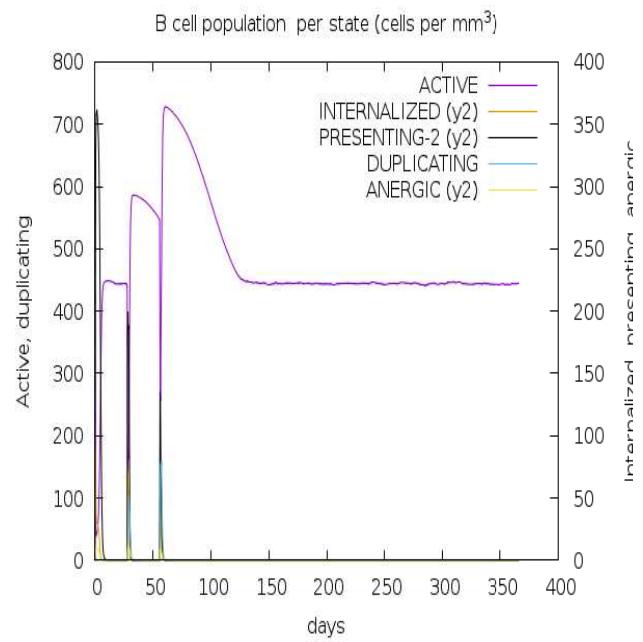
793

794

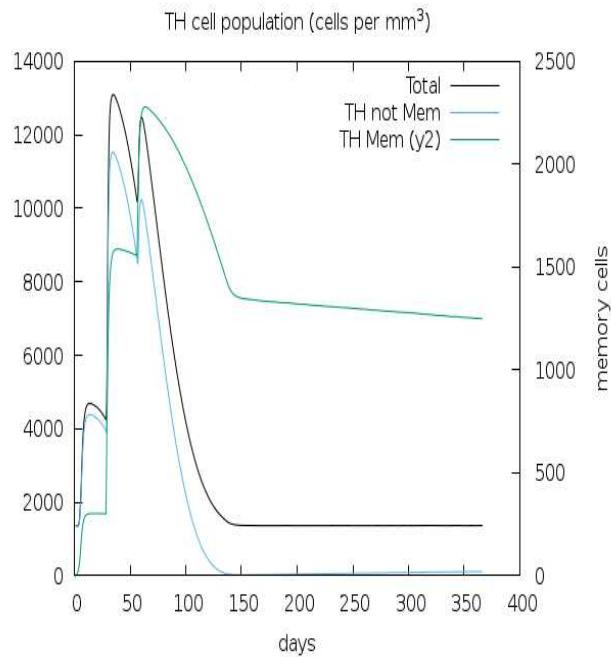
8b



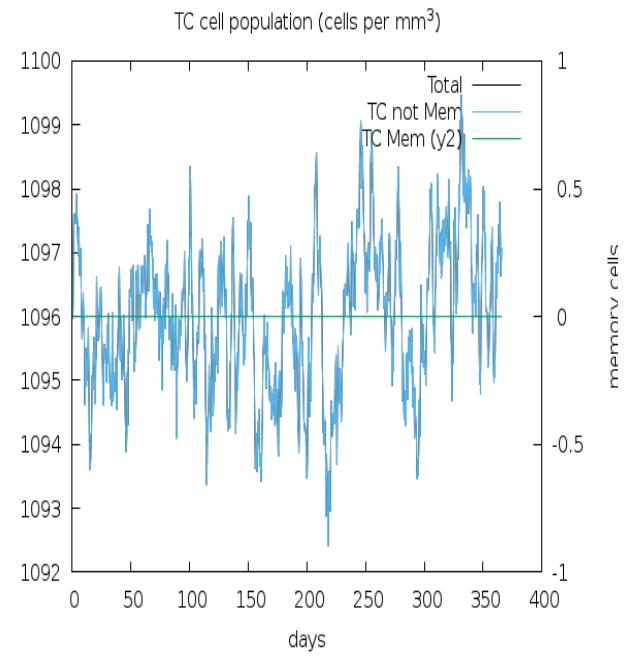
8c



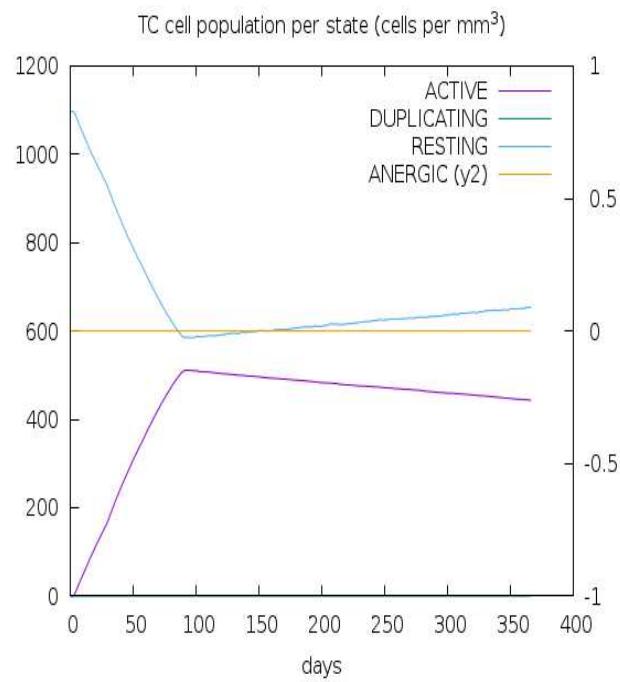
8d



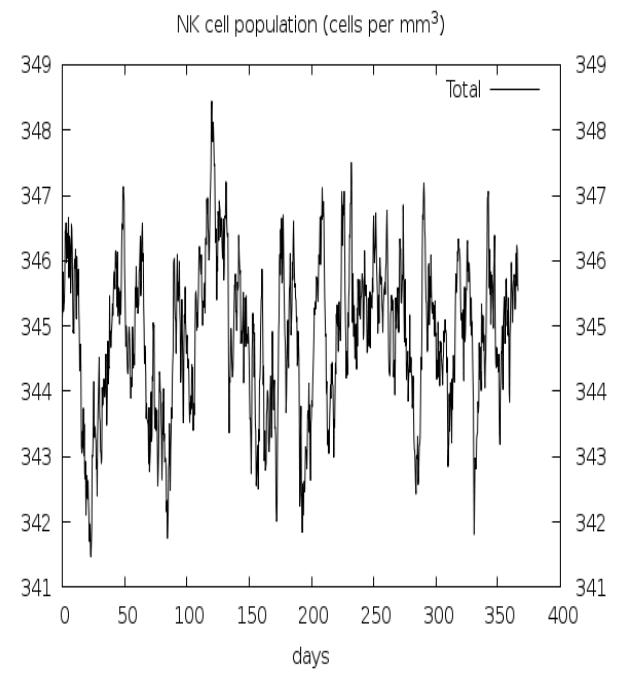
8e.



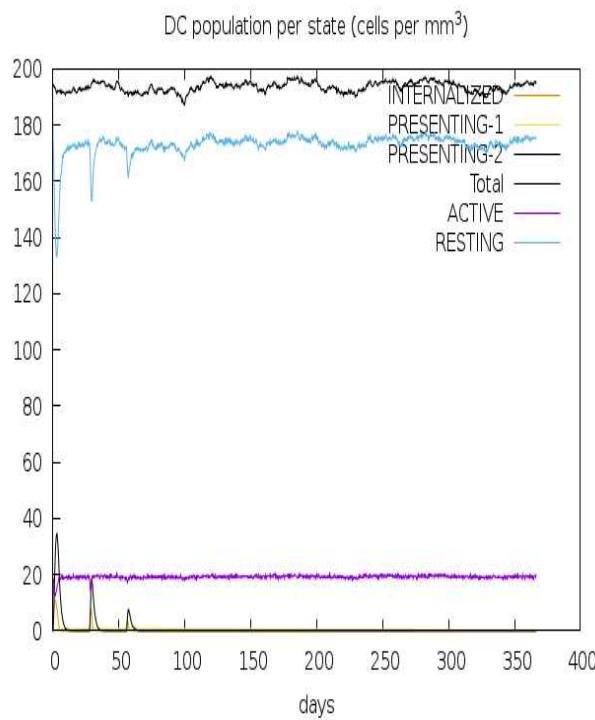
8f.



8g.



8h.



8i.

795

796

797

798

799

800

801

802

803

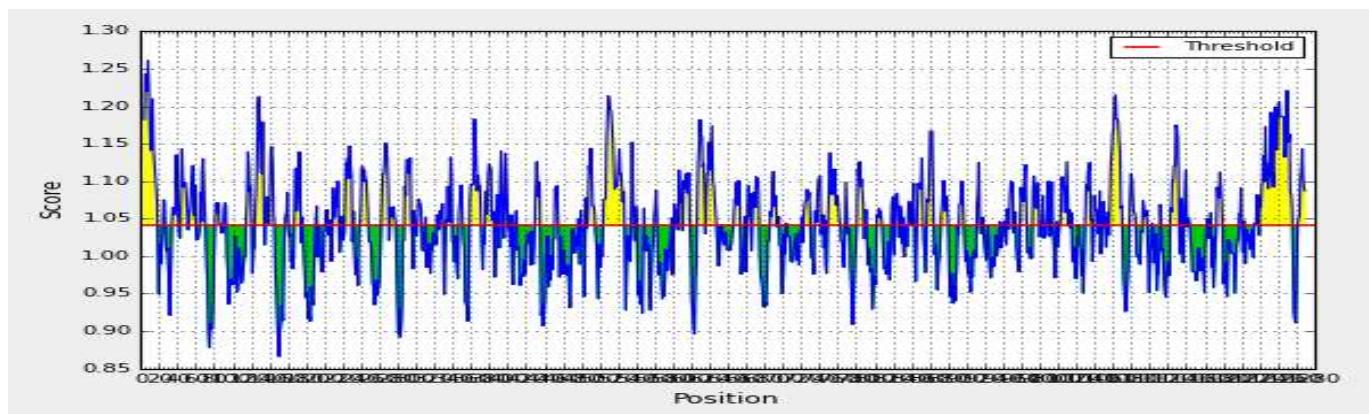
804

805

806

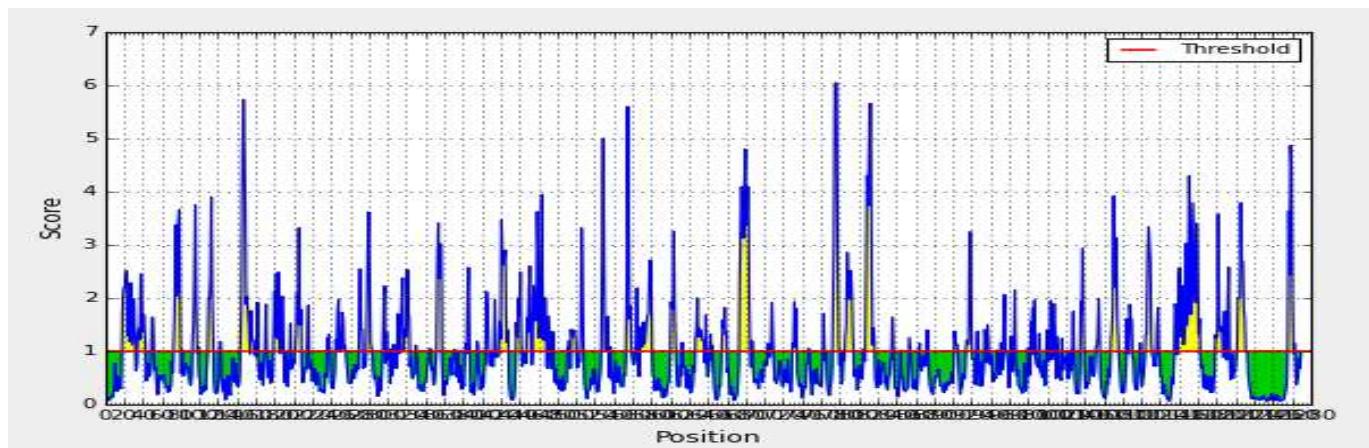
807

808 **Supplementary Data**



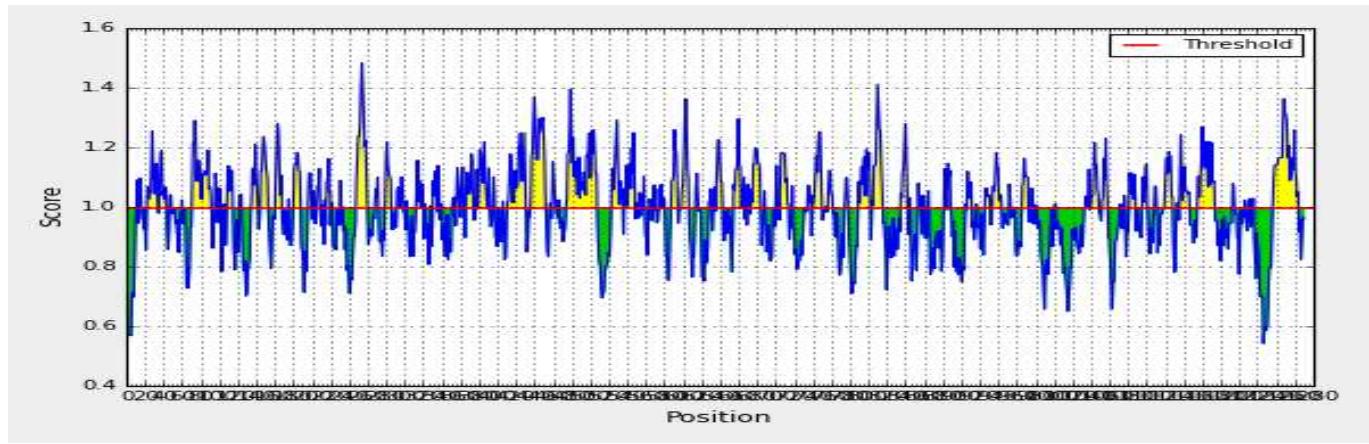
809

810 **a**



811

812 **b**



813

814 **c**

815 **Figure S1a-S1c:** Prediction of antigenic determinants. **a.** Kolaskar and Tongaonkar antigenicity scale. **b.** Emini
816 surface accessibility **c.** Chou and Fasman beta turns. Green regions under the threshold color denotes unfavourable

817 related to the properties of interest. Yellow colours are above the threshold sharing higher scores. Horizontal red
818 lines represent the threshold.

819

820 **Physiochemical analysis of the SARS-CoV-2 Spike Glycoprotein**

821 The sequence of the primary structure of SARS-CoV-2 spike glycoprotein was computed,
822 analyzed and tabulated [Table 1]. The molecular weight was estimated at 141178.47 Da. To
823 calculate the extinction coefficient, wavelengths of varying amount (276, 278, 279, 280 and 282
824 nm) were computed. But wavelength at 280nm is usually favored because of high protein
825 absorption. So, the extinction coefficient at 280nm is $148960 \text{ M}^{-1}\text{cm}^{-1}$ with respect to the
826 cysteine, trypsin and tyrosine concentrations. The spike glycoprotein is highly stable as the
827 instability index was 33.01, because protein instability index at or above 40 is considered not
828 stable. The isoelectric point or computed theoretical pI of the spike glycoprotein was acidic,
829 lower than 7. The information on the theoretical pI is useful in developing buffer system for the
830 purification of recombinant protein. The total number of negatively charged residues (Asp + Glu)
831 is 110, while that of positively charged residues (Arg + Lys) is 103. The positively charged
832 residues are lesser than the negatively charged counterparts which signifies that the protein is
833 intracellular. The half-life of the protein is 30hours, while the aliphatic index which is the
834 relative volume occupied by aliphatic side chains such as valine, isoleucine, alanine and leucine,
835 is 84.67. At such value, the protein has a high thermostability. The Grand Average hydropathy
836 (GRAVY) of a protein is calculated as the sum of hydropathy values of all amino acids, divided
837 by the number of residues in the sequence. The gravy value of the spike glycoprotein is -0.079,
838 depicting its hydrophilic nature, and better interaction with water. The individual amino
839 composition of the protein is summarized [Figure 1]. Every individual amino residue plays a
840 role in the protein function, structure and signaling, depending on their position. The four major
841 amino residues were leucine, serine, threonine and valine. Serine and threonine majorly perform

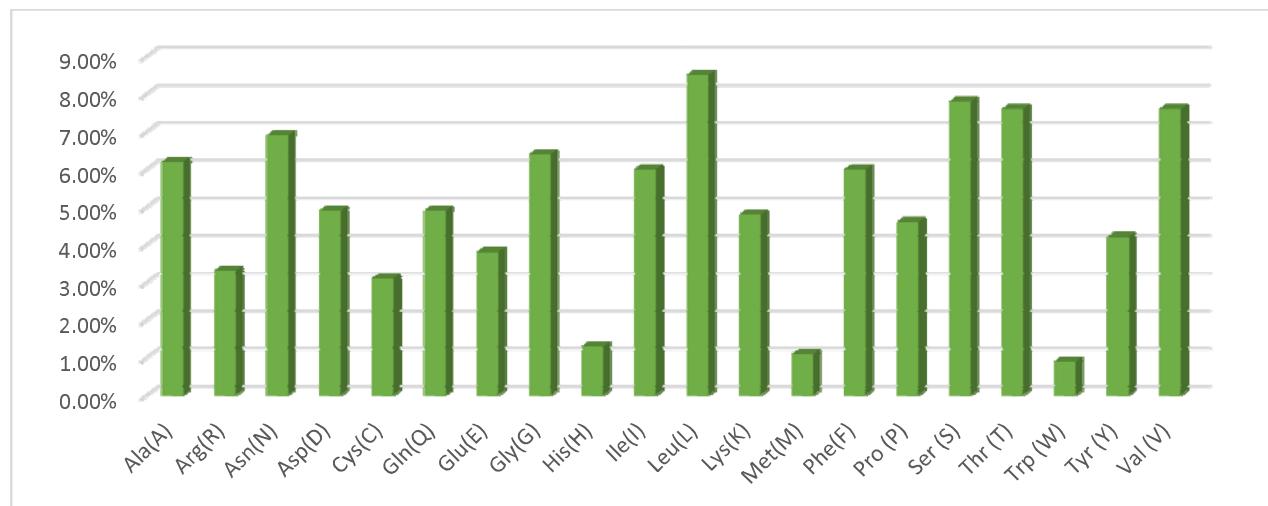
842 the phosphorylation function which is expedient for the protein's signaling pathway, as they have
843 hydroxyl functional group with affinity for phosphate group. The least amino residues were Trp
844 (0.90%), Met (1.10%) and His (1.3%).

845 **S-Table 1:** Physicochemical properties of SARS-CoV-2 Spike glycoprotein

Organism	m. wt.	Seq. length	Pi	EC (assuming all pairs of Cys residues from cysteine)	EC (assuming all cys residues are reduced)	Half-life (hrs)	II	Gravy	-R	+R	AI
SARS-CoV-2	141178.47	1273	6.24	148960	146460	30	33.01	-0.079	110	103	84.67

846 **NB:** (m.wt.: molecular weight, pI: isoelectric point, EC: extinction coefficient at 280nm, -R: number of
847 negatively charged residues, +R: number of positively charged residues, Gravy: Grand average
848 hydropathy, AI: Aliphatic index, II: instability index)

849



850
851 **Figure S2:** Percentage of amino acids present in SARS-CoV-2 Spike glycoprotein.
852

853
854
855
856
857
858
859

860

S-Table 2: Discontinuous B cell epitope contact numbers in the spike glycoprotein.

Chain ID	Residue ID	Residue Name	Contact Number	Propensity Score	Discotope score
A	809	PRO	8	2.079	-2.909
A	810	SER	7	2.066	-2.786
A	794	ILE	7	2.211	-2.745
A	793	PRO	7	2.483	-2.727
A	560	LEU	7	0.591	-2.39
A	1145	LEU	7	0.714	-1.921
A	491	PRO	6	0.255	-1.434
A	562	PHE	7	0.773	-1.289
A	1146	ASP	6	0.61	-1.017
B	1144	GLU	8	0.918	-3.082
B	811	LYS	9	1.491	-3.009
B	560	LEU	7	0.591	-2.909
B	1145	LEU	7	0.714	-2.786
B	1146	ASP	6	0.61	-2.39
B	809	PRO	8	2.079	-1.921
B	810	SER	7	2.066	-1.434
B	794	ILE	7	2.211	-1.289
B	793	PRO	7	2.483	-1.017
C	792	PRO	10	1.669	-3.331
C	1144	GLU	8	0.918	-3.082
C	811	LYS	9	1.491	-3.009
C	1145	LEU	7	0.714	-2.786
C	1146	ASP	6	0.61	-2.39
C	810	SER	8	1.996	-2.004
C	809	PRO	8	2.079	-1.921
C	794	ILE	7	2.211	-1.289
C	793	PRO	7	2.483	-1.017

861 Residues are shown in three-letter code, and number of contacts shows the connection of amino acid with
862 neighbouring groups

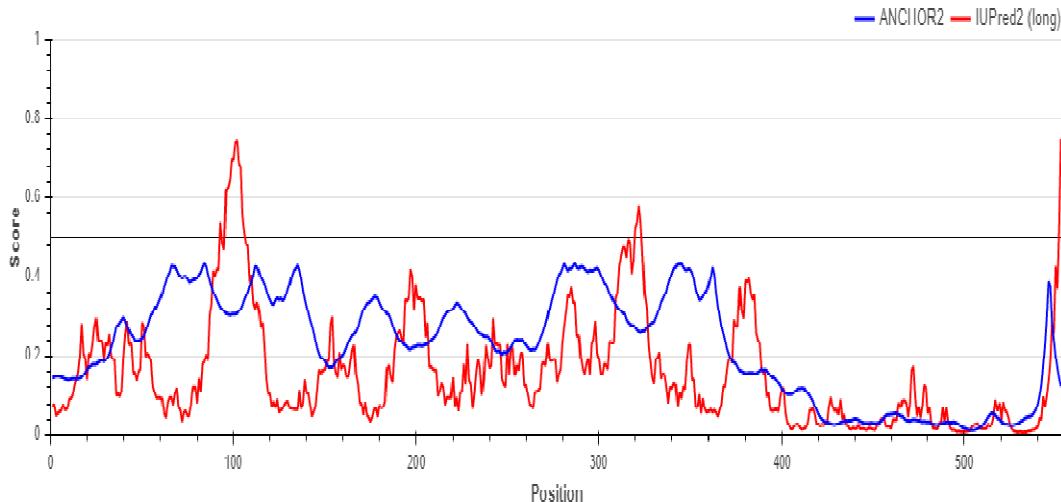
863



864

865 **Figure S3:** Diagrammatic representation of secondary structure prediction of the vaccine construct. Here, the β -
866 strands, α -helix and random coils are indicated by yellow, pink and blue colour, respectively.

867



868

869 **Figure S4:** Disordered region of the peptide vaccine construct. Regions (lines) exceeding the threshold of 0.5,
870 was considered disordered.

871

872

873

874