Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2 etiologic agent of global pandemic COVID-19: an *in silico* approach

M. Shaminur Rahman [Equal first author, 1] , **M. Nazmul Hoque** [Equal first author, 2] , **M. Rafiul Islam** [1] , **Salma Akter** [3] , **A. S. M. Rubayet-Ul-Alam** [4] , **Mohammad Anwar Siddique** [1] , **Otun Saha** [1] , **Md. Mizanur Rahaman** [Corresp., 1] , **Munawar Sultana** [1] , **Keith A Crandall** [5] , **M. Anwar Hossain** [Corresp. 1]

1 Department of Microbiology, University of Dhaka, Dhaka, Bangladesh

2 Department of Gynecology, Obstetrics and Reproductive Health, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur, Bangladesh

3 Department of Microbiology, Jahangirnagar University, Savar, Bangladesh

4 Department of Microbiology, Jashore University of Science and Technology, Jashore, Bangladesh

5 Computational Biology Institute, Milken Institute School of Public Health, George Washington University, Washington, Washington D.C., United States

Corresponding Authors: Md. Mizanur Rahaman, M. Anwar Hossain
Email address: razu002@du.ac.bd, hossaina@du.ac.bd

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of the ongoing pandemic of coronavirus disease 2019 (COVID-19), a public health emergency of international concern declared by the World Health Organization (WHO). An immuno-informatics approach along with comparative genomic was applied to design a multi-epitope-based peptide vaccine against SARS-CoV-2 combining the antigenic epitopes of the S, M and E proteins. The tertiary structure was predicted, refined and validated using advanced bioinformatics tools. The candidate vaccine showed an average of ≥ 90.0% world population coverage for different ethnic groups. Molecular docking and dynamics simulation of the chimeric vaccine peptide with the immune receptors (TLR3 and TLR4) predicted efficient binding. Immune simulation predicted significant primary immune response with increased IgM and secondary immune response with high levels of both IgG1 and IgG2. It also increased the proliferation of T-helper cells and cytotoxic T-cells along with the increased INF-γ and IL-2 cytokines. The codon optimization and mRNA secondary structure prediction revealed the chimera is suitable for high-level expression and cloning. Overall, the constructed recombinant chimeric vaccine candidate demonstrated significant potential and can be considered for clinical validation to fight against this global threat, COVID-19.

1 **Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2**

2 **etiologic agent of global pandemic COVID-19: an *in silico* approach**

4 M. Shaminur Rahman[1*], M. Nazmul Hoque[1,2*], M. Rafiul Islam[1], Salma Akter[1,3], A. S. M.

5 Rubayet-Ul-Alam[1,4], Mohammad Anwar Siddique[1], Otun Saha[1], Md. Mizanur Rahaman[1**],

6 Munawar Sultana[1], Keith A. Crandall[5], M. Anwar Hossain[1,6**]

7 [1]Department of Microbiology, University of Dhaka, Dhaka-1000, Bangladesh

8 [2]Department of Gynecology, Obstetrics and Reproductive Health, Bangabandhu Sheikh Mujibur

9 Rahman Agricultural University, Gazipur-1706, Bangladesh

10 [3]Department of Microbiology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh

11 [4]Department of Microbiology, Jashore University of Science and Technology, Jashore-7408,

12 Bangladesh

13 [5]Computational Biology Institute, Milken Institute School of Public Health, The George

14 Washington University, USA

15 [6]Present address: Jashore University of Science and Technology, Jashore-7408, Bangladesh

16 *Equal contribution

17 **Corresponding to:
18
19 M. Anwar Hossain, PhD
20 Professor
21 Department of Microbiology
22 University of Dhaka, Dhaka, Bangladesh
23 E-mail: hossaina@du.ac.bd
24 Or,
25 Dr. Md. Mizanur Rahaman
26 Assistant Professor
27 Department of Microbiology
28 University of Dhaka, Dhaka, Bangladesh
29 E-mail: razu002@du.ac.bd

31 **Abstract**

32   Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of the

33 ongoing pandemic of coronavirus disease 2019 (COVID-19), a public health emergency of

34 international concern declared by the World Health Organization (WHO). An immuno-informatics

35 approach along with comparative genomic was applied to design a multi-epitope-based peptide

36 vaccine against SARS-CoV-2 combining the antigenic epitopes of the S, M and E proteins. The

37 tertiary structure was predicted, refined and validated using advanced bioinformatics tools. The

38 candidate vaccine showed an average of $\geq 90.0\%$ world population coverage for different ethnic

39 groups. Molecular docking and dynamics simulation of the chimeric vaccine peptide with the

40 immune receptors (TLR3 and TLR4) predicted efficient binding. Immune simulation predicted

41 significant primary immune response with increased IgM and secondary immune response with

42 high levels of both IgG1 and IgG2. It also increased the proliferation of T-helper cells and

43 cytotoxic T-cells along with the increased INF-γ and IL-2 cytokines. The codon optimization and

44 mRNA secondary structure prediction revealed the chimera is suitable for high-level expression

45 and cloning. Overall, the constructed recombinant chimeric vaccine candidate demonstrated

46 significant potential and can be considered for clinical validation to fight against this global threat,

47 COVID-19.

51 **Keywords:** SARS-CoV-2, Muti-epitope, Chimeric Peptide Vaccine, B-cell Epitope, T-cell

52 Epitope.

Introduction

Emergence of the SARS-CoV-2, which was first reported in Hubei Province of Wuhan, China in December 2019, is responsible for the ongoing global pandemic of coronavirus disease 2019 (COVID-19) spreading across 216 countries, areas or territories with 6,992,274 active infection cases and 403,128 deaths until June 9th, 2020 (WHO, 2020). This SARS-CoV-2 is the third coronovairus (CoV) belonging to Genus *Betacoronavirus* that can infect human after the two previously reported coronavirus- severe acute respiratory syndrome (SARS-CoV) (Almofti et al., 2018; Wu et al., 2020), and Middle East respiratory syndrome (MERS-CoV) (Badawi et al., 2016; Pallesen et al., 2017; ul Qamar et al., 2019). The SARS-CoV-2 has a positive-sense, single-stranded, and ~30 kilobase long RNA genome showing 79.0% and 50.0% identity to the genomes of SARS-CoV and MERS-CoV, respectively (Abdelmageed et al., 2020; Tang et al., 2020; Lu et al., 2020; Hoque et al., 2020a). Among multiple encoded proteins (structural, non-structural and accessory), four major structural proteins are the spike (S) glycoprotein, small envelope protein (E), membrane protein (M), and nucleocapsid protein (N) (Ahmed et al., 2020).

The S glycoprotein, because of its higher antigenicity and surface exposure (Almofti et al., 2018; Zhou et al., 2019; Shang et al., 2020), plays the most crucial role for the attachment and entry of viral particles into the host cells through the host angiotensin-converting enzyme 2 (ACE2) receptor (Gralinski et al., 2020; Shang et al., 2020). It is noteworthy that E and M proteins also have important functions in the viral assembly, budding and replication of virus particles, as well as play role in augmenting the immune response against SARS-CoV (Shi et al., 2006; Schoeman & Fielding, 2019; Shang et al., 2020). Monoclonal antibodies (mAbs) primarily target the trimeric S glycoprotein of the virion consisting of three homologous chains (A, B and C), and this protein is composed of two major domains, the receptor-binding domain (RBD) and the N-

76   terminal domain (NTD) (Pallesen et al., 2017; Song et al., 2018; Zhou et al., 2019; Wrapp et al.,

77   2020). The NTD is located on the side of the spike trimer and has not been observed to undergo

78   any dynamic conformational changes (Shang et al., 2020). Thus, this specific region might play

79   role in viral attachment, inducing neutralizing antibody responses and stimulating a protective

80   cellular immunity (Almofti et al., 2018; ul Qamar et al., 2019; Shang et al., 2020).

81   Most of the recent vaccine candidates induce neutralizing antibodies against the different forms

82   and/or variants of the spike protein of the SARS-CoV-2 (Le et al. 2020). However, the immune

83   responses generated from using single protein have generally been inadequate to warrant their use

84   in the development of an effective prophylactic tool (Shi et al., 2015; Shey et al., 2019). On this

85   note, multi-epitope vaccine candidates have already been designed against several viruses,

86   including MERS-CoV and SARS-CoV, and their efficacies have been further reported (Almofti et

87   al., 2018; ul Qamar et al., 2019; Yong et al., 2019). Two related studies have reported the in-silico

88   design of epitope based chimeric vaccine candidates targeting E, M, S and N proteins of SARS-

89   CoV-2, albeit not peer-reviewed (Yazdani et al., 2020; Akhand et al., 2020). Additionally, Kibria

90   et al. (2020) performed an immunoinformatic approach to design a 70 aa long multi-epitope

91   vaccine focusing on the the virion outer surface proteins (E, M, and S) (Chan et al., 2020).

92   Scientists are racing over the clock to develop effective vaccine for controlling and preventing

93   COVID-19 based on the genomics, functional structures, and host-pathogen interactions, however

94   the ultimate results of these efforts is yet uncertain. Currently, 10 candidate vaccines against

95   SARS-CoV-2 are in the clinical trial, and 121 more under preclinical evaluation (WHO, 2020).

96   Strikingly, researchers are trialing different technologies, albeit targeting spike protein mainly,

97   some of which have not been used in a licensed vaccine before (Le et al., 2020). Their vaccine

98   appears to be effective and safe based on a limited data and application of the vaccine within a

99 relatively tiny group of individuals. Nevertheless, there are several uncertainties, for example,

100 whether it will relate to antibody responses in the general population, be safe within a specific sub-

101 population (children, pregnant women, and elder people), as well as lack of a standardized virus

102 neutralization assay and accurate vaccine titer are complicating data interpretation. Moreover, only

103 half of the medium dose receiver developed neutralizing antibody and the T-cell response is not

104 particularly impressive (Sheridan, 2020).

105 Thus, mimicking of a more natural state of the virus where surface exposed proteins, or the

106 immunodominant epitopes of those proteins influence the immune response might be a solution,

107 if those candidates ultimately cannot meet the final goal. Furthermore, excluding the nucleocapsid

108 (N) protein, embedded within the structure and attached to the viral genome, will amplify the

109 chance of developing a more pseudo-virus state for the expressed chimera that may produce

110 specific antibody as well as T-cell responses. Besides, peptide-based chimeric vaccines are

111 biologically safe as they do not need *in vitro* virus culture, and their selectivity might ensure

112 accurate activation of specific immune responses (Dudek et al., 2010; Wang et al., 2019).

113 Considering the facts, we have proposed the development of a multi-epitope vaccine candidate,

114 which differs from all the previous studies in the aspect of containing whole RBD and NTD regions

115 of the spike protein alongwith specific epitopes of M and E proteins giving an excellent chimeric

116 conformation and might lead to the generation of a more potent protective immune response since

117 smaller epitopes have less ability to give better immune protection. Hence, we can assume that

118 chimeric vaccine targeting multiple epitopes on the RBD and NTD segments of the S protein, M

119 and E proteins would be a potentially effective vaccine candidate in combatting SARS-CoV-2,

120 and therefore, could be used against the globally pandemic highly pathogenic SARS-CoV-2.

121   **Results**

122   **Comparative structural analysis of SARS-CoV, MERS-CoV and SARS-CoV-2**

123   Multiple sequence alignment revealed that the S protein of SARS-CoV-2 shares 77.38%

124   and 31.93% sequence identity with the S proteins of the SARS-CoV and MERS-CoV, respectively

125   (Fig. S1). The structural (validated using the Ramachandran plot as in Fig. S2) alignment of the

126   coronavirus S proteins reflects high degree of structural heterogeneity in the receptor-binding

127   domain (RBD) and N-terminal domain (NTD) of the chain A and chain C compared to that of

128   chain B (Fig. S3). Divergence of individual structural domains, NTD and RBD of 2019-nCoV

129   spike protein from both of the SARS-CoV and MERS-CoV warrants the domains for epitope-

130   based chimeric vaccine development, particularly against SARS-CoV-2.

131   **Screening for B-cell epitopes**

132   Linear epitopes prediction (ElliPro) based on solvent-accessibility and flexibility revealed

133   15, 18, and 19 epitopes within the chain A, B and C of S protein, respectively wherein score > 0.8

134   was the threshold for the highly antigenic epitopes (Table 1). The amino acid (aa) residues in 56-

135   194 and 395-514 position of the detected epitopes belonged to RBD and NTD regions of the S

136   protein, respectively. However, the epitopes with aa position of 1067-1146 were not selected as

137   the potential epitope candidate because of their presence in viral transmembrane domains (Fig.

138   S4). The tertiary structures of the RBD and NTD illustrate their surface-exposed position on the S

139   protein (Fig. 1). Using IEDB analysis resource and Bepipred linear epitope prediction 2.0 tools,

140   we predicted eight and six B-cell epitopes in RBD and NTD regions out of total 22 epitopes in S

141   protein, while the E and M proteins had 2 and 6 epitopes, respectively (Fig. 2, Table S1). However,

142 only 5 epitopes were exposed on the surface of the virion, and had a high antigenicity score (>

143 0.4), indicating their potentials in initiating immune responses (Table 2).

144   Among the five annotated epitopes having antigenicity score of ≥ 0.5 (VaxiJen 2.0 tool),

145 RBD and NTD regions each possessed two highly antigenic epitopes while the envelope (E)

146 protein contained only one highly antigenic epitope and membrane (M) protein has none (Table

147 2). Furthermore, the Kolaskar and Tongaonkar antigenicity profiling found five highly antigenic

148 epitopes in RBD region with an average (antigenicity) score of 1.042 (minimum = 0.907,

149 maximum = 1.214), and seven highly antigenic epitopes in NTD with an average (antigenicity)

150 score of 1.023 (minimum = 0.866, maximum = 1.213) (Fig. S5, Table S2). The average Kolaskar

151 scores for envelope protein B-cell epitope (EBE) and membrane protein B-cell epitope (MBE)

152 were 0.980 and 1.032, respectively (Table S2). However, through ABCPred analysis, we further

153 verified 18 and 11 B-cell epitopes in RBD and NTD regions with average antigenicity score of

154 0.775 and 0.773 in the associated domains, respectively (Table S3).

155 **Selection of T-cell and IFN-γ inducing epitopes**

156   The IEDB MHC-I prediction tool retrieved 77 T-cell epitopes in RBD that interacted with

157 21 possible MHC-I alleles whereas the NTD domain possessed 35 T-cell epitopes with 17 possible

158 MHC-I alleles (Supplemental Data 1). Similarly, the IEDB MHC-II prediction tool generated 13-

159 mer 124 peptides from the RBD, and 10-mer 73 peptides in the NTD segments of the S protein

160 that showed interaction with many different and/or common MHC-II alleles with an IC_{50} value

161 ranging from 1.4 to 49.9 nM (Supplemental Data 1). Furthermore, the analysis tool of the IEDB

162 generated an overall scores for proteasomal processing, TAP transport, and MHC-binding

163   efficiency indicating the intrinsic potential of the epitopes to be recognized by immunoreactive T-

164   cells (Supplemental Data 1).

165   The findings of IFNepitope program suggests that, both the target RBD and NTD regions

166   of S protein, and membrane protein B-cell linear epitope (MBE) had great probability to release

167   of IFN-γ with a positive score. A total of 56 potential positive IFN-γ inducing epitopes (15-mer)

168   were predicted for the RBD domain with an average epitope prediction score of 0.255 and the

169   maximum SVM score of 0.625. On the other hand, a total of 33 potential positive epitopes were

170   predicted for the NTD domain with an average epitope prediction score of 0.312 and the maximum

171   SVM score of 0.811. Moreover, the M protein also possessed several IFN-γ inducing epitopes

172   having an average epitope prediction score of 0.980 (Table S4).

173   **Design-construction, antigenicity and physiochemical properties of the chimeric vaccine**

174   **candidate**

175   The selected epitope-sequences for designing of chimeric construct were PADRE (13 aa),

176   MBE (20 aa), NTD (139 aa), RBD (200 aa), EBE (15 aa), and Invasin (16 aa), and the construct

177   was named as CoV-RMEN (417 aa) as shown in Fig. 3a. These segments were connected with a

178   repeat of hydrophobic (glycine; G) and acidic aa (glutamic acid; E) linkers for making the final

179   vaccine construct more flexibile with balanced ratio of acidic and basic amino acids. The molecular

180   weight (MW) of the CoV-RMEN was 46.8 kDa with a predicted isoelectric point (pI) of 8.71. The

181   projected half-life was 4.4 hours in mammalian reticulocytes in vitro, and >20 hours in yeast and

182   >10 hours in *E. coli in vivo*. The protein was predicted to be less soluble upon expression with a

183   solubility score of 0.330. An instability index (II) value 29.74 predicted the protein as stable (II of

184   > 40 indicates instability). The estimated aliphatic index was 66.59, indicating thermostability of

185   the final chimera. The predicted grand average of hydropathicity (GRAVY) was −0.300. The

186   antigenicity score of 0.450 was predicted by the VaxiJen 2.0 server with a virus model at a

187   threshold of 0.4, and further verified by ANTIGENpro showing score of 0.875 (maximum

188   expected score ranking is 1.0) indicating the high antigenic nature of the designed vaccine, CoV-

189   RMEN . Moreover, the vaccine was also predicted to be non-allergenic on both the AllerTOP v.2

190   and AllergenFP servers.

191   **Structural characterization of the CoV-RMEN**

192   The CoV-RMEN peptide was predicted to contain 43.2% alpha helix, 67.4% beta sheet,

193   and 12% turns (Fig. 3b, Fig. S6) using CFSSP:Chou and Fasman secondary structure prediction

194   server. Additionally, regarding the solvent accessibility of aa residues, 34%, 30% and 34% were

195   predicted to be exposed, medium exposed and buried respectively (Fig. S7). The RaptorX Property

196   server predicted only two aa residues in the disordered domains. The Phyre2 server predicted the

197   tertiary structure model of the designed chimeric protein in 5 templates (c5x5bB, c2mm4A,

198   c6vsbB, c5x29B and c6vybB) based opon heuristics to maximize confidence, alignment coverage

199   and percent identity. The final 3D structure of the CoV-RMEN peptide modelled at 82% with more

200   than 90% confidence (Fig. 3c). Moreover, 65 residues were modelled by *ab initio*. The selected

201   structural model has parameters of RMSD (0.414), GDT-HA (0.9538), and MolProbity (2.035).

202   The Ramachandran plot analysis of the finally modelled protein exhibited 94.7% of the aa residues

203   in favored regions (Fig. 3d), consistent with the 94.0% score predicted by the GalaxyRefine

204   analysis. Additionally, 4.8% of the residues located in allowed regions, and only 0.5% in

205   disallowed regions (Fig. 3d). The chosen model after refinement had an overall quality factor of

206   74.45% with ERRAT (Fig. S8) and a ProSA-web based Z-score of − 6.17(Fig. 3e).

207   **Molecular docking and dynamics simulation analysis**

208   Among the selected epitopes from the RBD and NTD segments, top five based on IC_{50}

209   score (Table 2) revealed highly favorable molecular interaction for stable binding with their

210   respective HLA alleles. Docking complexes thus formed have significantly negative binding

211   affinity (ΔG always remained $\leq$ -8.2 kcal mol^{-1}, average = -9.94 kcal mol^{-1}), and most of the amino

212   acid (aa) residues of the epitopes were involved in molecular interactions with their respective

213   HLA alleles (Fig. 4, Supplemental data 1). The immune responses of TLR2, TLR3 and TLR4

214   against vaccine construct (CoV-RMEN) were estimated by analyzing the overall conformational

215   stability of vaccine protein-TLRs docked complexes. The active interface aa residues of refined

216   complexes of CoV-RMEN and TLRs were predicted (Fig. 5, Table 3).

217   The relative binding free energies (ΔG) of the protein-TLRs complexes were significantly

218   negative (Table 3) which suggest that the interaction of the chimeric protein might favor

219   stimulation of the TLR receptors. Consistently, the number of contacts made at the interface (IC)

220   per property (ICs charged-charged: 5, ICs charged-polar: 2, ICs charged-apolar: 17, polar-polar:

221   1, ICs polar-apolar: 7 and apolar-apolar: 16) for the vaccine protein-TLR2 complex. Interface

222   contacts (IC) per property (ICs charged-charged: 16, ICs charged-polar: 22, ICs charged-apolar:

223   26, polar-polar: 6, ICs polar-apolar: 25 and apolar-apolar: 29) were for the vaccine protein-TLR3

224   complex. Also, vaccine protein-TLR4 complex showed similar (ICs) per property (ICs charged-

225   charged: 5, ICs charged-polar: 11, ICs charged-apolar: 30, polar-polar: 4, ICs polar-apolar: 31 and

226   apolar-apolar: 39). Furthermore, the molecular dynamics (MD) simulation analysis of the docked

227   CoV-RMEN-TLR3 and CoV-RMEN-TLR4 complexes showed soundly stable RMSD values

228   between ~4.35 and ~5.4 nm for a specified time frame of 100 ps at the reasonably consistent

229   temperature (~300 K) and pressure (1bar), whereas CoV-RMEN-TLR2 complex showed RMSD

230   value between 5.5 and 6.2 with same cut-off parameters. These data validated that the docked

231   complexes (CoV-RMEN-TLR3 and CoV-RMEN -TLR4) are more stable than CoV-RMEN-TLR2

232   (Fig. 5).

233   **Immune simulation**

234   The cumulative results of immune responses after three times antigen exposure with four-

235   week interval each time revealed that the primary immune response against the antigenic fragments

236   was elevated indicated by gradual increase of IgM level after each antigen exposure (Fig. 6a).

237   Besides, the secondary immune response, crucial for immune stability, have been shown as

238   increased with adequate generation of both IgG1 and IgG2. Also, the elevated level of all

239   circulating immunoglobulins indicates the accuracy of relevant clonal proliferation of B-cell and

240   T-cell population. The level of cytokines after antigen exposure increased concomitantly reflected

241   by escalation of IFN-γ and IL-2, which are most significant cytokines for anti-viral immune

242   response and clonal selection (Fig. 6b). The abundance of different types of B-cells and T-cells,

243   like antigen processing B-cells, resting memory B- and T- cells, B-cells with IgM and IgG remains

244   significantly higher indicating development of immune memory and consequently increased

245   clearance of antigen after exposure (Fig. 6c,d). Additionally, T-helper cells and cytotoxic T-cells

246   were found with a drastic up-regulation of Th1 concentration enhancing the B-cell proliferation

247   and immune memory development (Fig. 6e,f). The high level of immunoglobulin IgG1 + IgG2,

248   active B-cell and T-helper cell population reflected the development of strong immune response

249   reinforcing the indelible and peerless antigenicity of the CoV-RMEN vaccine candidate.

250   **Population coverage analysis**

251   The selected CTL and HTL epitopes covered 94.9% and 73.11% of the world population,

252   respectively. Importantly, CTL and HTL epitopes showed 98.63% population coverage worldwide

253   when considered in combination. The highest population coverage was found to be 99.99% in the

254   Latin American country, Peru (Fig. 7, Supplemental Data 2). In China, where the viral strain

255   (SARS-CoV-2) first appeared and had more devastating outbreaks, the population coverage for

256   CTL and HTL epitopes was 92.67% and 53.44%, respectively with a combined coverage of

257   96.59%. SARS-CoV-2 is currently causing serious pandemics in different continents of the globe

258   including Italy, England, Spain, Iran, South Korea and United States of America where the

259   combined population coverage was found to be 98.8%, 99.44%, 95.35%, 98.48%, 99.19% and

260   99.35%, respectively (Fig. 7a, Supplemental Data 2). In addition to geographical distribution, the

261   ethnic groups also found to be an important determinant for good coverage of the CTL and HTL

262   epitopes (Fig. 7b). Of the studied 147 ethnic groups, the Peru Amerindian had highest population

263   coverage for CTL (99.98%) while the HTL epitopes had highest population coverage for Austria

264   Caucasoid (88.44%) (Fig. 7b, Supplemental Data 2). Furthermore, 53.06% of the ethnic groups

265   had a combined population coverage of more than 90.0% for both CTL and HTL epitopes.

266   **Expression prediction of the CoV-RMEN**

267   The length of the optimized codon sequence of the vaccine construct CoV-RMEN in *E.*

268   *coli* (strain K12) was 1,251 nucleotides. The optimized nucleotide sequence had a Codon

269   Adaptation Index (CAI) of 0.87, and the average GC content of 50.26% showing the possibility of

270   good expression of the vaccine candidate in the *E. coli* host (Fig. 8a-c). Moreover, the evaluation

271   of minimum free energy for 25 structures of chimeric mRNA through Mfold'server showed that

272   ΔG of the best predicted structure for the optimized construct was $\Delta G = -386.50$ kcal/mol. The

273   first nucleotides at 5′ did not have a long stable hairpin or pseudoknot. Therefore, the binding of

274   ribosomes to the translation initiation site, and the following translation process can be readily

275   accomplished in the target host. These outcomes were in the agreement with data obtained from

276   the 'RNAfold'web server (Fig. 8 d,e) where the free energy was -391.37 kcal/mol.

277   After codon optimization and mRNA secondary structure analysis, the sequence of the

278   recombinant plasmid was designed by inserting the adapted codon sequences into pETite vector

279   (Lucigen, USA), which contains SUMO (Small Ubiquitin-like Modifier) tag and 6x-His tag

280   facilitating both the solubilization and affinity purification of the recombinant protein using

281   SnapGene software (Fig. 9). As alternative to *E. coli* for the expression system, HEK-293

282   eukaryotic cell line found promising for CoV-RMEN expression. The codon adaption index (CAI),

283   GC content for this system were 1.0 and 61.60 respectively, which indicate high level of expression

284   of the vaccine construct in the HEK-293 cell line as well.

285   **Discussion**

286   SARS-CoV-2, the virus with high zoonotic importance and transmission rate, has spread

287   rapidly around the world and causes life-threatening COVID-19 (Gorbalenya et al., 2020). The

288   number of SARS-CoV-2 infections, and subsequent deaths are increasing day by day (Tai et al.,

289   2020), and thus, COVID-19 outbreak was declared as a public health emergency by the

290   International Concern (HuiDS et al., 2020; Zhou et al., 2020a). Scientific community from across

291   the world on its own way is trying to find out and/or develop an effective and safe vaccine against

292   this rapidly emerging pandemic SARS-CoV-2 (Abdelmageed et al., 2020). Although a good

293   number of vaccine candidates for COVID-19 are now under trials, some of them are advanced to

294   human trials (Lane, 2020), none has yet been declared to be effective and safe for prevention of

295   SARS-CoV-2 infections. Strikingly, no effective therapeutic drugs or vaccines are yet to be

available for the treatment of SARS-CoV-2 patients (Hoque et al., 2020a). Through a

297 comprehensive genomic and proteomic study, we endeavor to desing an antigenic multi-epitope

298 (immunodominant) chimeric vaccine for SARS-CoV-2, named as CoV-RMEN (417 aa), which

299 will nullify the involvement of lab-escape viral transmission, reduce the cost, and may elicit

300 immunity by selectively stimulating antigen-specific B- and T-cells.

301   The novel approach of multi-epitope based (includes conserved multiple epitopes) vaccines

302 designing represents inducing specific cellular immunity, and highly potent neutralizing antibodies

303 against infections (Dawood et al., 2019; Yong et al., 2019; Gralinski & Menachery, 2020; Kibria

304 et al., 2020). These epitope-based vaccines also provide increased safety and have the ability to

305 focus on sustainable immune responses because of including conserved multiple epitopes. Unlike

306 the full-length S protein, the RBD and NTD segments possess critical neutralizing domains

307 without any non-neutralizing immunodominant region (ul Qamar et al., 2019; Gralinski &

308 Menachery, 2020; Shang et al., 2020; Wrapp et al., 2020). Mutations on the RBD may enable the

309 new strains to escape neutralization by established RBD-targeting antibodies, hence other

310 functional regions, especially the NTD, should be considered for developing an effective vaccine

311 as well (Wang et al., 2019; Zhou et al., 2019). Besides, combined administration of RBD and NTD

312 proteins induced highly potent neutralizing antibodies and long-term protective immunity in

313 animal models (Song et al., 2018). Considering the safety and effectiveness perspectives, the RBD

314 and NTD are more promising candidates in the development of SARS-CoV-2 vaccines over the

315 full-length S protein. The presence of E and M proteins on the envelope can augment the immune

316 response against SARS-CoV (Millet & Whittaker, 2015; Almofti et al., 2018) and thus, considered

317 for suitable candidate for vaccine development (Yong et al., 2020; Ahmed et al., 2020; Gralinski

318 & Menachery, 2020). Thus, antibodies against the immunologically substantial epitopes of S, M

319 and E proteins of SARS-CoV-2 would provide protective immunity to the infection (Yong et al.,

320 2020; Ahmed et al., 2020; Gralinski & Menachery, 2020; Shang et al., 2020). Therefore, the

321 immune response targeting the RBD and/or NTD of the S glycoprotein, M and E proteins of SARS-

322 CoV-2 would be an important prophylactic and therapeutic interventions, which can be tested

323 further in suitable models before clinical trials (Chan et al., 2020).

324   Effective immunity to viral infections is significantly dependent on activation of both B-

325 and T-cells (Shi et al., 2015; Shey et al., 2019). Therefore, inducing specific humoral or cellular

326 immunity against pathogens, an ideal vaccine should contain both B-cell and T-cell epitopes. Our

327 analyses revealed that selected RBD and NTD regions of the CoV-RMEN contain ample amount

328 of high-affinity B-cell, MHC Class I, MHC Class II and interferon-γ (IFN-γ) epitopes with high

329 antigenicity scores. Moreover, membrane B-cell epitope (MBE) and envelope B-cell epitope

330 (EBE) enhanced the overall stability, immunogenicity and antigenicity of the CoV-RMEN. The

331 development of memory B-cells and T-cells was evident, with memory in B-cells lasting for

332 several months. These finding opposed to several earlier reports where T-cell mediated immune

333 response was considered a long-lasting response compared to B-cells (Abdelmageed et al., 2020;

334 Wrapp et al., 2020). Another engrossing finding of this study was the development of Th1 response

335 which enhances the growth and proliferation of B- cells augmenting the adaptive immunity

336 (Carvalho et al., 2002). If a strong B-cell response occurred in animal trials (mice or rabbit), these

337 antibodies could be used in diagnostic purposes, as they should recognize the prominent antigens

338 on the viral surface (Kibria et al., 2020). Moreover, CD8+ and CD4+ T-cell responses play major

339 role in antiviral immunity (Abdelmageed et al., 2020). Another crucial fact is that Toll-Like

340 Receptors (TLRs) can effectively bind with spike protein of the CoV (Totura et al., 2015; Zander

341 et al., 2017), and might play an important role in the innate immune response to SARS-CoV-2

342 infection (Shahabi et al., 2020).

343   The physicochemical properties also revealed the chimera as a basic or alkaline protein

344 (pI= 8.71) and would be thermostable upon expression, and thus, our proposed vaccine CoV-

345 RMEN would be best suited for worldwide use in different endemic areas (Shey et al., 2019; ul

346 Qamar et al., 2019). The structural forms (secondary and tertiary) of the CoV-RMEN, when tested

347 as the synthetic peptides, showed the ability to fold into their native structure, hence could mimic

348 the natural infection by SARS-CoV-2 (Almofti et al., 2018). The refined tertiary (3D) structure of

349 the final vaccine construct markedly presented the desirable structural features based on

350 Ramachandran plot predictions (Shey et al., 2019; Srivastava et al., 2019). Molecular docking

351 analysis showed that predicted chimeric protein can establish stable protein-protein interactions

352 with TLRs (TLR-2, TLR-3, TLR-4) (Totura et al., 2015). An efficient activation of surface

353 molecules of the CoV-RMEN is very crucial for immune activation of dendritic cells, and

354 subsequent antigen processing and presentation to CD4+ and CD8+ T-cells via MHC-II and MHC-

355 1, respectively (Shi et al., 2015; Shey et al., 2019; Shang et al., 2020). The molecular dynamics

356 simulation also revealed that the docked CoV-RMEN-TLRs complexes were stable, and had more

357 binding affinity TLR-3 and TLR-4 (Abraham et al., 2015). Furthermore, the CoV-RMEN showed

358 good antigenicity scores on Vaxijen v2.0 and ANTIGENpro indicating that these peptide

359 sequences are supposed to be highly antigenic in nature (Shey et al., 2019). The non-allergenic

360 properties of the CoV-RMEN further strengthens its potential as a vaccine candidate (Shey et al.,

361 2019; ul Qamar et al., 2019).

362   Immune simulation of the CoV-RMEN exhibited expected results consistent with typical

363 immune responses, and there was a growing immune responses after the recurrent antigen

364   exposures (Fig. 6). The antiviral cytokine IFN-γ and cell stimulatory IL-2 level significantly

365   increased, which also contribute to the subsequent immune response after vaccination in host

366   (Almofti et al., 2018). This indicates high levels of helper T-cells and consequently efficient

367   antibody production, supporting a humoral response (Shey et al., 2019). A lower IC_{50} value

368   indicates higher binding affinity of the epitopes with the MHC class I and II molecules. While

369   most of the previous studies (Sakib et al., 2014; Adhikari et al., 2018) reported that a binding

370   affinity (IC_{50}) threshold of 250 nM identifies peptide binders recognized by T-cells, and this

371   threshold can be used to select peptides, we kept binding affinity within 50 nM to get better

372   confidence level in predicting epitopes for MHC-I and MHC-II alleles.The Simpson index

373   estimated clonal specificity suggested a possible diverse immune response and this is plausible

374   considering the generated chimeric peptide is composed of sufficient B- and T-cell epitopes (Fig.

375   6).

376   The interaction between T-cell epitopes, and their respective HLA alleles revealed

377   significant binding affinity reflecting the immune activation of B- and T-cells as supported by

378   other reports (Srivastava et al., 2019; Jaimes et al., 2020). T-cell epitopes from RBD and NTD

379   regions showing high interaction with HLA alleles covered more than 98% of the world population

380   with different ethnic groups, and these findings corroborated with many of earlier studies (Huang

381   et al., 2007; Jaimes et al., 2020; ul Qamar et al., 2019). The incorporation of GG and EGGE linkers

382   between the predicted epitopes of the CoV-RMEN produced sequences with minimized junctional

383   immunogenicity, and allowed the rational design construction of a potent multi-epitope vaccine

384   (Huang et al., 2007; Badawi et al., 2016; Shey et al., 2019; Srivastava et al., 2019). The

385   glycosylation of the surface antigens helps the enveloped viruses evade recognition by the host

386   immune system, and can influence the ability of the host to raise an effective adaptive immune

387   response (Pereira et al., 2018) or even be exploited by the virus to enhance infectivity (Wolfert et

388   al., 2013). Moreover, some antibodies such as mAb 8ANC195 have evolved to recognize peptide

389   epitope with no dependence on glycan binding (Kong et al., 2015). However, there is no data

390   available for antibodies specific to spike glycoproteins of SARS-CoV-2, whether their recognition

391   is interfered by the glycosylation of spike or may either be strengthened by sugars close to the

392   peptide epitope, or not interfered by sugar modification (Zhou et al., 2020b). Most of the epitopes

393   of the CoV-RMEN harbor no glycoside except NTD region (Watanabe et al, 2020). Most of the

394   glycans of the NTD epitopes present at the terminii of the putative HLA antigens, and may not

395   interfere with antigen presentation in an HLA complex (Grant et al., 2020). Integrity of the ACE2

396   receptor of the RBD, envelope protein B-cell epitope (EBE) and membrane protein B-cell epitope

397   (MBE) in the CoV-RMEN suggests that the vaccine may maintain efficacy despite antigenic drift

398   and glycosylation phenomena, as long as the virus continues to target the same host receptor (Grant

399   et al., 2020).

400   One of the first steps in validating a candidate vaccine is to screen for immunoreactivity

401   through serological analysis. This requires the expression of the recombinant protein in a suitable

402   host. As we focused on the epitopes without the glycosylation or non-significant glycosylation,

403   high-level expression of the vaccine was optimized into well-established and cost-effective

404   prokaryotic expression system *E. coli* K-12 strain as the first choice using the plasmid pETite

405   containing SUMO (Small Ubiquitin-like Modifier) tag and 6x-His tag facilitated both the

406   solubilization and affinity purification of the recombinant protein (Biswal et al., 2015). Codon

407   optimization of the CoV-RMEN revealed its high-level expression in *E. coli* (strain K12). Stable

408   mRNA structure, codon adaptability index (0.87), and the GC content (50.26%) were favourable

409   for high-level expression of the protein in the bacterium. After successful cloning of the gene, the

410 recombinant plasmid can be propagated efficiently using *E. coli* cells, and subsequent protein

411 expression can be performed in *E. coli* K-12 strain using IPTG (Isopropyl β-d-1-

412 thiogalactopyranoside) induction, and cultivation at 28 °C as also reported earlier (Biswal et al.,

413 2015).

414 As alternative to *E. coli*, eukaryotic cell line, HEK-293 was considered for CoV-RMEN

415 expression. The codon adaption index (CAI), GC content were 1.0 and 61.60 respectively, wihch

416 indicate high level of expression of vaccine construct in the HEK-293 cell line. In this case, pSec

417 Tag2 Mammalian expression vector could be used, which have secretion signal from the V-J2-C

418 region of the mouse Ig kappa-chain for efficient secretion of the recombinant protein, C-terminal

419 poly-histidine (6xHis) tag for rapid purification with C-termnal c-myc epitope for detection with

420 an Anti-myc antibody. To remove the affinity and detection tags (His tag and c-myc epitope) after

421 purification, Factor Xa clevage site (LVPR↓GS) could be added to the C-terminal of the CoV-

422 RMEN (Waugh et al., 2011).

423 **Conclusions**

424 This multi-epitope peptide vaccine candidate, CoV-RMEN possesses potential epitopes

425 form the RBD and NTD segments of spike (S), M and E proteins retaining potential antigenicity

426 and non-allergenicity properties. This chimera, suitable for high-level expression and cloning,

427 includes potential CTL, HTL and B-cell epitopes ensuring humoral and cell mediated immunity,

428 as well as predicted immune-simulation refers to increased production of immunoglobulins and

429 cytokines. Molecular docking and dynamic simulation of the CoV-RMEN with the immune

430 receptors (TLRs) predicted strong binding affinity, in particular with TLR3 and TLR4.

431 Remarkably, the CoV-RMEN had more than 90.0% world population coverage for different ethnic

432   groups. The limitations posed by fewer number of SARS-CoV-2 geneome sequence data which

433   tends to mutate frequently may not affect our analysis since we included four peptides from three

434   major proteins of SARS-CoV-2 genome in a multi-epitope vaccine with the high conservancy.

435   However, future *in vitro* and *in vivo* studies are required to assess the potentiality of the designed

436   vaccine candidate to induce a positive immune response against SARS-CoV–2 infections, and also

437   to validate the results obtained herein through immuno-informatics analyses.

438   **Methods**

439   **Sequence retrieval and structural analysis**

440   A total of 250 partial and complete genome sequences of SARS-CoV-2 were retrieved

441   from   NCBI   (National   Center   for   Biotechnology   Information,

442   https://www.ncbi.nlm.nih.gov/protein) (Table S5). We aligned these sequences through MAFFT

443   online server (https://mafft.cbrc.jp/alignment/server/) using default parameters, and Wu-Kabat

444   protein   variability   was   analyzed   (Fig.   S9)   in   protein   variability   server

445   (http://imed.med.ucm.es/PVS/) for SARS-CoV-2 NCBI reference genome (Accession no :

446   NC_045512.2). We retrieved the S protein sequences of the SARS-CoV and MERS-CoV from the

447   whole genome reference sequences of the respective three viruses from the NCBI database.

448   Moreover, the S proteins of SARS-CoV (GenBank accession no: NC_004718.3), MERS-CoV

449   (GenBank accession no: NC_019843.3) and SARS-CoV-2 were structurally compared using

450   SWISS homology modeling (Waterhouse et al., 2018) based on the protein databank (PDB)

451   templates 6acd, 5w9h and 6vsb, respectively aligned using PyMOL (Faure et al., 2019), and

452   observed for heterogeneous domains in the conformations. The models were optimized by energy

453   minimization using the GROMOS96 program (Van Gunsteren et al., 1996) implemented within

454   the Swiss-PdbViewer, version 4.1.0 (Guex & Peitsch, 1997). The Ramachandran plots of the

455   derived models were evaluated using a PROCHECK (version 3.5)-server to check the

456   stereochemical properties of the modeled structures (Laskowski et al., 1993). All further analyses

457   including epitopes selection, antigenicity and allergenicity profiles, molecular docking and final

458   vaccine construct were performed based on the NCBI reference genome of SARS-CoV-2.

459   **Screening for B and T cell epitopes**

460   Conformational B-cell epitopes on the S protein were predicted by ElliPro

461   (http://tools.iedb.org/ellipro/) with the minimum score value set at 0.4 while the maximum distance

462   selected as 6 Å (Kringelum et al., 2012). The linear B-cell epitopes of RBD and NTD regions of

463   S protein, full length E and M proteins were predicted by "BepiPred Linear Epitope Prediction"

464   (Larsen et al., 2006), and ABCPred with default parameters (Saha & Raghava, 2006). To find out

465   the most probable peptide-based epitopes with better confidence level, antigenecity of the

466   predicted peptides were further verified using VaxiJen antigenicity scores (Kringelum et al., 2012).

467   The Kolaskar and Tongaonkar antigenicity profiling from IEDB analysis resource was also used

468   for the RBD and NTD segments, E and M proteins (Kolaskar & Tongaonkar, 1990).

469   CTL epitopes, proteasomal cleavage and transporter associated with antigen processing

470   (TAP), and HTL epitopes of the SARS-CoV-2 S, E and M proteins were predicted using IEDB

471   resource tool Proteasomal cleavage/TAP transport/MHC class I combined predictor

472   (http://tools.iedb.org/main/tcell/) with all default parameters . Moreover, the HTL epitopes of the

473   proteins were screened using the IEDB tool "MHC-II Binding Predictions"

474   (http://tools.iedb.org/mhcii/). Threshold for strong binding peptides (IC$_{50}$) was set at 50 nM to

475   determine the binding and interaction potentials of helper T-cell epitopes and both major

476 histocompatibility complex (MHC) class I and alleles (Shi et al., 2015). Top five HLA epitopes

477 for each RBD and NTD segments were docked against the respective HLA (MHC-Ⅰ and MHC-

478 Ⅱ) allele binders by interaction similarity-based protein-peptide docking system GalaxyPepDock

479 of the GalaxyWeb, docked HLA-epitope complexes were refined in GalaxyRefineComplex and

480 binding affinity (ΔG) was determined PROtein binDIng enerGY prediction (PRODIGY) tool (Xue

481 et al., 2016).

IFN-γ-inducing epitope prediction

483   Potential IFN-γ epitopes of all the selected antigenic sites of RBD, NTD, envelope protein

484 B-cell epitope (EBE), and membrane protein B-cell epitope (MBE) were predicted by

485 "IFNepitope" server (http://crdd.osdd.net/raghava/ ifnepitope/scan.php). To identify the set of

486 epitopes associated with MHC alleles that would maximize the population coverage, we adopted

487 the "Motif and SVM hybrid" (MERCI: Motif-EmeRging and with Classes-Identification, and

488 SVM) approach. The prediction is based on a dataset of IFN-γ-inducing and IFN-γ-noninducing

489 MHC allele binders (Dhanda et al., 2013).

Design and construction of multi-epitope vaccine candidate (CoV-RMEN**)**

491   The candidate vaccine (denoted as 'CoV-RMEN') design and construction method follows

492 previously published peptide vaccine development protocol for different emerging infectious

493 diseases like SARS and MERS (Shi et al., 2015; Badawi et al., 2016; Almofti et al., 2018; Shey et

494 al., 2019; Srivastava et al., 2019. Ul Qamar et al., 2019). The multi-epitope protein was constructed

495 by positioning the selected RBD, NTD, MBE and EBE aa sequences linked with short, rigid and

496 flexible linkers GG. To develop highly immunogenic recombinant proteins, two universal T-cell

497 epitopes were used, namely, a pan-human leukocyte antigen DR-binding peptide (PADRE)

498 (Agadjanyan et al., 2005), and an invasin immunostimulatory sequence taken from *Yersinia*

499 (Invasin) (Li et al., 2015) were used to the N and C terminal of the vaccine construct respectively,

500 linked by EGGE.

501 **Secondary and tertiary structure of the CoV-RMEN**

502   Chou   and   Fasman   secondary   structure   prediction   server   (CFSSP:

503 https://www.biogem.org/tool/chou-fasman/),   and   RaptorX   Property

504 (http://raptorx.uchicago.edu/StructurePropertyPred/predict/) web-servers were used for secondary

505 structure predictions (Källberg et al., 2014). The tertiary structure of the CoV-RMEN was built in

506 homology/analogy   recognition   engine   (Phyre2)

507 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) web-server. The 3D model was

508 refined in a three-step process, initially energy minimization using the GROMOS96 program

509 implemented within the Swiss-PdbViewer (version 4.1.0). After energy minimization, the model

510 was refined using ModRefiner (https://zhanglab.ccmb.med.umich.edu/ModRefiner/) and then

511 GalaxyRefine server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE).

512   Furthermore, the local structural quality of the CoV-RMEN was refined with GalaxyRefine

513 server, and ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) was used to calculate

514 overall   quality   score   for   the   refined   structure.   The   ERRAT   server

515 (http://services.mbi.ucla.edu/ERRAT/) was also used to analyze non-bonded atom-atom

516 interactions compared to reliable high-resolution crystallography structures. A Ramachandran plot

517 was obtained through the RAMPAGE server (Lovell et al., 2003).

518 **Physiochemical properties prediction of CoV-RMEN**

519   Moreover, the online web-server ProtParam (Gasteiger et al., 2005) was used to assess

520   various physiochemical parameters of the CoV-RMEN including aa residue composition,

521   molecular weight, theoretical pI, instability index, *in vitro* and *in vivo* half-life, aliphatic index,

522   and grand average of hydropathicity (GRAVY). The solubility of the multi-epitope vaccine peptide

523   was evaluated using the Protein-Sol server (https://protein-sol.manchester.ac.uk/).

524   **Antigenicity, allergenicity and Immune simulation of the CoV-RMEN**

525   VaxiJen 2.0 and ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) web-servers were

526   used to predict the antigenicity of the CoV-RMEN while the AllerTOP 2.0 (http://www.ddg-

527   pharmfac.net/AllerTOP) and AllergenFP (http://ddg-pharmfac.net/AllergenFP/) web-servers were

528   used to predict vaccine allergenicity. Further, the immunogenicity and immune response profile of

529   the CoV-RMEN were characterized by *in silico* immune simulations using the C-ImmSim server

530   (http://150.146.2.1/C-IMMSIM/index.php) under default parameters with time steps set at 1, 84,

531   and 170 (each time step is 8 hours and time step 1 is injection at time = 0). Therefore, three shots

532   were given at four weeks apart.

533   **Molecular docking and dynamics of the CoV-RMEN with TLRs**

534   Molecular docking of the CoV-RMEN with the TLR2 (PDB ID:3D3I), TLR3 (PDB ID:

535   1ZIW) and TLR4 (PDB ID: 4G8A) receptors was performed using the High Ambiguity Driven

536   DOCKing (HADDOCK, version 2.4) (De Vries et al., 2010) web-server to evaluate the interaction

537   between ligand (CoV-RMEN) and receptors (TLRs) leading to an enhanced immune response.

538   CPORT (https://milou.science.uu.nl/services/CPORT/) was used to predict active interface

539   residues between the CoV-RMEN and TLRs. For the analysis of the stable complex formation, all

540   the complexes (TLRs with CoV-RMEN) were subjected to molecular dynamics (MD) simulation

541   by Gromacs 2020.2 using OPLS-AA (Abraham et al., 2015; Jorgensen et al., 1996). Finally, the

542   binding affinities of the docked chimeric protein-TLRs complexes were predicted using the

543   PRODIGY (PROtein binDIng enerGY prediction) (https://nestor.science.uu.nl/prodigy/) web-

544   server.

545   **Analysis of cDNA and mRNA for cloning and expression of CoV-RMEN**

546   Reverse translation and codon optimization were performed using the GenScript Rare

547   Codon Analysis Tool (https://www.genscript.com/tools/rare-codon-analysis) in order to express

548   the CoV-RMEN in the *E. coli* (strain K12). Stability of the mRNA was verified using two different

549   tools, namely RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) and the

550   mfold (http://unafold.rna.albany.edu/?q=mfold) web-servers. The optimized gene sequence

551   of CoV-RMEN will be artificially synthesized having N-terminal recombinant human rhinovirus

552   (HRV 3C) protease site (LEVLFQ↓GP) and cloned the final construct into pETite vector (Lucigen,

553   USA) through enzyme-free method (Waugh et al., 2011). Finally, the sequence of the recombinant

554   plasmid was designed by inserting the adapted codon sequences into pETite vector using

555   SnapGene software (from Insightful Science; available at snapgene.com). As an alternative to *E.*

556   *coli*, eukaryotic expression system HEK-293 was optimized using similar analysis for the vaccine

557   production.

558   **Population coverage by CTL and HTL epitopes**

559   The predicted T-cell epitopes were shortlisted based on the aligned Artificial Neural

560   Network (ANN) with half-maximal inhibitory concentration ($annIC_{50}$) values up to 50 nM. The

561   IEDB "Population Coverage" tool (http://tools.iedb.org/population/) was used to determine the

562   world human population coverage by the shortlisted CTL and HTL epitopes (Bui et al., 2006). We

563 used OmicsCircos to visualize the association between world population and different ethnic

564 groups (Hoque et al., 2020b).

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580 **Competing Interest**

581   The authors declare no competing interests.

582 **Author contributions**

583   MSR, MNH, and MRI done the overall study and also draft the manuscript. SA drafted

584 some parts of results and discussion. MNH finally compiled the manuscript. OS, and MAS helped

585 in final discussion and reference indexing. ASMRUA, MMR, MS and MAH contributed

586 intellectually to the interpretation and presentation of the results. MS and MAH developed the

587 concept of peptide vaccine development. Finally, all authors have approved the manuscript for

588 submission.

589 **Acknowledgements**

590   The authors thank Joynob Akter Puspo, Masuda Akter and Israt Islam (MS student), and

591 DR. Kazi Alamgir Hossain (PhD Fellow) of the Microbial Genetics and Bioinformatics

592 Laboratory, Department of Microbiology, University of Dhaka for their cooperation, suggestions

593 and overall encouragement during the preparation of the manuscript. The authors also extend their

594 thanks to those who made their sequence data available in NCBI.

595 **Supplemental Information**

596   Supplemental information supporting the findings of the manuscript could be found from

597 the publishers (PeerJ), and from the corresponding author on request.

598 **Data Availability**

599 This *in silico* study utilizes genomic sequence data available in NCBI database (Accession no :

600 NC_045512.2, NC_004718.3, and NC_019843.3).

601

602

References

Abdelmageed MI, Abdelmoneim AH, Mustafa MI, Elfadol NM, Murshed NS, Shantier SW, Makhawi AM. . 2020. Design of a Multiepitope-Based Peptide Vaccine against the E Protein of Human COVID-19: An Immunoinformatics Approach. *BioMed Research International* **2683286**, https://doi.org/10.1155/2020/2683286.

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2(2)**, 19–25.

Adhikari UK, Tayebi M, Rahman MM. 2018. Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging oropouche virus. *Journal of Immunology Research* **6718083**, https://doi.org/10.1155/2018/6718083.

Agadjanyan MG, Ghochikyan A, Petrushina I, Vasilevko V, Movsesyan N, Mkrtichyan M, Tommy S, Cribbs DH. 2005. "Prototype Alzheimer's disease vaccine using the immunodominant B cell epitope from β-amyloid and promiscuous T cell epitope pan HLA DR-binding peptide." *The Journal of Immunology* **174(3)**, 1580-1586.

Ahmed SF, Quadeer AA, McKay MR. 2020. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* **12(3)**, 254.

Akhand MRN, Azim KF, Hoque SF, Moli MA, Joy BD, Akter H, Ibrahim KA, Nadim A, Hasan M. 2020. Genome based Evolutionary study of SARS-CoV-2 towards the Prediction of Epitope Based Chimeric Vaccine. *bioRxiv* doi: https://doi.org/10.1101/2020.04.15.036285.

626   Almofti YA, Abd-elrahman KA, Gassmallah SAE, Salih MA. 2018. Multi Epitopes Vaccine

627   Prediction against Severe Acute Respiratory Syndrome (SARS) Coronavirus Using

628   Immunoinformatics Approaches. *American Journal of Microbiological Research* **6(3)**,

629   94-114.

630   Badawi MM, SalahEldin MA, Suliman MM, AbduRahim SA, elghafoor Mohammed AA,

631   SidAhmed ASA, Marwa MO, Salih MA. 2016. In Silico Prediction of a Novel

632   Universal Multi-epitope Peptide Vaccine in the Whole Spike Glycoprotein of MERS

633   CoV. *American Journal of Microbiological Research* **4(4)**, 101-21.

634   Biswal JK, Bisht P, Mohapatra JK, Ranjan R, Sanyal A, Pattnaik B. 2015. "Application of a

635   recombinant capsid polyprotein (P1) expressed in a prokaryotic system to detect

636   antibodies against foot-and-mouth disease virus serotype O." *Journal of Virological*

637   *Methods* **215,** 45-51.

638   Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. 2006. Predicting population

639   coverage of T-cell epitope-based diagnostics and vaccines. BMC *Bioinformatics*

640   **7(1)**,153.

641   Carvalho LH, Sano GI, Hafalla JC, Morrot A, De Lafaille MAC, Zavala F. 2002. IL-4-

642   secreting CD4+ T cells are crucial to the development of CD8+ T-cell responses

643   against malaria liver stages. *Nature Medicine* **8**, 166–70.

644   Chakraborty S, Chakravorty R, Ahmed M, Rahman A, Waise TM, Hassan F, Mustafizur R,

645   Shamsuzzaman S. 2010. A computational approach for identification of epitopes in

646   dengue virus envelope protein: a step towards designing a universal dengue vaccine

647   targeting endemic regions. *In silico Biology* **10(5-6)**, 235-246.

648   Chan JFW, Kok KH, Zhu Z, Chu H, To KKW, Yuan S, Yuen KY. 2020. Genomic
649   characterization of the 2019 novel human-pathogenic coronavirus isolated from a
650   patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes and*
651   *Infections* **9**, 221–236.

652   Dawood RM, Moustafa RI, Abdelhafez TH, El-Shenawy R, El-Abd Y, El Din NGB, Jean D,
653   El Awady MK. 2019. A multiepitope peptide vaccine against HCV stimulates
654   neutralizing humoral and persistent cellular responses in mice. *BMC Infectious*
655   *Diseases* **19**, 932.

656   De Vries SJ, Van Dijk M, Bonvin AM. 2010. The HADDOCK web server for data-driven
657   biomolecular docking. *Nature Protocol* **5(5)**, 883.

658   Dhanda SK, Vir P, Raghava GP. 2013. "Designing of interferon-gamma inducing MHC class-
659   II binders." *Biology Direct* **8(1)**, 30.

660   Doytchinova IA, Flower DR. 2007. VaxiJen: a server for prediction of protective antigens,
661   tumour antigens and subunit vaccines. *BMC Bioinformatics* **8(1)**, 4.

662   Dudek LN, Perlmutter P, Aguilar I, Croft PN, Purcell WA. 2010. Epitope discovery and their
663   use in peptide based vaccines. *Current Pharmaceutical Design* **16(28)**, 3149-3157.

664   Faure G, Joseph AP, Craveur P, Narwani TJ, Srinivasan N, Gelly JC, Joseph R, de Brevern
665   AG. 2019. iPBAvizu: a PyMOL plugin for an efficient 3D protein structure
666   superimposition approach. *Source Code for Biology and Medicine* 14, 5.

667   Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. 2005. Protein
668   identification and analysis tools on the ExPASy server. The Proteomics Protocols
669   Handbook, Springer**:** 571-607.

670   Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL,
671   Lauber C, Leontovich AM, Neuman BW, Penzar D. 2020. The species Severe acute
672   respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it
673   SARS-CoV-2. *Nature Microbiology* **5(4)**, 536-544, doi:10.1038/s41564-020-0695-z.

674   Gralinski LE, Menachery VD. 2020. Return of the coronavirus: 2019-nCoV. Viruses **12**, 135.

675   Grant OC, Montgomery D, Ito K, Woods RJ. 2020. Analysis of the SARS-CoV-2 spike
676   protein glycan shield: implications for immune recognition. *bioRxiv* doi:
677   https://doi.org/10.1101/2020.04.07.030445.

678   Guex N, Peitsch MC. 1997. "SWISS-MODEL and the Swiss-Pdb Viewer: an environment for
679   comparative protein modeling." *Electrophoresis* **18(15)**, 2714-2723.

680   Hasan A, Hossain M, Alam J. 2013. A computational assay to design an epitope-based Peptide
681   vaccine against Saint Louis encephalitis virus. *Bioinformatics and Biology Insights* **7**,
682   BBI-S13402.

683   Hoque MN, Istiaq A, Clement RA, Sultana M, Crandall KA, Siddiki AZ, Hossain M A. 2019.
684   Metagenomic deep sequencing reveals association of microbiome signature with
685   functional biases in bovine mastitis. *Scientific Reports* **9**, 13536.

686   Hoque MN, Chaudhury A, Akanda MAM, Hossain MA, Islam MT. 2020a. Genomic Diversity
687   and Evolution, Diagnosis, Prevention, and Therapeutics of the Pandemic COVID-19
688   Disease. *Preprints* **2020040359**, DOI: 10.20944/preprints202004.0359.v1.

689   Hoque MN, Istiaq A, Clement RA, Gibson KM, Saha O, Islam OK, Abir RA, Sultana M,
690   Siddiki AMAM, Crandall KA, Hossain MA. 2020b. Insights into the Resistome of
691   Bovine Clinical Mastitis Microbiome, a Key Factor in Disease Complication.
692   *Frontiers in Microbiology* **11**, 860, doi: 10.3389/fmicb.2020.00860.

693   Huang H, Hao S, Li F, Ye Z, Yang J, Xiang J. 2007. CD4(+) Th1 cells promote CD8(+) Tc1

694   cell survival, memory response, tumor localization and therapy by targeted delivery of

695   interleukin 2 via acquired pMHC I complexes. *Immunology* **120**, 148–159.

696   HuiDS IAE, Madani TA, Ntoumi F, Koch R, Dar O. 2020. The continuing 2019-nCoV

697   epidemic threat of novel coronaviruses to global health—The latest 2019 novel

698   coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases*

699   **91**, 264-266.

700   Jaimes JA, André NM, Millet JK, Whittaker GR. 2020. Structural modeling of 2019-novel

701   coronavirus (nCoV) spike protein reveals a proteolytically-sensitive activation loop as

702   a distinguishing feature compared to SARS-CoV and related SARS-like

703   coronaviruses. *arXiv Preprint, arXiv:2002.06196.*

704   Källberg M, Margaryan G, Wang S, Ma J, Xu J. 2014. RaptorX server: a resource for

705   template-based protein structure modeling. In Protein Structure Prediction **(pp. 17-**

706   **27)**, Humana Press, New York, NY.2014.

707   Kibria, KM K, Ullah H, Miah M. 2020. The multi-epitope vaccine prediction to combat

708   Pandemic SARS-CoV-2, an immunoinformatic approach. *Preprints*, doi:

709   10.21203/rs.3.rs-21853/v1.

710   Kolaskar A, Tongaonkar PC. 1990. "A semi-empirical method for prediction of antigenic

711   determinants on protein antigens." *FEBS Letters* **276(1-2)**, 172-174.

712   Kong L, Torrents de la Peña A, Deller MC, Garces F, Sliepen K, Hua Y, Stanfield RL, Sanders

713   RW, Wilson IA. 2015. Complete epitopes for vaccine design derived from a crystal

714   structure of the broadly neutralizing antibodies PGT128 and 8ANC195 in complex

715   with an HIV-1 Env trimer. *Acta Crystallography D Biology Crystallography* **71(Pt**

716   **10)**, 2099-108.

717   Kringelum JV, Lundegaard C, Lund O, Nielsen M. 2012. Reliable B cell epitope predictions:

718   Impacts of method development and improved benchmarking. *PLOS Computational*

719   *Biology* **8(12)**, e1002829.

720   Lane R. 2020. Sarah Gilbert: carving a path towards a COVID-19 vaccine. *Lancet* (London,

721   England) **395(10232)**, 1247.

722   Larsen JEP, Lund O, Nielsen M. 2006. Improved method for predicting linear B-cell

723   epitopes. *Immunome Research* **2(1)**, 2.

724   Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993 "PROCHECK: a program to

725   check the stereochemical quality of protein structures." *Journal of Applied*

726   *Crystallography* **26(2)**, 283-291.

727   Le TT, Andreadakis Z, Kumar A, Roman RG, Tollefsen S, Saville M, Mayhew S. 2020. The

728   COVID-19 vaccine development landscape. *Nature Reviews Drug Discovery* **19(5)**,

729   305-6.

730   Li H, Ning P, Lin Z, Liang W, Kang K, He L, Zhang Y. 2015. Co-expression of the C-terminal

731   domain of Yersinia enterocolitica invasin enhances the efficacy of classical swine-

732   fever-vectored vaccine based on human adenovirus. *Journal of Biosciences* **40**, 79–

733   90.

734   Lovell SC, Davis IW, Arendall III WB, De Bakker PI, Word JM, Prisant MG, Richardson, JS,

735   Richardson DC. 2003. "Structure validation by Cα geometry: ϕ, ψ and Cβ deviation."

736   Proteins: Structure, Function, *Bioinformatics* **50(3)**, 437-450.

759   Schoeman D, Fielding BC. 2019. Coronavirus envelope protein: current knowledge. *Virology*

760   *Journal* **16**, 69.

761   Mosaddeghi P, Negahdaripour M, Dehghani Z, Farahmandnejad M, Moghadami M, Nezafat

762   N, Masoompour SM. 2020. Therapeutic approaches for COVID-19 based on the

763   dynamics of interferon-mediated immune responses. *Preprints* **2020030206**, doi:

764   10.20944/preprints202003.0206.v2.

765   Shang W, Yang Y, Rao Y, Rao X. 2020. The outbreak of SARS-CoV-2 pneumonia calls for

766   viral vaccines. *npj Vaccines* **5**, 18.

767   Sheridan C. 2020. Questions remain following first COVID-19 vaccine results. *Nature*

768   *Biotechnology* doi: https://www.nature.com/articles/d41587-020-00015-x.

769   Shey RA, Ghogomu SM, Esoh KK, Nebangwa ND, Shintouo CM, Nongley NF, Asa BF,

770   Ngale FN, Vanhamme L, Souopgui J. 2019. In-silico design of a multi-epitope vaccine

771   candidate against onchocerciasis and related filarial diseases. *Scientific Reports* **9(1)**,

772   1-18.

773   Shi J, Zhang J, Li S, Sun J, Teng Y, Wu M, Li J, Li Y, Hu N, Wang H, Hu Y. 2015. Epitope-

774   Based Vaccine Target Screening against Highly Pathogenic MERS-CoV: An In Silico

775   Approach Applied to Emerging Infectious Diseases. *PLoS One* **10(12)**, e0144475.

776   Shi SQ, Peng JP, Li YC, Qin C, Liang GD, Xu L, Yang Y, Wang JL, Sun QH. 2006. The

777   expression of membrane protein augments the specific responses induced by SARS-

778   CoV nucleocapsid DNA immunization. *Molecular Immunology* **43**, 1791–1798.

779   Song W, Gui M, Wang X, Xiang Y. 2018. Cryo-EM structure of the SARS coronavirus spike

780   glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathogens* **14(8)**,

781   e1007236.

782   Srivastava S, Kamthania M, Kumar Pandey R, Kumar Saxena A, Saxena V, Kumar Singh S,

783   Kumar Sharma R, Sharma N. 2019. Design of novel multi-epitope vaccines against

784   severe acute respiratory syndrome validated through multistage molecular interaction

785   and dynamics. *Journal of Biomolecular Structure and Dynamics* **37(16)**, 4345-4360.

786   Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, Zhou Y, Du L. 2020. Characterization of

787   the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for

788   development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular &*

789   *Molecular Immunology* **19,** 1-8, doi: 10.1038/s41423-020-0400-4.

790   Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J. 2020.

791   On the origin and continuing evolution of SARS-CoV-2. *National Science Review*

792   **nwaa036**, https://doi.org/10.1093/nsr/nwaa036.

793   Totura AL, Whitmore A, Agnihothram S, Schäfer A, Katze MG, Heise MT, Baric RS. 2015.

794   Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune

795   response to severe acute respiratory syndrome coronavirus infection. *MBio* **6(3)**,

796   e00638-15.

797   ul Qamar MT, Bari A, Adeel MM, Maryam A, Ashfaq UA, Du X, Muneer I, Ahmad HI, Wang

798   J. 2018. Peptide vaccine against chikungunya virus: immuno-informatics combined

799   with molecular docking approach. *Journal of Translational Medicine* **16(1)**, 298.

800   Ul Qamar MT, Saleem S, Ashfaq UA, Bari A, Anwar F, Alqahtani S. 2019. Epitope-based

801   peptide vaccine design and target site depiction against Middle East Respiratory

802   Syndrome Coronavirus: an immune-informatics study. *Journal of Translational*

803   *Medicine* **17(1)**, 362.

804   Van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Krüger PKHC, Mark AE, Scott

805   WRP, Tironi IG, Hünenberger PH, Huenenberger PH, Krueger P. 1996. The

806   GROMOS96 manual and user guide.

807   Wang N, Rosen O, Wang L, Turner HL, Stevens LJ, Corbett KS, Bowman CA, Pallesen J,

808   Shi W, Zhang Y, Leung K. 2019. Structural Definition of a Neutralization-sensitive

809   Epitope on the MERS-CoV S1-NTD. *Cell Reports* **28(13)**, 3395-405.

810   Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. 2020. Site-specific glycan analysis

811   of the SARS-CoV-2 spike. *Science* **eabb9983**, doi: 10.1126/science.abb9983.

812   Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer

813   TAP, Rempfer C, Bordoli L, Lepore R. 2018. "SWISS-MODEL: homology modelling

814   of protein structures and complexes." *Nucleic Acids Research* **46(W1)**, W296-W303.

815   Waugh DS. 2011. An overview of enzymatic reagents for the removal of affinity tags. *Protein*

816   *Expression and Purification* **80(2)**, 283-293.

817   WHO, 2020. Novel Coronavirus (2019-nCoV) situation reports - World Health Organization

818   (WHO).

819   Wolfert MA., Boons GJ. 2013. Adaptive immune activation: glycosylation does matter.

820   *Nature Chemical Biology* **9(12)**, 776.

821   Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan

822   JS. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.

823   *Science* **367(6483)**, 1260-1263, doi: 10.1126/science.abb2507.

824   Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan

825   ML. 2020."A new coronavirus associated with human respiratory disease in China."

826   *Nature* 1-5.

827   Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. 2016. "PRODIGY: a web

828   server for predicting the binding affinity of protein–protein complexes."

829   *Bioinformatics* **32(23)**, 3676-3678.

830   Yazdani Z, Rafiei A, Yazdani M, Valadan R. 2020. Design an efficient multi-epitope peptide

831   vaccine candidate against SARS-CoV-2: An in silico analysis. *bioRxiv* doi:

832   https://doi.org/10.1101/2020.04.20.051557.

833   Yong CY, Ong HK, Yeap SK, Ho KL, Tan WS. 2019. Recent Advances in the Vaccine

834   Development Against Middle East Respiratory Syndrome-Coronavirus. *Frontiers in*

835   *Microbiololgy* **10**, 1781.

836   Zander RA, Vijay R, Pack AD, Guthmiller JJ, Graham AC, Lindner SE, Vaughan AM, Kappe

837   SH, Butler NS. 2017. Th1-like plasmodium-specific memory CD4+ T cells support

838   humoral immunity. *Cell Reports* **21(7)**, 1839-1852.

839   Zhou D, Qi R, Zhang W, Tian X, Peng C. 2020b. Identification of 22 N-glycosites on Spike

840   Glycoprotein of SARS-CoV-2 and Accessible Surface Glycopeptide Motifs:

841   Implications on Vaccination and Antibody Therapeutics. *Preprints*, **2020020381**, doi:

842   10.20944/preprints202002.0381.v2.

843   Zhou H, Chen Y, Zhang S, Niu P, Qin K, Jia W. et al. 2019. Structural definition of a

844   neutralization epitope on the N-terminal domain of MERS-CoV spike glycoprotein.

845   *Nature Communications* **10**, 3068.

846   Zhou P, Yang XL, Wang XG, Hu B, Zhang L. et al. 2020a. Discovery of a novel coronavirus

847   associated with the recent pneumonia outbreak in humans and its potential bat origin.

848   *Nature* doi: 10.1038/s41586-020-2012-7.

Figure 1

Fig. 1

The three-dimensional (3D) structure of the N-terminal domains (NTDs) and receptor binding domains (RBDs) of the spike (S) proteins of SARS-CoV-2 (surface view). The orange, cyan, and yellow colored regions represent the potential antigenic domains predicted by the IEDB analysis resource ElliPro analysis.
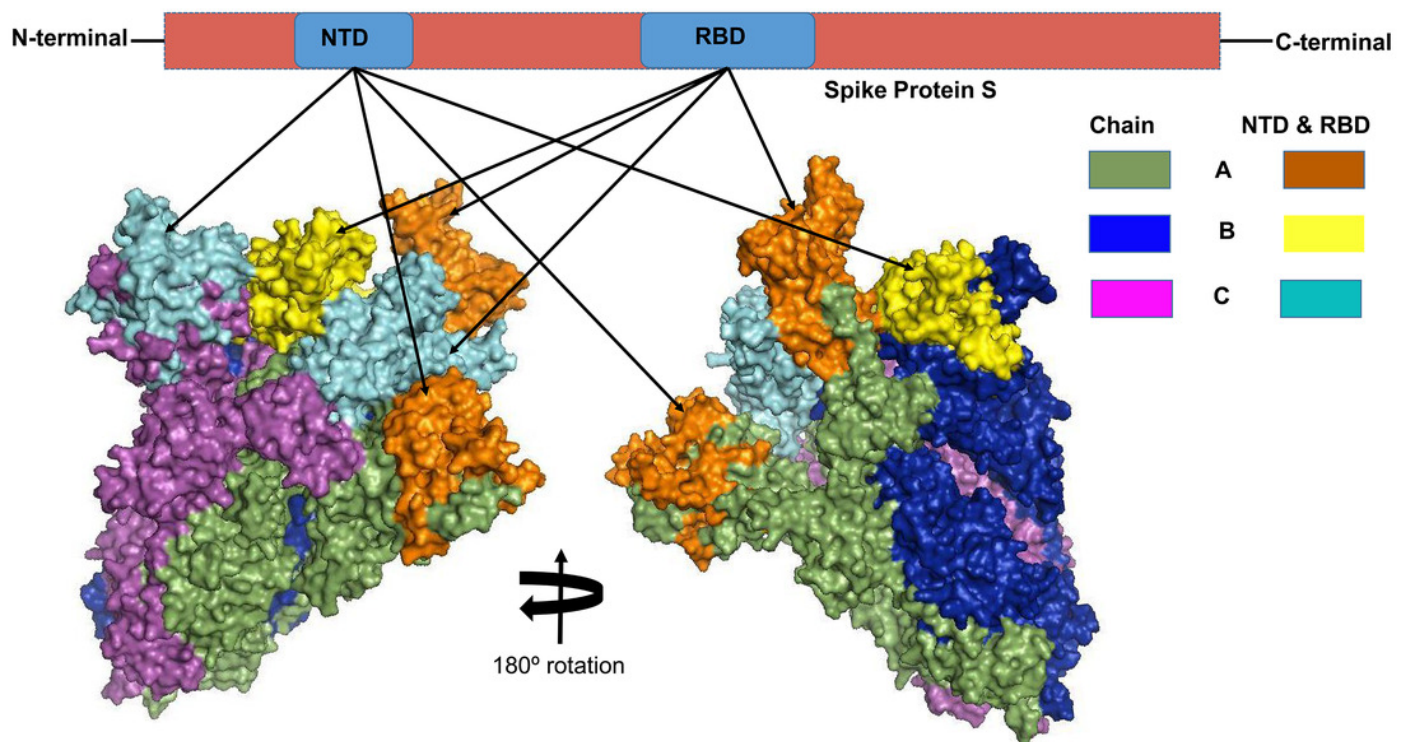
Figure 2

Fig. 2

Predicted B-cell epitopes using BepiPred-2.0 epitope predictor in IEDB-analysis resource web-based repository. Yellow areas above threshold (red line) are proposed to be a part of B cell epitopes in (a) RBD and (b) NTD regions of S protein, (c) envelop (E) and (d) membrane (M) proteins of SARS-CoV-2.
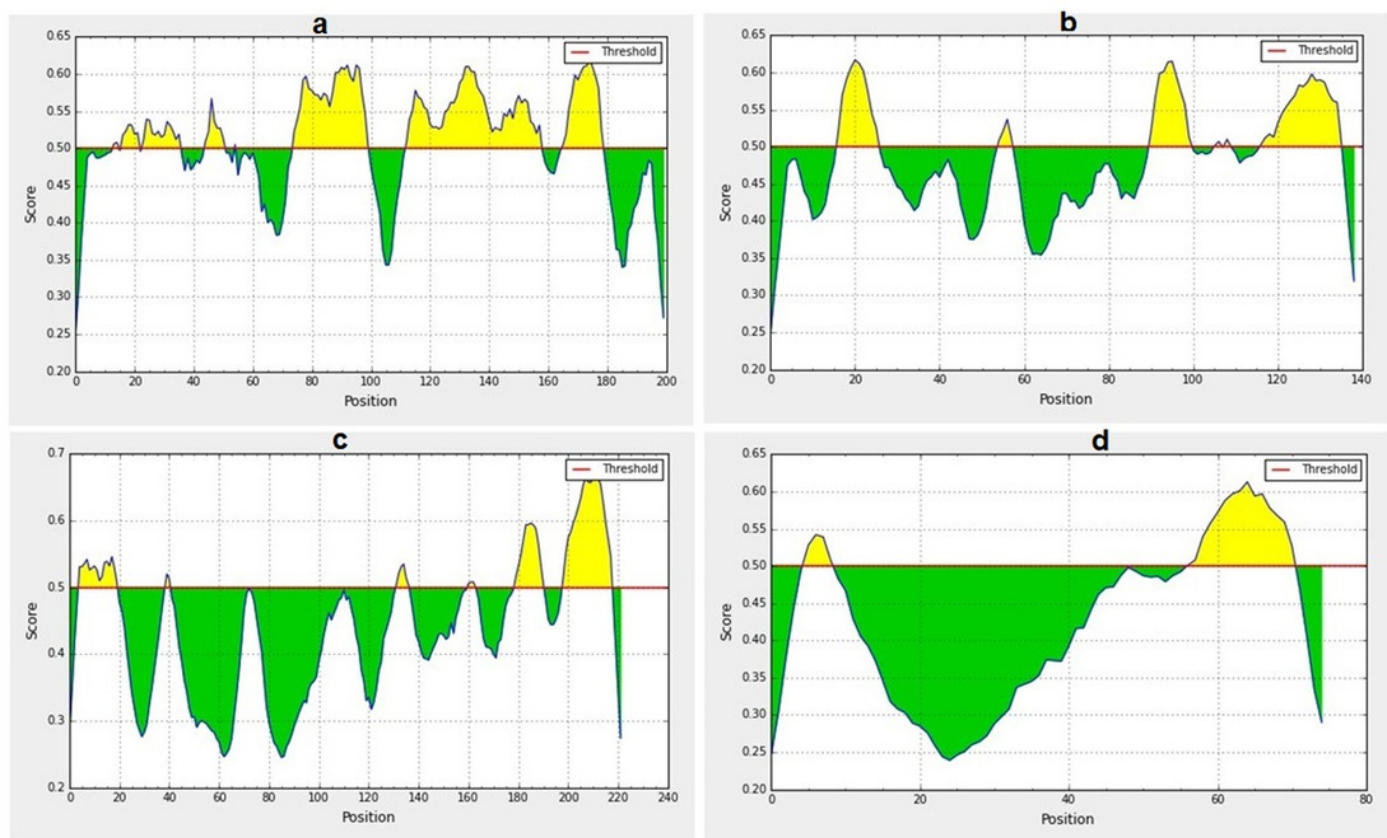
Figure 3

Fig. 3

Design, construction and structural validation of multi-epitope vaccine candidate (CoV-RMEN) for SARS-CoV-2. (a) Structural domains and epitopes rearrangement of CoV-RMEN, (b) secondary structure of CoV-RMEN as analyzed through CFSSP:Chou and Fasman secondary structure prediction server , (c) final tertiary structure of CoV-RMEN (surface view) obtained from homology modelling on Phyre2 in which domains and epitopes are represented in different colors (PADRE-smudge; membrane B-cell epitope, MBE-magenta; N-terminal domain, NTD-orange; receptor-binding domain, RBD-cyan; envelop B-cell epitope, EBE-blue; invasin-yellow), (d) validation of the refined model with Ramachandran plot analysis showing 94.7%, 4.8% and 0.5% of protein residues in favored, allowed, and disallowed (outlier) regions respectively, (e) ProSA-web, giving a Z-score of −6.17, and (f) the finally predicted primary structure of the CoV-RMEN.

>CoV-RMEN
AKFVAAWTLKAAAEGGEYRIGNYKLNTDHSSSSDNIAGGLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKV
CEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFGGRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSA
SFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEG
FNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPGGYVVYSRVKNLNSSRVPEGGETAKSKKFPSYTATYQF

Figure 4

Fig. 4

Molecular docking of top five MHC-I and MHC-II epitopes of RBD and NTD domains with respect to HLA allele binders. Figures (a-e) and (k-o) represent the top five MHC-I epitopes of RBD and NTD domains, respectively. Figures (f-j) and (p-t) represent the top five MHC-II epitopes binds of the same domains binds to their respective HLA alleles. The protein-peptide docking was performed in GalaxyWEB-GalaxyPepDock-server followed by the refinement using GalaxyRefineComplex and free energy (ΔG) of each complex was determined in PRODIGY server. Ribbon structures represent HLA alleles and stick structures represent the respective epitopes. Light color represents the templates to which the alleles and epitopes structures were built. Further information on molecular docking analysis is also available in Supplemental Data 1.

ATRFASVYA-A*30:01 ΔG/-10.1(kcal mol-1)	**NSASFSTFK -A*68:01** ΔG/-8.4(kcal mol-1)	**KLNDLCFTNV -A*02:03** ΔG/-10.8(kcal mol-1)	**FASVYAWNR -A*68:01** ΔG/-11.1(kcal mol-1)	**RLFRKSNLK -A*33:01** ΔG/-11.9(kcal mol-1)

MHC-I of RBD

a b c d e

VLSFELLHAPATVCG - DRB1*01:01 ΔG/-8.6(kcal mol-1)	**LSFELLHAPATVCGP - DRB1*01:01** ΔG/-11.8(kcal mol-1)	**VVLSFELLHAPATVC - DRB1*01:01** ΔG/-11.4(kcal mol-1)	**VVVLSFELLHAPATV - DRB1*01:01** ΔG/-9.8(kcal mol-1)	**GNYNYLYRLFRKSNL - DRB1*11:01** ΔG/-9.8(kcal mol-1)

MHC-II of RBD

f g h i j

FQFCNDPFL -A*02:06 ΔG/-9.2(kcal mol-1)	**LPFFSNVTW -B*15:01** ΔG/-11.0(kcal mol-1)	**SANNCTFEY -B*35:01** ΔG/-8.4(kcal mol-1)	**KQGNFKNLR -A*31:01** ΔG/-10.7(kcal mol-1)	**LPFNDGVYF -B*35:01** ΔG/-10.3(kcal mol-1)

MHC-I of NTD

k l m n o

QSLLIVNNATNVVIK - DRB1*13:02 ΔG/-8.2(kcal mol-1)	**SLLIVNNATNVVIKV - DRB1*13:02** ΔG/-9.6(kcal mol-1)	**TQSLLIVNNATNVVI - DRB1*13:02** ΔG/-10.4(kcal mol-1)	**KTQSLLIVNNATNVV - DRB1*13:02** ΔG/-8.4(kcal mol-1)	**LLIVNNATNVVIKVC - DRB1*13:02** ΔG/-8.8(kcal mol-1)

MHC-II of NTD

p q r s t

Figure 5

Fig. 5

Molecular docking and dynamics of CoV-RMEN vaccine with immune receptors (TLR2, TLR3 and TLR4). Docked complexes for (a) CoV-RMEN and TLR2, (b) CoV-RMEN and TLR3, and (c) CoV-RMEN and TLR4. Magnified interfaces of the respective complexes are figured to (d), (e) and (f) respectively. Active residues of CoV-RMEN colored magenta, and of TLRs colored orange with stick view. ΔG represents the binding affinity of the complexes. Molecular dynamics simulation study of (g) CoV-RMEN and TLR2, (h) CoV-RMEN and TLR3, and (i) CoV-RMEN and TLR4 complexes across the time window of 100 ps. The reasonably invariable RMSD value indicates a stable complex formation.
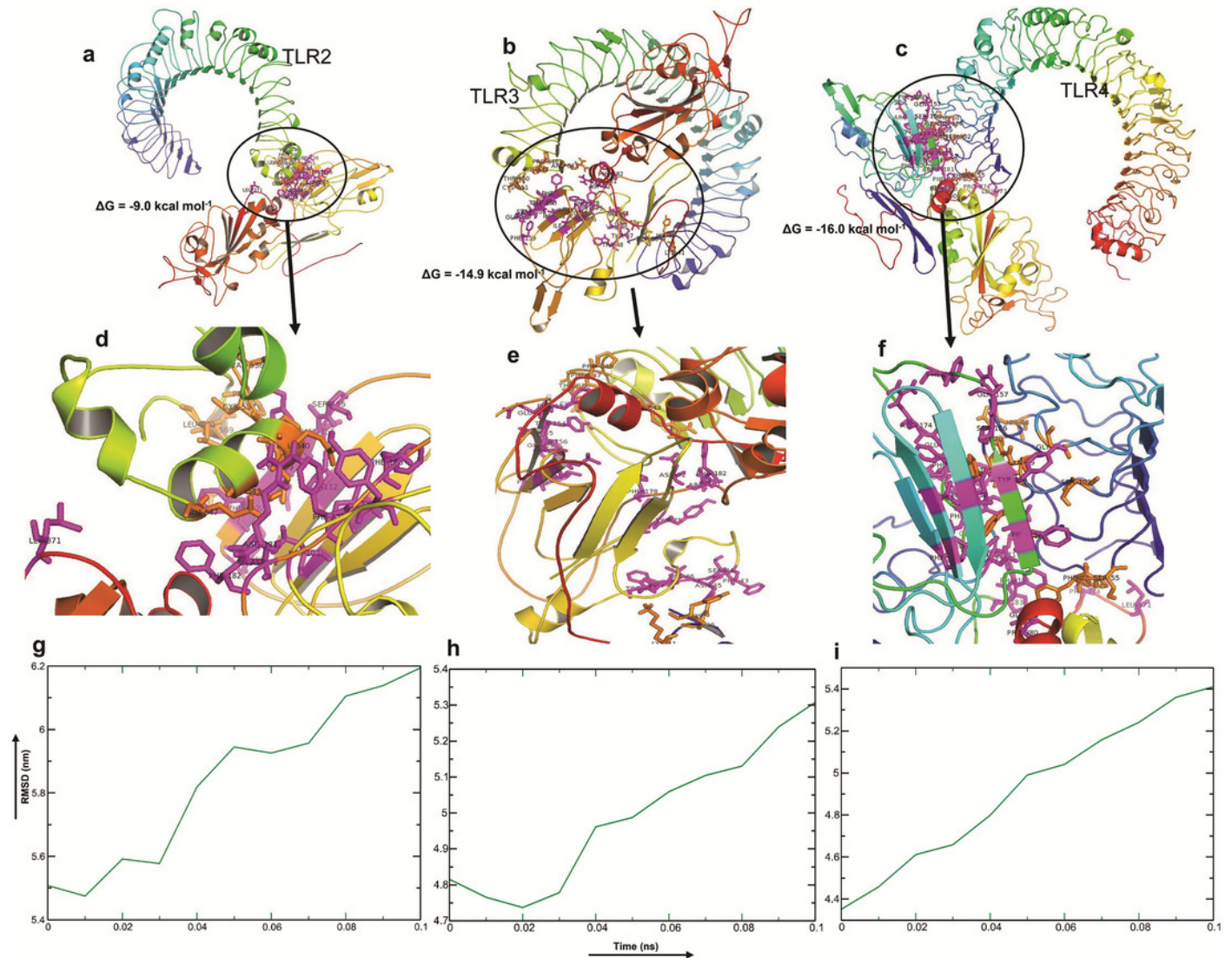
Figure 6

Fig. 6

C-ImmSim presentation of an *in silico* immune simulation with the chimeric peptide. (a) The immunoglobulins and the immunocomplex response to antigen (CoV-RMEN) inoculations (black vertical lines); specific subclasses are indicated as colored peaks, (b) concentration of cytokines and interleukins, and inset plot shows danger signal together with leukocyte growth factor IL-2, (c) B-cell populations after three injections, (d) evolution of B cell, (e) T-helper cell populations per state after injections, and (f) evolution of T-helper cell classes with the course of vaccination.
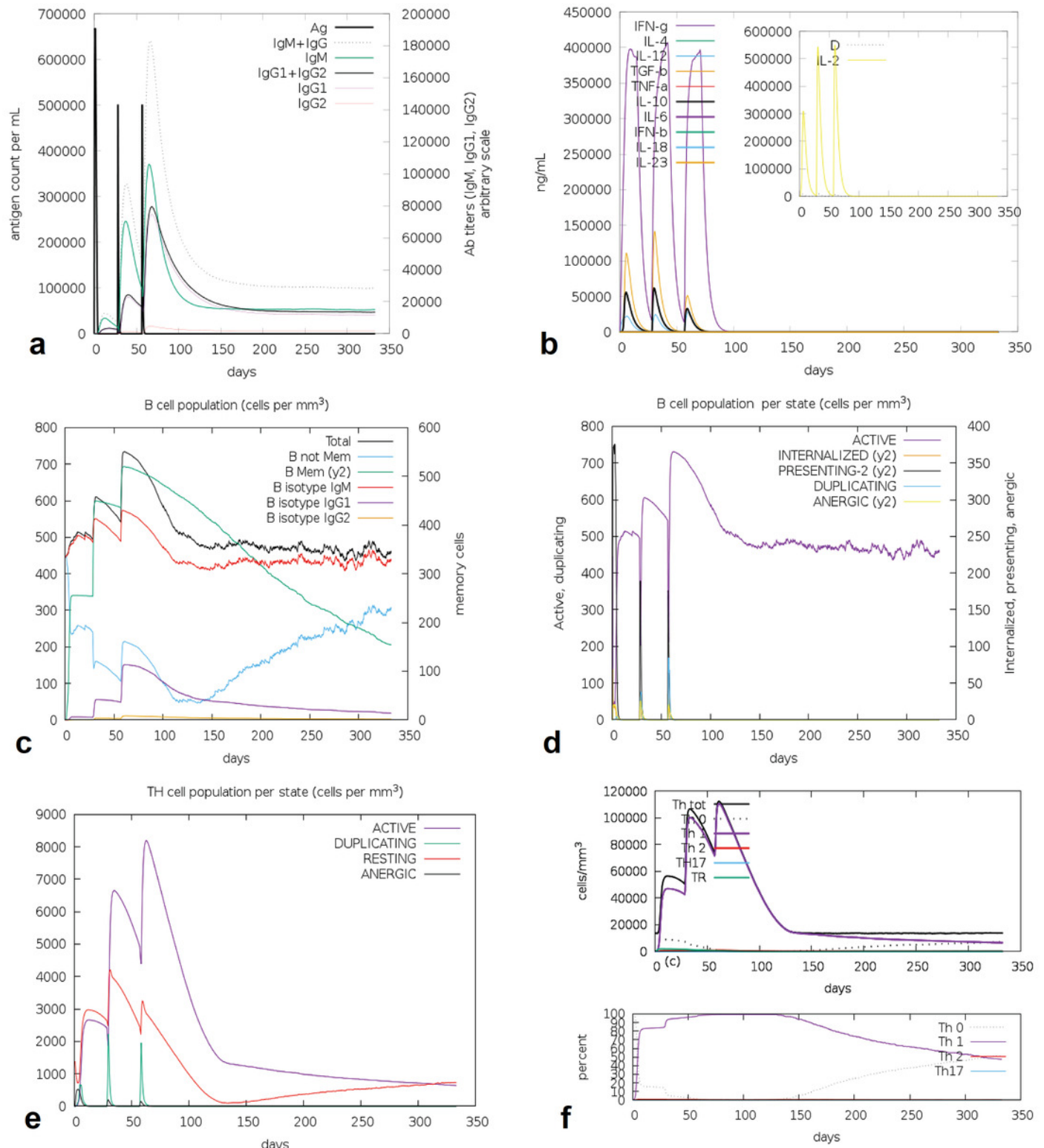
Figure 7

Fig. 7

Population coverage of the selected T-cell epitopes and their respective HLA alleles. The circular plot illustrates the relative abundance of the top 70 geographic regions and ethnic groups for selected CTL and HTL epitopes, which were used to construct the vaccine and their corresponding MHC HLA alleles were obtained for population coverage analysis both individually (either MHC-I or MHC-II) and in combination (MHC-I and MHC-II). (a) Population coverage of top seventy geographical regions out of 123 regions. (b) Population coverage of top seventy ethnic groups selected from 146 ethnic groups. Regions and ethnic groups in the respective MHC-I and MHC-II epitopes are represented by different colored ribbons, and the inner blue bars indicate their respective relative coverages. Further information on population coverage analysis is also available in Supplemental Data 1.
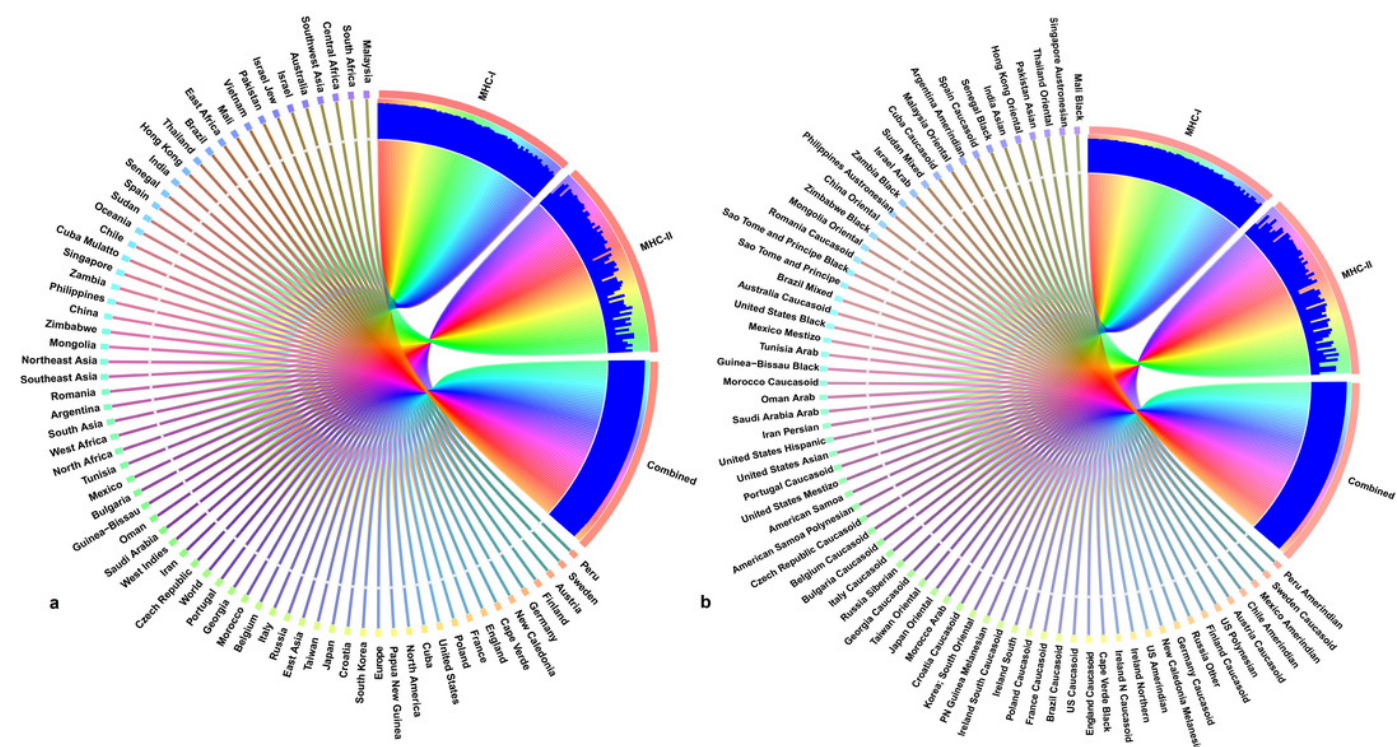
Figure 8

Fig. 8

**Codon optimization and mRNA structure of CoV-RMEN gene for expression in *E.*
*coli*** (a) GC curve (average GC content: 50.26%) of the optimized CoV-RMEN gene, (b)
percentage distribution of codons in computed codon quality groups, (c) relative distribution
of codon usage frequency along the gene sequence to be expressed in *E. coli*, and codon
adaptation index (CAI) was found to be 0.87 for the desired gene, (d) secondary structure
and stability of corresponding mRNA, and (e) resolved view of the start region in the mRNA
structure of CoV-RMEN.
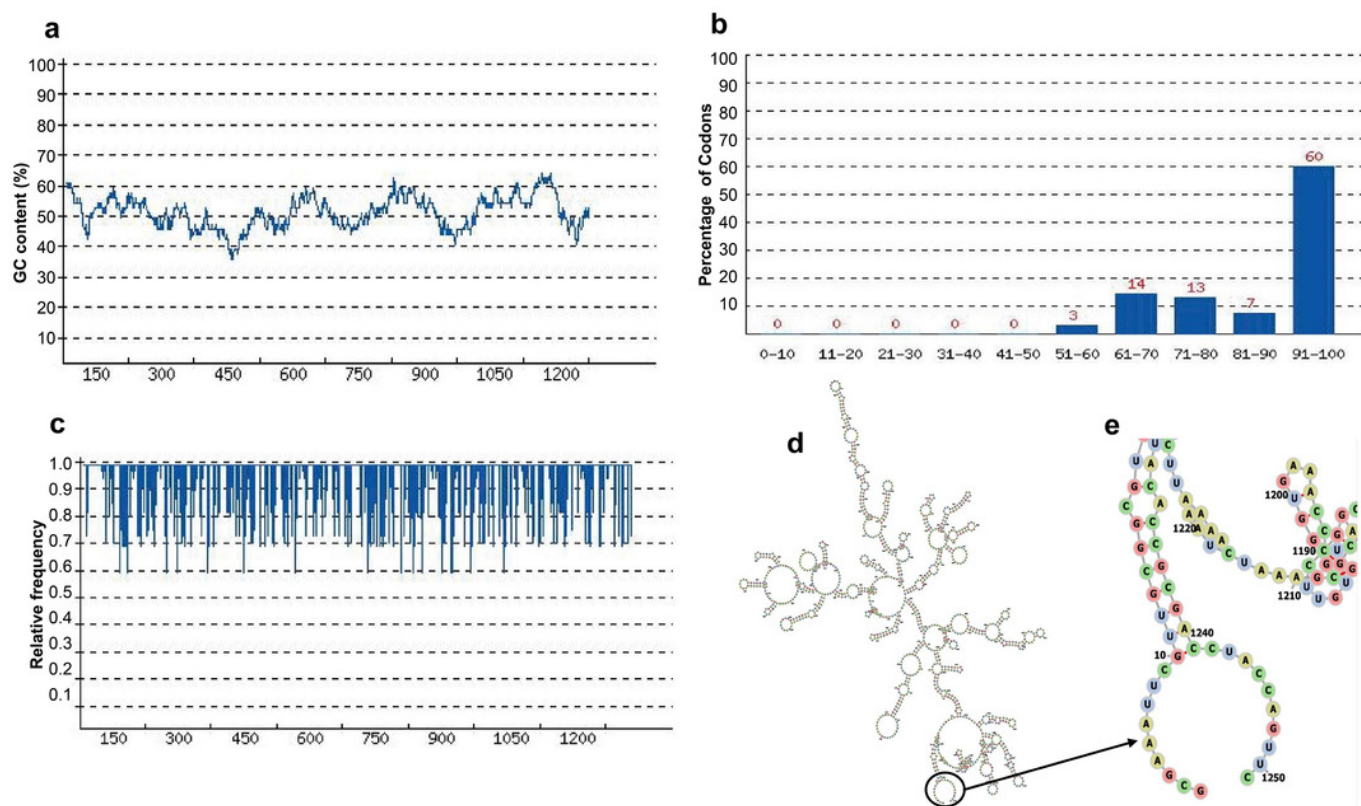
Figure 9

Fig. 9

***In silico* fusion cloning of the CoV-RMEN.** The final vaccine candidate sequence was inserted into the pETite expression vector where the red part represents the gene coding for the predicted vaccine, and the black circle represents the vector backbone. The six His-tag and SUMU-tag are located at the Carboxy-terminal end.
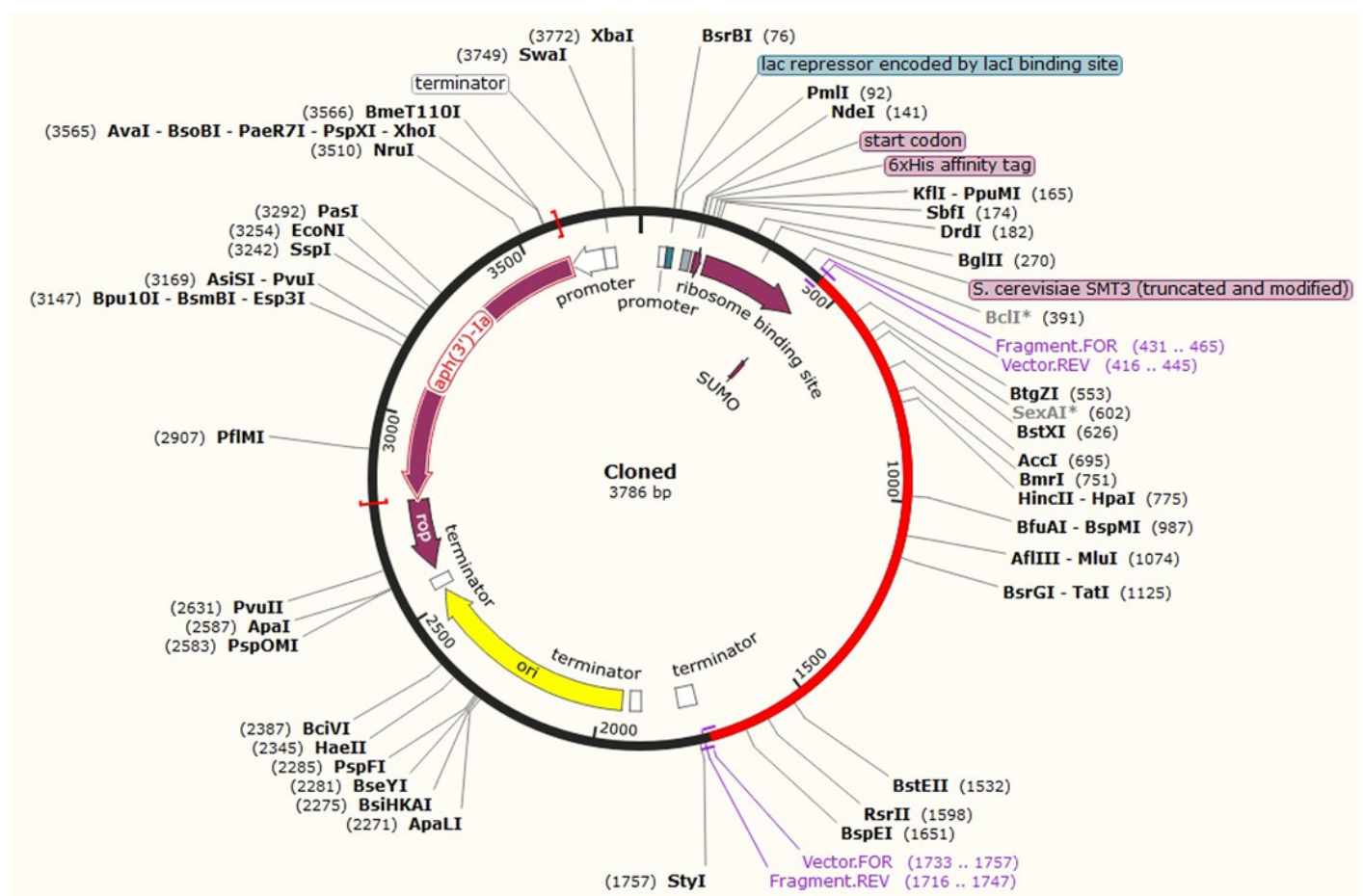
Table 1

Linear epitopes present on spike (S) glycoprotein surface predicted through ElliPro in IEDB-analysis resource based upon solvent-accessibility and flexibility are shown with their antigenicity scores. The highlighted green coloured regions were the potential antigenic domains while the yellow coloured region represents the trans-membrane domain of the S protein.

1   **Epitope-based chimeric peptide vaccine design against S, M, and E proteins of SARS-CoV-**

2   **2 etiologic agent of global pandemic COVID-19: an *in silico* approach**

3   M. Shaminur Rahman, M. Nazmul Hoque, M. Rafiul Islam, Salma Akter, A.S.M. Rubayet- Ul-

4   Alam, Mohammad Anwar Siddique, Otun Saha, Md. Mizanur Rahaman, Munawar Sultana, Keith

5   A. Crandall, M. Anwar Hossain

6   **Table 1.** Linear epitopes present on spike (S) glycoprotein surface predicted through ElliPro in

7   IEDB-analysis resource based upon solvent-accessibility and flexibility are shown with their

8   antigenicity scores. The highlighted green coloured regions were the potential antigenic domains

9   while the yellow coloured region represents the trans-membrane domain of the S protein.

No.	Chain	Start	End	Peptide	Residues	Score
1	A	395	514	VYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLS	120	0.837
2		58	194	FFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVF	137	0.835
3		1067	1146	YVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELD	80	0.83
4		201	270	FKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYL	70	0.76
5		331	381	NITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYG	51	0.706
6		700	720	GAENSVAYSNNSIAIPTNFTI	21	0.668
7		27	35	AYTNSFTRG	9	0.66
8		909	936	IGVTQNVLYENQKLIANQFNSAIGKIQD	28	0.633
9		789	813	YKTPPIKDFGGFNFSQILPDPSKPS	25	0.6
10		623	642	AIHADQLTPTWRVYSTGSNV	20	0.598
11		891	907	GAALQIPFAMQMAYRFN	17	0.591
12		579	583	PQTLE	5	0.551
13		687	692	VASQSI	6	0.55
14		653	659	AEHVNNS	7	0.539
15		679	684	NSPRRA	6	0.521
16	B	1067	1146	YVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELD	80	0.826
17		89	194	GVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVF	106	0.816
18		58	87	FFSNVTWFHAIHVSGTNGTKRFDNPVLPFN	30	0.81
19		203	270	IYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYL	68	0.748
20		465	509	ERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYR	45	0.727
21		436	458	WNSNNLDSKVGGNYNYLYRLFRK	23	0.672
22		700	720	GAENSVAYSNNSIAIPTNFTI	21	0.671
23		27	35	AYTNSFTRG	9	0.666
24		909	9036	IGVTQNVLYENQKLIANQFNSAIGKIQD	28	0.641
25		624	643	IHADQLTPTWRVYSTGSNVF	20	0.617
26		328	365	RFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADY	38	0.608
27		891	907	GAALQIPFAMQMAYRFN	17	0.602
28		577	583	RDPQTLE	7	0.598
29		790	817	KTPPIKDFGGFNFSQILPDPSKPSKRSF	28	0.595
30		673	693	SYQTQTNSPRRARSVASQSII	21	0.567
31		526	537	GPKKSTNLVKNK	12	0.553
32		653	661	AEHVNNSYE	9	0.548
33		554	563	ESNKKFLPFQ	10	0.52
34	C	56	194	LPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVF	139	0.84
35		1067	1146	YVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELD	80	0.822
36		201	270	FKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYL	70	0.77
37		27	35	AYTNSFTRG	9	0.676
38		465	509	ERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYR	45	0.675
39		700	720	GAENSVAYSNNSIAIPTNFTI	21	0.658
40		909	936	IGVTQNVLYENQKLIANQFNSAIGKIQD	28	0.633
41		437	458	NSNNLDSKVGGNYNYLYRLFRK	22	0.629
42		673	684	SYQTQTNSPRRA	12	0.619
43		790	817	KTPPIKDFGGFNFSQILPDPSKPSKRSF	28	0.602
44		891	907	GAALQIPFAMQMAYRFN	17	0.597
45		578	583	DPQTLE	6	0.589
46		620	631	VPVAIHADQLTP	12	0.579
47		329	362	FPNITNLCPFGEVFNATRFASVYAWNRKRISNCV	34	0.571

48		687	692	VASQSI	6	0.566
49		835	845	KQYGDCLGDIA	11	0.567
50		653	659	AEHVNNS	7	0.559
51		527	536	PKKSTNLVKN	10	0.546
52		635	642	VYSTGSNV	8	0.51

Table 2

B-cell epitopes predicted using Bepipred linear epitope prediction 2.0 in IEDB analysis resource web-server along with their start and end positions, average score, and VaxiJen 2.0 determined antigenicity scores.

1   **Table 2:** B-cell epitopes predicted using Bepipred linear epitope prediction 2.0 in IEDB analysis

2   resource web-server along with their start and end positions, average score, and VaxiJen 2.0

3   determined antigenicity scores.

Domain/proteins	Position	Sequences	Average Score	Antigenicity
RBD	341-342	VF	0.502	-
	344-349	ATRFAS	0.520	-0.151
	351-363	YAWNRKRISNCVA	0.522	0.394
	372-378	ASFSTFK	0.527	0.087
	382	V	0.464	-
	402-427	IRGDEVRQIAPGQTGKIADYNYKLPD	0.575	0.932
	440-485	NLDSKVGGNYNYLYRLFRKSNLKPFER DISTEIYQAGSTPCNGVEG	0.554	0.210
	493-516	QSYGFQPTNGVGYQ	0.535	0.670
NTD	72-81	GTNGTKRFDN	0.573	0.667
	110-113	LDSK	0.511	-
	146-155	HKNNKSWMES	0.573	0.174
	161-162	SS	0.503	-
	164	N	0.499	-
	172-191	SQPFLMDLEGKQGNFKNLRE	0.553	0.749
MBE	199-218	YRIGNYKLNTDHSSSSDNIA	0.614	0.222
EBE	57-71	YVYSRVKNLNSSRVP	0.565	0.449

Table 3

Active interface amino acid residues and binding scores among Toll Like Receptors (TLRs)
and the constructed vaccine CoV-RMEN.

1   **Epitope-based chimeric peptide vaccine design against S, M, and E proteins of SARS-CoV-**

2   **2 etiologic agent of global pandemic COVID-19: an *in silico* approach**

3   M. Shaminur Rahman, M. Nazmul Hoque, M. Rafiul Islam, Salma Akter, A.S.M. Rubayet- Ul-

4   Alam, Mohammad Anwar Siddique, Otun Saha, Md. Mizanur Rahaman, Munawar Sultana, Keith

5   A. Crandall, M. Anwar Hossain

7   **Table 3.** Active interface amino acid residues and binding scores among Toll Like Receptors
8   (TLRs) and the constructed vaccine CoV-RMEN.

Active residues of TLRs		Active residues of CoV-RMEN	HADDOCK score	ΔG (kcal mol⁻¹)
TLR-2	V536, C537, S538, C539, E540, S543, E547, P567, R569, L570	D72, Y75, L101, I103, I112, C150, F152, E153, Y154, V155, S156, F176, F178, R181, F182, L371	-30.4	-9.0
TLR-3	D36, H39, K41, R643, F644, P646, F647, T650, C651, E652, S653, I654, W656, F657, V658, N659, W660, I661, N662, E663	F43, S44, N45, V46, T47, W48, D72, Y75, F76, L101, I103, I112, F152, E153, Y154, V155, S156, Q157, F159, F178, R181, F182, L371	-47.2	-14.9
TLR-4	P53, F54, S55, H68, G70, Y72, S73, F75, S76, Q99, S102, G124,	G39, L40, D72, Y75, F76, F90, L101, I103, I112, F152, Y154, S156, Q157, F159, R174, E175, F176, F178, R181, F182, P183, L371, P374, P380, G381	-52.1	-16.0