

1 **E156G and Arg158, Phe-157/del mutation in NTD of spike protein in B.1.617.2 lineage of**  
2 **SARS-CoV-2 leads to immune evasion through antibody escape**

3 **Armi M Chaudhari<sup>1</sup>, Dinesh Kumar<sup>1</sup>, Madhvi Joshi<sup>1</sup>, Amrutlal Patel<sup>1</sup>, and Chaitanya**  
4 **Joshi<sup>1\*</sup>**

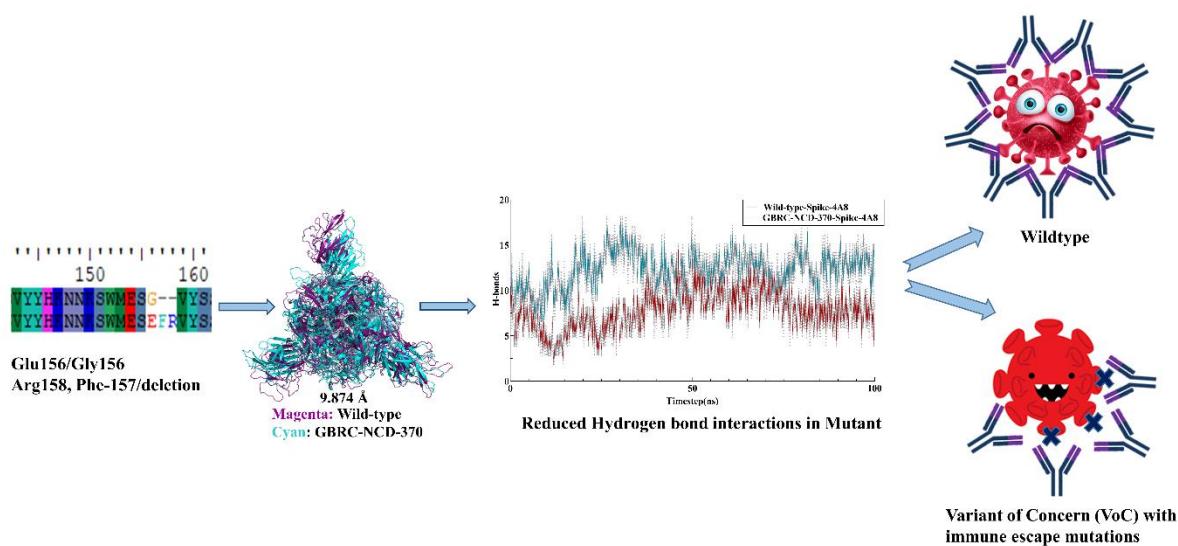
5 <sup>1</sup>Gujarat Biotechnology Research Centre (GBRC), Department of Science and Technology,  
6 Government of Gujarat, Gandhinagar-382011

7 **\*Corresponding author:** Chaitanya Joshi, Gujarat Biotechnology Research Centre (GBRC),  
8 Department of Science and Technology, Government of Gujarat, 6th Floor, Block B&D, MS  
9 Building, Gandhinagar-382011. **Tel/Fax:** 079-232-58680; **E-mail:** [director@gbrc.res.in](mailto:director@gbrc.res.in)

10 **Abstract**

11 Emerging variants of SARS-CoV-2 with better immune escape mechanisms and higher  
12 transmissibility remains a persistent threat across the globe. B.1.617.2 (Delta) variant was  
13 first emerged from Maharashtra, India in December, 2020. This variant is classified to be a  
14 major cause and concern of the second wave of COVID-19 in India. In the present study, we  
15 explored the genomic and structural basis of this variant through computational analysis,  
16 protein modelling and molecular dynamics (MD) simulations approach. B.1.617.2 variant  
17 carried E156G and Arg158, Phe-157/del mutations in NTD of spike protein. These mutations  
18 in N-terminal domain (NTD) of spike protein of B.1.617.2 variant revealed more rigidity and  
19 reduced flexibility compared to spike protein of Wuhan isolate. Further, docking and MD  
20 simulation study with 4A8 monoclonal antibody which was reported to bind NTD of spike  
21 protein suggested reduced binding of B.1.617.2 spike protein compared to that of spike  
22 protein of Wuhan isolate. The results of the present study demonstrate the possible case of  
23 immune escape and thereby fitness advantage of the new variant and further warrants  
24 demonstration through experimental evidence. Our study identified the probable mechanism  
25 through which B.1.617.2 variant is more pathogenically evolved with higher transmissibility  
26 as compared to the wild-type.

27 **Keywords:** SARS-CoV-2, COVID-19, Spike protein, Antibody, MD simulations and  
28 Docking



29

30 **1. Introduction**

31 India is witnessing the peak of another COVID-19 wave with over 0.32 million causalities  
32 since 2020, and more than 27.4 million confirmed positive cases as per WHO reports  
33 accessed on 27<sup>th</sup> May, 2021. Genome surveillance is a powerful tool to study the viral  
34 genomic profile, variants of concern and their epidemiological significance in disease  
35 outbreak outcome of the patients. All coronaviruses are positive-sense RNA viruses  
36 belonging to the order *Nidovirales* and family *Coronaviridae*. They are characterized by  
37 crown-like spikes on their surfaces and large enveloped genome of ~30 kilobases size. The  
38 SARS-CoV-2 genome contains four major structural proteins: spike (S), membrane (M),  
39 envelope (E), and nucleocapsid (N) protein. The spike (S) protein mediates entry and  
40 attachment of the coronavirus to the host cell surface receptors resulting in fusion and viral  
41 entry in the hosts. The membrane (M) protein defines the shape of the viral envelope while  
42 the envelope (E) protein and nucleocapsid (N) protein participates in viral assembly and  
43 budding of the virion complex in the infected cells [1,2]. SARS-CoV-2 uses ACE2 receptor  
44 for host cell entry and the spike protein of SARS-CoV-2 is primed by TMPRSS2 while the  
45 role of several other host receptors is partially explored with limited information that may  
46 determine the altered virulence and pathogenicity of the evolving SARS-CoV-2 lineages  
47 around the globe. SARS-CoV-2 possesses highly efficient and evolved strategies for  
48 proteolytic activation of spike, and host proteases have been shown to proteolytically process  
49 the spike protein. These include, but are not limited to, endosomal cathepsins, cell surface  
50 trans-membrane protease/serine (TMPRSS) proteases, furin, and trypsin that are critical  
51 determinants of the virus entry and pathogenesis in humans [3,4]. SARS-CoV-2, in  
52 comparison to SARS-CoV, contains a polybasic sequence motif, Arg-Arg-Ala-Arg (RRAR),  
53 at the S1/S2 boundary, furin-type cleavage site in its spike protein, which when cleaved can  
54 bind and activate neuropilin receptors. Further, research studies indicate that NRP1 enhances  
55 SARS-CoV-2 infectivity and is highly expressed in respiratory and olfactory epithelium [5].

56 Under the prevailing circumstances, the immune response of the patients plays a significant  
57 role in determining the disease fighting ability of the body. A myriad of various cell types  
58 such as macrophages, alveolar epithelial cells, lymphoid cells, and dendritic cells (DCs) have  
59 a major role in the first line of defense. Once the immune system is triggered for the entry of  
60 foreign viral pathogens inside the body and that breached the first lines of defense system,  
61 several specific molecular and inter-cellular signaling cascades ensure the establishment of  
62 the body's immune response [6,7]. When the invading respiratory viruses evolve mechanism  
63 that either circumvent or suppress the innate immune responses to create a window of  
64 opportunity for efficient virus replication, thereby often causing disease. The affected innate  
65 immune response also impacts subsequent adaptive immune responses, and therefore viral  
66 innate immune evasion often undermines fully protective immunity such as lack of virus  
67 neutralizing antibodies [8–11]. Further, genetic makeup and evolution in virus also enable  
68 them to develop mutations that can cause immune escape and immune evasion in the hosts,  
69 thereby increasing the chances of severity and virulence of the pathogenic variants [12].  
70 These variants, which are observed with features such as higher transmittability are  
71 categorized as Variants of Concern (VoC) by Public Health England (PHE), UK; CDC, USA,  
72 and World Health Organization (WHO) based on their risk assessment criteria of infection  
73 severity, susceptibility geographical prevalence, and transmission in humans. Therefore,  
74 genomic surveillance studies are essential in monitoring these variants that may even arise in  
75 the future pandemics.

76 New variants of SARS-CoV-2 are emerging challenge for scientific community and public  
77 health system in the different geographical regions across the globe. These variants, have  
78 been designated as Variant of Concern (VoCs) which has noticeable higher transmissibility  
79 and probably more virulent compared to other variants. Genomic structure of such variants  
80 suggests that they were evolved to escape the immune system of the host thereby giving them  
81 the fitness advantage and thus increased spread among the population. Further, research is  
82 needed to establish the mechanism of escape and potential host genetic factor that might help  
83 in these evolved pathogenic viral strains of SARS-CoV-2.

84 Furthermore, understanding of the role of virus-host interactions and immune response during  
85 these SARS-CoV-2 infections will be pivotal to ultimately meet these evolving challenges.  
86 Eventually, efficacy of the combined innate and adaptive responses is on the host's side,  
87 while the virulence and its capacity to evade the host's immune responses is on the virus' side,  
88 together, the balance between them dictate the disease outcome in the context of the host-  
89 virus interactions. Recent studies on spike protein interactions with monoclonal antibodies  
90 4A8 suggests that N- terminal domain is essential binding site for 4A8 [13]. Some prominent  
91 mutations (>99.7% frequency) in virus favors the virus like D614G [14] and some favors the  
92 host like C241T [15]. To find the same, this research focus on the mutations in N-terminal  
93 domain in B.1.617.2 (now delta) lineage of SARS-CoV-2 and its impact on protein structural  
94 changes and antibodies binding using molecular modeling and dynamics approach.

## 95 2. MATERIAL AND METHODS

### 96 2.1 Protein complexes used for this study

97 Variants of Spike protein from SARS-CoV-2 were taken into study. Mutated spike from  
98 SARS-CoV-2 used in this study were derived from amino acid sequence submitted in  
99 GAISAD with accession number EPI\_ISL\_2001211. Reference protein with PDB id 7KRQ  
100 was used for homology modelling.

101 **2.2 Protein modelling and Molecular dynamics simulations**

102 Homology modelling panel implemented in Schrodinger suite release 2021-1 was used to  
103 build mutated spike protein with reference protein 7KRQ. Sequence was imported and  
104 homology blast search was performed. Crystal structure of 7KRQ was imported in to maestro  
105 and protein complex refinement was performed using protein preparation wizard [16].  
106 Missing side chains were added through PRIME and pKa refinement was done with epik  
107 [17]. Molecular dynamics simulation for spike protein and Spike-antibody complexes were  
108 performed in Schrodinger suite 2021-1 implemented DESMOND till 100 nanoseconds (ns)  
109 [18]. Protein structures were refined using OPLS4 force field and altered hydrogen bonds  
110 were refilled using structure refinement panel implemented in Schrodinger[19][20]. Particle  
111 mesh Ewald method is applied to calculate long rage electrostatic interactions. [21]. The  
112 trajectories were recorded at every 1.2 ps intervals for the analysis. TIP3P water molecules  
113 were added and 1.5 M Salt concentration was added to neutralize the system [22]. The  
114 Martyna–Tuckerman–Klein chain coupling scheme with a coupling constant of 2.0 ps was  
115 used for the pressure control and the Nosé–Hoover chain coupling scheme for temperature  
116 control [23]. MD simulations were performed at 310.3K temperature. The behaviour and  
117 interactions between the protein and protein were analyzed using the Simulation Interaction  
118 Diagram tool implemented in Desmond MD package. The stability of complex was  
119 monitored by examining the RMSD of the protein and protein atom positions in time.  
120 PYMOL was used for obtaining high resolution images [24]. Protein modelling and  
121 Molecular dynamics simulations were performed into duplicates.

122 **2.3 Molecular docking of Spike protein with monoclonal antibodies using spike and  
123 affinity prediction**

124 Variants of spike proteins Wildtype (7KRQ) and Mutant (GBRC-NDC-370) were docked  
125 with monoclonal antibody 4A8. Protein structures were prepared using protein-preparation  
126 wizard [16]. After structure refinement of protein, PIPER was used for the protein-protein  
127 docking [25]. For binding residues (as shown in figure 2A) detection among both receptor  
128 (spike) and ligand (antibody-4A8) attraction forces were applied with <3Å cut off. 70000  
129 docking poses were checked for fulfilling the criteria of distance restrains applied for the  
130 binding sites residues. Recently deposited crystal structure of spike protein binding with  
131 monoclonal antibody was taken for applying the restraint file showing list of spike residues  
132 binding with residues of 4A8. Top 30 poses were generated and a pose with highest free glide  
133 energy was used for the MD analysis.

134 Alanine residue scanning was performed for the binding affinity prediction in PDB deposited  
135 spike antibody complex with id 7C2L [13]. Binding site residues were mutated to alanine in  
136 order to bind the pivotal residues involved into direct binding with antibody. Positive value of  
137  $\Delta$  affinity indicated that while mutating binding sites residues to alanine, binding is hindered

138 due to small side chains of alanine and which in terms implies that those important residues  
139 were essential for direct affinity with antibody [26,27]. Residue alanine scanning panel  
140 present in Biologics of Schrodinger 2021-1 is used to perform the above task.

141 **2.4 Binding energy calculation**

142 Binding energy for protein-protein complex was calculated in the form of Prime Molecular  
143 Mechanics-Generalized Born Surface Area (MMGBSA) using thermal\_mmgbasa.py  
144 implemented in PRIME module of Schrodinger [28–30].  $\Delta G$  of protein-protein complex was  
145 calculated using following equation.

146 
$$\Delta G_{\text{Bind}} = \Delta G_{\text{SA}} + \Delta G_{\text{Solv}} + \Delta E_{\text{MM}}$$

147 VSEB solvation model and OPLS4 force-field were used for calculation of MMGBSA.  
148 Protein-protein complex system seems to have stable RMSD pattern after 60ns. These frames  
149 were used to calculate MMGBSA. First energy minimized structure out of 30 was used to  
150 find dominant interacting residues among spike. Interaction image was taken in new version  
151 of Schrodinger 2021-2 where protein-protein interaction images can be taken in Biologics.

152 **2.5 Dynamics cross-correlation matrix (DCCM) and Principal Component analysis  
(PCA)**

154 Correlative and anti-correlative motions play a vital role in recognition and binding in a  
155 biological-complex system which can be prevailed by MD simulation trajectory by  
156 determining the covariance matrix about atomic fluctuations [31]. The extent of correlative  
157 motion of two residues (or two atoms or proteins) can be symbolized by the cross-correlation  
158 coefficient,  $C_{ij}$ . It is defined by following equation:

159 
$$C_{ij} = \frac{\langle \Delta r_i * \Delta r_j \rangle}{(\langle \Delta r_i \rangle^2 \langle \Delta r_j \rangle^2)^{1/2}} \quad \dots \text{eq: 1}$$

160 From above equation,  $i$  ( $j$ ) explains ith (jth) two residues (or two atoms or proteins),  $\Delta r_i$  ( $\Delta r_j$ )  
161 is the displacement vector corresponding to ith (jth) two residues (or two atoms or proteins),  
162 and  $\langle \dots \rangle$  stand for the ensemble average. The value of  $C_{ij}$  is from 1 to -1.  $+C_{ij}$  implies  
163 positively correlated movement (the same direction) indicated into blue color, and  $-C_{ij}$   
164 implies anti-correlated movement (opposite direction) indicated into red color. The higher the  
165 absolute value of  $C_{ij}$  is, the more correlated (or anti-correlated) the two residues (or two  
166 atoms or proteins).

167 PCA is an implicit tool to unsheathe the essential information from MD trajectories by  
168 pulling out global slow motions from local fast motions[32]. To perform PCA, the covariance  
169 matrix  $C$  was calculated initially. The elements  $C_{ij}$  in the matrix  $C$  are defined as:

170 
$$C_{ij} = \langle (r_i - \langle r_i \rangle) * (r_j - \langle r_j \rangle) \rangle \quad \dots \text{eq: 2}$$

171 From equation 2,  $r_i$  and  $r_j$  are the instant coordinates of the  $i^{\text{th}}$  or  $j^{\text{th}}$  atom,  $\langle r_i \rangle$  and  $\langle r_j \rangle$  and  
172 mean the average coordinate of the  $i^{\text{th}}$  or  $j^{\text{th}}$  atom over the ensemble. The principal  
173 components (PCs) were calculated by diagonalization and obtaining the eigenvectors and  
174 eigenvalues for the covariance matrix  $C$ . The principal components (PCs) are the projections

175 of a trajectory on the principal modes, of which usually the first few ones are largely  
176 responsible for the most important motions. DCCM and PCA both were analyzed using  
177 Schrodinger 2021-1 implemented python script *run trj\_essential\_dynamics.py* script of  
178 Desmond [18].

179 **3. Result and discussion**

180 Spike protein of SARS-CoV-2 is known to bind ACE-2 receptor mediating virus entry. Spike  
181 protein is more prone to mutations. For better penetrance viral spike had gone through several  
182 mutations like D614G for increasing spike density and infectivity, E484K for reducing the  
183 antibody neutralization , N501Y and K417N for altering spike interacting with human ACE  
184 and human derived antibody [33–35]. Our study focuses on major deletion occurred in NTD  
185 of spike protein at nucleotide position 22029-22035 (6bp) which in-terms induce 2 amino-  
186 acid deletions Arg158, Phe-157/del and one amino acid mutation E156G.

187 **3.1 Mutational landscape of spike protein and its effects with respect to B.1.617.2 lineage**

188 Mutations listed in table 1, were impacting major structural change or not were studied  
189 through structural alignment of both variants of spike proteins in Pymol. Superimposed  
190 structure of wild-type and GBRC-NCD-370 spike with alignment RMSD 6.905 Å was shown  
191 in figure 1A. RMSD value higher than 1Å suggests that these mutations were imparting  
192 significant structural change in both variants of spike [36]. Among the list of overall  
193 mutations, unique mutations E156G and Arg158, Phe-157/del were falling in NTD of spike.  
194 Figure 1D, explains the effect of these mutations in changing amino acids conformation in  
195 ball and stick form. One can visualize the difference in alignment of amino acids in NTD  
196 within both variants which in terms effects the change in intermolecular contacts within spike  
197 (Figure 1E). Due to these mutations intra-atomic contacts within amino acids of mutant  
198 (GBRC-NCD-370) is drastically increased with respect to the wild-type. In figure 1D, these  
199 mutations change intra-atomic contacts as such that in wild-type Glu-156 is interacting with  
200 Phe-140 and Arg158 by forming only one hydrogen bond and other hydrophobic interactions,  
201 while same cavity in mutant form four hydrogen bonds (Glu154-Ala123, Glu154-Arg102,  
202 Ser155-Asp142, and Val157-Gly156) and higher hydrophobic interactions. Higher intra-  
203 atomic contacts leads to the decrease in flexibility (increase in rigidity) by  $\Delta S_{vib}$  ENCoM: -  
204 **0.500 kcal.mol<sup>-1</sup>.K<sup>-1</sup>** ( $\Delta$  Vibrational Entropy Energy between Wild-Type and Mutant). To  
205 correlate these findings with MD simulations RMSD and RMSF were analyzed to further  
206 comment on flexibility of GBRC-NCD-370.

207 RMSD (Root mean square deviation) for wild-type and GBRC-NCD-370 complex observed  
208 were  $5.89 \pm 0.026$  and  $2.54 \pm 0.018$ , respectively (figure1G). RMSD graph is clearly narrating  
209 that mutations in GBRC-NCD-spike protein are enhancing its stability compared to the wild-  
210 type trimeric complex (figure1H). RMSF (Root mean square fluctuation) was 3.6 Å lower in  
211 NTD of GBRC-NCD-370 spike compared to the wild-type complex. Decreased RMSF in  
212 NTD explains reduced flexibility of amino acid residues within the region. Aurélie Bornot et  
213 al 2010 precisely explains the protein flexibility in terms of RMSF and B-factors, where  
214 increasing in RMSF values is related to increased change in protein conformation. In some  
215 cases, amino acid residues are flexible though RMSF can be rigid through B-factors [37]. In

216 our case, protein seems to be flexible in both cases through RMSF and B-factors  
217 (Supplementary figure S2). While no major change was observed with respect to other part of  
218 protein. As NTD is binding site of wide variety of monoclonal antibodies, rigidization in that  
219 region further affects the binding of antibody within both variants of spike. Principle  
220 component analysis was performed where first dominant dynamic mode PC1 among the  
221 trajectories were analyzed in VMD. Porcupines plots showing the projection of mode vectors  
222 based on the residue fluctuation throughout the trajectories were shown in Figure 2C. Length  
223 of mode-vectors in wild-type complex was higher compared to GBRC-NCD-370, which  
224 suggests that overall NTD flexibility is decreased in B.1.617.2 lineage. Increase in  
225 intermolecular contacts in mutated region further support the rigidization of GBRC-NCD-370  
226 (figure1E).

227 Mutations in NTD were covering the binding domain for monoclonal antibodies (Figure 2A).  
228 PDB id 7CL2 was chosen as a wild-type complex of spike with 4A8 monoclonal antibody.  
229 GBRC-NCD-370 variant was docked with 4A8 was chosen as mutant complex. Glide energy  
230 of wild type A48 and GBRC-NCD-4A8 were -115.64 and -68.74 respectively. More  
231 negative energy score was showing enhance binding among protein-protein complex.  
232 MEDUSA five class predictions narrate the decrease in flexibility by 2% for the GBRC-  
233 NCD-370 compared to wild-type (Figure 2B & 2C). Spike flexibility and rigidity were  
234 analyzed in term of its binding with monoclonal antibodies by performing alanine residue  
235 scanning (Figure 2D). Wild-type-4A8 was processed through residue scanning with alanine  
236 mutagenesis to investigate the important residue for the binding with 4A8. Amino acid  
237 residues Y145, K147, 150K, 152W, 156E, 157F, and 158R showed positive binding affinity  
238 values with 4A8 upon mutating these residues to alanine (Figure 2D). These results clearly  
239 indicate that the mutations in the NTD domain of spike caused decrease in binding of 4A8  
240 antibody in B.1.617.2 lineage. To further explore the effect, the mutations in NTD with  
241 respect to affinity with 4A8, MD simulations were performed in duplicates and binding  
242 energies among both variants were analyzed.

243 **3.2 Effect of Arg158, Phe-157/del and E156G in N-terminal domain with respect to the**  
244 **host immunity**

245 MD simulations of both mutant and wild type spikes with 4A8 antibody explored the effect  
246 of mutation Arg158, Phe-157/del and one amino acid mutation E156/G in terms binding with  
247 monoclonal antibodies and depict the case of immune evasion. RMSD of GBRC-NCD-370-  
248 4A8 and wildtype-4A8 was  $20.147 \pm 0.526 \text{ \AA}$  and  $16.142 \pm 0.453 \text{ \AA}$ , respectively (Figure 3A). In  
249 GBRC-NCD-370-4A8 plateau was reached after 65 ns and jumps were observed during the  
250 MD simulations, while wildtype-4A8 was found to be stable after 20ns only. These major  
251 difference among the both trajectories shows that GBRC-NCD-370-4A8 complex is having 4  
252  $\text{\AA}$  less RMSD, leads to more stable than wildtype-4A8. Hydrogen bonds formation within  
253 GBRC-NCD-370-4A8 complex was lower (on an average 6 hydrogen bonds difference was  
254 observed) compared to wildtype-4A8 (figure 3B). Hydrogen bond formation clearly indicates  
255 reduced interaction of antibodies in B.1.617.2 lineage.

256 Binding energy among the complex was analyzed through MMGBSA. Major energies  
257 contributing to the complex formation were elucidated in table 2 with bold text. Spike-4A8

258 complex formation is driving through the major electrostatic, covalent, ionic-interactions,  
259 lipophilic (hydrophobic) and Vander-Waals interactions within both complexes. Overall free  
260 energy binding  $\Delta G$  in wild-type and GBRC-NCD-370 complex is  $-119.086 \pm 19.42$  and  $-$   
261  $55.496 \pm 14.57$ , respectively. Interaction in energy minimized structure obtained through  
262 MMGBSA approach is shown in figure 3E & 3F. In wild-type complex overlapping strong  
263 interaction between charged negative (orange) and charge positive residues (blue) is way  
264 higher compare to mutant. For example A-Lys147: B-Glu72, A-Lys150: B-Glu57 & B-Glu55  
265 were forming hydrogen bonds and salt bridges in spike (A) and 4A8 (B). Jason E Donald and  
266 group suggests that salt bridges were geometric specific and designable interactions [38,39].  
267 Lys150 is forming salt bridge and hydrogen bonds with two negative charged amino acids  
268 Glu57 and Glu55 (Figure 3E & 3F). These kind of favorable interactions are formed in wild-  
269 type spike but absent in GBRC-NDC-370, leads to conclude that geometry of NTD in spike  
270 had changed as such that it is reducing the strong interaction with 4A8 in GBRC-NCD-370  
271 (mutant) (Table:2). This kind of salt-bridges are favorable exist in hydrophobic environment  
272 [40], leads to higher lipophilic energy in wild type (-30.334 Kcal/mole) compare to GBRC-  
273 NCD-370 (-11.9757Kcal.mole). Overall Wildtype spike seems to have better binding with  
274 monoclonal antibodies compared to the GBRC-NCD-370, which leads to conclude that there  
275 is possible case of immune evasion among B.1.617.2 lineage.

276 Dynamics cross-correlation matrix (DCCM) of wild-type and GBRC-NCD-370 spikes with  
277 4A8 is shown in figure 3. In DCCM wild-type-spike-4A8 is showing higher intensity of blue  
278 color compared to the GBRC-NCD-370-spike-4A8. Positive  $C_{ij}$  values indicate blue colors  
279 leads to better interaction profile between those residues. NTD region covers 17-305 amino  
280 acid residues where major residue contributing in direct contact was shown in figure 2. In the  
281 region covering orange arrow (antibody 4A8) intensity of blue color is higher, indicating  
282 more positive cross correlation with respect to NTD region of spike wild-type compare to  
283 GBRC-NCD-370. In wild-type complex NTD residues have showing higher negative cross-  
284 correlation compare to GBRC-NCD-370 (Figure 3C &3D). Results of negative cross-  
285 correlations in some regions were completely correlating with change in flexibility of NTD.  
286 Higher intensified positive cross-correlation showed structural compactness among the NTD  
287 in GBRC-NCD-370, which can be unfavorable for the antibody binding. Overall cross-  
288 correlation among 4A8 which were shown in box a and b with respect to spike, were positive  
289 in wild-type and negative in GBRC-NCD-370 supporting the case of antibody escape.

290 Mutant have increased rigidity which is concluded using Insilco methods. Increase in  
291 interatomic contacts have enhanced the rigidity which is shown by two platforms pymol and  
292 Medusa. Further PCA analysis and porcupine plot were also showing the decreased in  
293 flexibility (increased rigidity). Andrey Karshikoff and group had shown binding site residues  
294 of protein have flexible tendency, to have better interaction with ligand (here the case of  
295 antibody) [41]. As binding site residues are more engaged with intra-atomic contacts itself,  
296 tendency of sharing contacts with outsider protein will be less [42]. Based on results obtained  
297 through higher RMSF and Porcupine plots it was hypothesized that rigidization among NTD  
298 leads to antibody escape.

299

300 **4. Conclusion**

301 The present study addressed the critical structural and genomic determinants of the SARS-  
302 CoV-2 (B.1.617.2/Delta) variant which is most dominant in India during the second wave and  
303 spreading quickly in different geographical regions of the globe. The E156G and Arg158,  
304 Phe-157/del mutations in NTD of spike protein of SARS-CoV-2 (B.1.617.2/Delta) variant  
305 showed more rigidity and reduced flexibility compared to Wuhan isolate. Further, our study  
306 showed possible case of immune escape by demonstrating reduce binding of mutant spike  
307 compared to Wuhan isolate with reported antibody known to bind NTD of spike protein  
308 thereby providing insights into the structural basis and highlight the impact of the key  
309 mutations for the higher transmissibility, pathogenicity and virulence. Therefore, it is  
310 important to better monitor and identify the new emerging variants of SARS-CoV-2 using  
311 genome sequencing and surveillance that may have increased transmission, virulence and  
312 altered antigenicity evolved over time for epidemiological significance.

313 **Credit authorship and contribution statement**

314 AC performed all Insilco experiments, analysis, writing original draft preparation, data  
315 curation. DK analyzed the sequences, write and edited the manuscript. AP, MJ, CJ writing  
316 review and editing, supervision, project administration, funding acquisition

317 **Acknowledgement**

318 Authors would like to acknowledge Department of Science and Technology (DST),  
319 Government of Gujarat for infrastructure support for the research work. We also thank  
320 Gujarat University for providing us GPU accelerated computer facility to enhance the speed  
321 of our work.

322 **Funding**

323 Funding was provided by Department of Science and Technology (DST), Government of  
324 Gujarat, Gandhinagar, India.

325 **Declaration of Competing Interest**

326 The authors declare that they have no known competing financial interests or personal  
327 relationships that could have appeared to influence the work reported in this paper.

328

329

330

331

332

333

334

335 **Tables and Figures**

336 **Table 1:** List of spike protein mutation present in GBRC-NCD-370

| <b>Sr. No</b> | <b>Amino Acid Change</b> |
|---------------|--------------------------|
| <b>1</b>      | T19R                     |
| <b>2</b>      | G142/D                   |
| <b>3</b>      | del157/158/ R156G        |
| <b>4</b>      | R452G                    |
| <b>5</b>      | A222V                    |
| <b>6</b>      | L452R                    |
| <b>7</b>      | T478K                    |
| <b>8</b>      | D614G                    |
| <b>9</b>      | P681R                    |
| <b>10</b>     | D950N                    |

337

338 **Table 2:** Differences in energy components contributing to Complex formation within Wild-  
339 type (7KRQ) and Mutated GBRC-NCD-370 monoclonal antibodies (4A8).

| <b>Energy components</b>                          | <b>Wildtype-4A8</b> | <b>GBRC-NCD-370-4A8</b> |
|---|---------------------|-------------------------|
| <b>Glide energy</b>                               | <b>-115.64</b>      | <b>-68.74</b>           |
| $\Delta G$ Binding                                | -119.0869622        | -55.4968                |
| <b><math>\Delta G</math> Electrostatic energy</b> | -791.09011          | -357.715                |
| <b><math>\Delta G</math> Covalent energy</b>      | -14.19260182        | -7.50638                |
| <b><math>\Delta G</math> Hbonds energy</b>        | -8.385258773        | -5.17996                |
| <b><math>\Delta G</math> Lipophilic energy</b>    | -30.33398583        | -11.9757                |
| $\Delta G$ pi piinteraction energy                | -2.106536604        | 1.156437                |
| $\Delta G$ selfcontactcorrelation                 | -0.023560517        | -0.15669                |
| $\Delta G$ Solv_GB                                | 781.5183498         | 434.7153                |
| <b><math>\Delta G</math> vdw energy</b>           | -82.85846217        | -108.834                |

340

341

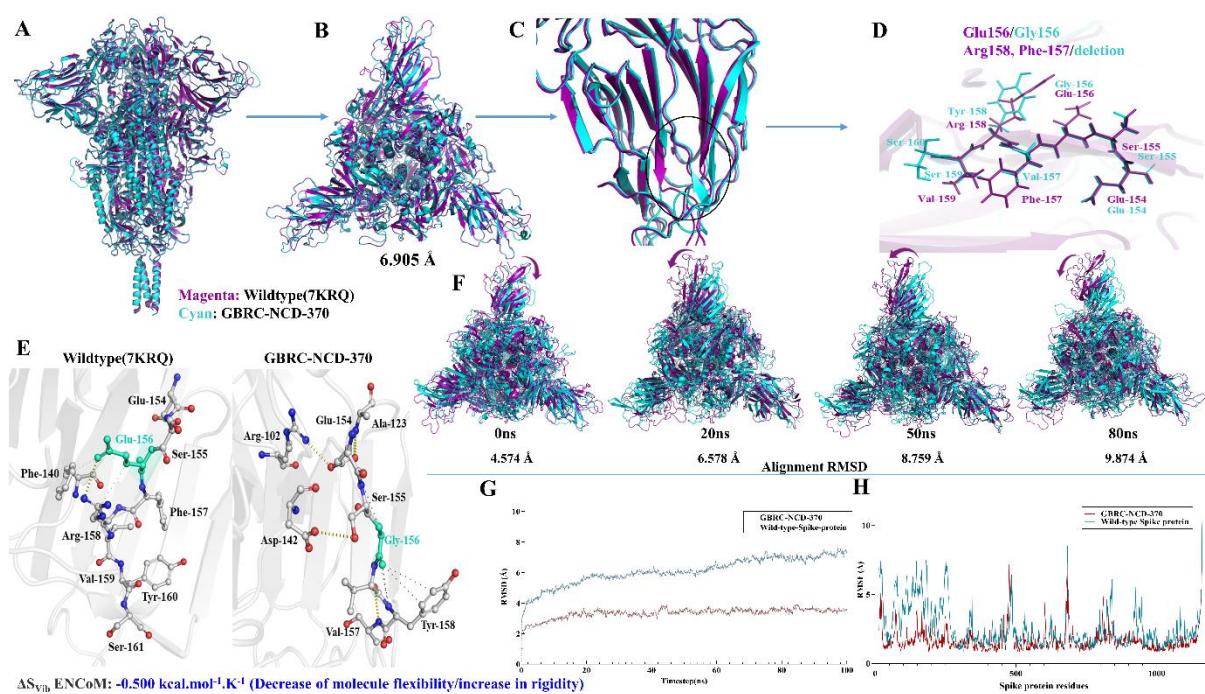
342

343

344

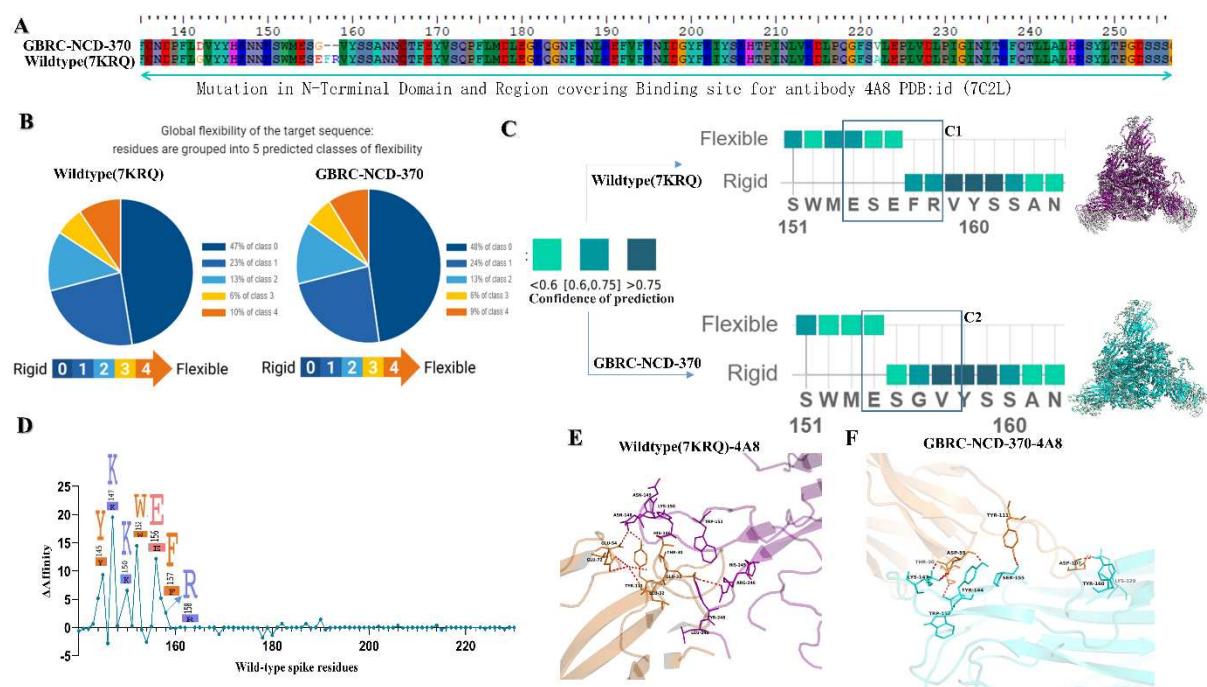
345

346 **Figures**



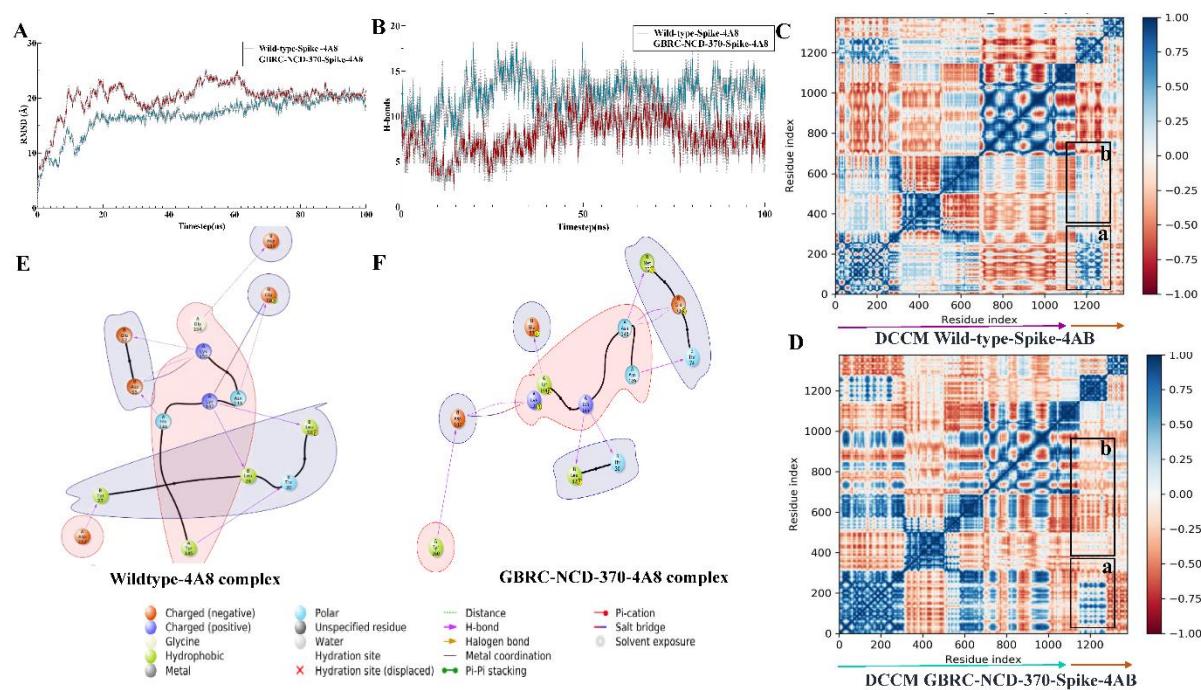
347  $\Delta S_{\text{vib}} \text{ ENCoM: } -0.500 \text{ kcal.mol}^{-1}.K^{-1}$  (Decrease of molecule flexibility/increase in rigidity)

348 **Figure 1: Rigidization and reduce in flexibility of N-Terminal domain of spike protein.**  
349 **1A:** 3D Structural alignment of wild-type [7KQR] and GBRC-NCD-370 trimetric spike  
350 proteins with superimposition RMSD value: 6.356. Wild-type protein is shown in magenta  
351 and GBRC-NCD-370 is shown in cyan color. **1B:** Top view of trimetric spike protein. **1C &**  
352 **1D:** Focusing structural difference in NTD of wild-type and GBRC-NCD-370 spike protein.  
353 **1E:** Intermolecular contacts between wild-type and mutant spike protein. **1F:** Frame  
354 superimposition of wild-type and GBRC-NCD-370 spike proteins for visualization of  
355 dynamics modes depicting difference in NTD. Magenta colored arrow showing dynamic  
356 moments of wild-type spike. **1G & 1H:** RMSD and RMSF plot generated from MD-  
357 Simulation respectively. Wild-type protein is shown in deep teal color and GBRC-NCD-370  
358 is shown in red color.



359

360 **Figure 2: Reduce flexibility influence the binding of monoclonal antibodies (4A8) with**  
 361 **spike protein.** **2A:** Amino acid residues falling in the binding region NTD of spike.  
 362 Wildtype and mutated sequences were annotated with NCBI reference/accession id number  
 363 MN986947.3 and GBRC-NCD-370 respectively. **2B:** Output generated from MEDUSA to  
 364 determine change rigidity and flexibility among both variants. 5 predicted classes were  
 365 generated in range of 0-4, where blue region explains rigid regions and yellow to orange  
 366 regions explains the flexible regions among proteins. **2C:** Flexible and rigid regions in region  
 367 covering mutation. Cyan to deep teal color represents the flexible to rigid region with COP  
 368 (confidence of prediction) with  $<0.6$  and  $>0.75$  respectively. Porcupines plots generated from  
 369 PCA analysis were also supporting the same were shown red tube conformation with mode  
 370 vectors. **2D:** Alanine residues scanning of wildtype-4A8 complex. Residues important in bind  
 371 with monoclonal antibodies were shown in logo plot with positive binding affinity. **2E:**  
 372 Binding pose of Wildtype-4A8 complex. Spike is shown in magenta color while 4A8 is  
 373 shown in orange color. Residues involved in pivotal contacts like hydrogen bonds (red) were  
 374 shown in ball and stick conformation with black colored labels. **2F:** Binding pose of GBRC-  
 375 NCD-370-4A8 complex. Spike is shown in cyan color while 4A8 is shown in orange color.  
 376 Residues involved in pivotal contacts like hydrogen bonds (red) were shown in ball and stick  
 377 conformation with black colored labels.



378

379

380 **Figure 3:** MD analysis of Spike-antibodies complexes. **3A:** RMSD (root mean square deviation) within Wildtype-4A8 (cyan) and gbrc-ncd-370-4A8 (mutant) complex. **3B:** 381 Hydrogen bonds formation within Wildtype-4A8 (cyan) and gbrc-ncd-370-4A8 (mutant) 382 complex. **3C:** Dynamics cross-correlation matrix obtained from trajectories analysis of 383 wild-type-4A8 complex. Spike protein shown in magenta arrow and orange arrow is indicating 384 4A8. **3D:** Dynamics cross-correlation matrix obtained from trajectories analysis of 385 GBRC-386 NCD-370-4A8 complex. Spike protein shown in cyan arrow and orange arrow is indicating 387 4A8. Blue to red color represents the  $c_{ij}$  values between 1 to -1. No cross correlation was 388 shown by white color. **3E & 3F:** Energy minimized structure obtained through MMGBSA 389 for Wildtype-4A8 and GBRC-NCD-370-4A8 respectively. Positive chagrined, negative 390 charged amino acid residues were shown in orange and blue color respectively

391

392

393

394

395

396

397

398

399

400 **References**

- 401 1. Gordon DE, Hiatt J, Bouhaddou M, et al. Comparative host-coronavirus protein interaction  
402 networks reveal pan-viral disease mechanisms. *Science* 2020; 370:
- 403 2. Rohaim MA, El RF, Clayton E, et al. Since January 2020 Elsevier has created a COVID-  
404 19 resource centre with free information in English and Mandarin on the novel coronavirus  
405 COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s  
406 public news and information . 2020;
- 407 3. Millet JK, Whittaker GR. Host cell proteases: Critical determinants of coronavirus tropism  
408 and pathogenesis. *Virus Research* 2015; 202:120–134
- 409 4. Hikmet F, Méar L, Edvinsson Å, et al. The protein expression profile of ACE2 in human  
410 tissues. *Molecular Systems Biology* 2020; 16:1–16
- 411 5. Daly JL, Simonetti B, Antón-Plágaro C, et al. Neuropilin-1 is a host factor for SARS-CoV-  
412 2 infection. *bioRxiv* 2020; 865:861–865
- 413 6. Melms JC, Biermann J, Huang H, et al. A molecular single-cell lung atlas of lethal  
414 COVID-19. *Nature* 2021;
- 415 7. Reizis B. Plasmacytoid Dendritic Cells: Development, Regulation, and Function.  
416 *Immunity* 2019; 50:37–50
- 417 8. Delorey TM, Ziegler CGK, Heimberg G, et al. COVID-19 tissue atlases reveal SARS-  
418 CoV-2 pathology and cellular targets. *Nature* 2021;
- 419 9. Voss WN, Hou YJ, Johnson N V., et al. Prevalent, protective, and convergent IgG  
420 recognition of SARS-CoV-2 non-RBD spike epitopes in COVID-19 convalescent plasma.  
421 *bioRxiv* 2020; 5268:1–11
- 422 10. Garcia-Beltran WF, Lam EC, Astudillo MG, et al. COVID-19-neutralizing antibodies  
423 predict disease severity and survival. *Cell* 2021; 184:476-488.e11
- 424 11. Deng X, Garcia-knight MA, Khalid MM, et al. Transmission, infectivity, and  
425 neutralization of a spike L452R SARS-CoV-2 variant. *Cell* 2021;
- 426 12. Hoffmann M, Arora P, Groß R, et al. SARS-CoV-2 variants B.1.351 and P.1 escape from  
427 neutralizing antibodies. *Cell* 2021; 2384–2393
- 428 13. Chi X, Yan R, Zhang J, et al. A neutralizing human antibody binds to the N-terminal  
429 domain of the Spike protein of SARS-CoV-2. *Science* 2020; 369:650–655
- 430 14. Zhang L, Jackson C, Mou H, et al. The D614G mutation in the SARS-CoV-2 spike  
431 protein reduces S1 shedding and increases infectivity. *bioRxiv* : the preprint server for  
432 biology 2020;
- 433 15. Chaudhari A, Chaudhari M, Mahera S, et al. In-Silico analysis reveals lower transcription  
434 efficiency of C241T variant of SARS-CoV-2 with host replication factors MADP1 and  
435 HNRNP-1. *bioRxiv* 2020;
- 436 16. Madhavi Sastry G, Adzhigirey M, Day T, et al. Protein and ligand preparation:  
437 parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-  
438 Aided Molecular Design* 2013; 27:221–234
- 439 17. Shelley JC, Cholleti A, Frye LL, et al. Epik: a software program for pKaprediction and

440 protonation state generation for drug-like molecules. *Journal of Computer-Aided Molecular*  
441 *Design* 2007; 21:681–691

442 18. . SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. 2006;

443 19. van Zundert GCP, Moriarty NW, Sobolev O V, et al. Macromolecular refinement of X-  
444 ray and cryoelectron microscopy structures with Phenix/OPLS3e for improved structure and  
445 ligand quality. *Structure* 2021;

446 20. Steinbrecher T, Abel R, Clark A, et al. Free Energy Perturbation Calculations of the  
447 Thermodynamics of Protein Side-Chain Mutations. *Journal of Molecular Biology* 2017;  
448 429:923–929

449 21. Toukmaji AY, Board JA. Ewald summation techniques in perspective: A survey.  
450 *Computer Physics Communications* 1996; 95:73–92

451 22. Zielkiewicz J. Structural properties of water: Comparison of the SPC, SPCE, TIP4P, and  
452 TIP5P models of water. *Journal of Chemical Physics* 2005; 123:

453 23. Martyna GJ, Klein ML, Tuckerman M. Nosé-Hoover chains: The canonical ensemble via  
454 continuous dynamics. *The Journal of Chemical Physics* 1992; 97:2635–2643

455 24. . pymol.

456 25. Kozakov D, Brenke R, Comeau SR, et al. PIPER: an FFT-based protein docking program  
457 with pairwise potentials. *Proteins* 2006; 65:392–406

458 26. Harlow GR, Halpert JR. Alanine-scanning Mutagenesis of a Putative Substrate  
459 Recognition Site in Human Cytochrome P450 3A4: ROLE OF RESIDUES 210 AND 211 IN  
460 FLAVONOID ACTIVATION AND SUBSTRATE SPECIFICITY\*. *Journal of Biological*  
461 *Chemistry* 1997; 272:5396–5402

462 27. Simonsen SM, Sando L, Rosengren KJ, et al. Alanine Scanning Mutagenesis of the  
463 Prototypic Cyclotide Reveals a Cluster of Residues Essential for Bioactivity\*. *Journal of*  
464 *Biological Chemistry* 2008; 283:9805–9813

465 28. Lyne PD, Lamb ML, Saeh JC. Accurate prediction of the relative potencies of members  
466 of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *Journal of*  
467 *Medicinal Chemistry* 2006; 49:4805–4808

468 29. Greenidge PA, Kramer C, Mozziconacci JC, et al. MM/GBSA binding energy prediction  
469 on the PDBbind data set: Successes, failures, and directions for further improvement. *Journal*  
470 *of Chemical Information and Modeling* 2013; 53:201–209

471 30. Beard H, Cholleti A, Pearlman D, et al. Applying Physics-Based Scoring to Calculate  
472 Free Energies of Binding for Single Amino Acid Mutations in Protein-Protein Complexes.  
473 *PLOS ONE* 2013; 8:null

474 31. Kormos BL, Baranger AM, Beveridge DL. A study of collective atomic fluctuations and  
475 cooperativity in the U1A-RNA complex based on molecular dynamics simulations. *Journal of*  
476 *Structural Biology* 2007; 157:500–513

477 32. Chang S, Hu J, Lin P, et al. Substrate recognition and transport behavior analyses of  
478 amino acid antiporter with coarse-grained models. *Molecular bioSystems* 2010; 6:2430—  
479 2438

480 33. Fratev F. The SARS-CoV-2 S1 spike protein mutation N501Y alters the protein

481 interactions with both hACE2 and human derived antibody: A Free energy of perturbation  
482 study. bioRxiv 2020; 2020.12.23.424283

483 34. Zhang L, Jackson CB, Mou H, et al. SARS-CoV-2 spike-protein D614G mutation  
484 increases virion spike density and infectivity. Nature Communications 2020; 11:1–9

485 35. Jangra S, Ye C, Rathnasinghe R, et al. SARS-CoV-2 spike E484K mutation reduces  
486 antibody neutralisation. The Lancet Microbe 2021; 1–2

487 36. Kufareva I, Abagyan R. Methods of protein structure comparison. Methods in Molecular  
488 Biology 2012; 857:231–257

489 37. Bornot A, Etchebest C, de Brevern AG. Predicting protein flexibility through the  
490 prediction of local structures. Proteins 2011; 79:839–852

491 38. Donald JE, Kulp DW, Degrado WF. Define salt bridge. Biochemistry 2012; 79:898–915

492 39. Meuzelaar H, Vreede J, Woutersen S. Influence of Glu/Arg, Asp/Arg, and Glu/Lys Salt  
493 Bridges on  $\alpha$ -Helical Stability and Folding Kinetics. Biophysical Journal 2016; 110:2328–  
494 2341

495 40. Lee JI, Pung Pung Hwang, Wilson TH. Lysine 319 interacts with both glutamic acid 269  
496 and aspartic acid 240 in the lactose carrier of Escherichia coli. Journal of Biological  
497 Chemistry 1993; 268:20007–20015

498 41. Karshikoff A, Nilsson L, Ladenstein R. Rigidity versus flexibility: The dilemma of  
499 understanding protein thermal stability. FEBS Journal 2015; 282:3899–3917

500 42. Manuscript A, Proteins FS. Pace\_proteín stability\_2014.pdf. 2015; 588:2177–2184

501