Computational investigation reveals that the mutant strains of SARS-CoV2 are highly infectious than wildtype

Rakesh Kumar[#], Rahul Kumar[#], Harsh Goel and Pranay Tanwar*

Dr.B.R.A.-Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, INDIA-110029

*Corresponding author:

Dr. Pranay Tanwar
Dr.B.R.A.-Institute Rotary Cancer Hospital,
All India Institute of Medical Sciences, New Delhi, INDIA-110029
E-mail: pranaytanwar@gmail.com

[#]**Equal contribution**

Total main figures   : 7

Total main tables   : 2

Supplementary figures   : 11

Supplementary tables   : 7

List of abbreviations

ACE2	Angiotensin converting enzyme2
COVID-19	Coronavirus disease-2019
DSSP	Dictionary of secondary structure of protein
ED	Essential dynamics
GROMACS	Groningen machine for chemical simulation
HADDOCK	High ambiguity driven docking
MDS	Molecular dynamics simulation
MT	Mutant
PCA	Principle component analysis
PDB	Protein data bank
ProSA	Protein structure analysis
QMEAN	Qualitative model energy analysis
RBD	Receptor binding domain
Rg	Radius of gyration
RIN	Residues interaction network
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SARS-CoV2	Severe acute respiratory syndrome-coronavirus 2
SASA	Solvent accessible surface area
SAVES	Structure analysis and verification server
WT	Wildtype

65   **ABSTRACT**

66   Remarkable infectivity of severe acute respiratory syndrome-coronavirus 2 (SARS-CoV2) is

67   due to the rapid emergence of various strains, thus enable the virus to rule the world. Over the

68   course of SARS-CoV2 pandemic, the scientific communities worldwide are responding to

69   newly emerging genetic variants. However, the mechanism behind the persistent infection of

70   these variants is still not known due to the paucity of study of these variants at molecular

71   level. In this scenario, computational methods have immense utility in understanding the

72   molecular and functional properties of different variants. Therefore, in this study various

73   mutants (MTs) of SpikeS1 receptor binding domain (RBD) of highly infectious SARS-CoV2

74   strains were carried and elucidated the protein structure and dynamics using molecular

75   dynamics (MD) approach. MD simulation study showed that all MTs exhibited stable

76   structures with altered functional properties. Furthermore, the binding strength of different

77   MTs along with WT (wildtype) was revealed through protein-protein docking and observed

78   that MTs showed high binding affinities than WT. Hence, this study shed light on the

79   molecular basis of infection caused by different variants of SARS-CoV2, which might play

80   an important role in to cease the transmission and pathogenesis of virus and also implicate in

81   rational designing of a specific drug.

82

83

84

85   **KEYWORDS:**

86   SARS-CoV2, mutant strains, molecular dynamics, network analysis, protein-protein

87   interaction

88

89

90

91

92

1. INTRODUCTION

93

94   Near the end of 2019, new public health crises arise from Wuhan city of China due to the

95   outbreak of nCoV (novel coronavirus) or SARS-CoV2, which spread worldwide and named

96   COVID-19 (coronavirus disease 2019) (Wang et al., 2020). Till now (February 2021), the

97   cumulative number of total confirmed cases and deaths of COVID-19 reached up to ~115 and

98   ~2 million (https://www.worldometers.info/coronavirus/), respectively (Fig. S1A&B). During

99   the first wave of SARS-CoV2, most of the confirmed cases are observed in the USA (~2.64

100   millions) (Fig. S1C; Table S1), while during the second wave, its number reached ~20.06

101   million (Fig. S1E) (Roser et al., 2020). To date, ~28.41 million confirmed cases were

102   observed in the USA alone, which are the highest number of cases arisen in any country in

103   entire world (Roser et al., 2020). However, the cumulative confirmed COVID-19 deaths

104   during first (~575.02 per million deaths) and second waves (~1,226.54 deaths per million) of

105   infections were reported highest in Italy (Fig. S1D&F) (Bontempi, 2021). While in UK,

106   ~1,801.49 deaths per million were observed, which are the highest number of deaths per

107   million estimated by any country throughout the world (Table S2). Rapid global spread and

108   transmission of COVID-19 provide the virus with substantial opportunities for the natural

109   selection of favourable mutations (Wu et al., 2020). Mutation variants in SARS-CoV2

110   genomes have raised serious concerns about changes in infectivity (Young et al., 2020). The

111   emergence of new strains driven by the adaptive mutation indicates its instability over a

112   period of time and can contribute towards its virulence. This rapid evolution seems to be a

113   common problem that shows the gravity of coronavirus infection as a communicable disease

114   and attract the attention of the scientific community as a matter of serious concern worldwide

115   (Menachery et al., 2016). Mostly, different strains are arising due to mutations in the receptor

116   binding domain (RBD) of SpikeS1 protein (Korber et al., 2020). SpikesS1 protein of virus

117   interacts with host angiotensin converting enzyme2 (ACE2), thus enable the virus to enter

118   into the host cell (Yang et al., 2020). V367F (Valine 367 to Phenylalanine) mutations in the

119   RBD of SpikeS1 protein was reported from the strains found in Wuhan, Shenzhen, Hong

120   Kong and France (Starr et al., 2020; Tang et al., 2020). R408I (Arginine 408 to Isoleucine)

121   mutant isolates from India (Saha et al., 2020). In addition, mutations such as G476S (Glycine

122   476 to Serine) and V483A (Valine 483 to Alanine) are reported in the United States (Wang et

123   al., 2021). Recently, a new strain N501Y in which Asparagine positioned at 501 of RBD is

124   substituted by Tyrosine arises in United Kingdom and South Africa (Tegally et al., 2020;

125   Wise, 2020). These mutations (V367F, R408I, G476S, V483A and N501Y) were mainly

126   located in the SpikeS1 RBD domain. Of these five, three mutations like G476S, V483A and

127   N501Y occur at the binding interface of RBD (C-terminal side) and PD (peptidase domain)

128   of the host ACE2 receptor (Lan et al., 2020). The rest of mutations V367F and R408I, mostly

129   affected the overall topology and stability of RBD as these mutations present at the loop

130   regions of N-terminal and turn that connect β-sheet3 and 4, respectively (Yan et al., 2020).

131   These mutations (V367F, R408I, G476S, V483A and N501Y) indirectly assist in the stable

132   binding of RBD to the host ACE2 receptor (Lan et al., 2020). In addition to the above

133   mutations, there are various other mutations found in the SpikeS1 protein beyond the RBD

134   domain that also contributed to the virus infectivity (Yan et al., 2020). Since targeting the

135   mutations in RBD domain provide the first step to halt the virus infection, therefore the

136   current study is mainly focused on the RBD mutations. To unravel the stability and dynamics

137   of highly infectious strains or RBD mutants (MTs), long run molecular dynamics (MD)

138   simulations were performed followed by elucidating the key residues involved in signalling

139   through residues interaction network (RIN) approach. Furthermore, the infectiveness of

140   above MTs were assessed through protein-protein (ACE2-RBD) docking of both WT and MT

141   RBD proteins. We observed that all MTs showed similar dynamic but differential structural

142   properties. Moreover, RBD MTs showed maximum binding affinities to the host ACE2

143   receptor protein. Hence, the work in the present study will help in the creation of a more

144   specific inhibitor or drug by taking the above residues in consideration.

2. MATERIALS AND METHODS

2.1 3D structures preparation and validation

147   The tertiary structure of RBD of Spike S1 protein was taken from RCSB-PDB (Research

148   collaboration for structural bioinformatics-protein data bank), which was experimentally

149   resolved in complexed form (PDB id: 6M17) (Rose et al., 2011; Yan et al., 2020). The

150   structure was prepared in PyMOL (The PyMOL Molecular Graphics System, Version 1.3

151   Schrodinger, LLC) and all other structures except RBD were excluded. Distinct missense

152   mutations which represented different strains of SARS-CoV2 such as V367F (Valine to

153   Phenylalanine), R408I (Arginine to Isoleucine), G476S (Glycine to Serine), V483A (Valine

154   to Alanine), N501Y (Asparagine to Tyrosine) were prepared in PyMOL as previously

155   described (Ningombam et al., 2021). 3D structures of all mutants (V367F, R408I, G476S,

156   V483A and N501Y) along with wildtype were energy minimised by using GROMACS (Van

157   Der Spoel et al., 2005). Energy minimised protein models were subjected to validation by

158   different quality estimation tools. Stereochemical properties, as well as local and global

159   qualities of all models were monitored through PROCHECK module of SAVES (structure

160 analysis and verification server), ProSA (Protein structure analysis;

161 https://prosa.services.came.sbg.ac.at/prosa.php) and QMEAN (Qualitative model energy

162 analysis; https://swissmodel.expasy.org/qmean/) web servers, respectively (Benkert et al.,

163 2008; Luthy et al., 1992; Wiederstein & Sippl, 2007).

2.2 Molecular dynamics simulation

165 Molecular dynamics (MD) simulations of 150ns for each WT and MTs were conducted using

166 GROMACS in conjunction with Amber force field (Oostenbrink et al., 2004; Van Der Spoel

167 et al., 2005). Initially, GROMACS readable coordinates and topology files were generated

168 through "pdb2gmx" module and all systems (WT and MTs) were solvated using "editconf"

169 and solvate scripts. Systems solvation were accomplished by immersing all WT and MT

170 proteins into triclinic box with TIP3P water model. After that, all systems were neutralized by

171 adding sodium and chloride counterions using "genion" module followed by energy

172 minimisation through the steepest-descent method. Hence after, two equilibration steps as

173 NVT (constant number of particles, volume, and temperature) and NPT (constant number of

174 particles, pressure, and temperature) ensembles were performed for 100 and 500ps at a

175 temperature of 300K and pressure of 1bar, respectively. The temperature and pressure of all

176 the systems were carried by v-rescale (a modified Berendsen method) and Parrinello-Rahman

177 methods, respectively. Long range electrostatic interactions were preserved by using Particle

178 Mesh Ewald (PME) methods and all bonds were constraint using Linear constraint solver

179 (LINCS) algorithm. Finally, the production run of 150ns was performed in which time steps

180 of 2fs were applied. Resultant MD trajectories were analysed by using gmx energy, gmx rms,

181 gmx rmsf, gmx gyrate, gmx sasa and gmx hbond modules of GROMACS. Secondary

182 structures analysis was done through the dictionary of secondary structure of protein (DSSP)

183 approach using do_dssp script (Kabsch & Sander, 1983).

2.3 Essential dynamics

185 Essential dynamics (ED) is an important parameter, used to generate coordinates of subspace

186 in which the motion of a protein is likely to take place that determines its biological function

187 (Amadei et al., 1993). A series of eigenvectors with respective eigenvalues were generated

188 after the construction and diagonalization of the covariance matrix of backbone atoms from

189 all WT and MT systems. We restricted our interpretation with backbone atoms in order to

190 avoid statistical noise and utilised stabled trajectories of last 25ns of all WT and MT systems

191 for our ED analysis. Simultaneously, to avoid the biologically misinterpretation of PCs we

192 calculated cosine values of first three PCs that should be ≤0.2 (Hess, 2000). First three

193 principle components (PC1, PC2 and PC3) or eigenvectors were used which usually
194 accompanied with maximum motions that are biologically significant (Kumar & Saran,
195 2018). ED analysis was performed by gmx covar, gmx analyze and gmx anaeig modules of
196 GROMACS.

2.4 Residues network analysis

198 Residues interacting networks (RIN) of WT and MTs were predicted on MD optimised
199 protein structure using NAPS (Network Analysis of Protein Structure) webserver
200 (Chakrabarty, & Parekh, 2016). The server facilitated qualitative and quantitative analyses of
201 RIN and protein contact network was constructed based on physiochemical properties of
202 residues in which different centralities were measured. We restrict our analysis with Cα
203 atoms where the residue is treated as node with an unweighted edge drawn within distance of
204 7Å connecting other residues in the given network. NAPS administered different centrality
205 measures, such as degree (C_D), closeness (C_C), and betweenness (C_B). C_D is defined as the
206 number of node link with the central node as mentioned in equation (1), while C_C is the
207 average farness from the central node to other nodes within the network as mentioned in
208 equations (1) and (2). C_B is depending upon the shortest path between the nodes. It is the ratio
209 of shortest routes from one node to other passes through u and the entire shortest path within
210 the network as described in equation (3).

$$C_d(u) = \sum_{\upsilon \epsilon V} A_{u\upsilon} \qquad \dots (1)$$

$$C_{cl}(u) = (n-1) \sum_{\upsilon \epsilon V} dist(u,\upsilon) \qquad \dots (2)$$

$$C_b(u) = \sum_{s \neq u \epsilon V} \sum_{t \neq u \epsilon V} \sigma_{st}(u)/\sigma_{st} \qquad \dots (3)$$

214 Whereas, A = adjacency matrix; V= set of all nodes; N= number of nodes in the network; dist
215 (u,υ) = shortest path distance between nodes u and υ; σst = shortest path between s and t; σst
216 (u) = shortest path between s and t passing through u.

2.5 Protein-protein docking

218 Protein-protein (p-p) docking of ACE2 and all WT and MTs SpikeS1 RBD were executed
219 through HADDOCK (High Ambiguity Driven protein-protein DOCKing) webserver (van
220 Zundert et al., 2016). Active residues which were identified through available experimental
221 complex structures from PDB, used as constraint to generate AIRs (Ambiguous Interaction
222 Restraints) during docking (Yan et al., 2020). HADDOCK carried p-p docking in 3
223 successive stages as previously described (Kumar & Mukherjee, 2021). Briefly, energy

224 minimization steps were performed initially, followed by simulated annealing refinement and

225 finally MD refinement with solvent. P-p complex structures obtained from the final stage

226 were clustered and HADDOCK scores were calculated as given in equations (4-6).

227 $E = 0.01\ E_{vdw} + 0.1\ E_{elec} + 0.01\ E_{AIR} - 0.01\ BSA + 1.0 E_{desolv} + 0.1 E_{data}$ $\qquad$ … (4)

228 $E = 1.0\ E_{vdw} + 1.0\ E_{elec} + 0.1\ E_{AIR} - 0.01\ BSA + 1.0 E_{desolv} + 0.1 E_{data}$ $\qquad$ … (5)

229 $E = 1.0\ E_{vdw} + 0.2\ E_{elec} + 0.1\ E_{AIR} + 1.0 E_{desolv} + 0.1 E_{data}$ $\qquad$ … (6)

230 In the above equations, E_{vdw}, E_{elec}, E_{AIR}, E_{desolv}, E_{data}, BSA, represented van der Waals,

231 electrostatic, AIR, desolvation energies and energy of other restraint data and buried surface

232 area, respectively. Best docked complexes with the lowest HADDOCK scores from all ACE-

233 WT and -MT complexes were finally analysed. Protein-protein interaction visualisations at

234 3D and 2D levels and H-bonding were analysed in PyMOL (The PyMOL Molecular

235 Graphics System, Version 1.3 Schrodinger, LLC) and DIMPLOT module of LigPlot+ v2.2.4,

236 respectively (Laskowski & Swindells, 2011). Furthermore, the binding strength of various

237 ACE2-WT and -MTs complexes were also verified by using mCSM-PPI2 server (Rodrigues

238 et al., 2019). mCSM-PPI2 is novel computational tool that accurately predicts the impacts of

239 missense mutations in p-p binding affinity. It used graph-based structural signatures (mCSM)

240 to represent the wild-type residue environment, inter-residue non-covalent interaction

241 networks, and evolutionary information, complex network metrics, and energetic terms to

242 generate an optimized predictor.

3. RESULTS

3.1 WT and MTs demonstrated similar dynamic behaviours

245 The tertiary structure of RBD comprised about 181 amino acids long polypeptide and having

246 seven β-sheets (Fig. 1A). An extended loop was present near C-terminal that facilitated its

247 interaction with protein partner (ACE2). Various MT models were generated in PyMOL by

248 replacing Valine367 with Phenylalanine (V367F), Arginine408 with Isoleucine (R408I),

249 Glutamine476 with Serine (G476S), Valine483 with Alanine (V483A) and Asparagine501

250 with Tyrosine (N501Y) in WT 3D structure. WT and MT model structures were energy

251 minimised and found that all models were exhibited the lowest values of minimum energies

252 (data not shown). V367F and R408I MTs were present in between β-sheets 1 and 2, β-sheet 3

253 and 4, respectively (Fig. 1A). Additionally, an extended loop was present near to C-terminal

254   harboured 2 mutations such as G476S and V483A, while N501Y mutation was located near

255   to the end of C-terminal. The extended loop and the region near C-terminal of SpikeS1 RBD

256   interacts with the peptidase domain (PD) of ACE receptor[15]. The structure validations of WT

257   and newly constructed MT models were inspected by measuring the Phi/Psi angles in the

258   Ramachandran plot and found that all 3D structures exhibited 68.5%, 29.5% and 2% of

259   residues to be placed in favourable, allowed and disallowed regions, respectively (Table S3).

260   Moreover, ProSA and QMEAN servers were also executed to analyse the local and global

261   geometry of WT and MT structures and found that all WT and MT proteins exhibited ProSA

262   and QMEAN values ranging from -5.15 to -5.25 and -7.33 to -8.15, respectively (Table S3).

263   The negligible number of residues in disallowed regions and lowest values of ProSA and

264   QMEAN scores indicated that all WT and MT protein structures had better stereochemical

265   and geometrical properties.

266   Preliminary analyses of effects of different mutations on 3D structure were predicted through

267   MUpro (Cheng et al., 2006) and I-mutant (Capriotti et al., 2005) servers and found that all

268   mutant except N501Y showed decreased structure stabilities (Tables S4 & S5). Further,

269   detailed structural stability and dynamics were examined through MD simulations. To assess

270   the structural stabilities of WT and MTs (V367F, R408I, G476S, V483A and N501Y), root

271   mean square deviation (RMSD) of protein backbone with respect to equilibrium structures

272   were calculated and found that average RMSD values of ~0.37 for WT, ~0.47 for V367F,

273   ~0.48 for R408I, ~0.35 for G476S, ~0.43 for V483A and ~0.40nm for N501Y MTs were

274   existed (Fig.1B). Further, RMSDs of all WT and MTs were equilibrated after 100ns time and

275   showed similar pattern (Fig. S2A). Moreover, V367F and R408I MTs displayed higher

276   deviations, whereas G476S MT exhibited lower deviation then WT, V483A and N501Y MTs

277   (Fig.1B). RMSD of all MTs and WT showed a similar and consistent pattern. Protein

278   compactness or globularity was measured by employing the radius of gyration (Rg) and

279   found that average Rg values of ~1.76, ~1.74, ~1.72, ~1.79, ~1.77 and ~1.78nm were

280   obtained for WT, V367F, R408I, G476S, V483A and N501Y MTs, respectively (Fig.1C).

281   Rgs were slightly reduced in V367F and R408I MTs, whereas G476S, V483A and N501Y

282   MTs showed slightly higher Rg then WT. Further, the Rg patterns of all MTs and WT were

283   stabilised after ~125ns time and showed similar behaviour (Fig. S2B). Taken together, all

284   MTs and WT showed similar pattern of RMSD and Rg, indicating that both WT and MT

285   proteins exhibited similar dynamic behaviours.

286   **3.2 Root mean square fluctuations analyses of WT and MTs**

287   Flexibility and stability of both WT and MTs were further accomplished by measuring root

288   mean square fluctuation (RMSF) at residues level. RMSF is a crucial parameter use to

289   monitor the structural flexibility due to side chain of residues. All MTs and WT showed mean

290   RMSFs in the range of 0.14 to 0.2, excluding extended loop segment WT, V367F, R408I,

291   G476S, V483A and N501Y MTs, respectively (Fig.2A). The extended loop segment present

292   at the C-terminal side displayed higher fluctuation as compared to rest of the protein

293   structures. It comprised ~26 amino acid long segment (470-495) showed an average RMSFs

294   of ~0.37, ~0.70, ~0.39, ~0.38, ~0.32 and ~0.46nm for WT, V367F, R408I, G476S, V483A

295   and N501Y MTs, respectively (Fig.2A). V367F, R408I and N501Y MTs displayed slightly

296   higher RMSFs, whereas V483A MT displayed lower RMSF then WT, indicating that V367F,

297   R408I and N501Y MTs were more flexible and V483A MT was less flexible or more rigid

298   than WT. Residues level inspection showing that Gly476, Ser477, Thr478, Asn481, Gly482

299   and Val483 remained highly mobile in WT (Fig.2B) whereas Cys480, Asn481 and Gly482 in

300   V367F MT (Fig.2C), Ser477 in R408I (Fig.2D), Cys480 and Asn481 in G476S (Fig.2E),

301   Ala483 and Glu484 in V483A and Asn481, Gly482 (Fig.2F), Val483 and Glu484 in N501Y

302   MT (Fig.2G) residues remained to be highly mobile. RMSF results indicating that

303   fluctuations were more pronounced at the loop region in all MTs and WT.

3.3 WT and MTs showed differential structural properties

305   Different structural properties of both WT and MT proteins were assessed through measuring

306   the solvent accessible surface area (SASA), quantitative analysis of intra and inter hydrogen-

307   bonding (H-bond) and qualitative as well as quantitative analyses of secondary structures

308   formation during the entire simulation period. SASA was calculated through gmx sasa

309   module of GROMACS and revealed that total SASA was higher in all MTs as compared to

310   WT. An average SASA (total) values of WT, V367F, R408I, G476S, V483A and N501Y

311   MTs were 103.2, 106.8, 106, 104.6, 108.7 and 108.3nm^2, respectively (Fig.3A). Moreover,

312   SASA behaviours remained constant and consistent during an entire simulation period in

313   almost both WT and MTs (Fig. S3). During SASA analysis, hydrophobic remained slightly

314   higher than the hydrophilic SASA in all WT and MT cases. Average values of hydrophobic

315   SASA were lied in the range of 54.2nm^2 for WT (low) to 58nm^2 for V483A (high). While

316   hydrophilic SASA values were in the range of 49nm^2 for WT (low) to 51.5nm^2 for N501Y

317   (high) (Fig. 3A). This is quite possible because RBD of SpikeS1 formed a core of 7 β-Sheets

318   which is largely hydrophobic in nature.

Structure integrity and interaction capacity of protein determined by various interaction forces out of which hydrogen bond (H-bond) plays a significant role in protein stability and functionality. We calculated both intra (protein-protein) and inter (protein-solvent) H-bond through gmx hbond module of GROMACS. WT, V367F, R408I, G476S, V483A, and N501Y MTs exhibited an average of 112.3, 105.5, 106.4, 115, 107.1, and 109.4 intra H-Bonds, respectively (Fig. 3B), while an average of 378.9, 388.5, 387.8, 373.4, 391.2, and 383.6 H-bonds of WT, V367F, R408I, G476S, V483A and N501Y were formed with the solvent molecule, respectively (Fig. 3C). Moreover, both intra and inter H-bonds remained consistent during an entire simulation period in WT and MT proteins (Fig. S4). During H-bond analysis, we observed that MTs exhibited high number of intra and inter H-bonds than WT except for G476S MT, which formed a lesser number of inter H-bonds. Further, secondary structures such as β-Sheets, helices, β-bridges, bends, turns and coils play an essential role in providing distinct shape and pattern to the protein structures. Each moiety (Sheet, helix, bridge, bend, turn and coil) was further elucidated independently using a dictionary of secondary structure of protein (DSSP) approach and found that helix and β-bridge contents were increased and decreased in MT and WT, respectively. Moreover, all secondary structure moieties remain altered in both WT and MT proteins (Fig. S5).

3.4 Essential motions of WT and MTs

Essential dynamics which exploited PCA was conducted to elucidate the collective or fundamental motions of WT and MTs. Collective movements are helping to study the conformational changes that existed in the protein, thus reflecting the biological function of that protein. We used stable MD simulation trajectories (125-150ns) to measure the protein dynamics and sketched the first 20 eigenvectors with their corresponding eigenvalues. Out of which, the first 3 eigenvectors accompanied with maximum motions and their cumulative percentage were 77.04, 87.4, 65.27, 67.33, 63.26 and 68.62% for WT, V367F, R408I, G476S, V483A, and N501Y MTs, respectively (Fig. 4A & B). Moreover, we found a similar pattern of fluctuations in PC1, PC2, and PC3 as we found during RMSF analysis (Fig. S6). To ensure the biological relevant motions of PCs, we calculated cosine content. We found that the first three PCs of both WT and MT proteins exhibited a low value of cosine (≤0.2), thus indicating that motions were not due to random diffusions (Table 1). Conformational sampling of WT and MTs were enquired by plotting and comparing the projection of the first three PCs in phase space in a manner such as PC1 vs PC2, PC1 vs PC3 and PC2 vs PC3. During the projection of PC1 vs PC2 analysis, N501Y occupied a larger sub space whereas the rest of

11

352   MTs and WT engaged smaller subspaces (Fig. 4C & D). Similar patterns were also observed

353   when projecting the PC1 vs PC3 and PC2 vs PC3 in phase space (Fig. S7). The above results

354   showed that N501Y mutant undergoes large conformational changes.

355   The conformational changes at the structural level were further elucidated by sequentially

356   superimposing the 30 frames of first PC for all WT and MTs (Fig. 5). The variations in the

357   motions were mostly found in extended loop regions of MTs as well as WT. Moreover,

358   N501Y MT exhibited additional motions near N-terminal besides the motion at extended loop

359   region (Fig. 5F). Furthermore, the direction and strength of motions were also examined

360   through plotting the porcupine plot of respective PCs (Fig. 5). In porcupine structure, the

361   arrows and length of cone represented direction and magnitude of motions. Different types of

362   motions were exhibited by the extended loop of WT and MT proteins. The upward motions

363   were found in the extended loop of WT and V483A MT (Fig. 5A & E), while downward

364   motions have existed in the loop regions of V367F, R408I and G476S MTs (Fig. 5B-D).

365   Moreover, mixed motions were found in N501Y MT, as downward and upward motions

366   exhibited by the extended loop and rotational motions were found in N-terminal region (Fig.

367   5F). ED results showed that N501Y MT exhibited a large conformational change at the

368   extended loop and other random moieties of protein, implying that these regions might play a

369   crucial role in acquiring the stable conformation.

370   **3.5 WT and MTs showed almost similar residues interaction networks**

371   To explore the essential residues involved in the interaction and signalling transduction in the

372   protein, residues interaction network (RIN) analysis was performed for all MTs and WT. MD

373   optimized 3D structures were employed to calculate the centrality and to decipher its

374   connectivity. Here, we computed betweenness centrality (C_B), which is a critical centrality

375   based on the shortest route between the nodes and considered the residues with a threshold

376   C_B value ≥ 0.2 (Fig. 6). WT and other mutants such as V367F, R408I, G476S, and V483A

377   showed only one residue (Tyrosine453) as par with threshold C_B value which is located in the

378   β-sheet region of the protein (Fig. 6A-E). In case of N501Y, four residues, namely Valine

379   350, Tyrosine451, Tyrosine453 and Tyrosine495 (V350, Y451, Y453 and Y495) were had

380   higher C_B values (Fig. 6F). Interestingly, all residues except Y453 lies at the loop region of

381   protein which indeed indicating an important region in the residual signalling of the protein.

382   **3.6 MTs had high binding affinities than WT**

383 Protein–protein (p-p) interaction is a physical association of two or more proteins due to

384 biochemical events driven by many biological forces like Van der Waals, electrostatics and

385 polar interactions. It plays a decisive role in the overall physiology of the cell and its

386 biological process, including signal transduction. Structure based molecular docking

387 approach was widely used to predict the orientation of one protein with respect to other and

388 the key residues involved in the interaction (Maurya et al., 2020). Here, we employed the

389 HADDOCK server to study the p-p docking of ACE2 (angiotensin-converting enzyme 2)

390 receptor and SpikeS1 RBD of WT and MTs. HADDOCK predicted different clusters for each

391 WT- and MT-protein complexes out of which cluster accompanied larger structures had the

392 lowest HADDOCK score, which indicated the p-p complex of high binding strength.

393 Topmost clusters size of ACE2-WT, -V367F, -R408I, -G476S, -V483A and -N501Y MT

394 complexes were had 168, 169, 127, 175, 176 and 62 structures, respectively (Table 2).

395 HADDOCK score of topmost clusters for ACE2-WT, -V367F, -R408I, -G476S, -V483A and

396 -N501Y MT complexes were -122.9 (±3.3), -122 (±2.2), -114.2 (±2), -121.4 (±2.6), -122.1

397 (±2.3) and -124.4 (±3.1), respectively (Fig. 7; Table 2). Various energies such as van der

398 Waals, electrostatic, desolvation and restraints violation energies significantly contributed to

399 determining the binding affinities of all WT and MTs p-p interactions (Figs. S8&S9). During

400 binding energy analysis, electrostatic energy had a major contribution in total binding energy

401 which were -212.3 (±27.8), -185.3 (±29.3), -213.5 (±14.2), -244.6 (±9.7), -214.6 (±18.2) and

402 -210.4 (±20.3) for ACE2-WT, -V367F, -R408I, -G476S, -V483A and -N501Y MT

403 complexes, respectively (Table 2). Additionally, all p-p complexes were stabilised by the

404 ample amount of hydrophilic as well as hydrophobic interactions (Figs. S10&S11; Table S6).

405 Furthermore, p-p binding free energies were also verified through the change in binding

406 affinities of ACE2-WT and -MT complexes by using mCSM-PPI server and we found that

407 N501Y MT exhibited a higher value of affinity changes ($\Delta\Delta G$= -2.147 Kcal/mol) as

408 compared to WT and other MTs (Table S7). The above results indicated that N501Y MT had

409 a high binding affinity to the host ACE2 protein, V367F and V483A have similar binding

410 affinities to that of WT, while R408I and G476S have lower binding affinity to an ACE1

411 receptor protein.

4. DISCUSSION

413 Viruses, particularly RNA viruses have a marvellous tendency to mutate themselves as

414 compare with DNA viruses, resulting in the rapid emergence of new strains (Elena &

415 Sanjuan, 2005). There are currently thousands of SARS-CoV2 variants reported in which the

13

416   speed and intensity of spreading some of them (variants) are explained, more than by their
417   genetic differences, by the habits of population, and the effectiveness of epidemiological
418   surveillance policies (Hodcroft et al., 2020). SARS-CoV2 is a novel epidemic strain of
419   *Betacoronavirus* responsible for the widespread viral pandemic COVID-19 (Wong & Saier,
420   2021). The virus gets into the cell with the help of spike glycoprotein (S1) present on the
421   surface of the virus envelope which interacts with the ACE2 receptor of host cell (Korber et
422   al., 2020). Receptor binding domain (RBD) of SpikeS1 protein interacts with ACE2 receptors
423   and are more prone to mutations resulting in the formation of new strains and also help the
424   viruses to escape from different therapies (Lan et al., 2020). The RBD consists of a core
425   region of β-sheets and an extended loop region known as receptor binding motifs that provide
426   a binding surface to the ACE2 protein (Yan et al., 2020). Before developing a specific and
427   effective drug, one must understand the molecular interaction of ACE2 and SpikeS1 protein
428   of different strains. Therefore, the present study has critically analysed the effect of various
429   mutations on the structural stability and dynamics of RBD of SpikeS1 protein by manifesting
430   different MTs or strains of SARS-CoV2 and assessed the binding affinities with host (human)
431   ACE2 receptor.

432   The tertiary structure of ACE2-SpikeS1 RBD complex was taken from PDB and used as a
433   template for creating various SARS-CoV2 mutants. Qualities of generated mutant models
434   were verified by inspecting the stereochemical and geometrical properties through different
435   model validation tools and found that all MTs (V367F, R408I, G476S, V483A and N501Y)
436   had good profiles of Ramachandran plots, acceptable scores of ProSA and QMEAN, similar
437   to WT protein, implying that all MTs along with WT had better stereochemical and
438   geometrical properties. To explore the structural and functional properties, the optimised
439   RBD MTs were subjected to MD simulations of 150ns simulation time. MD simulation is a
440   promising technique to study the stability and dynamics of macromolecules at an atomic level
441   in a realistic manner, which is hard to achieve through experimental methods (Kumar et al.,
442   2017; Kumar et al., 2020). Different parameters such as RMSD, RMSF and Rg were
443   measured to monitor the stability and dynamics of WT and MT proteins. RMSD results
444   suggested that WT as well as MT proteins displayed consistent behaviour with higher
445   RMSDs were found in V367F, R408I and V483A MTs as compared to WT, G476S and
446   N501Y MTs. RMSF analysis was performed to study the residue and structural level
447   alterations in WT and MTs and found residues like Gly476, Ser477, Thr478, Asn481, Gly482
448   and Val483 were highly fluctuated and located at loop regions of almost all MTs and WT.

14

449 Loops are essential moieties of protein that are highly mobile and act as a crucial site for

450 ligand interactions and help in acquiring the stable conformation of proteins (Baiesi et al.,

451 2019). Hence, the mobility of residues in the loop regions in WT and MTs help in the

452 interactions of RBD to partner protein such as ACE2.

453 Next, we have analysed the structural properties of all MTs along with WT by measuring

454 SASA, H-bonds and secondary structure units. The core region of protein mainly consists of

455 hydrophobic residues that act as a driving force during protein folding. The solvent

456 interaction intensity of amino acid and the core of protein is proportional to the total surface

457 area exposed to these environments (Kumar & Saran, 2021). Therefore, SASA of WT and

458 MTs were measured to monitor the surface area available for the surrounding environment

459 and found that all MTs displayed higher SASA then WT and hydrophobic SASA contributed

460 slightly more in comparison to hydrophilic SASA as RBD of protein mostly consist of core

461 followed by loop regions. Hydrogen bonding (H-bond) is an important parameter for

462 providing stability and strength of protein-ligand interaction properties of a particular protein.

463 The stability and protein-ligand interaction capacity were measured by calculating the intra

464 (protein-protein) and inter (protein-solvent) H-bonds (Pace et al., 2014). Our results

465 suggested that almost all MTs formed a high number of intra and inter H-bonds compared to

466 WT, demonstrating that MTs formed highly stable protein structures and have a higher

467 tendency of ligand interactions. Furthermore, secondary structure analyses of all MTs and

468 WT indicated that alterations in the random moieties of protein compromised with changing

469 in α-helix and β-sheet contents.

470 Essential dynamics (ED), a typical application of PCA, which is used to extract the biological

471 relevant motions by utilising the MD simulation trajectories (Berendsen & Hayward, 2000).

472 Equilibrated MD trajectories were harnessed for constructing the covariance matrix and

473 series of eigenvectors with associated eigenvalues were plotted for all MTs and WT. The first

474 three eigenvectors were used to study the conformational subspace of MTs and WT protein as

475 they are mostly encompassing the major motions. ED results suggested that, except for

476 N501Y MT, all other MTs were had similar conformational subspace as with WT, while

477 N501Y MT covered larger conformational space. Moreover, the structure level inspection

478 from the first eigenvector (PC1) revealed that the higher motions were accompanied by

479 extended loop region of the protein. The biological function of a protein is characterizing by

480 constituent residues and their chemical properties. These chemical versatilities of amino acids

481 determine the overall topology of protein and its behaviour in the surrounding environment

15

482 (Przulj et al., 2004). Different residues in the protein interact and help in the flow of signal,
483 which is better elucidated by the residues interaction network (RIN). In this analysis, an
484 entire protein considered as a network in which residues and interactions are plotted in the
485 form of nodes and edges, respectively (Davis et al., 2015). Three centralities were computed
486 out of which betweenness centrality (C_B) is most important concerning to protein function
487 (Kumar et al., 2020). RIN results suggested that Tyrosine453 was an important residue in WT
488 and V367F, R408I, G476S, and V483A MTs, while more than one amino acid had role in
489 signalling in N501Y MT. The above residues are mainly located on or near the 5th and 6th β-
490 sheets of RBD, demonstrating that this region might be crucial for the functioning of WT and
491 MT proteins.

492 SARS-CoV2 invades host cells through the interaction of RBD of SpikeS1 unit to the host
493 ACE2 receptor. Qualitatively, the infectivity of diverse SARS-CoV2 strains in the host is
494 proportional to the binding affinity of SARS-CoV2 SpikeS1 RBD of each strain to the ACE2
495 receptor of host cells (Mittal et al., 2020). Therefore, the assessment of binding affinity of
496 different MTs to host cells is vital for understanding the infectiousness of SARS-CoV2. The
497 binding affinity of WT and MTs RBD to ACE2 was elucidated by the HADDOCK protein-
498 protein (p-p) docking tool. P-p docking results suggested that V367F and V483A MTs have
499 similar binding affinities to WT, while R408I and G476S had fewer binding affinities.
500 Moreover, N501Y MT had a higher binding affinity than the rest of MT and WT proteins.
501 Binding affinity analysis indicated that N501Y MT binds to host ACE2 protein with
502 maximum binding strength, thus contributing highly infectious activity, as also reported
503 previously (Luan et al., 2021).

504 Proteins are having a dynamic personality and are central to the molecular function of a cell.
505 Any change in amino acid residue led to the changing in overall topology and hence affected
506 their functions. The molecular and functional consequences of change in amino acids are
507 better observed in the current pandemic (COVID-19) as the point mutations in the SpikeS1
508 RBD of SARS-CoV2 results in the generation of thousands of variants or strains (Chen et al.,
509 2020). All countries of the world are witnessed the infectiousness of these strains. The
510 alteration in amino acids showed both local and the global impact on the topology of protein
511 and in turn, altered the interacting residue, which ultimately affected the binding affinity of
512 two partner proteins (ACE2-SpikeS1 RBD). In V367F MT, a nonpolar hydrophobic amino
513 acid has aliphatic side chain (Valine) is mutated to an aromatic hydrophobic residue
514 (Phenylalanine) that enhanced the structural stability of the spike protein, thus facilitating

16

515 more efficient binding to the host ACE2 receptor. In case of R408I MT, there is change in
516 overall charge of protein residue (Arginine), which is mutated to the residues having a neutral
517 charge (Isoleucine). This also affects overall topology of protein as mutated residue have a
518 smaller size than WT residue, resulting in the decreasing of overall binding affinity as
519 reported during p-p interaction. In G476S MT, Glycine of WT residues is mutated to Serine
520 which affects the overall flexibility of protein as Glycine is more flexible than Serine. This
521 led to a decrease in the stability and binding affinity of G476S MT, as noticed during MD
522 simulation and p-p docking. In V483A MT Valine, a large hydrophobic protein residue is
523 mutated to Alanine, a small size hydrophobic residue that alters the overall dimension of the
524 protein, thus help in increasing the binding affinity. And, in case of N501Y MT, a
525 substitution of Asparagine to Tyrosine was found in the binding motif of SpikeS1 RBD,
526 which enhances the binding affinity to its protein partner and ultimately leads to a stable
527 virus-host relationship. Due to strong binding, this mutation also assists the virus in evading
528 antibody neutralization.

529 The present scenario of global pandemic, where the world keeps an eye towards the vaccine
530 at utmost level. In this regard, the computational approach has the potential to unravel the
531 mechanistic behaviour of RBD MTs and its interaction with ACE receptor at an atomic level.
532 However, certain limitations are associated with the current computational study, which may
533 be overcome by experimental methods. Therefore, experimental methods are highly urged to
534 support or validate the findings of the current research. Nevertheless, our study provides
535 computational evidence of the strength of highly infectious strains, which would help in
536 design a more specific drug to combat the current pandemic.

537 **CONCLUSION**
538 Noteworthy infectivity of severe acute respiratory syndrome-coronavirus 2 (SARS-CoV2) is
539 due to its fast-mutating ability thus led to extended infection. In this investigation, the
540 computational methods such as MDS and p-p docking have been applied and revealed that
541 mutant strains of SARS-CoV2 formed stabled protein structures and had higher binding
542 affinities to host protein. Hence, the present study provides the molecular evidence of this
543 endless pandemic which would help in designing a more specific inhibitor.

544 **CONFLICT OF INTEREST**
545 We have no conflict of interest

546 **ACKNOWLEDGEMENTS**

17

550 **AUTHOR CONTRIBUTION**

551 Rakesh K. and P.T. conceptualised and designed the study. Rakesh K. and Rahul K.
552 performed experiments and analysed the data. H.G. help in literature study. Rakesh K., Rahul
553 K. and H.G. wrote the manuscript. P.T. provided laboratory infrastructure. All authors read
554 and approved the final version of manuscript.

555 **DATA AVAILABILITY STATEMENT**

556 All datasets generated during this study are included in the article/supplementary material.

557 **REFERENCES**

558 Amadei, A., Linssen, A. B., & Berendsen, H. J. (1993). Essential dynamics of
559   proteins. *Proteins: Structure, Function, and Bioinformatics*, *17*(4), 412-425.

560 Baiesi, M., Orlandini, E., Seno, F., & Trovato, A. (2019). Sequence and structural patterns
561   detected in entangled proteins reveal the importance of co-translational folding. *Scientific*
562   *Reports*, *9*(1), 1-12.

563 Benkert, P., Tosatto, S. C., & Schomburg, D. (2008). QMEAN: A comprehensive scoring
564   function for model quality assessment. *Proteins: Structure, Function, and*
565   *Bioinformatics*, *71*(1), 261-277.

566 Berendsen, H. J., & Hayward, S. (2000). Collective protein dynamics in relation to
567   function. *Current Opinion in StructuralBbiology*, *10*(2), 165-169.

568 Bontempi, E. (2021). The europe second wave of COVID-19 infection and the Italy
569   "strange" situation. *Environmental Research*, *193*, 110476.

570 Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2. 0: predicting stability changes
571   upon mutation from the protein sequence or structure. *Nucleic Acids*
572   *Research*, *33*(suppl_2), W306-W310.

573 Chakrabarty, B., & Parekh, N. (2016). NAPS: network analysis of protein
574   structures. *Nucleic Acids Research*, *44*(W1), W375-W382.

575 Chen, J., Wang, R., Wang, M., & Wei, G. W. (2020). Mutations strengthened SARS-CoV-
576   2 infectivity. *Journal of Molecular Biology*, *432*(19), 5212-5226.

577   Cheng, J., Randall, A., & Baldi, P. (2006). Prediction of protein stability changes for
578   single-site mutations using support vector machines. *Proteins: Structure, Function, and*
579   *Bioinformatics*, *62*(4), 1125-1132.

580   Davis, D., Yaveroğlu, Ö. N., Malod-Dognin, N., Stojmirovic, A., & Pržulj, N. (2015).
581   Topology-function conservation in protein–protein interaction
582   networks. *Bioinformatics*, *31*(10), 1632-1639.

583   Elena, S. F., & Sanjuán, R. (2005). Adaptive value of high mutation rates of RNA viruses:
584   separating causes from consequences. *Journal of Virology*, *79*(18), 11555-11558.

585   Hess, B. (2000). Similarities between principal components of protein dynamics and
586   random diffusion. *Physical Review E*, *62*(6), 8438-8448.

587   Hodcroft, E. B., Zuber, M., Nadeau, S., Crawford, K., Bloom, J. D., Veesler, D., Vaughan,
588   T. G., Comas, I., Candelas, F. G., Stadler, T., & Neher, R. A. (2020). Emergence and
589   spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv : the*
590   *preprint server for health sciences*, 2020.10.25.20219063.

591   Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: pattern
592   recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original*
593   *Research on Biomolecules*, *22*(12), 2577-2637.

594   Korber, B., Fischer, W., Gnanakaran, S. G., Yoon, H., Theiler, J., Abfalterer, W., ... &
595   COVID, S. (2020). Spike mutation pipeline reveals the emergence of a more transmissible
596   form of SARS-CoV-2. *BioRxiv*.

597   Kumar, R., & Mukherjee, P. K. (2021). Trichoderma virens Bys1 may competitively
598   inhibit its own effector protein Alt a 1 to stabilize the symbiotic relationship with plant-
599   evidence from docking and simulation studies. *3 Biotech*, *11*(3), 1-14.

600   Kumar, R., & Saran, S. (2018). Structure, molecular dynamics simulation, and docking
601   studies of Dictyostelium discoideum and human STRAPs. *Journal of Cellular*
602   *Biochemistry*, *119*(9), 7177-7191.

603   Kumar, R., & Saran, S. (2021). Comparative modelling unravels the structural features of
604   eukaryotic TCTP implicated in its multifunctional properties: an in silico
605   approach. *Journal of Molecular Modeling*, *27*(2), 1-15.

606   Kumar, R., Kumar, R., Tanwar, P., Deo, S. V. S., Mathur, S., Agarwal, U., & Hussain, S.
607   (2020). Structural and conformational changes induced by missense variants in the zinc
608   finger domains of GATA3 involved in breast cancer. *RSC Advances*, *10*(65), 39640-
609   39653.

610   Kumar, R., Maurya, R., & Saran, S. (2017). Identification of novel inhibitors of the
611   translationally controlled tumor protein (TCTP): insights from molecular
612   dynamics. *Molecular BioSystems*, *13*(3), 510-524.

613   Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L.,
614   & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound
615   to the ACE2 receptor. *Nature*, *581*(7807), 215–220.

616   Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: multiple ligand-protein
617   interaction diagrams for drug discovery. *Journal of Chemical Information and*
618   *Modeling*, *51*(10), 2778–2786.

619   Luan, B., Wang, H., & Huynh, T. (2021). Molecular Mechanism of the N501Y Mutation
620   for Enhanced Binding between SARS-CoV-2's Spike Protein and Human ACE2
621   Receptor. *bioRxiv*, 2021-01.

622   Luthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-
623   dimensional profiles. *Nature*, *356*(6364), 83-85.

624   Maurya, R., Kumar, R., & Saran, S. (2020). AMPKα promotes basal autophagy induction
625   in *Dictyostelium discoideum*. *Journal of Cellular Physiology*, *235*(5), 4941-4953.

626   Menachery, V. D., Yount, B. L., Jr, Sims, A. C., Debbink, K., Agnihothram, S. S.,
627   Gralinski, L. E., Graham, R. L., Scobey, T., Plante, J. A., Royal, S. R., Swanstrom, J.,
628   Sheahan, T. P., Pickles, R. J., Corti, D., Randell, S. H., Lanzavecchia, A., Marasco, W. A.,
629   & Baric, R. S. (2016). SARS-like WIV1-CoV poised for human emergence. *Proceedings*
630   *of the National Academy of Sciences of the United States of America*, *113*(11), 3048–3053.

631   Mittal, A., Manjunath, K., Ranjan, R. K., Kaushik, S., Kumar, S., & Verma, V. (2020).
632   COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition
633   by SARS-CoV-2. *PLoS Pathogens*, *16*(8), e1008762.

634   Ningombam, S. S., Kumar, R., & Tanwar, P. (2021). Mutant strains of SARS-CoV-2 are
635   more prone to infect obese patient: a review. *Wiener Klinische Wochenschrift*, 1-10.

636   Oostenbrink, C., Villa, A., Mark, A. E., & Van Gunsteren, W. F. (2004). A biomolecular
637   force field based on the free enthalpy of hydration and solvation: the GROMOS force-field
638   parameter sets 53A5 and 53A6. *Journal of Computational Chemistry*, *25*(13), 1656-1676.

639   Pace, C. N., Fu, H., Lee Fryar, K., Landua, J., Trevino, S. R., Schell, D., ... & Grimsley, G.
640   R. (2014). Contribution of hydrogen bonds to protein stability. *Protein Science*, *23*(5),
641   652-661.

642   Przulj, N., Wigle, D. A., & Jurisica, I. (2004). Functional topology in a network of protein
643   interactions. *Bioinformatics*, *20*(3), 340-348.

20

644 Rodrigues, C. H., Myung, Y., Pires, D. E., & Ascher, D. B. (2019). mCSM-PPI2:
645 predicting the effects of mutations on protein–protein interactions. *Nucleic Acids*
646 *Research*, *47*(W1), W338-W344.

647 Rose, P. W., Beran, B., Bi, C., Bluhm, W. F., Dimitropoulos, D., Goodsell, D. S., Prlic,
648 A., Quesada, M., Quinn, G. B., Westbrook, J. D., Young, J., Yukich, B., Zardecki, C.,
649 Berman, H. M., & Bourne, P. E. (2011). The RCSB Protein Data Bank: redesigned web
650 site and web services. *Nucleic Acids Research*, *39*(Database issue), D392–D401.

651 Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2020). Coronavirus pandemic
652 (COVID-19). *Our world in data*.

653 Saha, P., Banerjee, A. K., Tripathi, P. P., Srivastava, A. K., & Ray, U. (2020). A virus that
654 has gone viral: amino acid mutation in S protein of Indian isolate of Coronavirus COVID-
655 19 might impact receptor binding, and thus, infectivity. *Bioscience Reports*, *40*(5),

656 Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K., Dingens, A. S., Navarro,
657 M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., King, N. P., Veesler, D., & Bloom, J.
658 D. (2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals
659 Constraints on Folding and ACE2 Binding. *Cell*, *182*(5), 1295–1310.e20.

660 Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., ... & Lu, J. (2020). On the origin and
661 continuing evolution of SARS-CoV-2. *National Science Review*, *7*(6), 1012-1023.

662 Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., ... &
663 de Oliveira, T. (2020). Emergence and rapid spread of a new severe acute respiratory
664 syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in
665 South Africa. *medRxiv*.

666 Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J.
667 (2005). GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, *26*(16),
668 1701-1718.

669 van Zundert, G., Rodrigues, J., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E.,
670 Melquiond, A., van Dijk, M., de Vries, S. J., & Bonvin, A. (2016). The HADDOCK2.2
671 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of*
672 *Molecular Biology*, *428*(4), 720–725.

673 Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus
674 outbreak of global health concern. *Lancet (London, England)*, *395*(10223), 470–473.

675 Wang, R., Chen, J., Gao, K., Hozumi, Y., Yin, C., & Wei, G. W. (2021). Analysis of
676 SARS-CoV-2 mutations in the United States suggests presence of four substrains and
677 novelvariants. *Communications Biology*, *4*(1), 1-14.

678   Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the
679   recognition of errors in three-dimensional structures of proteins. *Nucleic Acids*
680   *Research*, *35*(suppl_2), W407-W410.

681   Wise J. (2020). Covid-19: New coronavirus variant is identified in UK. *BMJ (Clinical*
682   *research ed.)*, *371*, m4857.

683   Wong, N. A., & Saier, M. H. (2021). The SARS-Coronavirus Infection Cycle: A Survey
684   of Viral Membrane Proteins, Their Functional Interactions and Pathogenesis. *International*
685   *Journal of Molecular Sciences*, *22*(3), 1308.

686   Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z.,
687   Wang, J., Sheng, J., Quan, L., Xia, Z., Tan, W., Cheng, G., & Jiang, T. (2020). Genome
688   Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in
689   China. *Cell Host & Microbe*, *27*(3), 325–328.

690   Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the
691   recognition of SARS-CoV-2 by full-length human ACE2. *Science*, *367*(6485), 1444-1448.

692   Yang, J., Petitjean, S., Koehler, M., Zhang, Q., Dumitru, A. C., Chen, W., Derclaye, S.,
693   Vincent, S. P., Soumillion, P., & Alsteens, D. (2020). Molecular interaction and inhibition
694   of SARS-CoV-2 binding to the ACE2 receptor. *Nature Communications*, *11*(1), 4541.

695   Young, B. E., Fong, S. W., Chan, Y. H., Mak, T. M., Ang, L. W., Anderson, D. E., Lee, C.
696   Y., Amrun, S. N., Lee, B., Goh, Y. S., Su, Y., Wei, W. E., Kalimuddin, S., Chai, L., Pada,
697   S., Tan, S. Y., Sun, L., Parthasarathy, P., Chen, Y., Barkham, T., … & Ng, L. F. (2020).
698   Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and
699   the inflammatory response: an observational cohort study. *Lancet (London,*
700   *England)*, *396*(10251), 603–611.

701

702   **LEGEND TO FIGURES**

703 **Fig. 1 Tertiary structure, RMSD and Rg analyses.** (A) Tertiary structure of SpikeS1 RBD of
704   SARS-CoV2, (B) Root mean square deviation at function of time and (C) Radius of
705   gyration at function of time in nanoseconds. 3D structure of RBD was shown in cartoon
706   mode with cyan colour and different mutations V367F, R408I, G476S, V483A and N501Y
707   were depicted in stick mode. Region interacting with host ACE2 of SpikeS1 RBD and
708   extended loop were highlighted in back and blue dotted circles, respectively. Different
709   sheets were labelled as β_1-β_7. WT, V367F, R408I, G476S, V483A and N501Y MTs were
710   labelled in black, red, green, blue, magenta and yellow colour, respectively.

711 **Fig. 2 Root mean square fluctuation (RMSF) analysis.** (A) RMSF plot at function of amino
712   acid residues, (B) Structure fluctuations of WT, (C) Structure fluctuations of V367F MT,

22

713   (D) Structure fluctuations of R408I MT, (E) Structure fluctuations of G476S MT, (F)
714   Structure fluctuations of V483A MT and (G) Structure fluctuation of N501Y MT. In 2D
715   RMSF plot, highly fluctuated regions were highlighted and static or mobile regions were
716   shown as blue and red colour, respectively. Amino acid residues were depicted in stick
717   mode with 3-letter code. WT, V367F, R408I, G476S, V483A and N501Y MTs were
718   labelled in black, red, green, blue, magenta and yellow colour, respectively.

719 **Fig. 3**   **Structural properties analyses of WT and MT proteins.** (A) Solvent accessible surface
720   area (SASA) plot, (B) Protein-protein hydrogen bonds (H-bond) and (C) Protein-water
721   hydrogen bonds (H-bond). WT, V367F, R408I, G476S, V483A and N501Y MTs were
722   labelled in black, red, green, blue, magenta and yellow colour, respectively.

723 **Fig. 4**   **Essential dynamics analysis.** (A) Plot of first 20 eigenvectors with associated eigenvalues
724   (solid lines) and their cumulative percentage (dotted lines) of WT, V367F, R408I and
725   G476S MTs, (B) Plot of first 20 eigenvectors with associated eigenvalues (solid lines) and
726   their cumulative percentage (dotted lines) of WT, V483A and N501Y MTs, (C) Projection
727   of PC1 vs PC2 of WT, V367F, R408I and G476S MTs and (D) Projection of PC1 vs PC2
728   of WT, V483A and N501Y MTs. WT, V367F, R408I, G476S, V483A and N501Y MTs
729   were labelled in black, red, green, blue, magenta and yellow colour, respectively.

730 **Fig. 5**   **Major motions analyses of first PC.** 30 frames of PC1 were sequentially superimposed
731   and associated motions were depicted in porcupine structures of (A) WT, (B) V367F, (C)
732   R408I, (D) G476S, (E) V483A and (F) N501Y. Intensities and nature of motions were
733   highlighted in dotted circles and arrows in the protein 3D and porcupine structures,
734   respectively.

735 **Fig. 6**   **Network centrality analysis.** Betweenness centrality (C_B) values of (A) WT, (B) V367F,
736   (C) R408I, (D) G476S, (E) V483A and (F) N501Y. Black dotted line represented the cut
737   off value used to filter the important residues. Right panel of each graph denoted 3D
738   structure displayed in cartoon modes in which important residues shown in sphere.

739 **Fig. 7**   **Protein-protein (ACE2-SpikeS1 RBD) interaction study.** (A) ACE2-WT complex, (B)
740   ACE2-V367F complex, (C) ACE2-R408I complex, (D) ACE2-G476S complex, (E)
741   ACE2-V483A complex and (F) ACE2-N501Y complex. ACE2 and SpikeS1 RBD were
742   shown in blue and green surface cartoon modes, respectively. Interacting residues of
743   ACE2 (Black) and SpikeS1 (Red) RBD were represented as stick modes with 3 letter
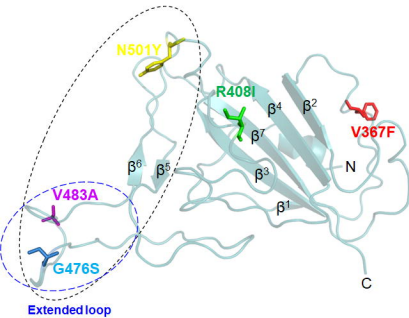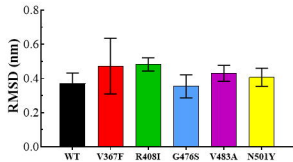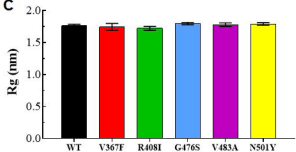744   amino acid code.

745   **Table 1. Cosine content analysis of first two principle components (PC1 and PC2)**

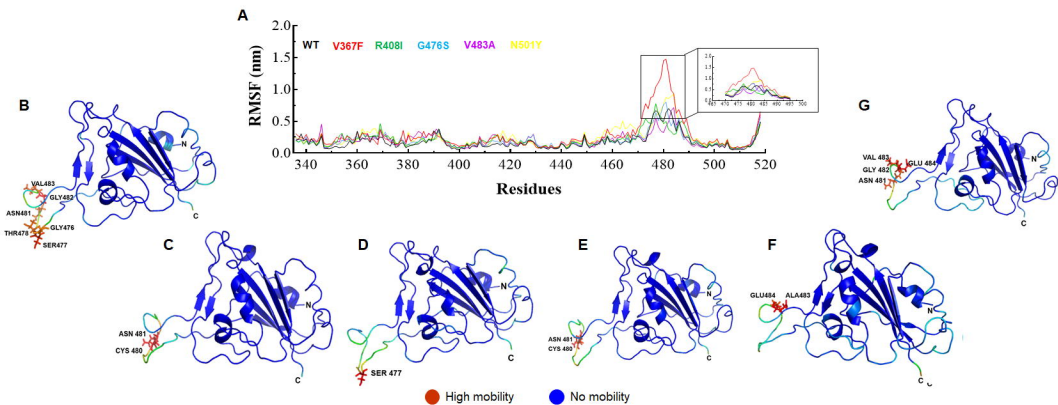|	50		100		125		150	
	PC1	**PC2**	**PC1**	**PC2**	**PC1**	**PC2**	**PC1**	**PC2**
WT	0.21	0.6	0.74	0.00014	0.54	0.64	0.09	0.002
V367F	0.0063	0.02	0.84	0.0034	0.007	0.596	4.7E-08	0.083
R408I	0.074	0.57	0.134	0.27	0.0001	0.441	0.0053	0.075
G476S	0.144	0.859	0.132	0.66	0.0041	0.007	0.09	0.12
V483A	0.427	0.062	0.0038	0.106	0.493	0.223	0.018	0.014
N501Y	0.021	0.514	0.682	0.054	0.00068	0.132	0.0272	0.0128

746

23

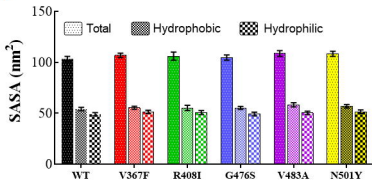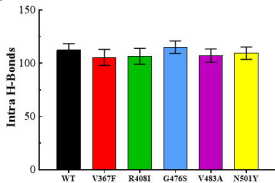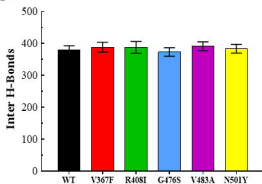747 **Table 2. Protein-protein (ACE2-RBD) interaction energies analyses**

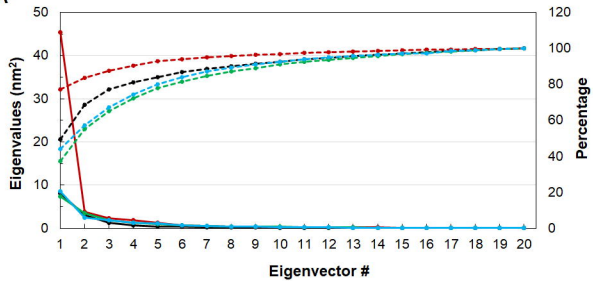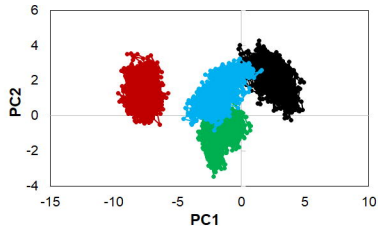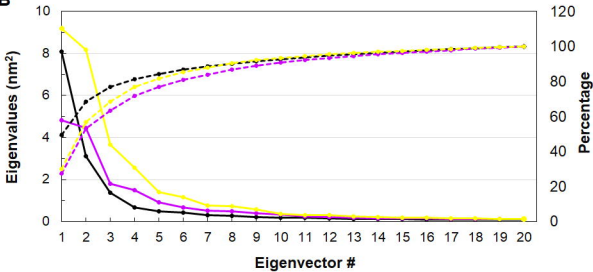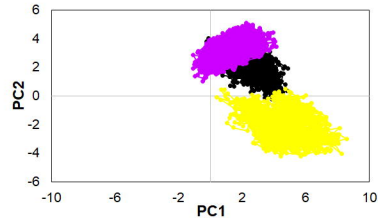|	Van der Waals energy (stdev.)	Electrostatic energy (stdev.)	Desolvation energy (stdev.)	Restraints violation energy (stdev.)	HADDOCK score (stdev.)
WT	-62.9 (±4.8)	-212.3 (±27.8)	-18.4 (±3.2)	8.3 (±11.8)	-122.9 (±3.3)
V367F	-62.8 (±4.4)	-185.3 (±29.3)	-23.8 (±4.2)	15.7 (±15.1)	-122 (±2.2)
G476S	-56.7 (±1.8)	-213.5 (±14.2)	-16.5 (±2.4)	17.1 (±15.5)	-114.2 (±2.0)
R408I	-59.5 (±2.3)	-244.6 (±9.7)	-15.4 (±3.3)	24.2 (±13.7)	-121.4 (±2.6)
V483A	-61.4 (±2.4)	-214.6 (±18.2)	-19.5 (±4.3)	16.9 (±14.8)	-122.1 (±2.3)
N501Y	-59.6 (±2.5)	-210.4 (±20.3)	-24.2 (±3.2)	14.1 (±13.0)	-124.4 (±3.1)

748

A

N501Y

R408I

β4

β2

V367F

N

β7

β5

β6

β3

β1

V483A

G476S

Extended loop

B

RMSD (nm)

WT V367F R408I G476S V483A N501Y

C

Rg (nm)

WT V367F R408I G476S V483A N501Y

WT **V367F** **R408I** **G476S** **V483A** **N501Y**

A

B

C

D

E

F

G

High mobility   No mobility