

1 **Designing a multi-epitope peptide-based vaccine against SARS-CoV-2**

2 **Abhishek Singh^{1,2}, Mukesh Thakur^{1*}, Lalit Kumar Sharma¹ and Kailash Chandra¹**

3 ¹Zoological Survey of India, New Alipore, Kolkata- 700053, India

4 ²Gujarat Forensic Sciences University, Sector 9, Gandhinagar, Gujarat 382007, India

5 *** Corresponding authors:**

6 Dr. Mukesh Thakur
7 Scientist C - Centre for DNA Taxonomy &
8 Coordinator - Centre for Forensic Sciences
9 Zoological Survey of India, New Alipore
10 Kolkata-700053, West Bengal, India
11 Mobile: +91-8171051282; FAX : 91-33-24008595
12 E. mail: thamukesh@gmail.com

13

14

Abstract

15 COVID-19 pandemic has resulted so far 14,395,16 confirmed cases with 85,711 deaths from the
16 212 countries, or territories. Due to multifacet issues and challenges in implementation of the
17 safety & preventive measures, inconsistent coordination between societies-governments and
18 most importantly lack of specific vaccine to SARS-CoV-2, the spread of Wuhan originated virus
19 is still uprising after taking a heavy toll on human life. In the present study, we mapped several
20 immunogenic epitopes (B-cell, T-cell, and IFN-gamma) over the entire structural proteins of
21 SARS-CoV-2 and by applying various computational and immunoinformatics approaches, we
22 designed a multi-epitope peptide based vaccine that predicted high immunogenic response in the
23 largest proportion of world's human population. To ensure high expression of the recombinant
24 vaccine in *E. coli*, codon optimization and in-silico cloning were also carried out. The designed
25 vaccine with high molecular affinity to TLR3 and TLR4, was found capable to initiate effective
26 innate and adaptive immune response. The immune simulation also suggested uprising high
27 levels of both B-cell and T-cell mediated immunity which on subsequent exposure cleared
28 antigen from the system. The proposed vaccine found promising by yielding desired results and
29 hence, should be tested by practical experimentations for its functioning and efficacy to
30 neutralize SARS-CoV-2.

31 **Keywords:** SARS-CoV-2; structural proteins; multi-epitope vaccine; immunoinformatics;
32 computational modelling

33 **Introduction**

34 Coronavirus disease 2019 (COVID-19) is one of the most potent pandemics characterized by
35 respiratory infections in Human. As per the latest report of World Health Organization (WHO)
36 dated 10th of April 2020, the worldwide coverage of COVID-19 exceeded 14,395,16 confirmed
37 cases with mortality of 85,711 deaths from 212 countries, areas or territories [1]. The outbreak of
38 COVID-19 emerged at Wuhan city of Hubei Province of P.R. China in the early December of
39 2019 [2], but the report of severe pneumonic patients with unknown etiology of symptoms was
40 released by the Chinese Center for Disease Control on December 31st, 2019 [3]. Later on, the
41 novel Betacoronavirus (SARS-CoV-2) was confirmed as the causative agent for the clusters of
42 pneumonic cases by the health professionals. Investigations on the origin of infections linked the
43 SARS-CoV-2 virus to a seafood market in Wuhan city, China which got spread to a large extent
44 of human population in China in a very short period of time. By the time, scientists/ health
45 professionals paid a serious note to understand the mode of transmission of COVID-19, it had
46 reached and infected several people in different countries with an insight into the escalated
47 transmission rate. This led the WHO to declare COVID-19 as a Public Health Emergency of
48 International Concern (PHEIC) on 30th of January 2020 [4].

49 Until the outbreak of Severe Acute Respiratory Syndrome (SARS) in 2002 and 2003,
50 coronaviruses were considered as a mild pathogenic virus causing respiratory and intestinal
51 infections in animals and humans [5-9]. With a decade long pandemic of SARS, another potent
52 pathogenic coronavirus, i.e., Middle East Respiratory Syndrome Coronavirus (MERS-CoV)
53 appeared specifically in the Middle Eastern countries [10]. Extensive studies on these two
54 coronaviruses have led to an understanding that they are genetically diverse but likely to be
55 originated from bats [11, 12]. Apart from these two viruses, four other coronaviruses are
56 commonly detected in Humans, viz., HKU1, 229E, OC43 and NL63 [13, 14]. Among all the
57 coronaviruses reported so far, SARS-CoV, MERS-CoV, and SARS-CoV-2 have relatively high
58 pathogenicity and escalated mortality rates. SARS CoV alone was responsible for 8422 positive
59 cases and 919 probable fatalities over a vast population of 32 countries [15]. With the first
60 mortality case reported in 2012, 2496 cases were found positive for MERS-CoV and 868 deaths

61 were reported from 27 countries [15]. Both of these pandemics were considered as highly
62 pathogenic in the last two decades until the outbreak of SARS-CoV-2 in December 2019. SARS-
63 CoV-2 has already spread in 212 countries and caused 85,522 deaths in just last three and half
64 months [1]. As in the current scenario, several countries are facing the most fatal problem of
65 community spread of SARS-CoV-2 e.g. China, Italy, Spain, France, Iran, USA, India, Germany
66 etc. Though, the estimated mortality rate due to SARS-CoV-2 is about 3.6% in China and about
67 1.5% outside China [16], the problem is grave and need to be respond urgently to curb the rapid
68 loss of human life.

69 Coronaviruses are broadly divided into four genera, i.e., Alpha-CoV, Beta-CoV, Gamma-CoV
70 and Delta-CoV, where Alpha-CoV and Beta-CoV are only reported to infect mammals but
71 Gamma-CoV and Delta-CoV are also capable to infect both mammals and birds [17]. Extensive
72 study of their biology suggests that they are non-segmented enveloped viruses with a positive-
73 sense single-stranded RNA of a size range of 26-32 kb [18]. Some of the recent research
74 suggested that SARS-CoV-2 has an identical genomic characterization as of Beta-CoV and thus
75 the genomic structure of SARS-CoV-2 follows 5'-leader-UTR-replicase-S (Spike)-E
76 (Envelope)-M(Membrane)-N (Nucleocapsid)-3' UTRpoly (A) tail with the subsequent species-
77 specific accessory gene at 3' terminal of the viral genome [19]. The structural protein includes
78 Spike/Surface protein (S), Envelop protein (E), Membrane protein (M) and Nucleocapsid protein
79 (N). These Spike/Surface proteins are known to have crucial roles in survival and viral
80 pathogenesis as the receptor binding capability and entry to the host cell is regulated by the
81 influence of Surface protein [20]. The Envelop and Membrane proteins are responsible for viral
82 assembly and Nucleocapsid protein is necessary for RNA genome synthesis [20]. This complex
83 genetic makeup and moderate mutation rate of SARS-CoV-2 requires the strategic development
84 of vaccine by targeting all the structural proteins. Unfortunately, there is no approved vaccine
85 available as on date against SARS-CoV-2. However, a few studies have predicted peptide-based
86 vaccine targets for specific glycoproteins [21, 22] but no study has covered all the structural
87 proteins for vaccine designing. With the recent development of integrated bioinformatics and
88 immunoinformatics approaches, vaccine designing and its specific application has become rapid
89 and cost effective as with the computational simulations and predictions, the targeted
90 immunogenic peptide can be designed and validated for its efficacy to work on human, even
91 before having the vaccine in hand for practical exposure and experiments. However,

92 conventional methods have simultaneously proven effective in vaccine designing with some
93 limitation, e.g. sacrificing the whole organism or large protein residues which unnecessarily
94 increases the antigenic load and probability of allergenicity [23]. We believe that an ideal multi-
95 epitope vaccine candidate can enhance immune response and subsequently lower down the risk
96 of re-infection by upraising the host immunogenicity. In this perspective, here we applied
97 integrated approaches of vaccine designing by mapping B-cell, T-cell and IFN-gamma epitopes
98 (humoral and adaptive immune response) of humans on the respective structural proteins based
99 on several essential criteria.

100 **Methodology**

101 Retrieval of SARS-CoV-2 protein sequences and antigenicity prediction

102 All the protein sequences of SARS-CoV-2 available on NCBI (GenBank: MN908947.3) was
103 retrieved and structural protein sequences were extracted for further analysis. We predicted
104 antigenicity from each protein using the online server VaxiJen v2.0 [24] and proteins that
105 showed antigenicity ≥ 0.4 in the virus category were subjected for further analysis.

106 **Prediction of physiochemical and secondary structural properties of the target proteins**

107 The physicochemical properties of target proteins were examined using ExPasy Protpram online
108 server [25] and the conformational states, e.g. Helix, sheet, turn and coil were predicted using the
109 online Secondary Structure Analysis Tool (SOPMA) [26] with default parameters.

110 **3 Dimensional homology modeling and validation**

111 Three-dimensional structure of target proteins was modeled using three homology modeling
112 tools,i.e. I Tasser, Raptor-X and Phyre2 [27-29] and post processing to reduce any distortions in
113 the modeled structure was done using the Galaxy refine server [30].The modeled structures were
114 subjected to RAMPAGE server for Ramachandran plot analysis for quality check and
115 validation[31 and visualized using Chimera 1.10.1 visualization system [32].

116 **Prediction of T-cell epitopes**

117 **Cytotoxic T-cell (CTL) epitopes**

118 Nine residues long CTL epitopes recognized by HLA class-I supertypes, i.e., A1, A2, A3, A24,
119 A26, B7, B8, B27, B39, B44, B58, B62 were predicted using NetCTL.1.2 server [33]
120 considering the default values of weight on C terminal cleavage, weight on TAP transport
121 efficiency and threshold for epitope identification. Further, another set of CTL epitopes
122 recognized by HLA class I alleles like A*02:02, B*07:02, C*04:01, E*01:01, etc were identified

123 by consensus method of Immune Epitope Database (IEDB) tool [34]. The identified epitopes
124 were scrutinized based on the consensus score i.e. ≤ 2 and only those epitopes were considered as
125 strong binders and selected for further analysis which were predicted by more than one allele of
126 both the method used.

127 **Helper T-cell (HTL) epitopes**

128 Fifteen residue long HTL epitopes recognized by HLA Class II DRB1 alleles were predicted
129 using Net MHC II pan 3.2 server [35]. The thresholds for strong binder and weak binder were set
130 as default. Further another set of HTL epitopes of 15 residues length recognized by HLA DR
131 alleles were identified using the consensus method as implemented in the IEDB server [34] The
132 epitopes with a percentile rank of ≤ 2 and predicted by more than one allele of both the methods
133 were considered as a strong binder and scrutinized for further analysis.

134 **Identification of overlapping T-cell epitopes**

135 Considering the fact that epitopes with affinity for multiple HLA alleles tend to induce relatively
136 more immune response in the host cell, we scrutinized overlapping epitopes with affinity to both
137 HLA class I and class II alleles for immunogenicity and allergen prediction using VaxiJen v2.0
138 server [24] and AllerTop v.2.0 [36] tool with an immunogenic threshold of 0.4. These epitopes
139 would have high potential to activate both CTL and HTL cells.

140 **Identification of Continuous and Discontinuous B-cell epitopes**

141 B-cell epitopes are short amino acid sequences recognized by the surface-bound receptors of B-
142 lymphocyte. Their identification plays a major role in vaccine designing, and thus two types-
143 Continuous and Discontinuous B- cell epitopes were predicted over the corresponding structural
144 protein sequences. The continuous B-cell epitopes, present on the surface of proteins, were
145 predicted using BCpred 2.0 [37] server and the discontinuous epitopes which were fairly large
146 were predicted using Ellipro server [38]. The prediction parameters like Minimum score and
147 Maximum distance (Anstrom) were set as 0.8 and 6, respectively.

148 **Identification of IFN- gamma epitopes**

149 Since, IFN gamma is the signature cytokine of both the innate and adaptive immune system and
150 epitopes with potency to induce IFN gamma could boost the immunogenic capacity of any
151 vaccine. Thus, IFN gamma epitopes were predicted for each structural protein by the IFNepitope
152 server [39].

153 **Characterization of predicted epitopes**

154 **Conservation analysis**

155 Predicted T-cell and B-cell epitopes were submitted to the IEDB conservancy analysis tool [40]
156 to identify the degree of conservancy in the structural protein sequences and the epitopes with
157 100% conservancy were selected for further analysis.

158 **Population coverage and autoimmunity identification**

159 Immunogenetic response of human population towards the selected overlapping CTL and HTL
160 epitopes against their respective HLA genotype frequencies was predicted using IEDB
161 population coverage analysis tool [41]. Those epitopes exhibiting 50% or more population
162 coverage were scrutinized and to reduce the probability of autoimmunity, all the selected
163 epitopes were subjected to BlastP search analysis against the Human proteome. Epitopes
164 showing similarity to any human protein were excluded from the further analysis.

165 **Interaction analysis of epitopes and their HLA alleles**

166 The sequence of scrutinized epitopes was submitted to an online server PEPFOLD 3 [42] for 3D
167 structure modeling. The structure of the most common HLA alleles i.e. HLA-DRB1 *01:01
168 (HLA class II) and HLA-A* 02:01 (HLA class I) in human population were retrieved from
169 Protein data bank (PDB) with a PDB ID of 2g9h and 1QEW. Any ligand associated with the
170 HLA allele's structure was removed and energy-minimization was carried out. Thereafter, the
171 modeled epitopes were docked with the corresponding HLA allele using a web server Patchdock
172 [43], to study the interaction pattern of receptor and ligand. The parameters like clustering
173 RMSD value and the complex type was kept as 0.4 and default. The result obtained from
174 Patchdock was forwarded to the Firedock server [44] for the refinement of the best 10 models.
175 Based on Global energy value, the complex with the lowest global energy was scrutinized and
176 the corresponding CTL and HTL epitopes were selected for the final vaccine construct.

177 **Construction and quality control of final multi-epitope vaccine**

178 For the construction of a multi-epitope vaccine against SARS-CoV-2, we followed Chauhan et
179 al. [23] with some modifications. The epitopes were finally selected based on meeting the
180 following criteria: 1). epitopes must be immunogenic, non-allergic and overlapping with affinity
181 to both HLA class I and class II alleles; 2).epitopes must be capable to activate both CTL and
182 HTL cells and must have a minimum 50% of the population coverage; 3).epitopes should not be
183 overlapping with any human gene and the predicted B-Cell epitopes should overlap with
184 finalized CTL and HTL epitopes and should be present on the surface of the target protein.

185 Based on these criteria, the HTL, CTL and IFN gamma epitopes were included in the final
186 construct of the multi-epitope vaccine. The HTL and IFN gamma epitopes were linked by
187 GPGPG linkers, whereas the CTL epitopes were linked by AAY linkers [23]. An adjuvant, 50S
188 ribosomal protein L7/L12 (Locus RL7_MYCTU) with an NCBI accession no: P9WHE3 was
189 added to the N terminal to enhance the immunogenicity of the constructed vaccine.

190 **Antigenicity, allergenicity and physiochemical properties of vaccine**

191 The antigenicity of the proposed vaccine was predicted using VaxiJen v2.0 server [24] and
192 allergenicity was predicted using Allertop v.2.0 prediction tool [36]. The physicochemical
193 property of the designed vaccine was predicted by submitting the final sequence of the vaccine to
194 the Exspasy Protpram online server [25].

195 **Prediction of secondary structure**

196 The secondary structure of the constructed vaccine was predicted using the SOPMA tool [26].
197 With the input of vaccine sequence and output width 70, the parameters like the number of
198 conformational states, similarity threshold, and window width were set as 4 (Helix, Sheet, Turn,
199 Coil), 8 and 17.

200 **Modeling, refinement, and validation of vaccine construct**

201 The final model of the constructed vaccine was prepared using a homology modeling tool
202 SPARKS-X [45]. Among the top 10 predicted models, the model with the highest Z-score was
203 selected and subjected for refinement using the Galaxy refine server [30]. Among all the
204 refinement tools, Galaxy refine is one of the most efficient server according to CASP10
205 evaluation. Then after, we utilized ProSA-webtool [46] to assess the overall and local quality of
206 the predicted model based on the Z-score. If the Z-score lies outside the range of native protein
207 of similar size, a chance of error increases in the predicted structure. Further to check the overall
208 quality of the refined structure of vaccine, Ramachandran plot analysis was carried out using the
209 RAMPAGE server [31]. The final model of the vaccine construct was visualized using Chimera
210 1.10.1 visulaization system [32].

211 **Prediction of Continuous and Discontinuous epitopes**

212 Continuous and Discontinuous B- cell epitopes in the vaccine construct were predicted using
213 BCpred 2.0 [37] and IEDB server [34], respectively. For effective designing of vaccine, the
214 constructed vaccine should posses' B-cell epitopes for potential enhancement of immunogenic
215 reactions.

216 **Molecular docking of constructed vaccine with TLR3 and TLR4 receptor**

217 The 3D structures of human TLR3 and TLR4 were retrieved from protein data bank (PDB ID:
218 2A0Z and 2z63). Any ligand attached to the retrieved structures was removed and the receptor
219 was subjected for model refinement using Galaxy refine server [30]. Molecular docking analysis
220 was performed using the Cluspro v.2 protein-protein docking server [47] to analyze the
221 interaction pattern of vaccine with TLR3 and TLR4. The server provides cluster scores based on
222 rigid docking by sampling billions of conformation, pairwise RMSD energy minimization. Based
223 on the lowest energy weight score and members, the final vaccine-TLR3 complex and TLR4
224 complex model was selected and visualized using Chimera 1.10.1 visualization system [32].

225 ***In silico* cloning and vaccine optimization**

226 Efficiency of the multi-epitope vaccine construct in cloning and expression is of the utmost
227 importance for vaccine designing. The final multi-epitope vaccine construct was submitted to the
228 Java Codon Adaptation Tool (JCat) [48], where codon optimization was performed in *E. coli*
229 strain K12. In addition to this, the parameters like Avoid rho-independent transcription
230 terminators, Avoid prokaryotic ribosome binding sites, Avoid Cleavage Sites of Restriction
231 Enzymes were selected for the final submission. The output of the JCat tool provided CAI value
232 and GC content of the optimized sequence. Ideally, the CAI value should be greater than 0.8 or
233 nearer to 1 [49], and the GC content should be in a range of 30 to 70% [50] to ensure that the
234 vaccine has potential translation, stability and transcription efficiency. Further, an *in silico*
235 cloning of a multi-epitope vaccine construct in *E. coli* pET- 28(+) vector was performed using
236 the SnapGene 4.2 tool [51].

237 **Immune simulations of vaccine construct**

238 To characterize the real-life immunogenic profiles and immune response of the multi-epitope
239 vaccine, C-ImmSim server [52] was utilized. Position-specific scoring matrix (PSSM) and
240 machine learning are the two methods, based on which C-ImmSim predicts immune epitopes and
241 immune interactions. It simultaneously simulates three different anatomical regions of mammals,
242 i.e. Bone marrow, Thymus and tertiary lymphatic organs. To conduct the immune simulation,
243 three injections at an interval of four weeks were administered with 1000 vaccine molecules per
244 injection. The parameters like Random seed, simulation volume, and simulation step were kept
245 as 12345, 10 μ l and 1050. Several studies suggested that the minimum interval between two
246 injections should be kept four weeks and thus the time steps followed for three injections were 1,

247 84 and 168 where each time step is equal to 8 hours of real-life [53]. Since, several patients
248 worldwide got recurrent infection of SARS-CoV-2, 12 consecutive injections were administered
249 four weeks apart for the assessment of after effects of vaccine exposure on SARS-CoV-2.

250 **Results**

251 **Target proteins and their antigenicity**

252 We scrutinized target structural protein sequences, i.e. Surface, Membrane, Envelop and
253 Nucleocapsid proteins from the whole-genome of SARS-CoV-2, 29903 bp ss- RNA (GenBank:
254 MN908947.3). Antigenicity prediction from the structural protein sequences revealed that
255 Envelop proteins has the highest antigenicity score 0.6025, followed by Membrane,
256 Nucleocapsid and Surface protein (table 1).

257 **Physiochemical and secondary structural properties**

258 Theoretical estimates of isoelectric point (PI) suggested that all the structural proteins were basic
259 in nature and their predicted instability indices were below 40 indicating their stable nature
260 except the Nucleocapsid protein with an instability index of 55.09 (table S1). The predicted
261 secondary structure characteristics of all the structural proteins showed variable percentage of
262 Alpha Helix, Extended Strand, Beta Turn and Random Coil (table S2).

263 **3 Dimensional structure of target proteins**

264 The refined finalized 3D structures obtained from Raptor-X was better than the models generated
265 from I Tasser and Phyre2 as Raptor-X models showed least percentage of outliers and therefore
266 selected for further analysis (figure 1). Details of various models of the structural proteins
267 through Ramachandran plot analysis are given in table S3.

268 **Prediction of CTL and HTL epitopes**

269 A total of 29 CTL epitopes in Envelop protein, 77 in Nucleocapsid, 340 in Surface and 89
270 epitopes in Membrane protein were predicted with strong binding affinity for multiple alleles.
271 Similarly, 35 HTL epitopes in Envelop protein, 51 in Nucleocapsid, 252 in surface and 23
272 epitopes in Membrane protein were predicted. All these epitopes were strong binders and
273 predicted percentile rank of ≤ 2 .

274 **Identification of overlapping T cell epitopes**

275 The CTL epitopes overlapping with HTL epitopes were scrutinized and subjected to
276 immunogenicity and allergenicity prediction. In total, 16 CTL overlapped epitopes were
277 predicted in Envelop protein, 8 in Nucleocapsid protein, 25 in Surface protein and 9 epitopes in
278 Membrane protein (Table S4-S11). All these selected epitopes were predicted to be non-allergic
279 with high antigenic scores.

280 **Identification of B-cell epitopes**

281 The identified continuous B cell epitopes present on the surface of their respective proteins were
282 selected based on their antigenic, non-allergic and overlapping nature in the finalized CTL and
283 HTL epitopes. With these criteria, five epitopes were obtained in Surface protein, three in
284 Nucleocapsid and one in Membrane protein (table S12). No epitope with such criteria was
285 identified in Envelop protein. Further, one discontinuous epitope in Envelop protein, seven in
286 Nucleocapsid, nine in Surface and four discontinuous epitopes were predicted in Membrane
287 protein (table S13).

288 **Identification of IFN- gamma epitopes**

289 The server IFNepitope predicted eight epitopes in Envelop protein, 39 in Membrane, 66 in
290 Nucleocapsid and 281 in Surface protein. Based on antigenicity and non-allergenicity, eight
291 epitopes were scrutinized in Envelop protein, 13 in Membrane, 43 in Nucleocapsid and 81 in
292 Surface protein. To obtain the highest immunogenicity and reduce the overload of epitopes in
293 vaccine construct, we selected one epitope from each protein with the highest antigenicity score
294 and non-allergenicity (table S14).

295 **Characterization of predicted epitopes**

296 **Conservation analysis, Population coverage, and autoimmunity identification**

297 The conservation prediction resulted that all the selected T cell and B cell epitopes were 100%
298 conserved among the target structural proteins and population coverage prediction showed that
299 five epitopes in Envelop protein, two in Membrane, two in Nucleocapsid and six epitopes in
300 Surface protein covered more than 50% of worldwide population. No selected epitope with
301 >50% population coverage showed homology with any human protein. Thus, altogether 30
302 epitopes, 15 each of CTL and HTL, met the above criteria were processed for the interaction
303 analysis with the commonly occurring HLA alleles in the Human population (table S15).

304 **Interaction analysis of epitopes and their HLA alleles**

305 Molecular interaction analysis of each selected CTL and HTL epitopes with their respective
306 HLA alleles resulted in 10 predicted docked complex. Each complex showed different Global
307 energy value. Out of the characterized 30 epitopes, the epitopes with highest negative value of
308 Global energy (set threshold -40) were considered as effectively interacting epitopes with the
309 corresponding HLA allele, and thus selected for final vaccine construct (table 2). This resulted in
310 two CTL and HTL epitopes in Envelop protein, one in Membrane, two in Nucleocapsid and two
311 in Surface protein. For an ideal vaccine, the epitopes included should have potential affinity
312 towards their respective alleles and thus greater the interaction between epitopes and their alleles,
313 more is the affinity.

314 **Construction of final multi-epitope vaccine**

315 We selected seven CTL, seven HTL and four IFN gamma epitopes for construction of the multi-
316 epitope vaccine (figure S1). The adjuvant was coupled by EAAAK linker with CTL epitope and
317 subsequently, AAY linker was used to couple CTL epitopes and GPGPG linker was used to
318 couple HTL and IFN gamma epitopes (figure 2). The final multi-epitope vaccine construct was
319 composed of 430 amino acid residues which was then validated for antigenic, allergenic and
320 physiochemical properties.

321 **Evaluation of antigenicity, allergenicity, physiochemical properties and secondary 322 structure of the vaccine construct**

323 The multi-epitope vaccine construct found to be immunogenic (Ag score- 0.627), non-allergic
324 with 45.131 KDa predicted molecular weight. The prediction suggested that vaccine construct
325 was stable and basic in nature with instability index 23.46 and theoretical PI 9.01. The half-life
326 of the vaccine in mammalian reticulocytes (in vitro) was 30 hrs, while it was about 20 hrs in
327 yeast (in-vivo) and 10 hrs in *E. coli* (in vivo).

328 The predicted secondary structure of the vaccine constructed consisted 42.33% alpha-helix,
329 22.09% extended strand, 4.42% beta-turn and 31.16% random coil.

330 **Modeling, refinement, and validation of vaccine construct**

331 Of the top 10 predicted models, the model with the Z-score of 7.61 with template 1dd3A was
332 selected for refinement. The ProSA-web tool assessed the overall and local quality of the refined
333 model with a Z-score of -0.47 which was close to the range of native protein of similar size
334 (figure 3). Further, the Ramachandran plot analysis revealed 92% of residues in the favored

335 region, 6.1% in the allowed region and 1.9% outliers. This signified that the quality of the
336 predicted vaccine construct was adequate for further analysis.

337 **Prediction of Continuous and Discontinuous epitopes**

338 BCpred 2.0 server predicted four continuous epitopes i.e. AKILKEYGLD,
339 EILDKSKEKTSFD, LKESKDLV and VPKHLKKGLSKEEAESLKKQLEEV on the surface of
340 predicted vaccine (figure S2), while IEDB predicted six discontinuous epitopes (figure S3) with
341 a score higher than 0.8 (Table S16).

342 **Molecular docking of constructed vaccine with TLR3 and TLR4 receptor**

343 Cluspro v.2 predicted 30 models each of vaccine receptor TLR3 complex and TLR4 complex
344 with their corresponding cluster scores (Table S17-S18). Among these models, the model
345 number 2 in TLR3 complex and model number 1 in TLR4 complex were selected as a best-
346 docked complex with the lowest energy score of -1199.1 with 46 members (TLR3) and lowest
347 energy score of -1229.9 with 79 members (TLR4). This signifies potential molecular interaction
348 between predicted vaccine construct with TLR3 and TLR 4 receptors (figure 4).

349 ***In silico* cloning and vaccine optimization**

350 The Java Codon Adaptation Tool optimized the codon usage of the vaccine and produced an
351 optimized codon sequence of length 1290 nucleotides. The CAI value of optimized sequences
352 was 0.95, close to 1 and the GC content was found to be 54.41 %, fall between the range of 30 to
353 70%. Therefore, the adaptation was satisfactory indicating potential expression of vaccine
354 construct in host *E.coli*. Later on, the restriction sequence of Xho I and Not I restriction enzymes
355 were added to N and C terminal of adapted codon sequence. Further, the adapted sequence was
356 also cloned in pET- 28(+) vector using SnapGene tool (figure 5).

357 **Immune simulations of vaccine construct**

358 C-ImmSim server-generated immune response was consistent with the actual immune response
359 as depicted by the results (figure 6 and S4). The primary response was generated as marked by
360 the increase in the level of IgM. Similarly, the secondary response was characterized by an
361 increased level of IgM+IgG, IgG1+IgG2, IgG1, IgG2, and B-cell populations. On the subsequent
362 exposure of vaccine, a decrease in the level of antigens was observed indicating the development
363 of immunogenic response in the form of immune memory. Both the T cell populations i.e. CTL

364 and HTL developed increased response corresponding to the memory cells signifying the
365 immunogenicity of T cell epitopes included in the vaccine construct. Increased activity of
366 macrophages was also observed at each exposure, whereas the activity of NK cells was observed
367 consistent throughout the period. A significant increase in the level of IFN gamma, IL-10, IL-23,
368 and IL-12 was also observed at subsequent exposure (figure 6). After the repeated exposure of
369 vaccine construct through 12 injections after a regular interval, a remarkable increase in the
370 level of IgM+IgG, IgG1+IgG2. Though, IgG1 increased slightly but IgM and IgG2 increased
371 consistently. A surge in the memory corresponding to B cell and T cells was observed
372 throughout the exposure while the level of IFN gamma was consistently high from first to last
373 exposure (figure S4). This signifies that the vaccine proposed in this study has generated a strong
374 immune response in case of short exposure and immunity increases even on subsequent repeated
375 exposure.

376 **Discussion**

377 The recent outbreak of SARS-CoV-2 in Wuhan city of China raised several questions regarding
378 the susceptibility of Humans against these novel pathogens. As predicted earlier about the re-
379 emergence of novel variants of SARS-CoV [54], SARS-CoV-2 proved to be more lethal as
380 compared to the previous variants in terms of wide spread of virus in a short period of time. This
381 escalated rate of transmission initiated a pursuit for a vaccine development against SARS-CoV-
382 2, which is currently a worldwide pandemic with over 1.43 million cases and 85, 711 deaths [1].
383 With strict and comprehensive measures, vaccine development and its application could play a
384 key role in eliminating the virus from the Human population or to restrain the spread among
385 individuals and different populations. Thus several efforts are being made to address the
386 challenge that appeared in the current scenario with appreciable advancements in the
387 understanding of virus biology and its etiology. The lack of knowledge regarding the response of
388 the immune system against viral infection is one of the major limitations in the path of vaccine
389 development for SARS-CoV-2.

390 This study is the piolt attempt in describing the potential immunogenic target over the structural
391 proteins and proposes a novel multi-epitope vaccine construct, by providing new rays of hope in
392 the initial phase of vaccine development. Certain criteria like poor antigenicity, allergenicity, low
393 affinity towards the immune cells, autoimmunity and oversize that could influence the

394 effectiveness, have been evaluated on the proposed vaccine construct following various
395 computational and immunoinformatics approach. The retrieved structural proteins and their
396 antigenicity score suggested that the Envelop protein is the most potent protein to generate
397 immune response considering its role in viral assembly. Since, immunity against any pathogen is
398 prominently dependent, how it gets recognized by B cells and T cells. We identified epitopes
399 corresponding to B cells and T cells in each structural protein so that both humoral and cellular
400 immunity can be induced with the exposure of vaccine construct. The proposed vaccine construct
401 is designed based on the epitopes that have been selected with the most robust criteria, e.g. their
402 nature of antigenic, non-allergic, 100% conserved among the target proteins, their affinity for
403 multiple alleles, no homology with any of the human proteins, their worldwide coverage of
404 human population (>50% of worldwide population covered) and effective molecular interaction
405 with their respective HLA alleles. Further,, IFN gamma epitopes which are equally effective in
406 immune regulatory, antiviral and antitumor activities, the final vaccine construct was designed
407 with also due considering of the IFN gamma epitopes to increase the immune recognition ability.
408 The designed vaccine of 430 amino acid residues has a molecular weight of 45.131K Da, also
409 falls in the defined range of average molecular weight of a multi-epitope vaccine. The instability
410 index and theoretical PI also suggested that the vaccine is stable and basic in nature and the
411 estimated half-life suggested that the recognized peptide does not possess a short half-life and
412 would remain viable for a span adequate to generate a potential immune response. Vaccine
413 model refinement and validation indicated that quality of the predicted model was good as more
414 than 90% residues were in the favored region [55]. The B-cell continuous epitopes predicted
415 over the surface of the vaccine suggested that the vaccine is capable to get recognized by B-cells
416 indicating its effectiveness. TLR 3 and TLR 4 have proven recognition capability in both SARS-
417 CoV and MERSCoV. Considering the similar genome characterization of SARS-CoV and
418 SARS- CoV-2, the molecular interaction of vaccine with TLR 3 and TLR 4 through docking
419 analysis suggested that the constructed vaccine has a significant affinity towards the toll-like
420 receptors to act as sensor for recognizing molecular patterns of pathogen and initiating immune
421 response. Thus the vaccine TLR complex is capable of generating an effective innate immune
422 response against SARS-CoV-2. Further, Codon adaptation improved the expression of the
423 recombinant vaccine in *E coli*. Strain K12 with significant codon adaptation index and GC
424 content indicating elevated expression level. After successful cloning of the recombinant vaccine

425 in the pET- 28(+) vector, the simulation based generated immune response suggested that the
426 primary and secondary immune response will be coherent with the actual expected response.
427 Subsequent exposure for three injections indicated the potential capability of vaccine for
428 generating memory cells and a higher probability of significant immune response on later
429 exposure to coronavirus. The consistent high level of IFN gamma also supported the activation
430 of humoral immunity. To understand the immunogenic potential of vaccine, repeated exposure in
431 the form of 12 injections was administered and the result suggested the consistent generation of a
432 strong immune response. With all these immunoinformatics approaches, the vaccine designed
433 against SARS-CoV-2 showed promising results in inducing immune response.

434 Without proper control methods or vaccines, it is difficult to curb coronavirus pandemic given
435 the current situation. However, several medications have been tested against SARS-CoV-2, but
436 none of them showed complete effectiveness. In this study, a vaccinomics approach was carried
437 out to design a multi-epitope vaccine against SARS-CoV-2 using several in silico and
438 immunoinformatic approaches. Based on the computational and immunoinformatics approaches,
439 we propose that the designed vaccine has all the potential to induce both the innate and adaptive
440 immune systems and can neutralize the SARS-CoV-2, and therefore, an experimental validation
441 of the designed vaccine must be undertaken by the professional for reaching a conclusion for its
442 safety, efficacy and success.

443 **Acknowledgements**

444 We thank Zoological Survey of India, Kolkata for providing access to the computational support,
445 working space and facilities to undertake the present study.

446 **References**

- 447 1. World Coronavirus disease 2019 (COVID-19) Situation Report – 77, 9th April 2020.
- 448 2. Zhang, Y.Z., Holmes, E.C. A Genomic Perspective on the Origin and Emergence of
449 SARS-CoV-2. *Cell*. 2020.
- 450 3. World-Health-Organization Coronavirus disease (COVID-19) outbreak. Available
451 online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed on
452 3rd April 2020).

- 453 4. World-Health-Organization Statement on the second meeting of the International Health
454 Regulations(2005) Emergency Committee regarding the outbreak of novel coronavirus
455 (2019-nCoV). Available online: [https://www.who.int/news-room/detail/30-01-2020-](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)))
456 [statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)))
457 [emergency-committee-regarding-the-outbreak-of-novelcoronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)))
458 (accessed on 29 March 2020).
- 459 5. Masters, P. S., Perlman, S. In *Fields Virology Vol. 2* (eds Knipe, D. M. &Howley, P. M.)
460 825–858 (Lippincott Williams & Wilkins, 2013).
- 461 6. Zhong, N. S. et al. Epidemiology and cause of severe acute respiratory syndrome (SARS)
462 in Guangdong, People’s Republic of China, in February, 2003. *Lancet*. **362**, 1353–1358
463 (2003).
- 464 7. Drosten, C. et al. Identification of a novel coronavirus in patients with severe acute
465 respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976 (2003).
- 466 8. Fouchier, R. A. et al. Aetiology: Koch’s postulates fulfilled for SARS virus. *Nature*. **423**,
467 240 (2003).
- 468 9. Ksiazek, T. G. et al. A novel coronavirus associated with severe acute respiratory
469 syndrome. *N. Engl. J. Med.* **348**, 1953–1966 (2003).
- 470 10. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. Fouchier, R. A.
471 Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J.*
472 *Med.* **367**, 1814–1820 (2012).
- 473 11. Lau, S. K. et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese
474 horseshoe bats. *Proc. Natl Acad. Sci.* **102**, 14040–14045 (2005).
- 475 12. Ithete, N. L. et al. Close relative of human Middle East respiratory syndrome coronavirus
476 in bat, South Africa. *Emerg. Infect. Dis.* **19**, 1697–1699 (2013).
- 477 13. Su, S. et al. Epidemiology, genetic recombination, and pathogenesis of
478 coronaviruses. *Trends Microbiol.* **24**, 490–502 (2016).
- 479 14. Forni, D., Cagliani, R., Clerici, M., Sironi, M. Molecular evolution of human coronavirus
480 genomes. *Trends Microbiol.* **25**, 35–48 (2017).
- 481 15. Meo, S A., et al. Novel Coronavirus 2019-nCoV: Prevalence, Biological and Clinical
482 Characteristics Comparison With SARS-CoV and MERS-CoV. *European Review for*
483 *Medical and Pharmacological Sciences.* **24**, 4, pp. 2012-2019 (2020).

- 484 16. Baud, D. et al. Real estimates of mortality following COVID-19 infection, *The Lancet*
485 *Infectious diseases*. (2020).
- 486 17. Woo, P. C. et al. Discovery of seven novel mammalian and avian coronaviruses in the
487 genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus
488 and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus
489 and deltacoronavirus. *J. Virol.* **86**, 3995–4008 (2012).
- 490 18. Su, S. et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses.
491 *Trends Microbiol.* **24**, 490-502 (2016).
- 492 19. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *New*
493 *England Journal of Medicine*. (2020).
- 494 20. Schoeman, D., Fielding, B.C. Coronavirus envelope protein: current knowledge. *Virol*
495 *J.* **16**, 69 (2019).
- 496 21. Abdelmageed, M.I. et al. Design of multi epitope-based peptide vaccine against E protein
497 of human 2019-nCoV: An immunoinformatics approach. *bioRxiv*. (2020).
- 498 22. Bojin, F., Gavriiliuc, O., Margineanu, M., Paunescu, V. Design of an Epitope-Based
499 Synthetic Long Peptide Vaccine to Counteract the Novel China Coronavirus (2019-
500 nCoV). *Preprints*. (2020).
- 501 23. Chauhan, V. et al. Designing a multi-epitope based vaccine to combat Kaposi Sarcoma
502 utilizing immunoinformatics approach. *Sci Rep.* **9**, 2517 (2019).
- 503 24. Doytchinova, I. A., Flower, D. R. VaxiJen: a server for prediction of protective antigens,
504 tumour antigens and subunit vaccines. *BMC Bioinformatics.* **8**, 4 (2007).
- 505 25. Wilkins, M. R. et al. Protein identification and analysis tools in the ExPASy
506 server. *Methods Mol. Biol.* **112**, 531–552 (1999).
- 507 26. Geourjon, C., Deleage, G. SOPMA: significant improvements in protein secondary
508 structure prediction by consensus prediction from multiple
509 alignments. *Bioinformatics.* **11**, 681–684 (1995).
- 510 27. Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nature*
511 *methods.* **12**, 7 (2015).
- 512 28. Källberg, M. et al. Template-based protein structure modeling using the RaptorX web
513 server. *Nature protocols.* **7**, 1511 (2012).

- 514 29. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., Sternberg, M. J. The Phyre2 web
515 portal for protein modeling, prediction and analysis. *Nature protocols*. **10**, 845 (2015).
- 516 30. Heo, L., Park, H., Seok, C. GalaxyRefine: Protein structure refinement driven by side-
517 chain repacking. *Nucleic acids research*. **41**, 384-388 (2013).
- 518 31. Lovell, S. C. et al. Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins:*
519 *Structure, Function, and Bioinformatics*. **50**, 437–450 (2003).
- 520 32. Pettersen, E. F. et al. UCSF Chimera- a visualization system for exploratory research and
521 analysis. *Journal of Computational Chemistry*. **13**, 1605–1612 (2004).
- 522 33. Larsen, M. V. et al. Large-scale validation of methods for cytotoxic T-lymphocyte
523 epitope prediction. *BMC bioinformatics*. **8**, 424 (2007).
- 524 34. Kim, Y. et al. Immune epitope database analysis resource. *Nucleic acids research*. **40**,
525 525–530 (2012).
- 526 35. Jensen, K. K. et al. Improved methods for predicting peptide binding affinity to MHC
527 class II molecules. *Immunology*. **154(3)**, 394-406 (2018).
- 528 36. Bangov, I., Doytchinova, I., Dimitrov, I., Flower, D. AllerTOP v.2 - A server for in silico
529 prediction of allergens. *J. Mol. Model*. **20**, 2278-2284 (2014)..
- 530 37. Manzalawy, Y.EL., Dobbs, D., Honavar, V. Predicting linear B- cell epitopes using
531 string kernels. *Journal of Molecular Recognition: An Interdisciplinary Journal*. **21**, 243–
532 255 (2008).
- 533 38. Ponomarenko, J. et al. ElliPro: a new structure-based tool for the prediction of antibody
534 epitopes. *BMC bioinformatics*. **9**, 514 (2008).
- 535 39. Dhanda, S. K., Vir, P., Raghava, G. P. Designing of interferon-gamma inducing MHC
536 class-II binders. *Biology direct*. **8**, 30 (2013).
- 537 40. Bui, H. H., Sidney, J., Li, W., Fusseder, N., Sette, A. Development of an epitope
538 conservancy analysis tool to facilitate the design of epitope-based diagnostics and
539 vaccines. *BMC bioinformatics*. **8**, 361 (2007).
- 540 41. Bui, H. H. et al. Predicting population coverage of T-cell epitope-based diagnostics and
541 vaccines. *BMC bioinformatics*. **7**, 153 (2006).
- 542 42. Lamiable, A. et al. PEP-FOLD3: faster de novo structure prediction for linear peptides in
543 solution and in complex. *Nucleic acids research*. **44**, 449–454 (2016).

- 544 43. Schneidman, D., Inbar, Y., Nussinov, R., Wolfson, H. PatchDock and SymmDock:
545 Servers for Rigid and Symmetric Docking. *Nucleic acids research*. **33**. 363-7 (2005).
- 546 44. Mashiach, E., Schneidman, D., Andrusier, N., Nussinov, R., Wolfson, Haim. FireDock:
547 A Web Server for Fast Interaction Refinement in Molecular Docking. *Nucleic acids*
548 *research*. **36**. 229-32 (2008).
- 549 45. Y. Yang., E. Faraggi., H. Zhao., Y. Zhou. Improving protein fold recognition and
550 template-based modeling by employing probabilistic-based matching between predicted
551 one-dimensional structural properties of query and corresponding native properties of
552 templates. *Bioinformatics*. **27(15)**, 2076–2082 (2011).
- 553 46. Wiederstein, M., Sippl, M. J. ProSA-web: interactive web service for the recognition of
554 errors in three-dimensional structures of proteins. *Nucleic acids research*. **35**, 407–410
555 (2007).
- 556 47. Kozakov, D. The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**, 255–
557 278 (2017).
- 558 48. Grote, A. et al. JCat: a novel tool to adapt codon usage of a target gene to its potential
559 expression host. *Nucleic Acids Research*. **33**, 526–531 (2005).
- 560 49. Morla, S., Makhija, A., Kumar, S. Synonymous codon usage pattern in glycoprotein gene
561 of rabies virus. *Gene*. **584**, 1–6 (2016).
- 562 50. Ali, M. et al. Exploring dengue genome to construct a multi-epitope based subunit
563 vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci*
564 *Rep.* **7**, 9232 (2017).
- 565 51. SnapGene software (from Insightful Science; available at snapgene.com).
- 566 52. Rapin, N., Lund, O., Bernaschi, M., Castiglione, F. Computational immunology meets
567 bioinformatics: the use of prediction tools for molecular binding in the simulation of the
568 immune system. *PLoS One*. **5**, e9862 (2010).
- 569 53. Castiglione, F., Mantile, F., De Berardinis, P., Prisco, A. How the interval between prime
570 and boost injection affects the immune response in a computational model of the immune
571 system. *Comput. Math. Methods Med.* 1–9 (2012).
- 572 54. Cui, J., Li, F., Shi, Z. Origin and evolution of pathogenic coronaviruses. *Nat Rev*
573 *Microbiol.* **17**, 181–192 (2019).

574 55. Wlodawer, A. Stereochemistry and validation of macromolecular structures. *Methods*
575 *Mol. Biol.* **1607**, 595–610 (2017).

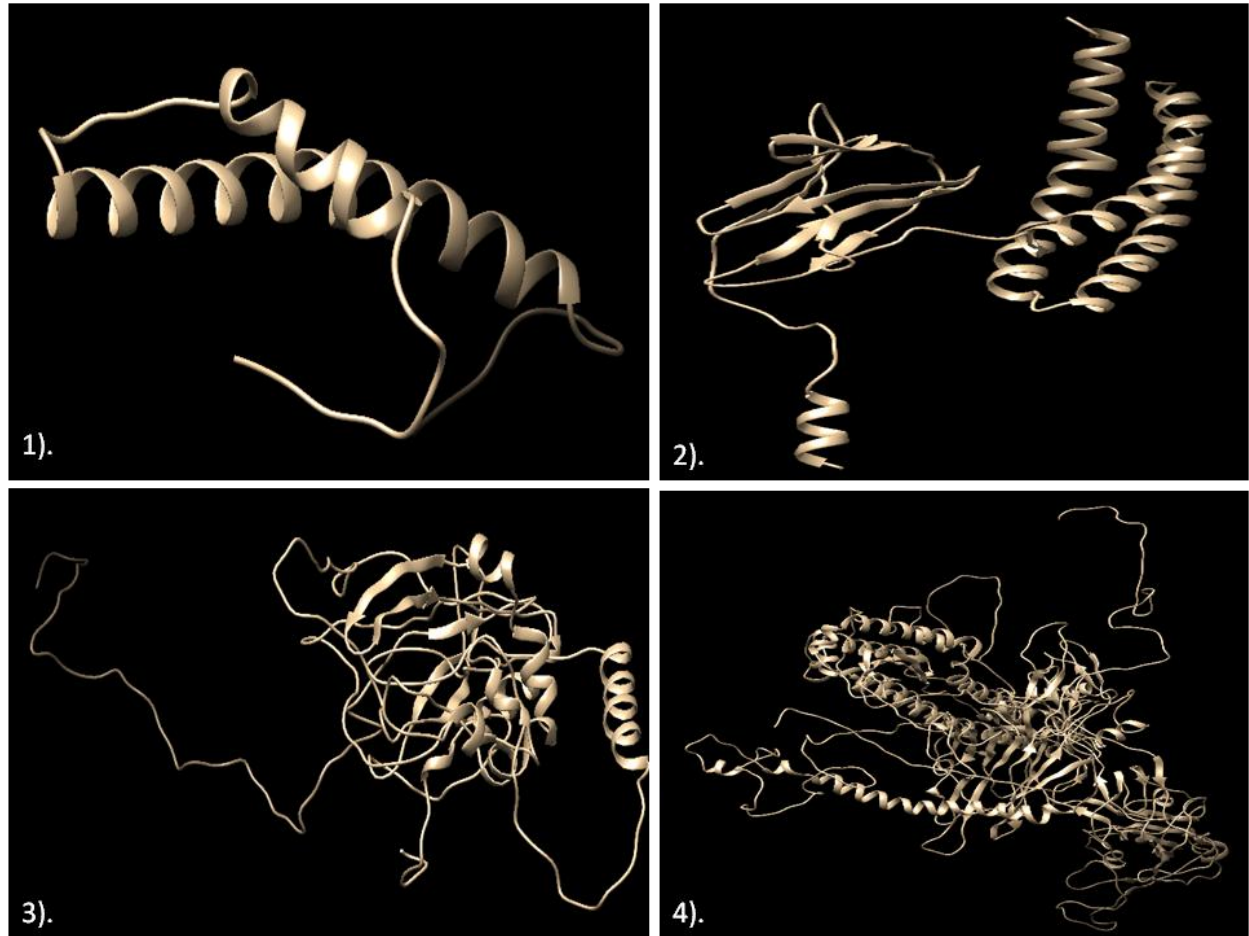


Figure 1). The 3D models of structural proteins of SARS- CoV-2.1). Envelop protein; 2). Membrane protein; 3). Nucleocapsid protein; 4). Surface protein

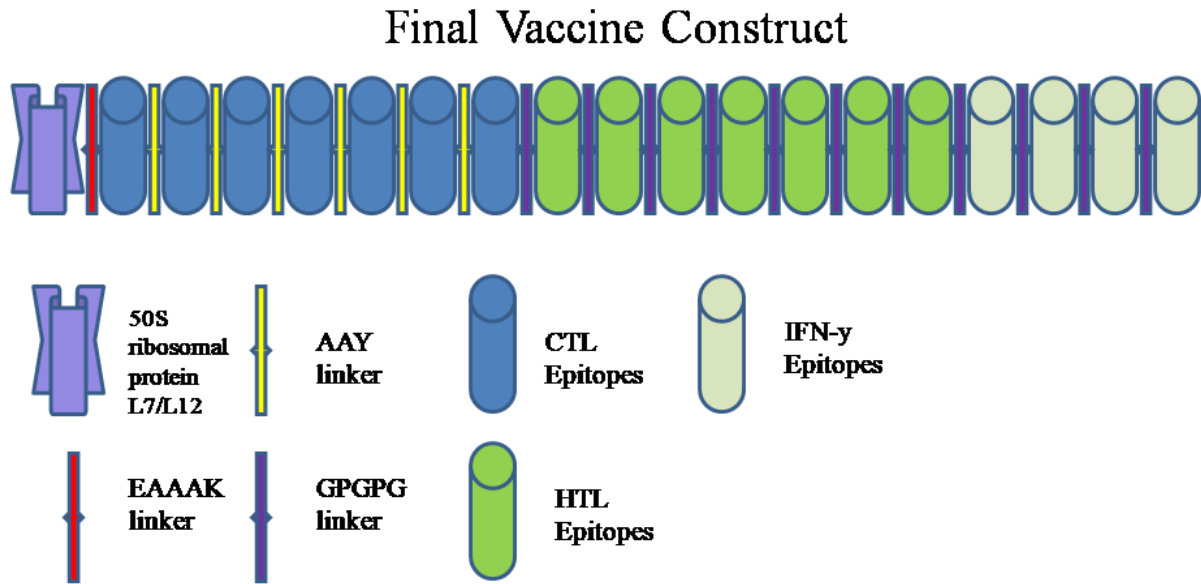


Figure 2. Schematic diagram of final vaccine construct

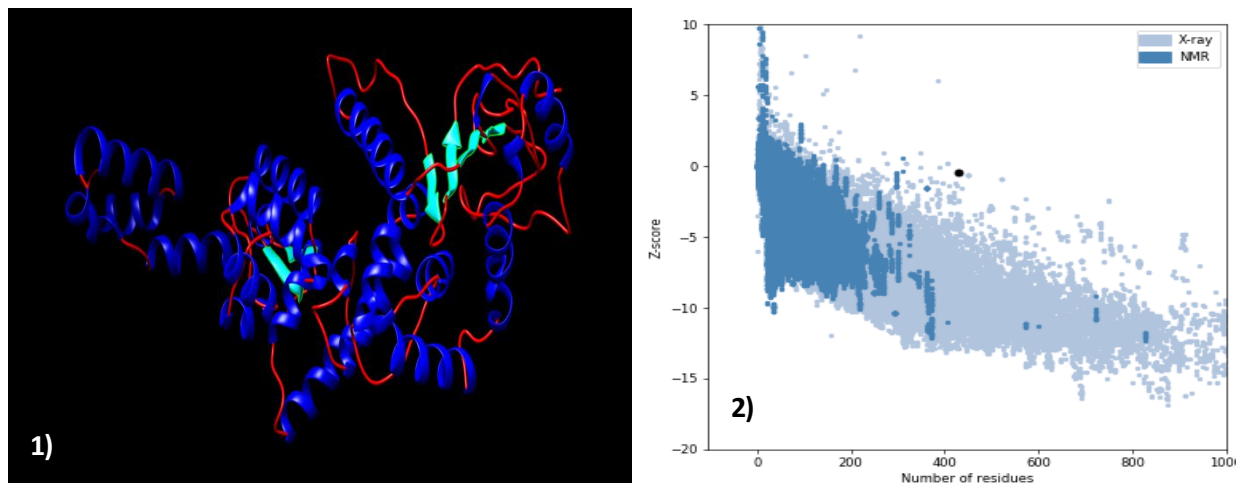


Figure 3. 1). 3 dimensional structure of final vaccine construct, 2). ProSA validation of predicted structure with Z score of -0.47.

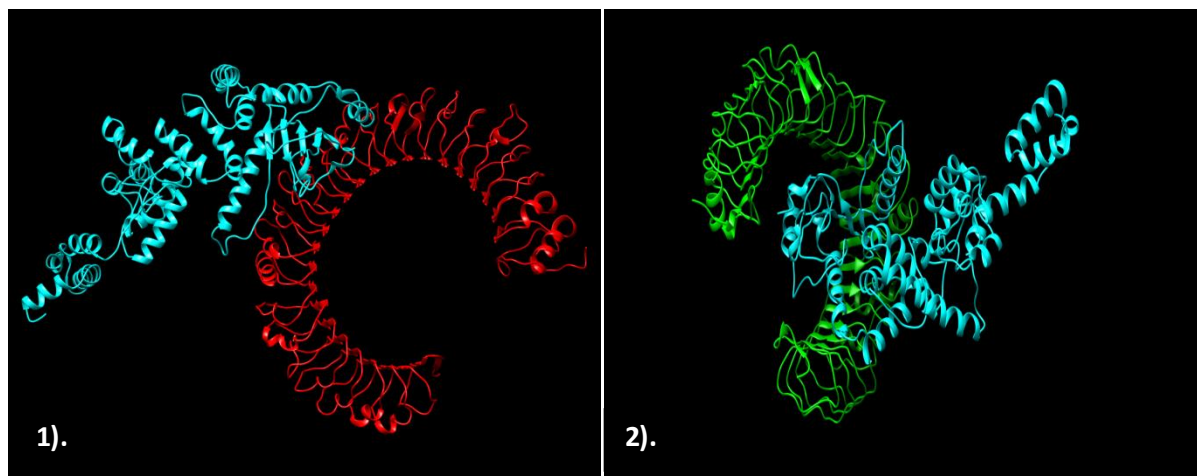


Figure 4. The interaction pattern of designed vaccine with TLR3 and TLR4. 1). Vaccine (Cyan) docked with receptor TLR3 (Red). 2). Vaccine (Cyan) docked with receptor TLR4 (Green).

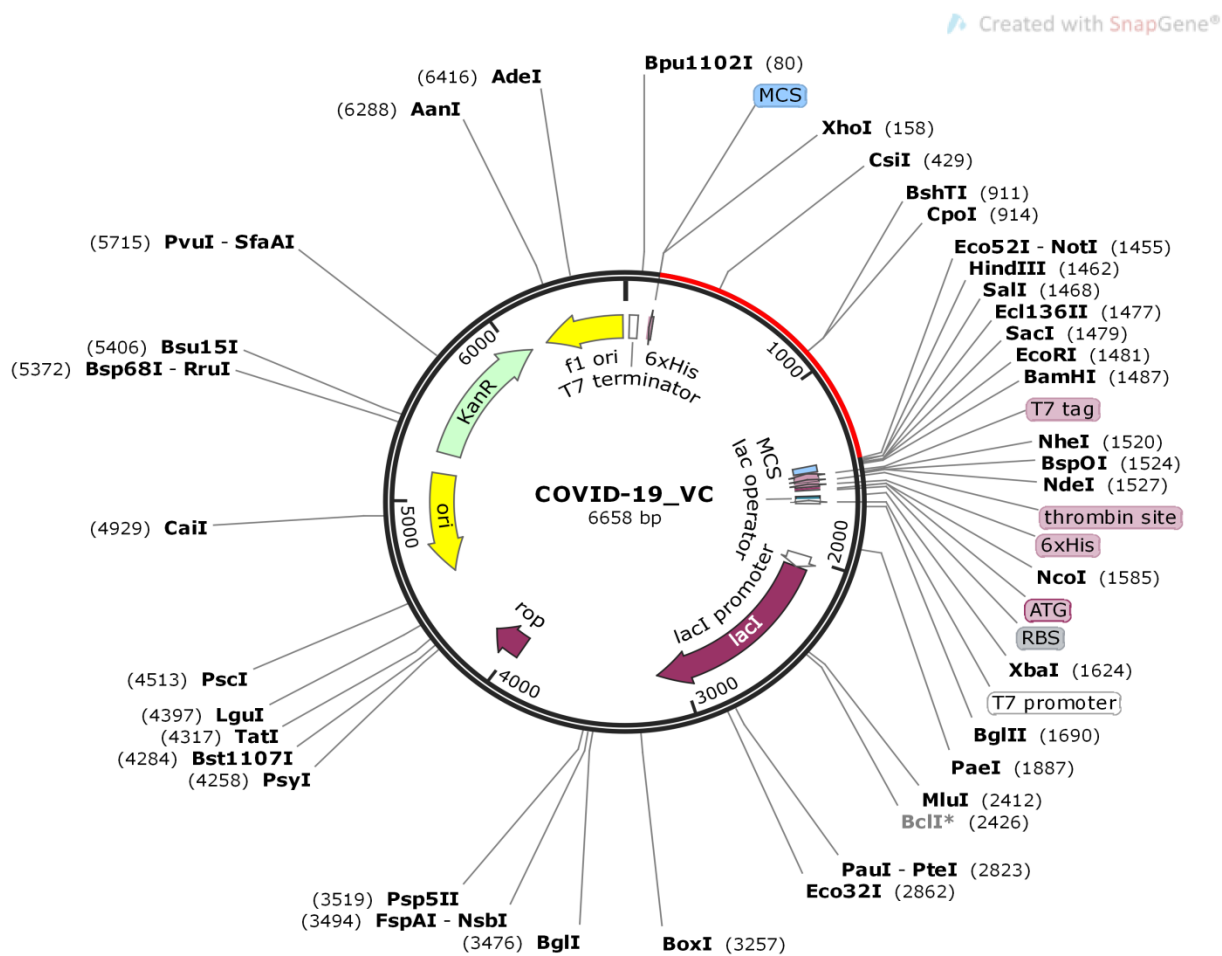


Figure 5. In silico cloning of vaccine. The segment represented in red is the multi-epitope vaccine insert in pET- 28(+) expression vector.

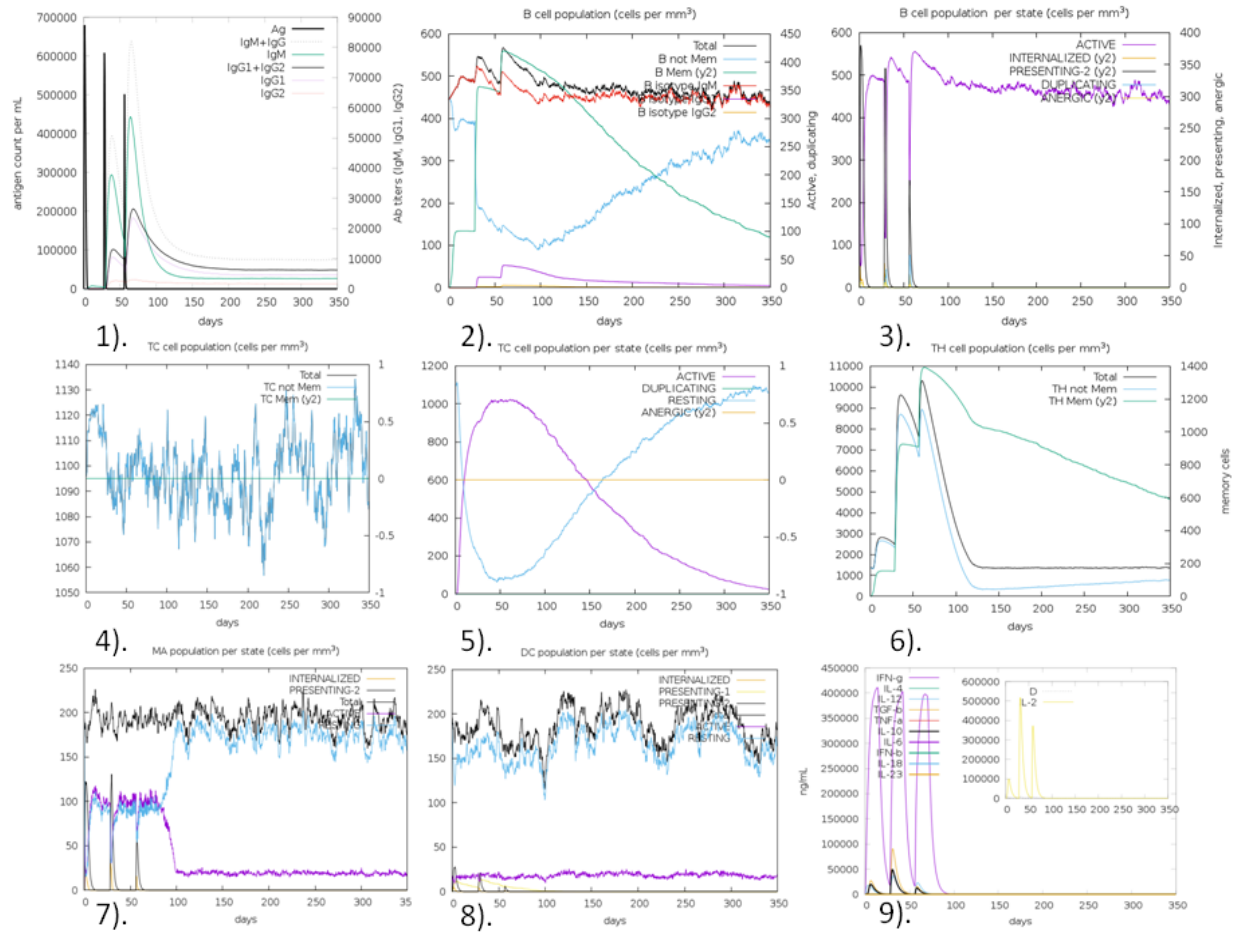


Figure. 6. In silico simulation of immune response using vaccine as an antigen after subsequent three injections. 1). Antigen and Immunoglobins. 2). B-cell population. 3). B-cell population per state. 4). Cytotoxic T-cell population. 5). Cytotoxic T-cell population per state. 6). Helper T-cell population. 7). Macrophages population per state. 8). Dendritic cell population per state. 9). Cytokine production.

Table 1. Details of target protein and their antigenicity score

Structural Protein	Protein ID (NCBI)	Size (aa)	Antigenicity score
Surface protein	QHD43416.1	1273	0.4646
Membrane protein	QHD43419.1	222	0.5102
Envelop protein	QHD43418.1	75	0.6025
Nucleocapsid protein	QHD43423.2	419	0.5059

Table 2. Final epitopes selected for multi-epitope vaccine construct after docking analysis with respective HLA alleles

S.No	Protein	HTL	Global_Energy	CTL	Global_Energy
1	Envelop	LLFLAFVVFLVTLA	-71.44	LLFLAFVVF	-41.45
2	Envelop	LAFVVFLVTLAILT	-64.39	FLLVTLAIL	-43.31
3	Membrane	TLACFVLAAYRINW	-53.12	TLACFVLAA	-54.85
4	Nucleocapsid	GDAALALLLDRLNQ	-87.95	GDAALALLL	-40.49
5	Nucleocapsid	AQFAPSASAFFGMSR	-58.74	AQFAPSASA	-43.29
6	Surface	IPFAMQMA YRFNGIG	-46.33	FAMQMA YRF	-41.37
7	Surface	FVFLVLLPLVSSQCV	-74.92	FVFLVLLPL	-55.15