

Preliminary Structural Data Revealed That the SARS-CoV-2 B.1.617 Variant's RBD Binds to ACE2 Receptor Stronger Than the Wild Type to Enhance the Infectivity

Abbas Khan,^{*,[a, b]} Dong-Qing Wei,^{*,[a, b, c]} Kafila Kousar,^[d] Jehad Abubaker,^[g] Sajjad Ahmad,^[e] Javaid Ali,^[f] Fahd Al-Mulla,^[h] Syed Shujait Ali,^[i] N. Nizam-Uddin,^[j] Abrar Mohammad Sayaf,^[k] and Anwar Mohammad^[g]

The evolution of new SARS-CoV-2 variants around the globe has made the COVID-19 pandemic more worrisome, further pressuring the health care system and immunity. Novel variations that are unique to the receptor-binding motif (RBM) of the receptor-binding domain (RBD) spike glycoprotein, i.e. L452R-E484Q, may play a different role in the B.1.617 (also known as G/452R.V3) variant's pathogenicity and better survival compared to the wild type. Therefore, a thorough analysis is needed to understand the impact of these mutations on binding with host receptor (RBD) and to guide new therapeutics development. In this study, we used structural and biomolecular

simulation techniques to explore the impact of specific mutations (L452R-E484Q) in the B.1.617 variant on the binding of RBD to the host receptor ACE2. Our analysis revealed that the B.1.617 variant possesses different dynamic behaviours by altering dynamic-stability, residual flexibility and structural compactness. Moreover, the new variant had altered the bonding network and structural-dynamics properties significantly. MM/GBSA technique was used, which further established the binding differences between the wild type and B.1.617 variant. In conclusion, this study provides a strong impetus to develop novel drugs against the new SARS-CoV-2 variants.

Introduction


Over the last two decades, Asia has been the epicentre of three highly pathogenic human coronavirus epidemics caused by severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome (MERS-CoV), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), respectively.^[1] SARS-CoV-2 was first identified in December 2019, in Wuhan, China, and acknowledged as a previously unknown β -coronavirus.^[2] SARS-CoV-2 has been demonstrated to transmit effectively among human populations because of its striking

potential for rapid transmission through aerosol and fomites, making the coronavirus disease-19 (COVID-19) a pandemic.^[3] SARS-CoV-2 is a single-stranded and positive-sense RNA genome whose size is approximately equal to ~ 30 kb.^[4] The genome codes for accessory, non-structural, and structural proteins (spike [S], envelope [E], membrane [M], and nucleocapsid [N]).^[3a,5]

SARS-CoV-2 infection begins with the virus binding to a host's cellular angiotensin-converting enzyme (ACE2) receptors.^[6] This binding event involves virus surface S protein. The protein comprises S1 and S2 subunits; the S1 subunits

- [a] A. Khan., D.-Q. Wei
 State Key Laboratory of Microbial Metabolism
 Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances
 Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences
 Ministry of Education and School of Life Sciences and Biotechnology
 Shanghai Jiao Tong University, Shanghai 200030 (P. R. China)
 E-mail: abbaskhan@sjtu.edu.cn
 dqwei@sjtu.edu.cn
- [b] A. Khan., D.-Q. Wei
 Department of Bioinformatics and Biological Statistics
 School of Life Sciences and Biotechnology
 Shanghai Jiao Tong University, Shanghai, 200240 (P. R. China)
- [c] D.-Q. Wei
 Peng Cheng Laboratory, Vanke Cloud City Phase I, Building 8
 Xili Street, Nashan District, Shenzhen, Guangdong, 518055 (P. R. China)
- [d] K. Kousar
 Atta ur Rahman School of Applied Biosciences
 National University of Science and Technology (NUST)
 Islamabad (Pakistan)
- [e] S. Ahmad
 Department of Health and Biological Sciences, Abasyn University
 Khyber Pakhtunkhwa (Pakistan)

- [f] J. Ali
 Swat Institute of Nuclear Medicine Oncology and Radiotherapy (SINOR) Hospital
 Saaidu Sharif, Khyber Pakhtunkhwa (Pakistan)
- [g] J. Abubaker, A. Mohammad
 Department of Biochemistry and Molecular Biology
 Dasman Diabetes Institute (Kuwait)
- [h] F. Al-Mulla
 Department of Genetics and Bioinformatics
 Dasman Diabetes Institute (Kuwait)
- [i] S. S. Ali
 Center for Biotechnology and Microbiology
 University of Swat, Swat, KP (Pakistan)
- [j] N. Nizam-Uddin
 Biomedical Engineering Department
 HITEC University, Taxila (Pakistan)
- [k] A. Mohammad Sayaf
 Department of Chemistry
 University of Swat, Khyber Pakhtunkhwa (Pakistan)

 Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cbic.202100191>

contain a receptor-binding domain (RBD) and facilitates virus binding to ACE2, while the S2 subunit allows membrane fusion as a sequel to the viral genome's integration into the host's genome, initiating infection.^[7] Reports from England demonstrated that mutations in the S protein led to a new contagious strain named "B.1.1.7".^[11] These mutations in this strain may enhance virus transmissibility and infectivity, including the deletion of residues 69–70 and 144 and the substitution of A570D, D614G, T716I, S982A, D1118H, P681H, K417N, K417T, E484 K and N501Y.^[8] From a vaccination perspective, these mutations may support the virus's immune evasion capability, rendering vaccines ineffective against the B.1.1.7 strain.^[9] More worrisome is the continuous spread of new SARS-CoV-2 strains in South Africa (B.1.351) and the United Kingdom (B.1.1.7), carrying mutations of N501Y and E484 K within the RBD domain. In the United Kingdom, the number of 501Y mutation-based cases has been observed to increase from 0.1% to 49.7% in just one month. This mutation has also been reported in co-occurrence with other mutations in orf8, orf1a, and N proteins.^[10] Both the B.1.351 and B.1.1.7 variants are more contagious than the initially reported strain. However, the B.1.1.7 strain is unlikely to evade the protective immunity developed through the currently developed vaccine, whereas the B.1.351 variant may overcome this immunity.^[9c,11] Very recently, a novel B.1.617 variant with two mutations L452R and E484Q on the RBD domain was identified in India. Around 10,787 samples were collected from 18 Indian states, uncovering 771 cases of known variants 34 cases of the B.1.351 variant, 736 cases of the B.1.1.7 variant, and one case of the P.1 variant. According to government officials, an increasing number of variants with mutations, specially E484Q and L452R, have been seen in samples collected from the western state of Maharashtra.^[12] Such double mutations may confer increased infectivity and facilitate evasion of the immune system. In time, the new SARS-CoV-2 B.1.617 variant may be more pathogenic, enhancing transmissibility; therefore, close analysis of such variants and predictions of dangerous mutations is crucial in controlling devastating effects of the COVID-19 in the future.

Thus, understanding SARS-CoV-2 mechanisms is essential to determine how the mutations help the virus to survive by protecting themselves from hosts' immune defence. Computational approaches in predicting the impact of these mutations on the S protein structure, function and binding to ACE2 offer great promise for devising therapeutic strategies.^[13] Accordingly, in this present study, our objective is to employ several computational techniques to decipher the impact of the newly emerged B.1.617 carrying L452R-E484Q mutations on the overall S protein structure and function, assessing the effect on binding with ACE2 and the overall impact on interactions network. These findings may aid in opening new avenues of curtailing the SARS-CoV-2.

Computational Section

B.1.617 variant modelling and protein-protein docking

Our study began with a retrieval of the SARS-CoV-2 spike protein sequence (accession number: P0DTC2) from the UniProt database^[14] to map mutations exact location. Homology modelling of the mutated sequence was performed using the Modeller tool. The wild-type spike protein sequence (PDB ID: 6M0J) was then accessed from the protein data bank (PDB)^[15] and, subsequently, used in the UCSF Chimera v.1.15 interactive visualization program^[16] to generate the desired E484Q-L452R mutant. The 6XDG (E) and 6M0J (E) structures reported experimentally were used as modelling templates. Then, molecular docking of the mutants with an ACE2 receptor molecule was conducted through a high ambiguity-driven protein-protein docking (HADDOCK) server,^[17] following the protocol used by Abbas et al. in 2021.^[11] The dissociation constant (K_D) for the complexes was determined using the PROtein binDing enerGY prediction (PRODIGY) server.^[18]

Dynamics of the spike wild type and B.1.617 mutant-ACE2 complexes

We examined the structural dynamics of the B.1.617 mutant-ACE2 complex vis-à-vis the wild spike-ACE2 complex through a repeated molecular dynamics simulation run of 500 ns. This was accomplished through AMBER20 simulation package using FF19SB to describe receptors-ligand backbone and sidechain parameters.^[19] Complete details of systems preparation and running MD simulation production run were followed from Abbas et al., 2021.^[11] Briefly, systems were solvated in a TIP3P solvation box and afterwards neutralized by adding an appropriate number of counter ions. Systems minimization was achieved via 6000 and 3000 steps of steepest descent and conjugate gradient algorithms. Heating of systems was done to 300 K, followed by the equilibration phase. The equilibrium for each system was attained at almost 50 ns and then the final stage production run for each system was run at time scale of 500 ns. Simulation trajectories were generated that were analysed through CPPTRAJ module of AMBER.^[20]

Estimation of binding free energy

The binding free energies of both the mutant and the wild-type complexes were estimated through the AMBER MMGBSA.py module.^[21] The systems were subjected to the MM/GBSA method, and energy terms such as electrostatic energy, van der Waals energy (vdW), and polar and non-polar solvation energies were determined and is widely used by various studies.^[22] The energy terms were also mathematically processed to estimate the system's net binding free energy, using the equation below:

$$\Delta G_{\text{net binding energy}} = \Delta G_{\text{complex binding energy}} - [\Delta G_{\text{receptor binding energy}} + \Delta G_{\text{ligand binding energy}}]$$

Each of the above components of net binding energy can be split as follows:

$$G = G_{\text{bonded}} + G_{\text{van der waals}} + G_{\text{polar solvation energy}} + G_{\text{non-polar solvation energy}}$$

The entropic calculation was not conducted since it is computationally expensive and time consuming process and also highly susceptible to significant inaccuracies.^[23]

Results and Discussion

The recent SARS-CoV-2 variants that have emerged in different countries of the world have developed a distressing situation. The appearance of this novel strain demonstrates that the wild-type strain has been exposed to increased genetic pressure and has developed genetic variants. Variations produced by such genetic changes may have resulted in significant changes in propagation, virulence, and clinical implications. Recently Abbas et al., employed protein-protein docking methods in a biophysical investigation to examine the effect of mutations reported in the United Kingdom, South Africa, Brazil, and other countries on the structure and binding of ACE2 and revealed that the B.1.351, P.1 variants are more dangerous and may seriously

threaten the efficacy of the already developed vaccine. Recently, a new variant known as B.1.617 of SARS-CoV-2 was reported in India, and it might have contributed to increased infectivity and immune evasion. These mutations are acquired by the RBD of the spike protein. The spike protein is a multi-domain protein that uses the RBD domain to bind to the human ACE2 protein. Because of the essential role of Spike protein in binding and pathogenesis, it has been deemed as a potential drug target to treat COVID-19 patients. These mutations may be related to functional heterogeneity and may particularly use a different approach to infection. The RBD-specific mutations (L452R-E484K) reported in India requires careful examination in the search for new therapeutic alternatives. Therefore, in the current study, we performed comparative binding and biophysical analysis of the wild type and B.1.617 mutant upon the interaction with ACE2. The primary sequence of the wild-type spike-RBD was obtained from UniProt, and the reported mutations were generated. Using Modeller, the structure of the B.1.617 variant was modelled. The multi-domain representation of the spike protein, the modelled RBD structure of the B.1.617 variant, and sequence alignment are given in Figure 1.

Using HADDOCK, protein-protein docking of ACE2 with the wild spike-RBD and B.1.617 variant was performed to explore the structural mechanism of the SARS-CoV-2 variants. Previously Abbas et al., reported that with 1 salt-bridge (between Glu30 and Lys417), 11 hydrogen bonds and 139 non-bonded interactions, the docking score for the wild type was reported to be -122.6 ± 0.7 kcal/mol. Using a similar approach, docking of

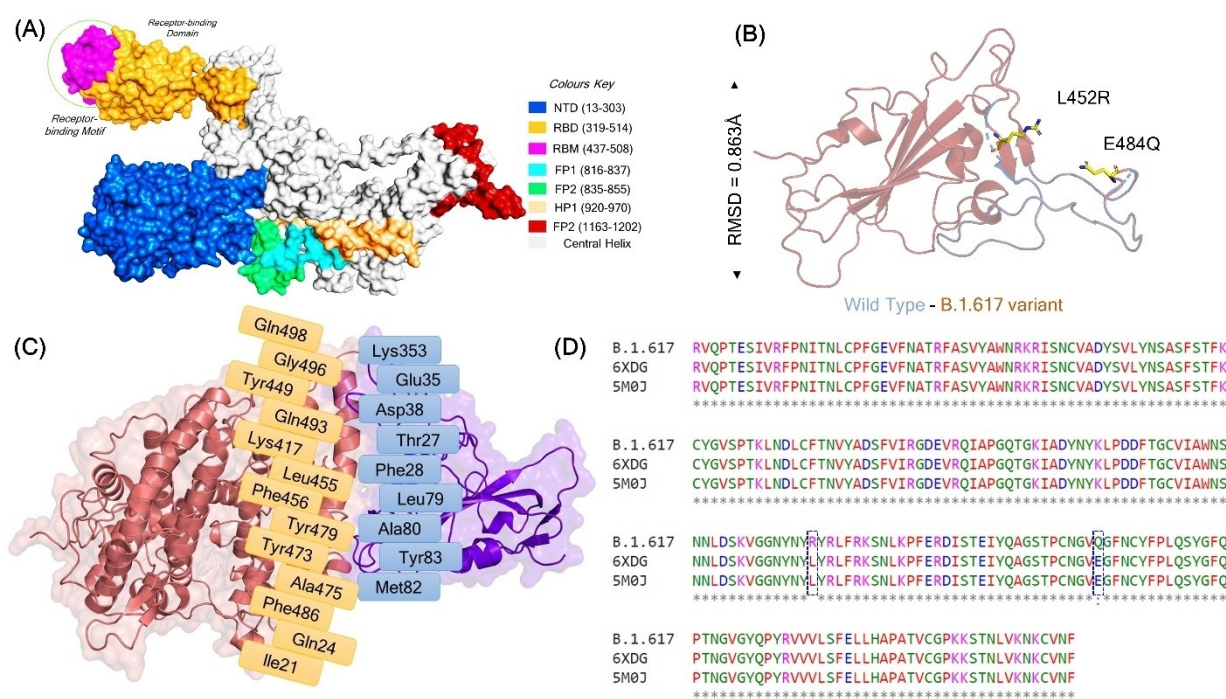


Figure 1. (A) The multidomain organization of the spike proteins from SARS-CoV-2. (B) The homology model structure of the B.1.617 variant superimposed on the wild-type spike RBD (RMSD = 0.863 Å). (C) The interacting residues on the ACE2-spike RBD interface. (D) The sequence alignment between the query sequence and templates used for homology modelling. The mutated residues are indicated with a circle.

Table 1. represent the docking scores along with the interaction energies and K_D values for the wild type and B.1.617 variant.

Parameter	Wild type	B.1.617 variant
HADDOCK score (kcal/mol)	-122.6 +/- 0.7	-123.0 +/- 3.5
Cluster size	64	53
RMSD (Å)	1.7 +/- 1.0	0.6 +/- 0.4
Van der Waals energy (kcal/mol)	-59.6 +/- 2.3	-59.4 +/- 4.1
Electrostatic energy (kcal/mol)	-181.4 +/- 15.5	-205.3 +/- 11.2
Desolvation energy (kcal/mol)	-27.1 +/- 3.4	-26.1 +/- 1.5
Restraints violation energy(kcal/mol)	4.7 +/- 3.8	36.4 +/- 25.6
Buried surface area (Å ²)	1965.3 +/- 120.6	1840.6 +/- 86.3
Z-score	-1.9	-1.3
K_D (dissociation constant) docking conformation	5.2×10^{-10}	2.0×10^{-08}
K_D (dissociation constant) MD equilibrated conformation	4.00×10^{-09}	1.4×10^{-09}

human ACE2 with the B.1.617 variant spike-RBD was performed. The resulting docking score was -123.0 ± 3.5 kcal/mol. Interaction analysis showed that the new variant altered the bonding network and formed distinct interactions. The mutated residues formed significant interactions with ACE2. Only one salt bridge (between Glu30 and the mutated Arg452), seven hydrogen bonds and 129 non-bonded contacts were reported. Similarly, only 7 hydrogen bonds were formed among Glu30-Arg452, Glu30-Arg452, Tyr34-Gln484, Glu35-Tyr453, Glu35-Gln493, Gln37-Tyr489 and Glu75-Asn501 residues. It can be seen that the bonding pattern is significantly altered by these mutations. Likewise, the new substituted residues formed three hydrogen bonds: two between Glu30 and Arg452 and another one between Tyr34 and Gln484. This shows that the binding pattern is significantly affected by the mutation. Interestingly the electrostatic energy contribution (-205.3 ± 11.2 kcal/mol) endured as the primary contributing influence in the B.1.617 variant. This notion of more electrostatic energy is supported by previous studies^[24] that state that evaluation of the binding of SARS-CoV-2 and SARS-CoV RBDs with ACE2 have revealed more hydrogen bonds and electrostatic interactions in SARS-CoV-2.^[24,25] However, the Van der Waals energy of the wild type and B.1.617 variant is comparable. These results obtained from docking only are further validated and explained in Hydrogen bonding analysis after MD simulation section. Decisively, this rationalizes that the binding of the mutant complex is principally due to the electrostatic contribution and further suggests that the notable differences in the binding pattern are noteworthy for differential infectivity.

We further determined the K_D (dissociation constant) of the wild and mutant complexes to provide convincing insights into the binding distinctions. It is used to estimate and order the strengths of bio-macromolecular association.^[26] The K_D kinetics are frequently used to foresee the affinity of antigen-antibody associations, protein-ligand interactions, and significant biological macromolecule interactions. The lower the K_D value, the stronger the interaction.^[27] The K_D values of diverse macromolecules association and their interactions have been calculated previously.^[28] We used PRODIGY (PROtein binDing energy prediction), an online server that computes the binding affinity, K_D , for various biological complexes, to gain a deeper understanding of the binding of the wild type and mutant complex. Previous study reported a K_D of 5.2×10^{-10} for the wild type,

while here in this study, the predicted K_D value for the B.1.617 variant was 2.0×10^{-08} , which shows a relatively tighter binding of the wild type than the B.1.617 variant for the docking conformation. Next, we extracted MD equilibrated average conformation to evaluate the binding strength. The K_D value for the wild-type MD equilibrated (Average structure) 4.0×10^{-09} while for the B.1.617 variant MD equilibrated structure the K_D value was predicted to be 1.4×10^{-09} . This shows that the B.1.617 variant equilibrated structure demonstrate stronger binding properties than the wild-type equilibrated structure. The lower K_D values obtained in this analysis are consistent with the results obtained in *in-vitro* binding assays of SARS-CoV-2 and SARS-CoV, which reported a tighter binding of the SARS-CoV-2 RBD to the host receptor ACE2 than the SARS-CoV.^[29] The docking scores along with the interaction energies and K_D values for the wild type and B.1.617 variant are given in Table 1.

Next, we determined the consequence of the fixed amino-acid replacements reported within the RBD of the SARS-CoV-2 spike protein by calculating its stability as root mean square deviation (RMSD). RMSD can be used to identify variations between proteins' initial and final structural conformations. These conformational changes can be estimated with an RMSD during simulation, determining a biological molecule's stability. A smaller deviation during such simulations indicates a highly stable structure. RMSD values were calculated for each protein-protein complex from the C α backbone, using 500 ns repeated trajectories. As Figure 2A shows, the wild system attained equilibration at and stable at RMSD of 2.0 Å. Almost all the systems attained the equilibrium position at 50 ns. The system remained stable during the simulation; however, a deviation in the RMSD was observed between 50–60 ns when it increased to 4.0 Å. This later decreased back to 2.0 Å and remained uniform for the wild-type complex until 500 ns with the exception of two very minor deviations at 225 ns and 390 ns. This pattern was also observed in a previous study, which reported the more stable behaviour of the wild-type system. The dynamic differences between the wild type and the mutated systems determine the binding differences and behaviour. The average RMSD for replicate 1 during the first 75 ns remained 4.0 Å; however, the RMSD abruptly increased and experienced significant convergence between 76–125 ns. The average RMSD between 76–125 ns was reported to be 6.0 Å. Afterwards, the RMSD value decreased and remained stable until 200 ns.

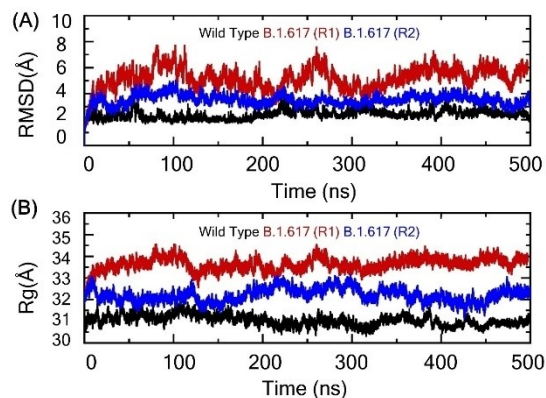


Figure 2. (A) The RMSDs of all the studied complexes for 500 ns each. The RMSD of the wild type is shown in blue, while the other mutants are shown in different colours. (B) The Rg values of all the studied complexes. The Rg of the wild type is shown in black, while the other mutants are shown in different colours (Replicate 1: red; Replicate 2: blue).

Afterwards from 220 ns to 300 ns, another period of the increased-decreased curve was noticed where maximum RMSD reaches ~ 8 Å at 260 ns. From 310 ns onward, the RMSD again starting to surge till 420 ns, followed by another small up-down surge and then later in the simulation gain stability. Similarly, Replicate 2 also reported a higher RMSD value compared to the wild type. The average RMSD for Replicate 2 was 4.0 Å. Replica 2, in comparison to Replica 1 is more stable in term of RMSD and considerably consistent. Previous studies have reported that fluctuations in RMSD values for these complexes are linked with the opening or closing motion of the claw-like structure in ACE2.^[30] The higher RMSD values obtained at different time intervals throughout a simulation were probably associated with the binding and unbinding of some destabilized interactions in the ACE2-spike interface. Consequently, the unusual instability of the RBD was associated with the Darwinian selection-driven epistasis in protein evolution.^[22g]

The radius of gyration (Rg) was calculated to evaluate the compactness during the simulation, as given in Figure 2B. The Rg values increased and decreased during the simulation, and the average Rg value for the wild type was 31.2 Å. In contrast, the average Rg values for the two replicates of the B.1.617 variant were 32.4 Å and 33.0 Å, respectively. Thus, the wild-type system seemed to remain more compact during the first 100 ns; however, this compactness was lost between 95 and 130 ns, and then it remained uniform again until 500 ns where it was seen in a very low Rg pattern. On the other hand, the two replicates remained less compact during the MD simulation. The structure of Replicate 1 remained substantially open. The initial Rg value continuously increased until 125 ns; however, the Rg then remained uniform until 200 ns. Similarly, the Replica 2 Rg values remained consistent with the RMSD results. No significant convergence was observed; however, the average Rg values remained higher than the wild type's. The fluctuation in Rg values during the simulations of all the studied systems was associated with the binding and unbinding of one or another end of the spike RBD domain. An increase and decrease

in the Rg pattern can also be pointed to 500 ns though these variations between the wild type and replica 2 are comparatively less than the replica 1. Additionally, the bonding rearrangement during the MD simulation holds and releases the two receptors, which causes a perturbation in the structural compactness.

The wild type exhibited minimal root mean square fluctuation (RMSF) compared to the reported mutant (replicates). In the B.1.617 variant, the residues 1–200 exhibited greater fluctuations than the wild type. Next, we found that the RMSF pattern for the wild type and Replicate 2 remained comparable (Figure 3); however, for replicate 1 the residual fluctuation remained higher during the simulation than the wild type. For the mutant (both replicas), higher fluctuations were observed between regions 100–200 compared to the wild type. These fluctuations were linked with the presence of three important loops involved in binding with ACE2, thus highlighting the functional importance for binding, which is possible due to essential conformational changes during Darwinian evolution and amino acids fixation. This higher fluctuation is resulted from the terminal loop residues Ala522, Thr523, Val523 and Cys525 which exhibit unusual flexibility. These overall results suggest that possible evolutionary changes in the mutants have led to variations in their dynamic function.

Furthermore, an intermolecular bonding network was explored for the wild type and B.1.617 variant by using an equilibrated average structure from MD trajectories to evaluate the differences in binding pattern accurately. As given in Figure 4A, the wild-type interaction pattern has been changed during the MD simulation while the B.1.617 variant formed extra interactions during the MD time period. A total of 11 hydrogen bonds and 1 salt bridge was detected in the wild-type complex, while 15 hydrogen bonds and two salt bridges were reported in the B.1.617 variant complex (Table 2). In the wild type, the two strong bonds between Tyr83-Asn487 and Glu35-Gln493 remained conserved and has also been reported in the previous study as sustained interactions. Similarly, the interaction Glu35-Gln493 plays an important role in locking the

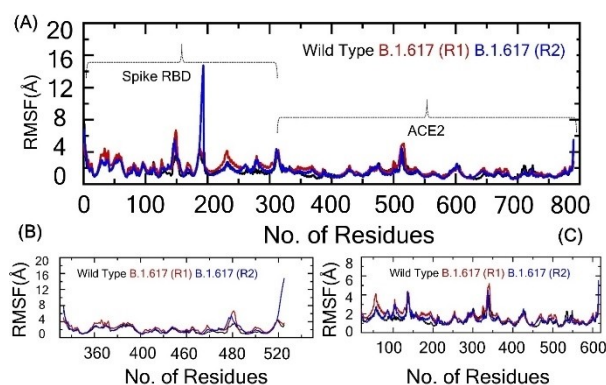


Figure 3. The residual flexibility index for the wild type and B.1.617 variant. (A) The RMSF of the wild type and B.1.617 variant (two replicas) complexes are shown in different colours (wild type: black; Replicate 1: Red; Replicate 2: blue). (B) Only shows the RMSF of the wild type and B.1.617 RBD domain while, (C) represent the RMSFs of the ACE2 receptor.

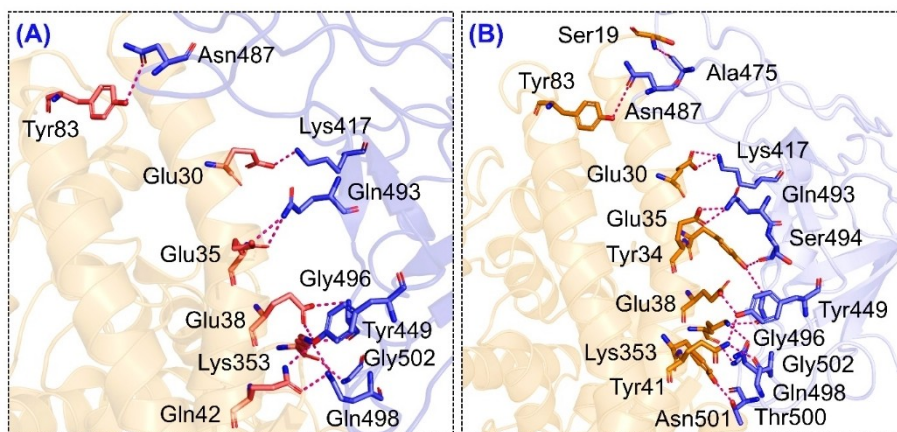


Figure 4. The interaction pattern of the wild type and B.1.617 variant in the MD equilibrated sampled structures. (A) shows the binding of the wild-type, while (B) show the B.1.617 variant interaction pattern. The ACE2 receptor is coloured orange, while the RBD domain is shown as blue.

Table 2. Binding (Interaction) differences between the wild type and B.1.617 variant in the MD equilibrated sampled structures.				
Complex	ACE2	RBD	Distance (Å)	Bond type
Wild type	Tyr83[HH]	Asn487[OD1]	2.46	hydrogen bond
	Lys353[HZ3]	Gly496[O]	1.89	hydrogen bond
	Lys353[HZ1]	Gln498[OE1]	1.89	hydrogen bond
	Glu30[OE2]	Lys417[HZ2]	1.74	hydrogen bond
	Glu38[OE1]	Tyr449[HH]	1.67	hydrogen bond
	Glu38[OE1]	Gln498[H]	2.02	hydrogen bond
	Glu35[OE1]	Gln493[HE22]	2.11	hydrogen bond
	Glu35[OE1]	Gln493[HE20]	1.92	hydrogen bond
	Glu38[OE2]	Gly496[H]	2.18	hydrogen bond
	Gln42[OE1]	Gln498[HE22]	2.40	hydrogen bond
	Lys353[O]	Gly502[H]	2.04	hydrogen bond
	Glu30[OE2]	Lys417[NZ]	2.66	salt bridge
	B.1.617 variant	Ser19[N]	Ala475[O]	3.10
Tyr83[OH]		Asn487[OD1]	3.55	hydrogen bond
Glu30[OE1]		Lys417[NZ]	2.79	hydrogen bond
Glu35[OE2]		Gln493[NE2]	2.83	hydrogen bond
Glu35[OE1]		Gln493[NE2]	2.75	hydrogen bond
His34[OH]		Ser494[O]	2.66	hydrogen bond
His34[OH]		Gly496[N]	2.90	hydrogen bond
Lys353[NZ]		Gln498[OE1]	2.86	hydrogen bond
Tyr41[OH]		Thr500[OG1]	2.68	hydrogen bond
Glu38[OE2]		Tyr449[OH]	2.57	hydrogen bond
Glu38[OE2]		Tyr440[OH]	2.97	hydrogen bond
Tyr34[OH]		Gly496[N]	3.49	hydrogen bond
Gln42[OE1]		Gln498[NE2]	2.92	hydrogen bond
Tyr41[OH]		Asn501[N]	3.54	hydrogen bond
Lys353[O]		Gly502[N]	2.82	hydrogen bond
Glu30[OE2]		Lys417[NZ]	2.79	salt bridge
Glu30[OE2]		Lys417[NZ]	2.90	salt bridge

correct orientation of Gln493 for interaction is essentially retained during the 500 ns simulation, which is also reported by previous studies.^[11,24b, 31] Interesting, a cluster of interactions by Lys353 was observed during the MD simulation. This residue, Lys353, formed a stronger network of hydrogen bonds with Gly496, Gln498 and Gly502. Interestingly, a previous study also reported the Lys353 interaction as a critical mediator in the attachment to the receptor.^[11,24b, 31] Moreover, a single salt bridge between Glu30-Lys417 was observed, which is an essential element of the interaction network. The interaction pattern of the wild type RBD and ACE2 is shown in Figure 4A.

Differences in the bonding network of the wild type and the B.1.617 variant was found significantly. In the docking conformations, only 7 hydrogen bonds were reported, while in the equilibrated MD structure, 15 hydrogen bonds and 2 salt bridges were established. Interestingly among the key interactions established between the ACE2 and B.1.617 variant, RBD formed shared contacts with the wild type. Among the 15 hydrogen bonds, the interactions formed by Tyr83, Glu35, Gln42 and Lys353 established by B.1.617 variant complex are conserved between the wild type and the variant structures. In the case of B.1.617 variant Ser19-Ala475, Glu30-Lys417, Tyr34-

Ser494, His34-Gly496, Tyr41-Asn501 established essential new hydrogen bonds while Glu30 maintained two salt bridges with Lys417. The three hydrogen bond pairs Tyr41-Asn501, Lys353-Gln498 and Lys353-Gly502 reported here are also previously observed as the key interactions that were limited in the SARS-CoV RBD-ACE2 complex.^[11,24b,31] Interestingly the interaction Glu38-Tyr449 established here was also reported in a previous study and concluded that this interaction might also play a

significant role in the enhanced binding. Additionally, the two salt bridges may justify the increased electrostatic energy of the B.1.617 variant complex. Conclusively this shows that during the simulation, both the structures are altered significantly and established several new contacts, which are essential for the binding differences. The interaction pattern of B.1.617 variant complex is shown in Figure 4B, while the details of the interactions of both complexes along with the bonding distances are given in Table 2.

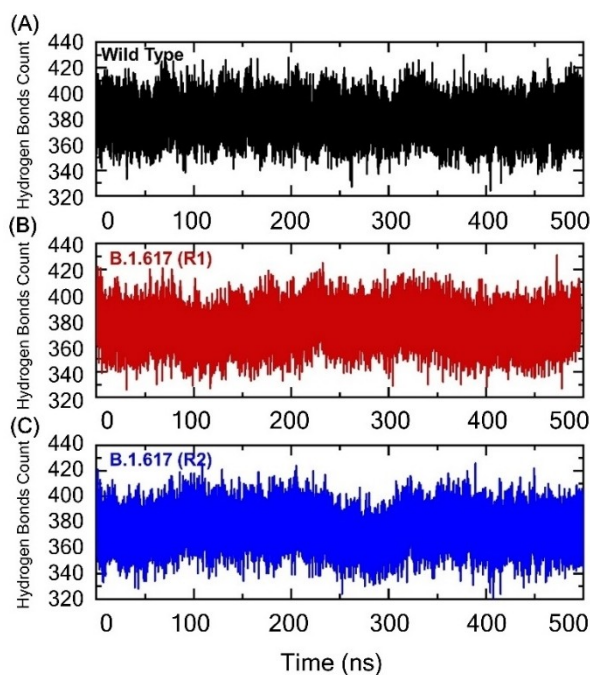


Figure 5. The hydrogen bonding index of the wild-type and mutant complexes. The hydrogen bonding graph of the wild type is shown in black, while the other mutants are shown in different colours (Replicate 1: Red; Replicate 2: blue).

Furthermore, we also performed the total hydrogen bonding analysis (both intra and intermolecular) as a function of the total number of frames in order to understand the variations in the bonding network induced by the substitutions in the RBD of the Spike protein. From each trajectory, 25000 frames were considered, and the total number of hydrogen bonds during the simulation was calculated and presented in Figure 5. In the wild type, the average hydrogen bonds during the 500 ns simulations were reported to be 388, while in replicate 1 the average hydrogen bonds were 394, and in replicate 2 the average hydrogen bonds were 396. These results are consistent with our docking results, which showed that the wild type has 11 hydrogen bonds while the mutant complex has only seven hydrogen bonds. This finding shows that the mutations (B.1.617 variant) have altered their hydrogen-bonding network and may use a different strategy than the B.1.1.7, B.1.351, and P.1 variants. Moreover, to justify our findings, we also calculated binding free energy using 500 ns simulation trajectories.

The MM/GBSA energies of both the wild type and the B.1.617 variant was compared and contrasted to extract ideas about the mutations induced conformation changes in the RBD region and their influence on ACE2 binding. Detail atomic-level energies at different time intervals demonstrated that and presented in Supporting Table S1 while the binding free energies for the last 200 ns and their averages were calculated and given in Table 3. The rationale behind the presentation of

Table 3. MM/GBSA binding free energies of the wild type and B.1.617 variant. All energy values are presented in kcal/mol.			
Wild Type	301-400 ns	401-500 ns	Averages
van der Waals	-92.65 ± 0.15	-87.51 ± 0.09	-90.08 ± 0.09
Electrostatic interactions	-687.7 ± 1.01	-698.97 ± 0.74	-693.33 ± 0.74
Generalized Born	729.74 ± 1.01	733.15 ± 0.01	731.44 ± 0.74
Nonpolar solvation energy	-12.21 ± 0.01	-11.78 ± 0.01	-11.99 ± 0.01
Total binding energy	-62.82 ± 0.13	-65.12 ± 0.09	-63.97 ± 0.09
B.1.617 (R1)	301-400 ns	401-500 ns	Averages
van der Waals	-94.75 ± 0.11	-99.06 ± 0.12	-96.90 ± 0.12
Electrostatic interactions	-931.52 ± 0.66	-1142.93 ± 0.74	-1037.22 ± 0.70
Generalized Born	978.4 ± 0.68	1173 ± 0.77	1075.70 ± 0.73
Non-polar solvation energy	-11.39 ± 0.01	-12.61 ± 0.01	-12.00 ± 0.01
Total binding energy	-59.26 ± 0.07	-81.59 ± 0.11	-70.42 ± 0.09
B.1.617 (R2)	301-400 ns	401-500 ns	Averages
van der Waals	-97.55 ± 0.17	-99.31 ± 0.09	-98.43 ± 0.13
Electrostatic interactions	-1142.03 ± 0.55	-1214.13 ± 0.30	-1178.08 ± 0.42
Generalized Born	1169.87 ± 0.55	1236.38 ± 0.27	1203.12 ± 0.41
Non-polar solvation energy	-13.50 ± 0.02	-13.63 ± 0.01	-13.56 ± 0.01
Total binding energy	-83.21 ± 0.16	-90.69 ± 0.12	-86.95 ± 0.15

the 200 ns MM/GBSA result is due to uniform equilibration state attained by all the complexes including the wild type and mutant replicas. In a particular case of Replica 2 the equilibration was also attained in the earlier time intervals therefore, the free energy results for the 500 ns simulation time intervals are presented in the Supporting Table S1. The data show that the wild type (van der Waals energy: -90.08 ± 0.09 kcal/mol), the B.1.617 variant had a higher van der Waals energy contribution (-96.90 ± 0.12 kcal/mol in Replica 1 and -98.43 ± 0.13 kcal/mol in Replica 2) upon ACE2 binding. However, the electrostatic energy is more favourable and seems to contribute heavily to the stability of the variant complex with ACE2. In both replicas, the B.1.617 variant showed double the electrostatic energy contribution of the wild type. This finding implies that the mutations in the B.1.617 variant may allow for a fitter conformation with ACE than the wild type. The electrostatic contribution to solvation energy for the wild type was (-693.33 ± 0.74 kcal/mol) and while for the B.1.617 variant (-1037.22 ± 0.70 kcal/mol in Replica 1 and -1178.08 ± 0.42 kcal/mol in Replica 2). The non-polar solvation energy estimated for the complexes favoured good stability for both the wild type and the variant. For wild type, the non-polar solvation energy was -11.99 ± 0.01 kcal/mol, versus -12.00 ± 0.01 kcal/mol and -13.56 ± 0.01 kcal/mol for Replica 1 and Replica 2, respectively. Overall, net binding energy revealed that the B.1.617 variant was somewhat stable in terms of complex formation with ACE2 (-70.42 ± 0.09 kcal/mol in Replica 1 and -86.95 ± 0.15 kcal/mol in Replica 2) compared to the wild type (-63.97 ± 0.09 kcal/mol). Unlike the results obtained through docking studies based on single structure prediction, the current results using 25000 structural frames concluded that the B.1.617 variant shows a stronger binding affinity with ACE2 than the wild type. Conclusively, these findings align with the above docking and simulation data are an indication that mutations' induced structural dynamics to support the RBD's increased binding to ACE2.

This study was conducted to explore the binding differences of the spike RBD of the wild and B.1.617 variant (L452R-E484Q) with the host ACE2 by using combined structural modelling and biophysical approaches. This investigation revealed that the B.1.617 variant possess a stronger binding affinity for the host ACE2 in both replicates and the bonding network is substantially different from the wild type. We concluded that L452R-E484Q mutations and other factors such as genetic variability in the other proteins might help the virus enforce infection and its severity. This study provides a strong impetus to develop novel drugs the new variants.

Data availability

All the data are available on RCSB, UniProt and any simulation data would be provided on reasonable demand. The accession numbers to access this data are given in the manuscript.

Acknowledgements

Dong-Qing Wei is supported by grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Science Foundation of China (Grant No. 32070662, 61832019, 32030063), the Science and Technology Commission of Shanghai Municipality (Grant No.: 19430750600), as well as SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2021ZD02). The computations were partially performed at the PengCheng Lab. and the Center for High-Performance Computing, Shanghai Jiao Tong University.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: ACE2-spike docking · B.1.617 variant · biophysical simulation · dissociation constant · SARS-CoV-2

- [1] Z. Zhu, X. Lian, X. Su, W. Wu, G. A. Marraro, Y. Zeng, *Respir. Res.* **2020**, *21*, 224.
- [2] a) T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, S.-H. Jung, *J. Med. Chem.* **2016**, *59*, 6595–6628; b) A. Kawana, *Nihon Rinsho. Jap. J. Clin. Med.* **2016**, *74*, 1967–1972.
- [3] a) H. Hilal El Idrissi, *Gene Rep.* **2020**, *20*, 100756; b) T. Acter, N. Uddin, J. Das, A. Akhter, T. R. Choudhury, S. Kim, *Sci. Total Environ.* **2020**, *730*, 138996–138996; c) N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, W. Tan, *New Engl. J. Med.* **2020**, *382*, 727–733.
- [4] a) D. Benvenuto, M. Giovanetti, A. Ciccozzi, S. Spoto, S. Angeletti, M. Ciccozzi, *J. Med. Virol.* **2020**, *92*, 455–459; b) T. Estola, *Avian Dis.* **1970**, *14*, 330–336; c) *Nat. Microbiol.* **2020**, *5*, 536–544.
- [5] a) R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R. Ehmann, K. Zwirgmaier, C. Drosten, C. Wendtner, *Nature* **2020**, *581*, 465–469; b) H. S. Rahman, M. S. Aziz, R. H. Hussein, H. H. Othman, S. H. Salih Omer, E. S. Khalid, N. A. Abdulrahman, K. Amin, R. Abdullah, *Int. J. Surgery Open* **2020**, *26*, 125–136; c) V. Stadnytskyi, C. E. Bax, A. Bax, P. Anfinrud, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 11875–11877; d) S. Perlman, *New Engl. J. Med.* **2020**, *382*, 760–762; e) P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, *Nature* **2020**, *588*, E6–E6.
- [6] a) S. M. Haque, O. Ashwaq, A. Sarief, A. K. Azad John Mohamed, *Fut. Virol.* **2020**, *15*, 625–648; b) W. Li, M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, M. Farzan, *Nature* **2003**, *426*, 450–454; c) M. Letko, A. Marzi, V. Munster, *Nat. Microbiol.* **2020**, *5*, 562–569.
- [7] a) W. Li, C. Zhang, J. Sui, J. H. Kuhn, M. J. Moore, S. Luo, S. K. Wong, I. C. Huang, K. Xu, N. Vasilieva, *EMBO J.* **2005**, *24*, 1634–1643; b) F. Li, *J. Virol.* **2015**, *89*, 1954–1964; c) A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Velesler, *Cell* **2020**, *181*, 281–292, e286; d) C. Burkard, M. H. Verheije, O. Wicht, S. I. van Kasteren, F. J. van Kuppeveld, B. L. Haagmans, L. Pelkmans, P. J. Rottier, B. J. Bosch, C. A. de Haan, *PLoS Pathog.* **2014**, *10*, e1004502.
- [8] a) T. Koyama, D. Platt, L. Parida, *Bull. W. H. O.* **2020**, *98*, 495; b) T. Kirby, *Lancet Respir. Med.* **2021**, *9*, e20–e21.
- [9] a) A. Muik, A.-K. Wallisch, B. Sängler, K. A. Swanson, J. Mühl, W. Chen, H. Cai, D. Maurus, R. Sarkar, Ö. Türeci, P. R. Dormitzer, U. Şahin, *Science* **2021**, eabg6105; b) C. M. Voloch, R. da Silva F, Jr., L. G. P. de Almeida, C. C. Cardoso, O. J. Brustolini, A. L. Gerber, A. P. de C. Guimarães, D. Mariani, R. Mirella da Costa, O. C. Ferreira, Jr., Covid19-UFRJ Workgroup, LNCC-Workgroup, A. C. Cavalcanti, T. S. Frauches, C. M. Braga de Mello,

- R. Mello Galliez, D. Souza Faffe, T. M. P. P. Castiñeiras, A. Tanuri, A. T. R. de Vasconcelos, *medRxiv* **2020**, <https://doi.org/10.1101/2020.12.23.20248598>; c) M. H. Cheng, J. M. Krieger, B. Kaynak, M. A. Arditi, I. Bahar, *bioRxiv* **2021**, <https://doi.org/10.1101/2021.01.10.426143>; d) A. A. Pater, M. S. Bosmeny, C. L. Barkau, K. N. Ovington, R. Chilamkurtthy, M. Parasrampur, S. B. Eddington, A. O. Yinusa, A. A. White, P. E. Metz, R. J. Sylvain, M. M. Hebert, S. W. Benzinger, K. Sinha, K. T. Gagnon, *bioRxiv* **2021**, <https://doi.org/10.1101/2021.01.11.426287>.
- [10] a) Y. Itan, L. Shang, B. Boisson, M. J. Ciancanelli, J. G. Markle, R. Martinez-Barricarte, E. Scott, I. Shah, P. D. Stenson, J. Gleeson, *Nat. Methods* **2016**, *13*, 109–110; b) D. A. Collier, I. A. T. M. Ferreira, R. Datir, B. Meng, L. Bergamaschi, CITIID-NIHR BioResource COVID-19 Collaboration, A. Elmer, N. Kingston, B. Graves, K. G. C. Smith, J. R. Bradley, P. A. Lyons, L. Ceron-Gutierrez, G. Barcenas-Morales, R. Doffinger, M. R. Wills, R. K. Gupta, *SSRN* **2021**, <https://dx.doi.org/10.2139/ssrn.3782450>; c) N. G. Davies, R. C. Barnard, C. I. Jarvis, A. J. Kucharski, J. Munday, C. A. B. Pearson, T. W. Russell, D. C. Tully, S. Abbott, A. Gimma, W. Waites, K. L. M. Wong, K. van Zandvoort, J. D. Silverman, CMMID COVID-19 Working Group, The COVID-19 Genomics UK (COG-UK) Consortium, K. Diaz-Ordaz, R. Keogh, R. M. Eggo, S. Funk, M. Jit, K. E. Atkins, W. J. Edmunds, *medRxiv* **2020**, <https://doi.org/10.1101/2020.12.24.20248822>.
- [11] A. Khan, T. Zia, M. Suleman, T. Khan, S. S. Ali, A. A. Abbasi, A. Mohammad, D.-Q. Wei, *J. Cell. Physiol.* **2021**, <https://doi.org/10.1002/jcp.30367>.
- [12] J. Singh, S. A. Rahman, N. Z. Ehtesham, S. Hira, S. E. Hasnain, *Nat. Med.* **2021**, <https://doi.org/10.1038/s41591-021-01397-4>.
- [13] A. Khan, S. Khan, S. Saleem, N. Nizam-Uddin, A. Mohammad, T. Khan, S. Ahmad, M. Arshad, S. S. Ali, M. Suleman, *Comput. Biol. Med.* **2021**, *104420*.
- [14] M. Magrane, *Database* **2011**, bar009.
- [15] P. W. Rose, B. Beran, C. Bi, W. F. Bluhm, D. Dimitropoulos, D. S. Goodsell, A. Prlić, M. Quesada, G. B. Quinn, J. D. Westbrook, *Nucleic Acids Res.* **2010**, *39*, D392–D401.
- [16] T. D. Goddard, C. C. Huang, T. E. Ferrin, *Structure* **2005**, *13*, 473–482.
- [17] C. Dominguez, R. Boelens, A. M. Bonvin, *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.
- [18] L. C. Xue, J. P. Rodrigues, P. L. Kastriitis, A. M. Bonvin, A. Vangone, *Bioinformatics* **2016**, *32*, 3676–3678.
- [19] a) R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888; b) R. Salomon-Ferrer, D. A. Case, R. C. Walker, *Wiley Interdiscipl. Rev. Comput. Mol. Sci.* **2013**, *3*, 198–210.
- [20] D. R. Roe, T. E. Cheatham III, *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
- [21] T. Hou, J. Wang, Y. Li, W. Wang, *J. Chem. Inform. Model.* **2011**, *51*, 69–82.
- [22] a) A. Khan, W. Heng, Y. Wang, J. Qiu, X. Wei, S. Peng, S. Saleem, M. Khan, S. S. Ali, D.-Q. Wei, *Phytother. Res.* **2021**, *35*, 2841–2845; b) A. Khan, M. Junaid, C.-D. Li, S. Saleem, F. Humayun, S. Shamas, S. S. Ali, Z. Babar, D.-Q. Wei, *Front. Mol. Biosci.* **2020**, *6*, 159; c) A. Khan, Z. Rehman, H. F. Hashmi, A. A. Khan, M. Junaid, A. M. Sayaf, S. S. Ali, F. U. Hassan, W. Heng, D.-Q. Wei, *Interdiscipl. Sci.* **2020**, *12*, 155–168; d) A. Khan, M. T. Khan, S. Saleem, M. Junaid, A. Ali, S. S. Ali, M. Khan, D.-Q. Wei, *Comput. Struct. Biotechnol. J.* **2020**, *18*, 2174–2184; e) A. Khan, M. Khan, S. Saleem, Z. Babar, A. Ali, A. A. Khan, Z. Sardar, F. Hamayun, S. S. Ali, D.-Q. Wei, *Interdiscipl. Sci. Comput. Life Sci.* **2020**, *12*, 335–348; f) A. Khan, M. Junaid, A. C. Kaushik, A. Ali, S. S. Ali, A. Mehmood, D.-Q. Wei, *PLoS One* **2018**, *13*; g) I. Hussain, N. Pervaiz, A. Khan, S. Saleem, H. Shireen, D.-Q. Wei, V. Labrie, Y. Bao, A. A. Abbasi, *Genes Immun.* **2020**, *21*, 409–419.
- [23] Z. Sun, Y. N. Yan, M. Yang, J. Z. Zhang, *J. Chem. Phys.* **2017**, *146*, 124124.
- [24] a) A. Spinello, A. Saltalamacchia, A. Magistrato, *J. Phys. Chem. Lett.* **2020**, *11*, 4785–4790; b) J. de Andrade, P. F. B. Gonçalves, P. A. Netz, *Chem-BioChem* **2021**, *22*, 865–875.
- [25] a) J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, *Nature* **2020**, *581*, 215–220; b) Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, K.-Y. Yuen, *Cell* **2020**, *181*, 894–904.e9.
- [26] J. P. Landry, Y. Fei, X. Zhu, *Assay Drug Dev. Technol.* **2012**, *10*, 250–259.
- [27] J. P. Landry, Y. Fei, X. Zhu, *Int. Drug Discovery* **2011**, 8–13.
- [28] J. Landry, Y. Sun, X. Guo, X. Zhu, *Appl. Opt.* **2008**, *47*, 3275–3288.
- [29] a) X. Tian, C. Li, A. Huang, S. Xia, S. Lu, Z. Shi, L. Lu, S. Jiang, Z. Yang, Y. Wu, *Emerg. Microb. Infect.* **2020**, *9*, 382–385; b) A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Velesler, *Cell* **2020**, *183*, 1735.
- [30] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, *Science* **2020**, *367*, 1444–1448.
- [31] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, F. Li, *Nature* **2020**, *581*, 221–224.

Manuscript received: April 22, 2021
Revised manuscript received: June 23, 2021
Accepted manuscript online: June 23, 2021
Version of record online: July 5, 2021