

# Critical Sequence Hot-spots for Binding of nCOV-2019 to ACE2 as Evaluated by Molecular Simulations

Mahdi Ghorbani<sup>1,3</sup>, Bernard R. Brooks<sup>3</sup>, Jeffery B. Klauda<sup>1,2</sup>

<sup>1</sup>Department of Chemical and Biomolecular Engineering, <sup>2</sup>Biophysics Graduate Program,  
University of Maryland, College Park, MD 20742, USA

<sup>3</sup>Laboratory of Computational Biology, National, Heart, Lung and Blood Institute, National  
Institutes of Health, Bethesda, Maryland 20824, USA

\*Corresponding Author: [jbklauda@umd.edu](mailto:jbklauda@umd.edu)

## Abstract:

The novel coronavirus (nCOV-2019) outbreak has put the world on edge, causing millions of cases and hundreds of thousands of deaths all around the world, as of June 2020, let alone the societal and economic impacts of the crisis. The spike protein of nCOV-2019 resides on the virion's surface mediating coronavirus entry into host cells by binding its receptor binding domain (RBD) to the host cell surface receptor protein, angiotensin converter enzyme (ACE2). Our goal is to provide a detailed structural mechanism of how nCOV-2019 recognizes and establishes contacts with ACE2 and its difference with an earlier coronavirus SARS-COV in 2002 via extensive molecular dynamics (MD) simulations. Numerous mutations have been identified in the RBD of nCOV-2019 strains isolated from humans in different parts of the world. In this study, we investigated the effect of these mutations as well as other Ala-scanning mutations on the stability of RBD/ACE2 complex. It is found that most of the naturally-occurring mutations to the RBD either strengthen or have the same binding affinity to ACE2 as the wild-type nCOV-2019. This may have implications for high human-to-human transmission of coronavirus in regions where these mutations have been found as well as any vaccine design endeavors since these mutations could act as antibody escape mutants. Furthermore, in-silico Ala-scanning and long-timescale MD simulations, highlight the crucial role of the residues at the interface of RBD and ACE2 that may be used as potential pharmacophores for any drug development endeavors. From an evolutionary perspective, this study also identifies how the virus has evolved from its predecessor SARS-COV and how it could further evolve to become more infectious.

## Introduction

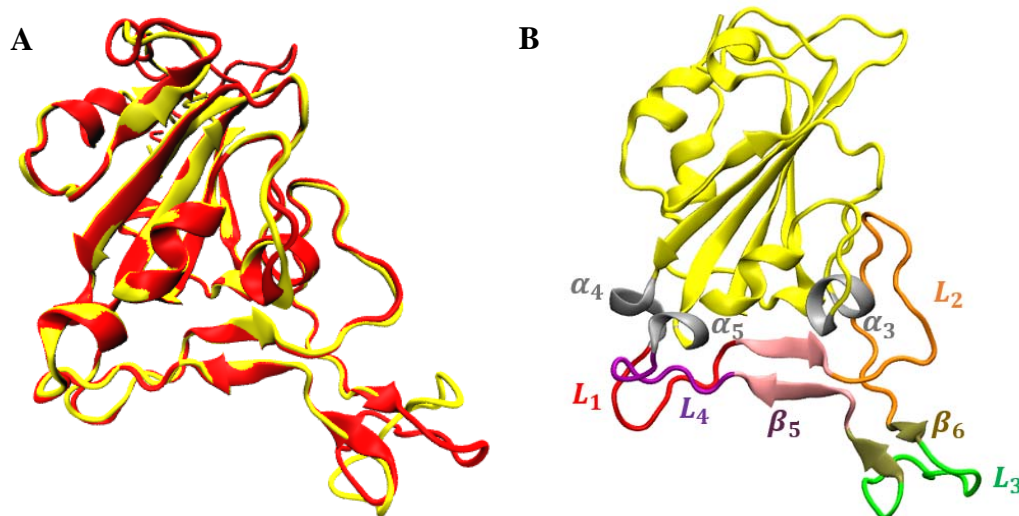
The novel coronavirus (nCOV-2019) outbreak emerging from China has become a global pandemic and a major threat for human public health. According to World Health Organization (WHO) as of June 27<sup>th</sup> 2020, there has been about 10 million confirmed cases and approaching 500,000 deaths due to coronavirus in the world.<sup>1-2</sup> Much of the human population including the United States of America were under lockdown or official stay-at-home orders to minimize the continued spread of the virus.

Coronaviruses are a family of single-stranded enveloped RNA viruses. Phylogenetic analysis of coronavirus genome has shown that nCOV-2019 belongs to the beta-coronavirus family, which also includes MERS-COV, SARS-COV and bat-SARS-related coronaviruses.<sup>3-4</sup> It is worth mentioning that SARS-COV, which was widespread in 2002 caused more than 8,000 cases and about 800 deaths and MERS-COV (middle east respiratory syndrome coronavirus) in 2012 also spread in more than 25 countries, causing about 2,500 cases and more than 850 deaths. ([www.who.int/health-topics/coronavirus](http://www.who.int/health-topics/coronavirus)).<sup>5</sup>

In all coronaviruses, a homotrimeric spike glycoprotein on the virion's envelope mediates coronavirus entry into host cells through a mechanism of receptor binding followed by fusion of viral and host membranes.<sup>3,6</sup> Coronavirus spike protein contains two functional subunits S1 and S2. The S1 subunit is responsible for binding to host cell receptor, and the S2 subunit is responsible for fusion of viral and host cell membranes.<sup>3,7</sup> The spike protein in nCOV-2019 exists in a meta-stable pre-fusion conformation that undergoes a substantial conformational rearrangement to fuse the viral membrane with the host cell membrane.<sup>7,8</sup> nCOV-2019 is closely related to bat coronavirus RaTG13 with about 93.1% sequence similarity in the spike protein gene. The sequence similarity of nCOV-2019 and SARS-COV is less than 80% in the spike genome.<sup>2</sup> S1 subunit in the spike protein includes a receptor binding domain (RBD) that recognizes and binds to the host cells receptor. The RBD of nCOV-2019 shares 72.8% sequence identity to SARS-COV RBD and the root mean squared deviation (RMSD) for the structure between the two proteins is 1.2 Å, which shows the high structural similarity.<sup>4,8,9</sup> Experimental binding affinity measurements using Surface Plasmon Resonance (SPR) have shown that nCOV-2019 spike protein binds its receptor human angiotensin converter enzyme (ACE2) with 10 to 20

fold higher affinity than SARS-COV binding to ACE2.<sup>7</sup> Based on the sequence similarity between RBD of nCOV-2019 and SARS-COV and also the tight binding between RBD of nCOV-2019 and ACE2, it is most probable that nCOV-2019 uses this receptor on human cells to gain entry into the body.<sup>3,6,7,10</sup>

The spike protein and specifically the RBD domain in coronaviruses have been a major target for therapeutic antibodies. However, no monoclonal antibodies targeted to RBD have been able to bind efficiently and neutralize nCOV-2019.<sup>7,11</sup> The core of nCOV-2019 RBD is a 5-stranded antiparallel  $\beta$ -sheet with connected short  $\alpha$ -helices and loops (Figure 1). Binding interface of nCOV-2019 and SARS-COV with ACE2 are very similar with less than 1.3 Å RMSD. An extended insertion inside the core containing short strands,  $\alpha$ -helices and loops called the receptor binding motif (RBM) makes all the contacts with ACE2. In nCOV-2019 RBD, the RBM forms a concave surface with a ridge loop on one side and it binds to a convex exposed surface of ACE2. The overlay of SARS and nCOV-2019 RBD proteins are shown in Figure 1A. The binding interface in nCOV-2019 contains loops  $L_1$  to  $L_4$  and short  $\beta$ -strands  $\beta_5$  and  $\beta_6$  and a short helix  $\alpha_5$ . The location of RBM in nCOV-2019 RBD as well as different helices, strands and loops is shown in Figure 1B.



**Figure 1.** A) superposition of the RBD of SARS-COV (yellow) and nCOV-2019(red). B) different regions in the binding domain of nCOV-2019 defining the extended loop (non-yellow).

The sequence alignment between SARS-COV in human, SARS civet, Bat RaTG13 coronavirus and nCOV-2019 in the RBM is shown in Figure 2. There is a 50% sequence similarity between the RBM of nCOV-2019 and SARS-COV. RBM mutations played an important role in the SARS epidemic in 2002.<sup>12,3</sup> Two mutations in the RBM of SARS-2002 from SARS-Civet were observed from strains of these viruses. These two mutations were K479N and S487T. These two residues are close to the virus binding hotspots in ACE2 including hotspot-31 and hotspot-353. Hotspot-31 centers on the salt-bridge between K31-E35 and hotspot-353 is centered on the salt-bridge between K353-E358 on ACE2. Residues K479 and S487 in SARS-Civet are in close proximity with these hotspots and mutations at these residues caused SARS to bind ACE2 with significantly higher affinity than SARS-Civet and played a major role in civet-to-human and human-to-human transmission of SARS coronavirus in 2002.<sup>3,13-15</sup> Numerous mutations in the interface of SARS-COV RBD and ACE2 from different strains of SARS isolated from humans in 2002 have been identified and the effect of these mutations on binding ACE2 have been investigated by surface plasmon resonance.<sup>16,17</sup> Two identified RBD mutations (Y442F and L472F) increased the binding affinity of SARS-COV to ACE2 and two mutations (N479K, T487S) decreased the binding affinity. It was demonstrated that these mutations were viral adaptations to either human or civet ACE2.<sup>16,17</sup> A pseudotyped viral infection assay of the interaction between different spike proteins and ACE2 confirmed the correlation between high affinity mutants and their high infection.<sup>16</sup> Further investigation of RBD residues in binding of SARS-COV and ACE2 was performed through ala-scanning mutagenesis, which resulted in identification of 10 residues that reduce binding affinity to ACE2 upon mutation to alanine. These residues are K390, R429, D429, D454, I455, N473, F483, Q492, Y494 and R495.<sup>18</sup> RBD mutations have also been identified in MERS-COV, which affected their affinity to receptor (DPP4) on human cells.<sup>17</sup> Multiple monoclonal antibodies have been developed for SARS since 2002 that neutralized spike glycoprotein on SARS-COV surface.<sup>19-23</sup> However, multiple escape mutations exist in the RBD of SARS-COV that affect neutralization with antibodies, which led to the use of a cocktail of antibodies as robust treatment.<sup>24</sup>

```

SARS-2002  424  NTRNIDATSTGNYNKYRYLRHGKLRPFERDISNVVFFSPDGKPCTP-PALNCYWPLNDYG 482
SARS-civet 424  NTRNIDATSTGNYNKYRYLRHGKLRPFERDISNVVFFSPDGKPCTP-PALNCYWPLKDYG 482
Bat RaTG13 446  NSKHIDAKEGGNFNYLYRLEFRKANLKPFFERDISTEIQAGSKPCNGQTGLNCYYPLYRYG 505
nCOV-2019 437  NSNNLDSKVGGNYNLYRLEFRKSNLKPFFERDISTEIQAGSTPCNGVEGENCFYFPLQSYG 496
*:.:.*:.  **:* ** :*:.:*****.  .. ...**  .:****** **

SARS-2002  483  FYTTTGIGYQPY 494
SARS-civet 483  FYTTSGIGYQPY 494
Bat RatG13 506  FYPTDGVGHQPY 517
nCOV-2019 497  FQPTNGVGYQPY 508
*  *  *:*:***

```

**Figure 2.** Sequence comparison of the receptor binding motif (RBM) in SARS-2002, SARS-civet, Bat RaTG13 and nCOV-2019. The mutations from SARS-2002 to nCOV-2019 are marked with blue. Important mutations in RBM are marked with yellow. Red color shows the 3-residue motif in SARS and Civet and 4-residue motif in RaTG13 and nCOV-2019.

Full genome analysis of nCOV-2019 in different countries and the receptor binding surveillance has shown multiple mutations in the RBD of glycosylated spike. The GISAID database<sup>25</sup> ([www.gisaid.org/](http://www.gisaid.org/)) contains genomes on nCOV-2019 from researchers across the world since December 2019. Latest report by the GISAID database on June 2020 have shown 25 different variants of RBD from strains of nCOV-2019 collected from different countries along with the number of occurrences in these regions which is listed below for the seven most occurring mutations:

213x **N439K** (211 Scotland, England, Romania), 65x **T478I** in England, 30x **V483A** (26 USA/WA, 2 USA/UN, USA/CT, England), 10x **G476S** (8 USA/WA, USA/OR, Belgium), 7x **S494P** (3 USA/MI, England, Spain, India, Sweden), 5x **V483F** (4x Spain, England), 4x **A475V** (2 USA/AZ, USA/NY, Australia/NSW).

It is not known whether these mutations are linked to the severity of coronavirus in these regions. The focus of this article is to elucidate the differences between the interface of SARS-COV and nCOV-2019 with ACE2 to understand with atomic resolution the interaction mechanism and hotspot residues at the RBD/ACE2 interface using long-timescale molecular dynamics (MD) simulation. An alanine-scanning mutagenesis in the RBM of nCOV-2019 helped to identify the key residues in the interaction, which could be used as potential pharmacophores for future drug development. Furthermore, we performed molecular simulations on the seven most common mutations found from surveillance of RBD mutations N439K, T478I, V483A,

G476S, S494P, V483F and A475V. From an evolutionary perspective this study shows the residues in which the virus might further evolve to be even more dangerous to human health.

## Methods

### Sequence comparison and mutant preparation

nCOV-2019 shares 76% sequence similarity with SARS-2002 spike protein, 73% sequence identity for RBD and 50% for the RBM.<sup>1</sup> Bat coronavirus RaTG13 seems to be the closest relative of nCOV-2019 sharing about 93% sequence identity in the spike protein.<sup>6</sup> The sequence alignment of SARS-2002 (accession number: AFR58742), SARS-civet (accession number: AY304486.1), Bat RaTG13 (accession number: MN996532.1) and nCOV-2019 (accession number: MN908947.1) are in Figure 2.<sup>26</sup> To investigate the roles of critical mutations on the complex stability of nCOV-2019 with ACE2, mCSM-PPI2 webserver<sup>27</sup> was used to the residues in nCOV-2019 that are at the interface with ACE2. This method uses a graph-based signature framework and predicts the effect of alanine substitution at interface residues on the binding energy of the complex. 21 different residues were identified to be in contact with ACE2 and were chosen to do further MD simulation. On the other hand, mutations are also observed in RBD domain from full genome analysis of different nCOV-2019 variants collected from different countries compiled in GISAID database.<sup>25</sup> The mutations selected are listed in Table S1 along with their location in RBD.

### Molecular dynamics simulations

The crystal structure of nCOV-2019 in complex with hACE2 (pdb id:6M0J)<sup>28</sup> as well as SARS-COV complex with human ACE2 (pdb id: 6ACJ)<sup>29</sup> were obtained from RCSB ([www.rcsb.org](http://www.rcsb.org)).<sup>30</sup> The RBD domain of nCOV-2019 comprises of 194 residues (333-526) and SARS-COV includes 190 residues (323-512). ACE2 protein contains 597 residues (19-615) in the complex structure for both systems. All the structures including nCOV-2019, SARS-COV and all 21 alanine substitutions of nCOV-2019 were prepared and solvated with the CHARMM-GUI webserver.<sup>31</sup> A TIP3P water model was used for the solvent and Param99SB-ILDN AMBER force field (FF)<sup>32</sup> was used for all the complexes. POT ions were added to all systems to neutralize the charges on the proteins. Each system contained about 260,000 atoms.

5000 steps of energy minimizations were done using the steepest descent algorithm. In all steps the LINCS algorithm was used to constraint all bonds containing hydrogen atoms and a

time step of 2 fs was used as the integration time step.<sup>33</sup> Equilibration of all systems were performed in three steps. In the first step, 100,000 steps of simulation were performed using a velocity-rescaling thermostat to maintain the temperature at 310K with a 0.1 ps coupling constant in NVT ensemble under periodic boundary conditions and harmonic restraints on the backbone and sidechain atoms of the complex.<sup>34</sup> The velocity rescaling thermostat was used in all other steps of simulation. In the next step, we performed 300,000 steps in the isothermal-isobaric NPT ensemble at temperature of 310K and pressure of 1bar using a Berendsen barostat.<sup>35</sup> This was done by decreasing the force constant of the restraint on backbone and side chain atoms of the complex from 1000 to 100 and finally to  $10 \frac{kJ}{mol nm^2}$ . Berendsen barostat was only used for the equilibration step due to usefulness in rapidly correcting density. In the next step the restraints were removed, and the systems were subjected to 1,000,000 steps of MD simulation under NPT ensemble.

In the production run, harmonic restraints were removed and all the systems were simulated using a NPT ensemble where the pressure was maintained at 1bar using the Parrinello-Rahman barostat<sup>36</sup> with a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  and a coupling constant of 0.5ps. It is important to note that all the Berendsen barostat was only used for the equilibration step as it was shown that this barostat can cause unrealistic temperature gradients.<sup>37</sup> The production run lasted for 500ns for SARS-COV and nCOV-2019 complexes and 300ns for all the mutants using with a 2fs timestep and the particle-mesh Ewald (PME)<sup>36</sup> for long range electrostatic interactions using GROMACS 2018.3 package.<sup>38</sup> All mutant systems were constructed as described before and all complexes ran for 300ns of production run.

*Gibbs free energy and correlated motions.* The last 400ns of simulation was used to explore the dominant motions in SARS-COV, nCOV-2019 and last 200ns for all its mutants using principal component analysis (PCA) as part of the quasiharmonic analysis method.<sup>39</sup> For this method the rotational and translational motions of RBD of all systems were eliminated by fitting to a reference (crystal) structure. Next, 4,000 snapshots from the last 400ns of SARS-COV and nCOV-2019 systems and 2,000 snapshots from last 200ns of all mutant systems were taken to generate the covariance matrix between  $C_{\alpha}$  atoms of RBD. Diagonalization of this matrix resulted in a diagonal matrix of eigenvalues and their corresponding eigenvectors.<sup>38,40</sup> The first eigenvector which indicate the first principal component was used to visualize the dominant

global motions of all complexes through porcupine plots using the (PorcupinePlot.tcl) script in vmd.

The principal components were used to calculate and plot the approximate free energy landscape (aFEL). We refer to the free energy landscape produced by this approach to be approximate in that the ensemble with respect to the first few PC's (lowest frequency quasi-harmonic modes) is not close to convergence, but the analysis can still provide valuable information and insight. `g_sham`, `g_covar` and `g_anaeig` functions in GROMACS<sup>41</sup> were used to obtain principal components and aFEL. In each aFEL the deep valleys represent the most stable conformations separated by some intermediate states.

The dynamic cross-correlation maps (DCCM) were obtained using MD\_TASK package to identify the correlated motions of RBD residues.<sup>42</sup> In DCCM the cross-correlation matrix  $C_{ij}$  is obtained from displacement of backbone  $C_{\alpha}$  atoms in a time interval  $\Delta t$ . The DCCM was constructed using the last 400ns of SARS-COV and nCOV-2019 systems and last 200ns of all mutant systems with a 100ps time interval.

*Binding free energy from MMPBSA method.* The Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) method was applied to calculate the binding energy between RBD and ACE2 in all complexes.<sup>43,44</sup> for SARS-COV and nCOV-2019 system, 200 snapshots of the last 400ns and for all other systems, 100 snapshots of the last 200ns simulation were used to for the calculation of binding free energies with an interval of 2ns. The binding free energy of a ligand-receptor complex can be calculated as:

$$\Delta G_{bind,aq} = \Delta H - T\Delta S = \Delta G_{complex} - [\Delta G_{protein} + \Delta G_{ligand}] \quad (1)$$

$$\Delta G_{bind,aq} = \Delta E_{MM} + \Delta G_{bind,solv} - T\Delta S \quad (2)$$

$$\Delta E_{MM} = \Delta E_{Covalent} + \Delta E_{Elect} + \Delta E_{VDW} \quad (3)$$

$$\Delta G_{bind,solv} = \Delta G_{polar} + \Delta G_{non-polar} \quad (4)$$

$$\Delta G_{non-polar} = \gamma \times SASA + b \quad (5)$$

In these equations,  $\Delta E_{MM}$ ,  $\Delta G_{bind,solv}$  and  $-T\Delta S$  are calculated in the gas phase.  $\Delta E_{MM}$  is the gas phase molecular mechanical energy changes which includes covalent, electrostatic and van der Waals. Based on previous studies the entropy change during binding is small and neglected in these calculations.<sup>44-46</sup>  $\Delta G_{bind,solv}$  is the solvation free energy which comprises the polar and non-polar components. The polar solvation is calculated using the MMPBSA method

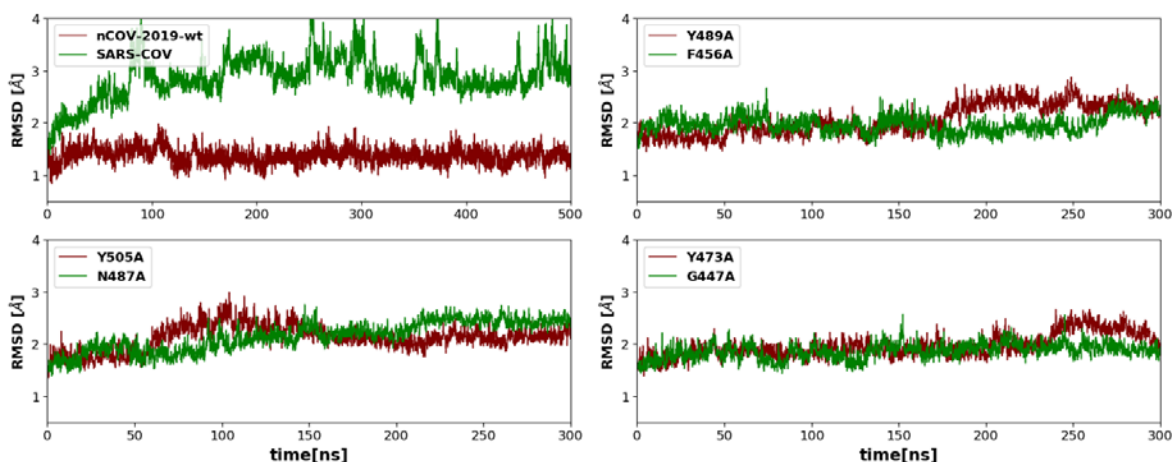


by setting a value of 80 and 2 for solvent and solute dielectric constants. The non-polar free energy is simply estimated from solvent accessible surface area (SASA) of the solute from equation 5.

## Results

### Structural dynamics

To compute the RMSD of systems, the rotational and translational movements were removed by first fitting the  $C_{\alpha}$  atoms of the RBD to the crystal structure and then computing the RMSD with respect to  $C_{\alpha}$  atoms of RBD in each system.



**Figure 3.** The  $C_{\alpha}$  RMSD plots for nCoV-2019 and SARS-COV and a few nCoV-2019 mutations.

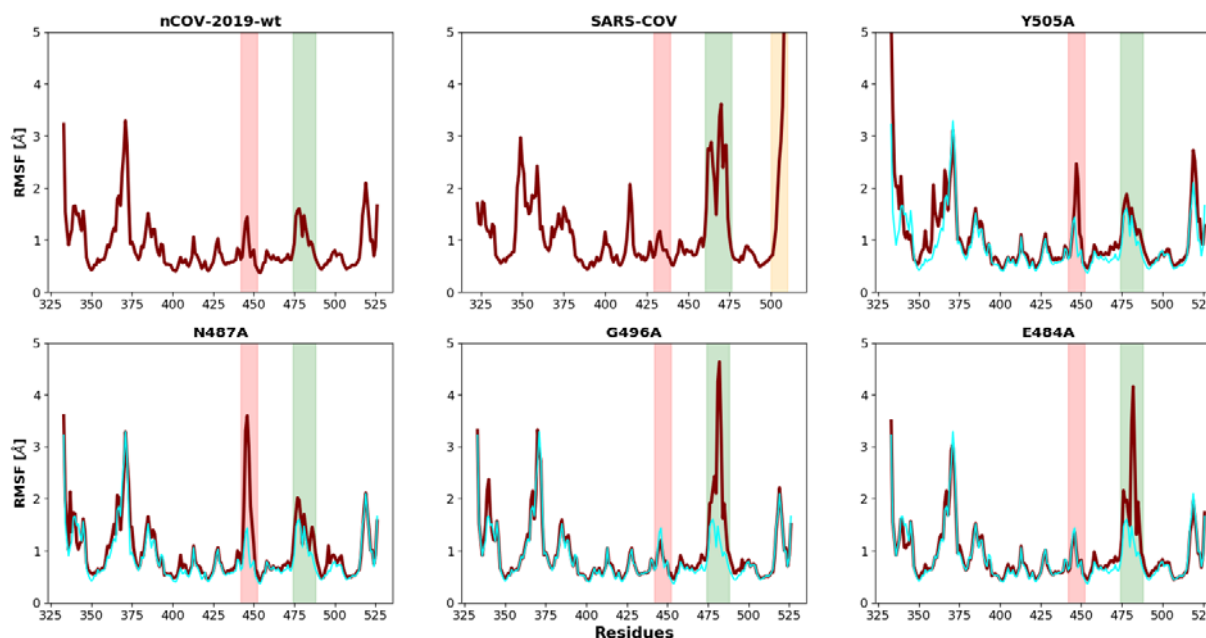
Figure 3 shows the RMSD plot in the RBD of SARS-COV, nCoV-2019 and some of its variants. Comparison of RMSD of SARS and nCoV-2019 RBD shows that SARS-COV has a larger RMSD throughout the 500ns simulation. In nCoV-2019, RMSD is very stable with a value of about 1.5 Å whereas in SARS-COV the RMSD increases up to ~4 Å after 100 ns and then fluctuates between 3 and 4 Å. The change in RMSD of SARS is partially related to the motion in the C-terminal which is a flexible loop.

The RMSD plots for the nCoV-2019 mutants show similar behaviors to nCoV-2019-wt. In most of the variants, the RMSD is stable during the 300ns simulation. However, a few mutations show some RMSD variance. In mutation Y489A, the RMSD increases from  $1.82 \pm$

0.06 Å to  $2.15 \pm 0.10$  Å after 150 ns. Mutation Y505A resulted in an increase in RMSD up to 100ns to a value to  $1.98 \pm 0.20$  Å and decreased afterwards. RMSD for mutation N487A shows an increasing behavior with a value of  $2.10 \pm 0.23$  Å. In mutation Y473A, the RMSD is  $1.94 \pm 0.17$  Å. Mutations N439K, V483A and V483F showed a stable RMSD of  $\sim 1.5$  Å. For mutations T478I, G476S, S494P and A475V RMSD increases up to  $\sim 2$  Å. These variations in the backbone RMSD show the involvement of these residues in the complex stability. RMSD plots for other mutations are shown in Figure S1.

Since the extended loop (residues 449 to 510 shown in figure 1B from  $\alpha_4$  to  $\alpha_5$ ) of the RBD makes all contacts with ACE2, the RMSD was computed by also fitting to the  $C_\alpha$  atoms of this region (Figure S2). The extended loop in nCOV-2019 is very stable with less deviation (RMSD =  $0.86 \pm 0.017$  Å) from the crystal structure compared to SARS-COV having an RMSD of  $2.79 \pm 0.05$  Å. Few of the mutants show an increase in the loop RMSD during the simulation. In mutants N487A and Y449A the loop RMSD jumps to a value of about  $2.00 \pm 0.04$  Å and  $1.63 \pm 0.09$  Å respectively after about 200 ns of simulation. Mutants G447A and E484A show a loop RMSD values of  $2.22 \pm 0.03$  Å and  $1.96 \pm 0.02$  Å during the last 100ns of simulation. The loop RMSD in mutations Y505A, Y473A and S494A is about 1.5 Å during simulation which is higher than nCOV-2019. Other mutants showed a stable extended loop (observed in loop RMSD) during simulation. The loop RMSD for some residues are high which indicates these residues are involved in the stability of the extended loop region.

To characterize of the dynamic behavior for each amino acid in the RBD, we analyzed the root mean square fluctuation (RMSF) of all systems. The RMSF plots for nCOV-2019, SARS-COV and four other mutations are shown in Figure 4. nCOV-2019 shows less fluctuations compared to SARS-COV.  $L_3$  in nCOV-2019 corresponding to residues 476 to 487 (shown in red in Figure 4) has smaller RMSF (1.5 Å) than SARS-COV  $L_3$  residues 463 to 474.  $L_1$  in both nCOV-2019 and SARS-COV (green) has small fluctuation (less than 1.5 Å). Moreover, the C-terminal residues of SARS-COV shows high fluctuation (Figure 4 shown in orange). Few mutants show higher fluctuation in the  $L_1$ . Mutants Y505A and S494A had a RMSF of 2.5 Å and mutation N487A has a RMSF of about 4 Å in the  $L_1$ . Mutation Y449A has a higher fluctuation RMSF of about 3 Å in the  $L_3$ . Mutants G496A, E484A and G447A show a high fluctuation of about 4.5 Å in the  $L_3$ . RMSF of other variants are shown in Figure S3.

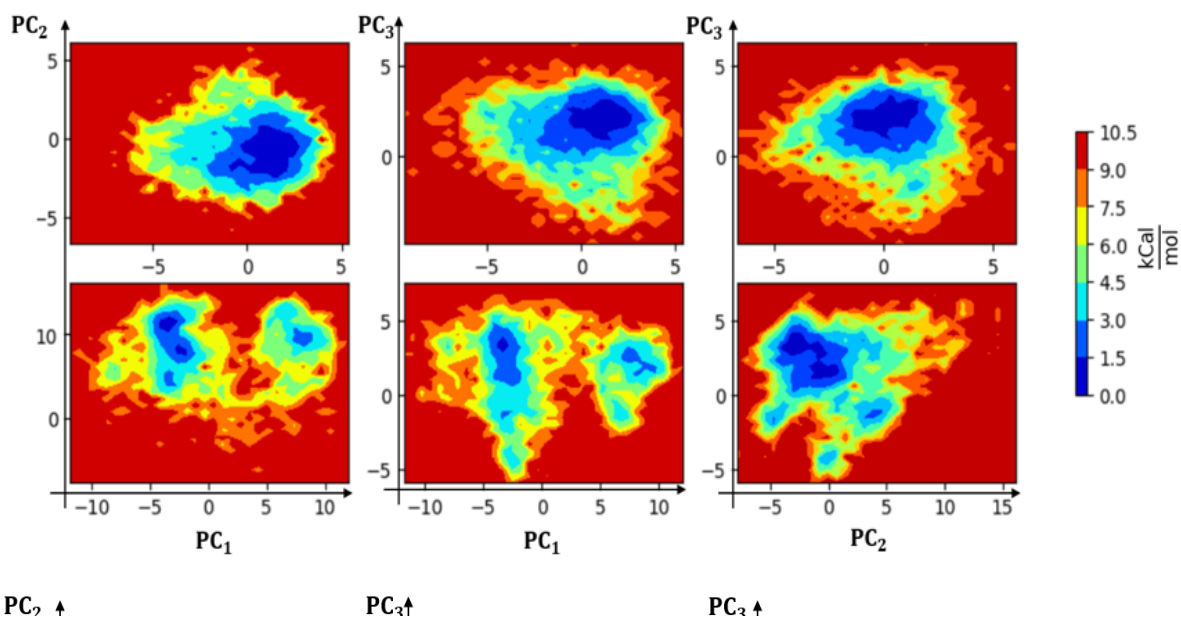


**Figure 4)** RMSF plots for nCOV-2019-wt, SARS-COV, Y505A, N487A, G496A and E484A. the red shaded region shows the fluctuation in  $L_1$  and the green shaded shows the fluctuation in  $L_3$ . The orange shaded region in SARS-COV shows the fluctuation in C-terminal. For comparison, RMSF of nCOV-2019-wt is shown as cyan in other plots.

### Principal component analysis (PCA) and approximate free energy landscape

To identify the dominant motions in the nCOV-2019, SARS-COV, and all the mutants, PCA was performed. Most of the combined motions were captured by the first ten eigenvectors generated from the last 400ns for SARS-COV and nCOV-2019 and the last 200ns for each nCOV-2019 mutant. The percentage of the motions captured by the first three eigenvectors was 51% for nCOV-2019 and 68% for SARS-COV. In all mutations, more than 50% of the motions were captured by the first three eigenvectors. The first few PC's describe the highest motions in a protein which are related to a functional motion such as binding or unbinding of protein from

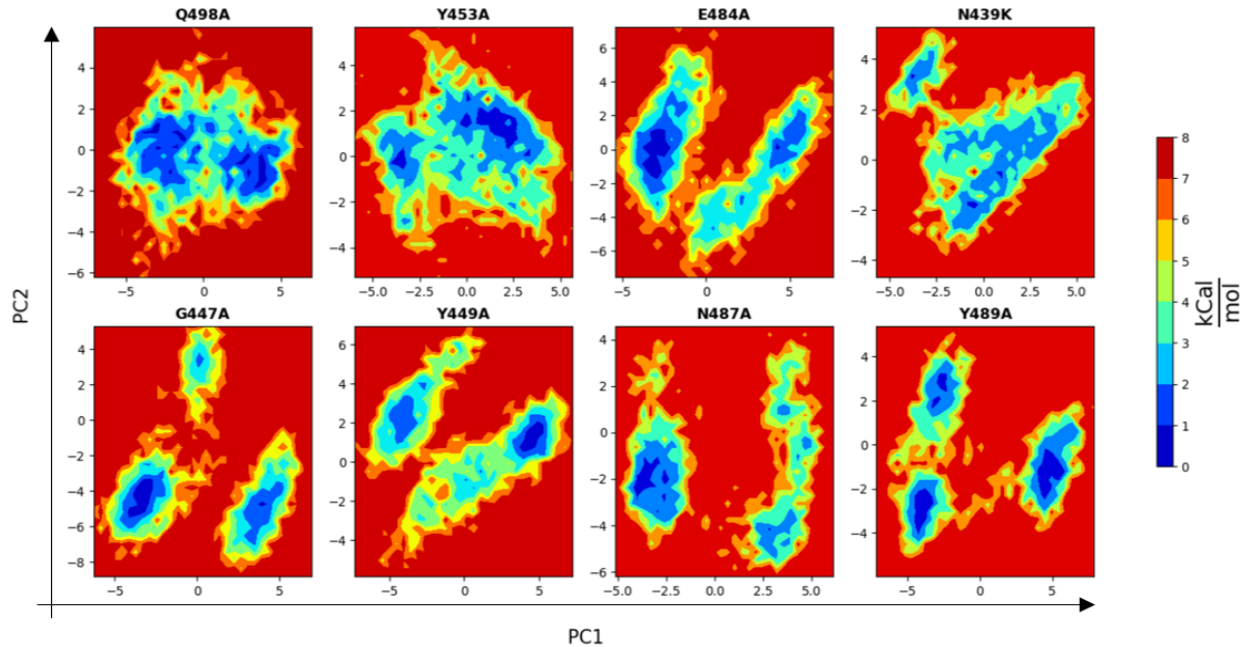
receptor. The first three eigenvectors were used to calculate the approximate FEL (aFEL) using the last 400ns of simulation for nCOV-2019 and SARS-COV shown in Figure 5, which displays the variance in conformational motion. SARS-COV showed two distinct low free energy states shown as blue separated by a metastable state. There is a clear separation between the two regions by a free energy barrier of about 6-7.5 kcal/mol. These two states correspond to the loop motions in the  $L_3$  as well as the motion in C-terminal residues of SARS-COV. The  $L_3$  motion in nCOV-2019 is stabilized by H-bond between N487 on RBD and Y83 on ACE2 as well as a  $\pi$ -stacking interaction between F486 and Y83. It is evident that the nCOV-2019 is more stable than SARS-COV and exist in one conformation whereas the SARS-COV interface fluctuates and the aFEL is separated into two different regions.



**Figure 5.** Mapping of the principal components of the RBD for the aFEL from the last 400ns of simulations for SARS-COV (top row) and nCOV-2019-wt (Bottom row). The color bar is relative to the lowest free energy state.

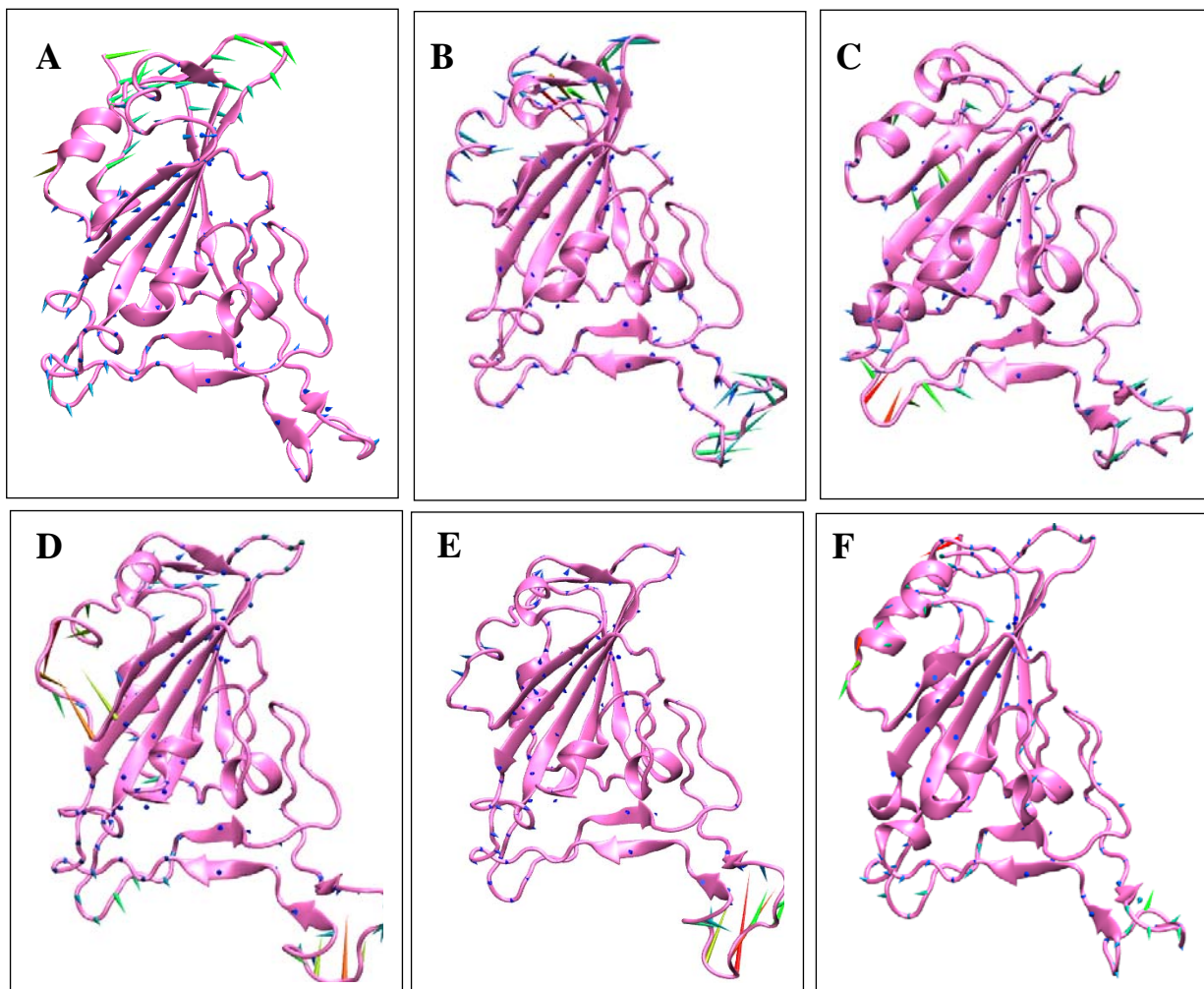
The first two eigenvectors were used to calculate and plot the aFEL as a function of first two principal components using the last 200ns of the simulation for mutant systems. aFEL for a few mutants are shown in Figure 6 and the rest of them are shown in Figure S4. In some mutants such as Q498A and Y453A the aFEL shows only one low energy state. However, aFEL for other mutants such as E484A, G447A, Y449A and N487A separate into multiple regions divided by metastable states. Mutations Y449A, G477A and E484A also showed high fluctuation in  $L_3$  and

multiple states in aFEL can be assigned to different states of the  $L_3$ . In mutation N487A the  $L_1$  showed large fluctuations and the two states in the aFEL of this mutant correspond to different states of  $L_1$ .



**Figure 6.** aFEL plots for nCOV-2019, SARS-COV and mutants by projecting onto the first two eigenvectors.

In each system, the first eigenvector was used to construct the porcupine plots to visualize the most dominant movements (Figure 7). nCOV-2019 showed a small motion in  $L_3$  and the core region and the extended loop region is very rigid showing small cones in the porcupine plot. The core structure of the RBD remains dormant as the cones are blue in most of the regions. In SARS-COV, the C-terminal region shows large motions (Figure 7B). Mutation N487A showed a large motion in  $L_1$  (Figure 7C). Mutations Y449A, G447A and E484A demonstrate large motions in  $L_3$  (Figure 7). Overall, these plots show the involvement of these residues in dynamic stability of RBD/ACE2 complex.

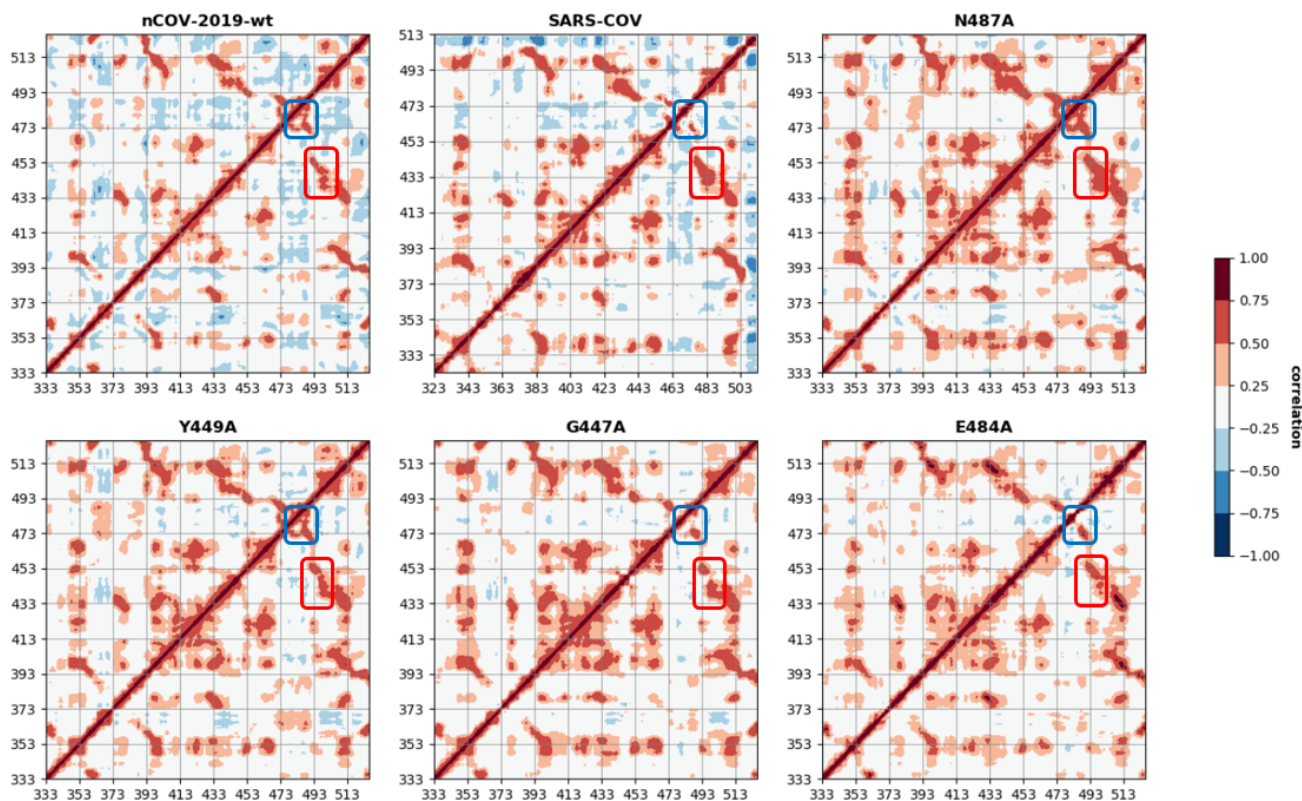


**Figure 7.** Porcupine plots for A) nCOV-2019 B) SARS-COV C) N487A D) Y449A E) G447A F) 18-Glu484. Arrows show the direction of movement for the  $C_{\alpha}$  atoms and length of arrows show the magnitude of the motion. longer arrows are red and shorter arrows are blue.

### Dynamic cross-correlation maps (DCCM)

The correlated motions of RBD atoms were also analyzed with the DCCM based on the  $C_{\alpha}$  atoms of RBD from the last 400ns of simulation for nCOV-2019 and SARS-COV and the last 200ns for the mutant systems (Figure 8). DCCM for nCOV-2019 showed a correlation between

residues 490-505 (containing  $\alpha_5$ ,  $L_4$  and  $\beta_5$  regions) and residues 440-455 (containing  $\alpha_4$ ,  $L_1$  and  $\beta_5$  regions) shown in red rectangle in Figure 8, which is referred to here as *correlation I*. This correlation showed the coordination of these regions for binding ACE2 effectively. Another important correlation that appears in DCCM of nCOV-2019 is between residues 473-481 with residues 482-491. These residues are in  $L_3$  and  $\beta_6$  regions and their correlation in nCOV-2019 is stronger than SARS-COV. This is due to the presence of  $\beta_6$  strand in nCOV-2019, whereas in SARS-COV these residues all belong to  $L_3$ . This indicates that  $L_3$  in nCOV-2019 has evolved from SARS-COV to adopt new a secondary structure, which causes strong correlation and make the loop act as a recognition region for binding. The correlation in  $L_3$  is shown as blue rectangle in Figure 8. Some of the mutations disrupted the patterns of correlation and anti-correlation in nCOV-2019-wt. Mutation N487A showed a stronger correlation in  $L_3$  and  $\beta_6$  strand than the wild-type RBD. In mutation E484A, *correlation I* is stronger than the nCOV-2019-wt. DCCM for other mutants are shown in Figure S5. It is worth mentioning that mutation F486A disrupts the DCCM of nCOV-2019 by introducing strong correlations in the core region of RBD as well as the extended loop region. Residue F486A resides in  $L_3$  and plays a crucial role in stabilizing the recognition loop by making a  $\pi$ -stacking interaction with residue Y83A on ACE2.



**Figure 8.** Dynamic cross correlation maps for nCOV-2019, SARS-COV and mutants with residue numbers of the RBD domain. Red box shows the correlation between  $\alpha_5$ ,  $L_4$  and  $\beta_5$  regions and  $\alpha_4$ ,  $L_1$  and  $\beta_5$ . Blue box shows the correlation in  $L_3$  and  $\beta_6$  regions.

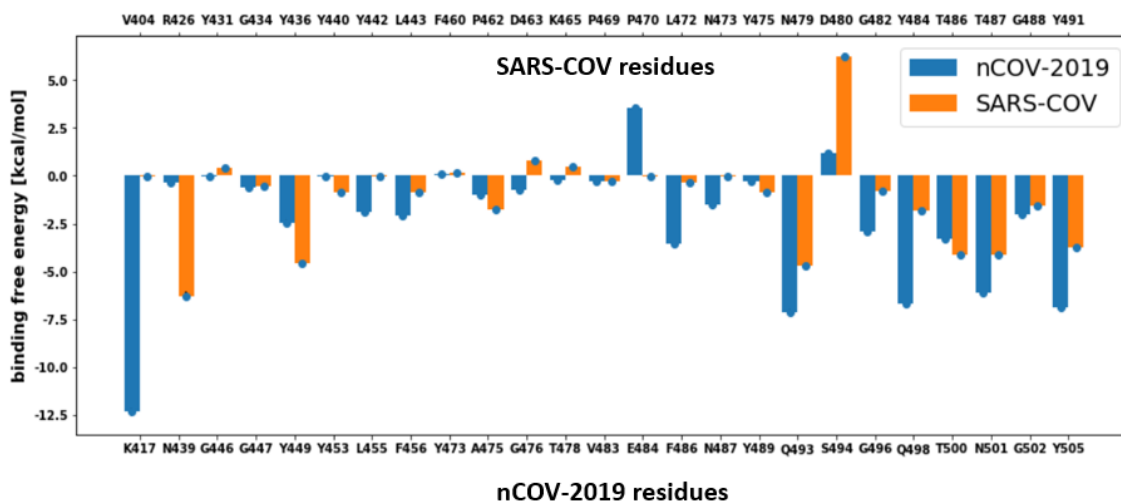
### Binding free energies

The binding energetics between ACE2 and the RBD of SARS-COV, nCOV-2019 and all its mutant complexes were investigated by the MMPBSA method.<sup>44</sup> The binding energy was partitioned into its individual components including: electrostatic, van der Waals, polar solvation and solvent accessible surface area (SASA) to identify important factors affecting the interface of RBD and ACE2 in all complexes. nCOV-2019 has a total binding energy of  $-50.22 \pm 1.93 \frac{kcal}{mol}$ , whereas SARS-COV has a much lower binding affinity  $-18.79 \pm 1.53 \frac{kcal}{mol}$ . Decomposition of binding energy to its components show that the most striking difference between nCOV-2019 and SARS-COV is the electrostatic contribution which is  $-746.69 \pm 2.66 \frac{kcal}{mol}$  for nCOV-2019 and  $-600.14 \pm 7.65 \frac{kcal}{mol}$  for SARS-COV. This high electrostatic contribution is compensated by a large polar solvation free energy which is  $797.30 \pm 3.12 \frac{kcal}{mol}$



for nCOV-2019 and  $659.61 \pm 8.98 \frac{kcal}{mol}$  for SARS-COV. nCOV-2019 also possess a higher van der Waals (vdw) contribution ( $-89.93 \pm 0.46 \frac{kcal}{mol}$ ) than SARS-COV ( $-70.07 \pm 1.22 \frac{kcal}{mol}$ ). Furthermore, the SASA contribution to binding for SARS-COV was  $-8.30 \pm 0.15 \frac{kcal}{mol}$  and  $-10.58 \frac{kcal}{mol}$  for nCOV-2019.

The binding free energy for nCOV-2019 and SARS-COV was decomposed into a per-residue based binding affinity to find the residues that contribute strongly to the binding and complex formation (Figure 9). Most of the investigated residues in the RBM of nCOV-2019 had a favorable contribution to total binding energy. Residues Q498, Y505, N501, Q493 and K417 contributed more than  $-5 \frac{kcal}{mol}$  and are crucial for binding. A few residues such as E484 and S494 contributed positively (unfavorable) to the total binding energy. Among all the interface residues K417 had the highest contribution to the total binding energy with  $-12.34 \pm 0.23 \frac{kcal}{mol}$ . The corresponding residue in SARS-COV is Val404 only had a  $-0.02 \pm 0.01 \frac{kcal}{mol}$  contribution, which points to the importance of this residue for nCOV-2019 binding to ACE2. Residue Q498 contributed to total binding by  $-6.72 \pm 0.18 \frac{kcal}{mol}$  and its corresponding residue in SARS-COV is a Y484 that contributed to total binding by  $-1.83 \pm 0.06 \frac{kcal}{mol}$ . Other important residues Y505 and N501 have more negative contribution to total binding energy than their counterparts in SARS-COV residues Y491 and T487 respectively. Residue D480 in SARS-COV contributed positively to binding energy by  $6.2 \pm 0.15 \frac{kcal}{mol}$  and the corresponding residue in nCOV-2019 which is a S494 residue lowered this positive contribution to only  $1.17 \pm 0.06 \frac{kcal}{mol}$ . In SARS-COV residue R426 had a high contribution to total binding energy ( $-6.27 \pm 0.22 \frac{kcal}{mol}$ ) although the corresponding residue in nCOV-2019 is N439 with a contribution of  $-0.32 \pm 0.02 \frac{kcal}{mol}$ . Mutation N439K which is the most observed among all the mutations of RBD can recover part of interaction at this position. These important mutations on RBM of nCOV-2019 from SARS-COV caused RBD of nCOV-2019 to bind ACE2 with much stronger (about  $30 \frac{kcal}{mol}$ ) affinity.



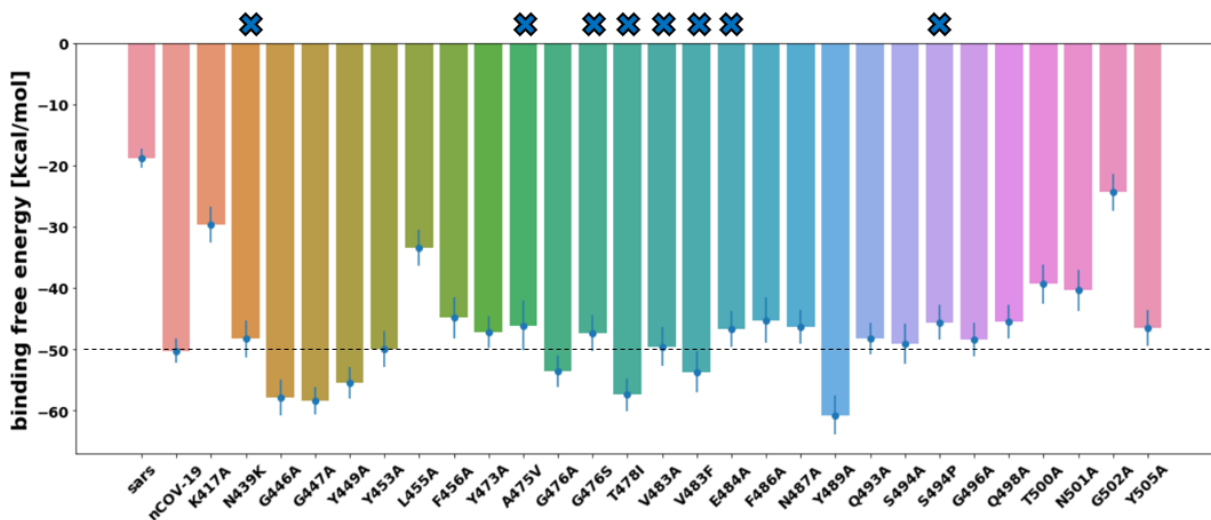
**Figure 9.** Binding energy decomposition per residue for the RBM of nCOV-2019 and SARS-COV.

Binding free energy decomposition to its individual components for all mutants is represented in Table S2. In all complexes, a large positive polar solvation free energy disfavors the binding and complex formation, which is compensated by a large negative electrostatic free energy of binding. All variants had similar solvent accessible surface area energies which is between  $-10$  and  $-11 \frac{kcal}{mol}$  for all complexes. The vdw free energy of binding ranged from  $-84.68 \pm 0.68 \frac{kcal}{mol}$  for mutant Q493A to  $-102.23 \pm 0.72 \frac{kcal}{mol}$  for N487A. Mutant K417A had the lowest electrostatic contribution to binding  $-415.67 \pm 5.07 \frac{kcal}{mol}$  and mutants N439K and E484A had the highest electrostatic binding contribution of  $-989.80 \pm 5.6 \frac{kcal}{mol}$  and  $-941.20 \pm 3.95 \frac{kcal}{mol}$  respectively.

Most alanine substitutions exhibited similar or lower total binding energy to nCOV-2019, however a few mutants had higher binding affinity than the wild type. Mutant Y489A had a total binding energy of  $-60.73 \pm 3.14 \frac{kcal}{mol}$  which was about  $10 \frac{kcal}{mol}$  higher than wild type binding energy. Mutants G446A, G447A, Y449A and T478I also demonstrated higher total binding energy than nCOV-2019 with respectively  $-57.79 \pm 2.92$ ,  $-58.37 \pm 2.32$ ,  $-55.38 \pm 2.58$  and

$-57.51 \pm 2.71 \frac{kcal}{mol}$  total binding energies. Other alanine substitutions had similar or lower total binding energy than nCOV-2019. Mutant G502A has the lowest binding affinity among all the mutants with a binding energy of  $-24.31 \pm 2.98 \frac{kcal}{mol}$ . Mutant systems K417A, L455A, T500A and N501A are the other mutant with total binding energies  $-29.56 \pm 2.95 \frac{kcal}{mol}$  and  $-33.47 \pm 2.93 \frac{kcal}{mol}$ ,  $-39.27 \pm 3.18 \frac{kcal}{mol}$  and  $-40.36 \pm 3.30 \frac{kcal}{mol}$  respectively which are significantly lower (less negative) than the wild type complex. The electrostatic component of binding energy contributes the most to this low binding energies for these mutants. The contribution of RBM residues to binding with ACE2 for nCOV-2019 were mapped to the RBD structure and is shown in Figure 11B.

Natural mutants exhibited similar or higher binding affinities compared to wild-type nCOV-2019. Importantly, mutation T478I which is one of the most occurred mutations in England based on GISAID database<sup>25</sup> increases the binding affinity to  $-57.41 \pm 2.71 \frac{kcal}{mol}$  which is about  $7 \frac{kcal}{mol}$  higher than wild-type binding energy. N439K demonstrated a high electrostatic energy which is compensated by large polar solvation energy and this mutant has a total binding energy of  $-48.27 \pm 3.07 \frac{kcal}{mol}$ . S494P showed a total binding energy of  $-45.58 \pm 2.83 \frac{kcal}{mol}$  which was lower than the wild type binding energy. Other natural mutants showed binding energies similar to wild-type.



**Figure 10.** Total free binding energy of SARS-COV, nCOV-2019 and mutants. Natural mutants are shown with X at the bar base.

### Hydrogen bond, Salt-bridge and Hydrophobic contact analysis:

Important hydrogen bonds (H-bonds) and salt bridges between nCOV-2019 RBD or SARS-COV RBD and ACE2 for the last 400ns of trajectory are shown in Table 1. nCOV-2019 RBD makes 10 H-bonds/1 salt bridge with ACE2 whereas SARS-COV makes only 5 H-bonds/1 salt bridge with ACE2 with more than 30% persistence.

#	nCOV-2019	ACE2	% occupancy	SARS-COV	ACE2	% occupancy
1	G502	K353	89	Y436	D38	96
2	Q493	E35	83	<b>R426</b>	<b>E329</b>	<b>87</b>
3	N487	Y83	80	T486	D355	83
4	Q498	D38	73	G488	K353	80
<b>5</b>	<b>K417</b>	<b>D30</b>	<b>55</b>	N479	K31	52
6	T500	D355	53	Y440	H34	47
7	Y505	E37	52			
8	Q498	K353	49			
9	Y449	D38	45			
10	G496	K353	37			
11	Q493	K31	32			

**Table 1.** H-bonds and salt-bridges between nCOV-2019 and ACE2 and SARS-COV and ACE2 that persist for >30%. Salt bridge is shown as bold.

The evolution of the coronavirus from SARS-COV to nCOV-2019 has reshaped the interfacial hydrogen bonds with ACE2. G502 in nCOV-2019 has the most persistent H-bond with residue K353 on ACE2. This residue was G488 in SARS-COV, which also makes H-bond with K353 on ACE2. Q493 in nCOV-2019 makes H-bond with E35 and a less persistent H-bond with K31 on ACE2. This residue was an N479 in SARS-COV, which only makes one H-bond with K31 on ACE2 with 52% occupancy. An important mutation from SARS-COV to nCOV-2019 is residue Q498, which was Y484 in SARS-COV. Q498 makes two H-bonds with residues D38 and K353 on ACE2, whereas Y484 in SARS-COV does not make any H-bonds. Importantly a salt bridge between K417 and D30 in nCOV-2019/ACE2 complex contributes to total binding

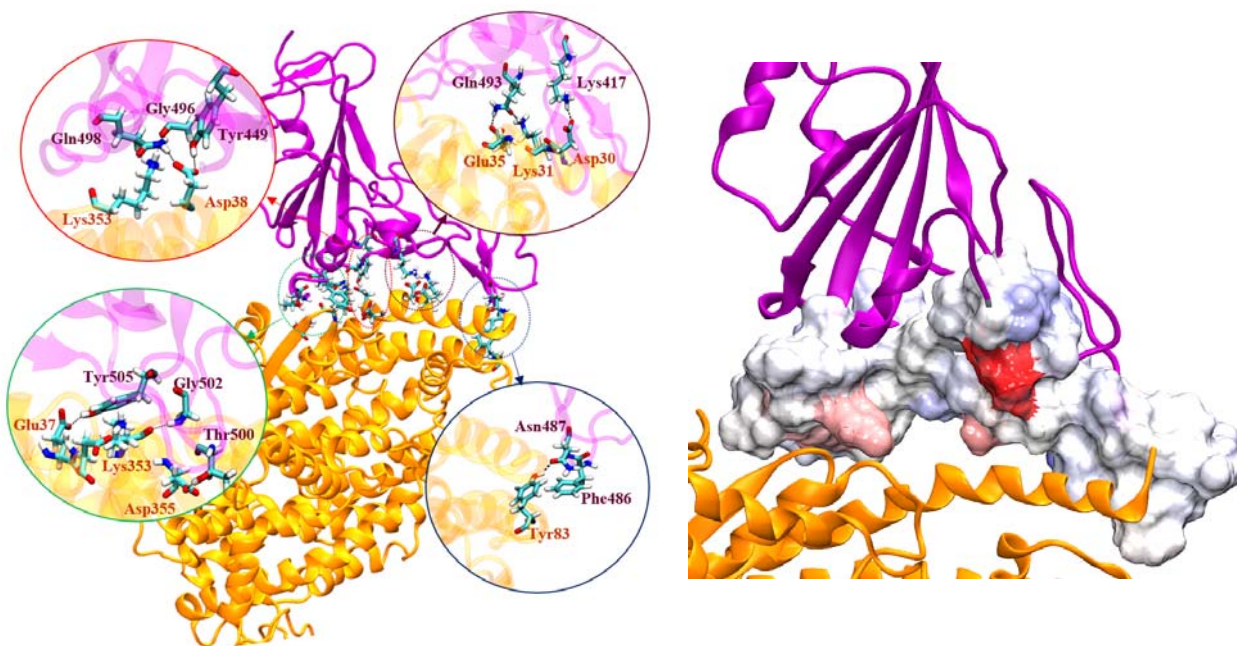
energy by  $-12.34 \pm 0.23 \frac{\text{kcal}}{\text{mol}}$ . This residue is V404 in SARS-COV which is not able to make any salt-bridge and does not make H-bond with ACE2. Wang et al.<sup>47</sup> used a free energy perturbation (FEP) approach and showed that mutation V404 to K417 increases the binding affinity of nCOV-2019 RBD to ACE2 by  $-2.2 \pm 0.9 \frac{\text{kcal}}{\text{mol}}$ . A salt bridge between R426 on RBD and E329 on ACE2 stabilizes the complex in SARS-COV/ACE2. This residue is N439 in nCOV-2019 and is unable to make salt-bridge with ACE2 residue E329. One of the most observed mutations in nCOV-2019 according to GISAID database<sup>25</sup> is N439K which recovers some of the electrostatic interactions with ACE2 at this position.

The evolution from SARS-COV to nCOV-2019 also has destabilized some specific ACE2/RBD interactions. An important H-bond in SARS-COV is between Y436 on RBD and D38 on ACE2. This residue is Y449 in nCOV-2019 and makes H-bond with D38 on ACE2 but with a lower occupancy than SARS-COV. The unchanged T486 in SARS-COV corresponds to T500 in nCOV-2019, both of which make consistent H-bonds with ACE2 residue D355.

Hydrophobic interactions also play an important role in stabilizing RBD/ACE2 complex in nCOV-2019. An important interaction between nCOV-2019 RBD and ACE2 is the  $\pi$ -stacking interaction between F486 (RBD) and Y83 (ACE2). This interaction helps in stabilizing  $L_3$  in nCOV-2019 compared to SARS-COV where this residue is L472. It was observed by Wang et al. that mutation Leu472 to Phe486 in nCOV-2019 results in a net change in binding free energy of  $-1.2 \pm 0.2 \frac{\text{kcal}}{\text{mol}}$ . Other interfacial residues in nCOV-2019 RBD that participate in hydrophobic interaction with ACE2 are L455, F456, Y473, A475 and Y489. It is interesting to note that all these residues except Y489 have mutated from SARS-COV. Spinello and co-workers<sup>48</sup> performed long-timescale ( $1\mu\text{s}$ ) simulation of nCOV-2019/ACE2 and SARS-COV ACE2 and found that  $L_3$  in nCOV-2019 is more stable due to presence of the  $\beta_6$  strand and existence of two H-bonds in  $L_3$  (H-bonds G485-C488 and Q474-G476). Importantly, an amino acid insertion in  $L_3$  makes this loop longer than  $L_3$  in SARS-COV and enable it to act like a recognition loop and make more persistent H-bonds with ACE2. L455 in nCOV-2019 RBD is important for hydrophobic interaction with ACE2 and mutation L455A lowers the vdW contribution of binding to  $-84.96 \pm 0.56 \frac{\text{kcal}}{\text{mol}}$  which is about  $5 \frac{\text{kcal}}{\text{mol}}$  smaller than the wild-type complex. The H-bonds between RBD of nCOV-2019 and SARS-COV are shown in Figure 11A.

The structural details discussed here are in agreement with other structural studies of nCOV-2019 RBD/ACE2 complex.<sup>4,49,50</sup>

H-bond analysis was also performed for the mutant systems and the results for H-bonds with more than 40% consistency are in Table S3. Few of the alanine substitutions increase the number of interfacial H-bonds between nCOV-2019 RBD and ACE2. Interestingly, the ala-substitution at Y489A increased the number of H-bonds in the wild-type complex. Mutation in some of the residues having consistent H-bonds in the wild type complex such as Q498A, Q493A and T500A stunningly increased the persistence of H-bonds in the neighboring residues and maintaining the number of H-bonds in the wild-type complex. This indicates that the RBM of nCOV-2019 can reshape the network of H-bonds and strengthen the H-bonds upon mutation in these locations. However, few mutations also decrease the number of H-bonds from the wild-type complex. Alanine substitution at residue G502 has a significant effect on the network of H-bonds between nCOV-2019 and SARS-COV. This residue locates at the end of  $L_4$  loop near two other important residues Q498 and T500. This mutation breaks the H-bonds at these residues. Mutation K417 decreases the number of H-bonds to only 5 where the H-bond at residue Q498 is broken. This indicates the delicacy of H-bond from residue Q498 which can easily be broken upon ala-substitution at other residues. Furthermore, mutation N487 also decreases the number of H-bonds by breaking the H-bond at Q498.



**Figure 11)** A) H bonds between RBD of A) nCOV-2019 and B) SARS-COV. B) mapping contribution of interface residues to structure in RBD of nCOV-2019. The RBD domain is purple and the ACE2 is yellow. The RBD in contact with AC2 is rendered in a surface format with more red being a favorable contribution to binding (more negative) and blue unfavorable contribution (positive).

## Discussion

In this work, we performed MD simulations to unveil the detailed molecular mechanism for receptor binding of nCOV-2019 and comparing it with SARS-COV. The role of key residues at the interface of nCOV-2019 with ACE2 were investigated by computational ala-scanning. A rigorous 500ns MD simulation was performed for nCOV-2019 and SARS-COV as well as 300ns MD simulation on each mutant. These simulations aiding in our understanding in the dynamic role of RBD/ACE2 interface residues and estimating the binding free energy of these variants, which shed light on crucial residues for the RBD/ACE2 complex stability. Moreover, numerous mutations have been identified in the RBD of different nCOV-2019 strains from all over the world not known to be critical for infection.<sup>51</sup> The role of these mutations on the stability of RBD/ACE2 complex were investigated to shed light on their role in the viral infection of coronavirus.

Changes in RBD structure of nCOV-2019, SARS-COV and mutants from their crystal structure were analyzed by RMSD and RMSF. nCOV-2019 showed a stable structure with a RMSD=1.5 Å, whereas SARS-COV had a larger RMSD value between 3-4 Å during the simulation. Most mutations of nCOV-2019 maintained similar stability to the wild-type. However, in few nCOV-2019 mutations resulted in larger deviations (>2 Å), i.e., Y489A, F456A, Y505A, N487A, K417A, Y473A, Y449A.. We further investigated the structure of the extended loop domain (Figure 1B) and discovered that nCOV-2019 is stable with RMSD of less than 1 Å, whereas the extended loop in SARS-COV shows an RMSD of about 3Å during simulation (figure S2). Some mutants showed high RMSD in this region. Alanine-substitution at residue N487 increased the extended loop RMSD to 2.5 Å. Other mutations that increased the extended loop RMSD (>2 Å)include Y449A, G477A and E484A. The dynamic behavior of RBD

was further investigated by analyzing the RMSF of all systems. As shown in Figure 4, nCOV-2019 shows less fluctuation in  $L_3$  than SARS-COV. This is due to the presence of a 4-residue motif (GQTQ) in nCOV-2019  $L_3$ , which forces the loop to adopt a compact structure by making 2 H-bonds (G485-C488 and Q474-G476) and thereby reducing the fluctuations in the loop. Residues F486 and N487 play major roles in stabilizing the recognition loop by making  $\pi$ -stacking and H-bond interactions with residue Y83 on ACE2. Alanine substitution at N487 introduced large RMSF to  $L_1$ . Mutation L472 to F486 was shown to favor binding by  $-1.2 \pm 0.2 \frac{kcal}{mol}$  using FEP.<sup>47</sup> In addition, this mutation was shown to be among the five mutations that produce a super affinity ACE2 binder based on SARS-COV RBD.<sup>6</sup> Alanine mutations at residues Y449, G447 and E484 increased the motion in  $L_3$  characterized by large RMSF in this region.

Using principal component analysis, the free energy landscape for nCOV-2019 and SARS-COV demonstrated that the former occupies only one low energy state whereas the latter forms two distinct low energy basins separated by a metastable state with a barrier of about 6-7.5  $\frac{kcal}{mol}$ . This confirms that the level of binding for the RBD domain is weaker in SARS-COV due to the presence of two basins. Similarly, alanine-substitution for a few residues ( Y489A, N487A, Y449A and E484A) caused the free energy landscape to degenerate into separate multiple low energy regions. Dominant motions in the RBD are visualized in Figure 7 using the first eigenvector of the PCA. nCOV-2019 and SARS-COV did not show any strong motion in the extended loop region. Porcupine plots of alanine-mutants demonstrated that mutant N487A shows large motion in  $L_1$  region and Y449A, G447A and E484A showed large motions in  $L_3$ .

To better characterize the functional motions of RBD, DCCM for all systems are constructed and demonstrated in Figure 8. nCOV-2019 showed large correlation between the  $\alpha_4$ - $L_1$ - $\beta_5$  and  $\alpha_5$ - $L_4$ - $\beta_5$  region. This correlation was stronger in SARS-COV and few mutants such as Y449A, G447A and E484A. Another important correlation in nCOV-2019 is inside  $L_3$  and  $\beta_6$ . This correlation is stronger in nCOV-2019 than SARS-COV due to the presence of  $\beta_6$  which makes the loop to adopt correlated motions. Few mutants impact the correlation in this region such as N487A. DCCM maps for other mutants are presented in Figure S5. Interestingly, mutant F486A which resides in  $L_3$  and participates in binding by  $\pi$ -stacking interaction with Y83



on ACE2, disrupts the DCCM of wild-type nCOV-2019 and introduces strong correlation in the extended loop region as well as core structure of RBD.

The details of hydrogen bond and salt-bridge pattern in nCOV-2019 and SARS-COV to ACE2 (Table 1) are key to the virus attachment to the host. nCOV-2019 residues participate in 10 H-bonds/1 salt bridge with ACE2, whereas SARS-COV only has 5 H-bonds/1 salt bridge with ACE2. This significantly contributes to the  $31 \frac{kcal}{mol}$  difference in total binding free energy of nCOV-2019 and SARS-COV. The binding energies calculated here for nCOV-2019 and SARS-COV ( $-50.22 \pm 1.93$  and  $-18.79 \pm 1.53 \frac{kcal}{mol}$ , respectively) are in good agreement with the binding energies calculated using Generalized Born method (GB) by Spinello et al.<sup>52</sup> Moreover, the patterns of H-bonds between nCOV-2019 and ACE2 was also already characterized by other groups<sup>47,52</sup>, which agrees with our work. An important H-bond between nCOV-2019 and ACE2 is between G502 on RBD and K353 of ACE2. G502 is in  $L_4$  region, which is populated by 5 H-bonds between RBD and ACE2. The contribution of this residue to the total binding energy is  $-2.03 \pm 0.04 \frac{kcal}{mol}$  and the Ala-substitution at G502 has the highest effect on the binding energy among all the residues by reducing the total binding energy to  $-24.31 \pm 2.98 \frac{kcal}{mol}$  which is the lowest among all mutants. This mutation breaks the other H-bonds in  $L_4$  such as H-bonds from residues Q498 and T500. This residue is preserved and corresponds to residue G488 in SARS-COV, which also makes a persistent H-bond with residue K353 on ACE2. Residue Q493 in nCOV-2019 participates in binding ACE2 by making two H-bonds with residues E35 and K31 on ACE2. Q493 corresponds to residue N479 in SARS-COV, which only makes one H-bond with residue Lys31 on ACE2 with low consistency (52%). This caused Q493 to have more contribution to total binding ( $-7.11 \pm 0.16 \frac{kcal}{mol}$ ) than its counterpart N479 ( $-4.73 \pm 0.10 \frac{kcal}{mol}$ ). However, alanine substitution at Q493 did not have a large effect on total binding energy and this mutant had a total binding energy of  $-48.19 \pm 2.64 \frac{kcal}{mol}$ . This mutation increases the persistence of H-bonds in neighboring residues and maintains the number of persistent H-bonds in the wild-type complex. Residues Q498 and T500 in nCOV-2019 are crucial for binding by making H-bonds with ACE2 residues D38, D355 and K353. Residue Q498 corresponds to residue Y484 in SARS-COV which does not make any H-bond in SARS-COV/ACE2 complex. Q498 contributes to binding by  $-6.72 \pm 0.18 \frac{kcal}{mol}$  which is more than the contribution of Y484

in SARS-COV ( $-1.83 \pm 0.06 \frac{kcal}{mol}$ ). Ala-substitution at Q498 did not show large impact on total binding energy. Residue T500 is conserved and corresponds to residue T486 which also makes a H-bond with Asp355 on ACE2. Mutation of T500 to Alanine lowers the binding energy to  $-39.27 \pm 3.18 \frac{kcal}{mol}$ , which is about  $10 \frac{kcal}{mol}$  lower than the wild-type complex. Residue N487 in nCOV-2019 locates in  $L_3$  and plays a crucial role in stabilizing the recognition loop by making H-bond with Y83 on ACE2. This residue contributes to total binding energy of nCOV-2019 by  $-1.52 \pm 0.06 \frac{kcal}{mol}$ , whereas its corresponding residue in SARS-COV does not show any contribution to binding energy ( $-0.02 \pm 0.05 \frac{kcal}{mol}$ ). This demonstrates that  $L_3$  in SARS-COV has evolved to be an important recognition loop in nCOV-2019, which participates in binding with ACE2. Residue K417 in nCOV-2019 has the most contribution to total binding energy ( $-12.34 \pm 0.23 \frac{kcal}{mol}$ ) by making a salt-bridge with residue D30 on ACE2. This residue is crucial for binding of RBD and ACE2 and alanine substitution lowers the total binding energy to  $-29.56 \pm 2.95 \frac{kcal}{mol}$ . This salt-bridge is found to be important for stability of the crystal structure of RBD/ACE2 complex in nCOV-2019.<sup>50</sup> This residue is Val404 in SARS-COV which does not participate in binding ACE2. Another important residue in nCOV-2019 is L455 which contributes to binding by  $-1.86 \pm 0.03 \frac{kcal}{mol}$ . This residue is important for hydrophobic interaction with ACE2 and mutating it to alanine lowers the total binding energy to  $-33.47 \pm 2.93 \frac{kcal}{mol}$ . The vdw energy for this mutant is about  $5 \frac{kcal}{mol}$  lower than the wild-type complex, which is part of the reason for lowering the total binding energy of this mutant.

Total binding energy calculation of all the variants showed that mutation Y489A has the strongest binding energy among all systems which is about  $10 \frac{kcal}{mol}$  stronger than that of nCOV-2019 complex. This residue is located in  $\beta_6$ , which is part of the recognition loop. This mutation causes the extended loop to move closer to ACE2 interface and all the H-bonds in the wild-type complex are more consistent after alanine substitution at this location. Per-residue binding energy decomposition showed that this residue participates in binding only by  $-0.27 \pm 0.04 \frac{kcal}{mol}$ . A high electrostatic interaction energy of  $-786.30 \pm 4.74 \frac{kcal}{mol}$  is the reason for higher binding energy of mutant Y489A than wild-type complex. It is interesting to note that among the 5 residues L455, F456, Y473, A475 and Y489 that make hydrophobic interactions with ACE2,

Y489 is the only residue that is conserved from SARS-COV. This shows the capacity of the virus to evolve even further in the RBD to bind the receptor ACE2 with stronger affinity. Other mutations that increase the binding energy are G446A, G447A, Y449A and T478I. Residues G446, G447 and Y449 all reside in  $L_1$  and mutation to alanine can make  $L_1$  take a more rigid form as shown in RMSF plot in Figure S3.

Important mutations found in naturally-occurring nCOV-2019 appear to influence to some extent the binding to ACE2. Mutation T478I which is one of the highly seen mutations in England according to GISAID database, increases the binding energy of nCOV-2019 to ACE2 by about  $7 \frac{kcal}{mol}$ . Mutation N439K has the highest occurrence among all strains of coronavirus in the GISAID database which demonstrated the highest electrostatic interaction among all studied systems. This residue corresponds to R426 in SARS-COV which exerts a salt-bridge interaction with E329 on ACE2. Mutation N439K recovers some of this ACE2 interaction. Mutant E484A which is also one of the observed mutations in Spain based on GISAID database, demonstrates a high electrostatic interaction with ACE2. E484 contributes to binding by  $3.56 \pm 0.15 \frac{kcal}{mol}$  whereas the corresponding residue in SARS-COV, P469 contributes to binding of SARS-COV to ACE2 by  $-0.27 \pm 0.01 \frac{kcal}{mol}$ . This residue is close to D30 on ACE2 and have electrostatic repulsion with this residue. High electrostatic energy in mutations N439K and E484A are compensated by a large polar solvation energy and these mutants show total binding energies similar to wild-type complex. This could be due to the use of MMPBSA approach for calculation of polar solvation and further studies are needed to study the effect of these mutations on viral infectivity of coronavirus.

Additional sequence differences between nCOV-2019 and SARS-COV influence RBD/ACE2 binding. Residue D480 in SARS-COV contributes negatively to total binding energy ( $6.25 \pm 0.14 \frac{kcal}{mol}$ ) and mutation of this residue to S494 in nCOV-2019 lowers this negative contribution to  $1.17 \pm 0.06 \frac{kcal}{mol}$ . D460 in SARS-COV is located in a region of high negative charge from residues E35, E37 and D38 on ACE2. Electrostatic repulsion between D480 on SARS-COV and the acidic residues on ACE2 is the reason for highly negative contribution of this residue to binding of SARS-COV to ACE2. Mutation to S494 in this location removes this highly negative contribution.

To our knowledge this is first detailed molecular simulation study on the effect of mutations on binding of nCOV-2019 to ACE2. Previous computational studies have found that nCOV-2019 binds to ACE2 with a total binding affinity which was about  $30 \frac{kcal}{mol}$  stronger than SARS-COV and is in fair agreement with the results here.<sup>52</sup> The critical role of interface residues and residues are computationally investigated here and in other articles and the results of all the studies indicate the importance of these residues for the stability of the complex and finding hot-spot residues for the interaction with receptor ACE2.<sup>47,52-54</sup> It is interesting to note the role of  $L_3$  in the stability of the RBD/ACE2 complex. The amino acid insertions in  $L_3$  for nCOV-2019 has converted as unessential part of RBD in SARS to a functional domain of the RBD. This loop participates in binding ACE2 by making H-bond as well as  $\pi$ -stacking interaction with ACE2, which makes this region to act as a recognition loop. Previous studies on SARS-COV have shown there is a correlation between higher binding affinity to receptor and higher infection rate by coronavirus.<sup>55-58</sup> High binding affinity for some mutants such as T478I could be the reason for higher human-to-human transmission rate in regions where these mutations are found. It is also alerting that mutations at other residues such as G446A, G449A, Y449A and Y489A increase the binding affinity considerably and should be monitored. Mutations of nCOV-2019 RBD that do not change the binding affinity and complex stability, could have implications for antibody design purposes since they could act antibody escape mutants. Escape from monoclonal antibodies are observed for mutations of SARS-COV in 2002 and these mutations should be considered for any antibody design endeavors against consider these escape mutations.

## Conclusion

In conclusion, this study unraveled key molecular traits underlying the higher affinity of nCOV-2019 for ACE2 compared to SARS-COV and unveiled critical residues for the interaction by *in-silico* alanine scanning mutations and binding free energy calculations. The higher affinity of nCOV-2019 to binding with ACE2 correlates with higher human-to-human transmissibility of nCOV-2019 compared to SARS-COV. Ala-scanning mutagenesis of the interface residues of nCOV-2019 RBM has shed light on the crucial interface residues and helped obtain an atomic-

level understanding of the interaction between coronavirus and the receptor ACE2 on the host cell. MD simulations on RBD mutations found in strains of nCOV-2019 from different countries aiding in the understanding of how these mutations can play and importance to in viral infection with ACE2 attachment. In addition to previously reported residues, it was found that residue F486 locating in  $L_3$  plays a crucial role in dynamic stability of the complex by a  $\pi$ -stacking interaction with ACE2. Per-residue free energy decomposition pinpoints the critical role of residues K417, Y505, Q498, Q493 in binding ACE2. Alanine scanning of interface residues in nCOV-2019 RBD showed that alanine substitution at some residues such as G502, K417 and L455 can significantly decrease the binding affinity of the complex. Moreover, mutation T478I, which is one of the most probable mutations in RBD of nCOV-2019 is found to binding ACE2 with about  $7 \frac{kcal}{mol}$  higher affinity than wild-type. Other occurring mutations N439K and E484A are found to increase the electrostatic interaction of RBD with ACE2. It is also alerting that some of the alanine substitutions at residues G446, G447 and Y489 substantially increased the binding affinity that may lead to a strongly RBD attachment to ACE2 and influence the infection virulence. On the other hand, most mutations are found not to impact the binding affinity of RBD with ACE2 in nCOV-2019 which could have implications for vaccine design endeavors as these mutations could act as antibody escape mutants. Receptor recognition is the first line of attack for coronavirus and this study gives novel insights to key structural features of interface residues for advancement of effective therapeutic strategies to stop the coronavirus pandemic.

### **Dedication:**

The authors would like to dedicate this article to the doctors and nurses who sacrificed their time, health and even their lives to fight COVID-19, particularly those in Iran and the United States. JBK would also like to dedicate this work to family friend Joe Kaplan (Silver Spring, MD) who passed away due to COVID-19 on April 22, 2020.

### **Acknowledgement:**

This work was supported by National Heart, Lung and Blood institute (NHLBI) at the National Institute of Health (NIH). The authors acknowledge the biowulf high performance computation center for providing the time and resources for this project.

## References:

- (1) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382* (8), 727–733. <https://doi.org/10.1056/NEJMoa2001017>.
- (2) Zhou, P.; Yang, X. Lou; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L.; Chen, H. D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R. Di; Liu, M. Q.; Chen, Y.; Shen, X. R.; Wang, X.; Zheng, X. S.; Zhao, K.; Chen, Q. J.; Deng, F.; Liu, L. L.; Yan, B.; Zhan, F. X.; Wang, Y. Y.; Xiao, G. F.; Shi, Z. L. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2020**, *579* (7798), 270–273. <https://doi.org/10.1038/s41586-020-2012-7>.
- (3) Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu. Rev. Virol.* **2016**, *3* (1), 237–261. <https://doi.org/10.1146/annurev-virology-110615-042301>.
- (4) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K. Y.; Wang, Q.; Zhou, H.; Yan, J.; Qi, J. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **2020**, *181* (4), 894-904.e9. <https://doi.org/10.1016/j.cell.2020.03.045>.
- (5) Chan, P. K. S.; To, W. K.; Ng, K. C.; Lam, R. K. Y.; Ng, T. K.; Chan, R. C. W.; Wu, A.; Yu, W. C.; Lee, N.; Hui, D. S. C.; Lai, S. T.; Hon, E. K. L.; Li, C. K.; Sung, J. J. Y.; Tam, J. S. Laboratory Diagnosis of SARS. *Emerg. Infect. Dis.* **2004**, *10* (5), 825–831. <https://doi.org/10.3201/eid1005.030682>.
- (6) Wan, Y.; Shang, J.; Graham, R.; Baric, R. S.; Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: An Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **2020**, *94* (7), 1–9. <https://doi.org/10.1128/jvi.00127-20>.
- (7) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science* (80-. ). **2020**, *367* (6483), 1260–1263. <https://doi.org/10.1126/science.aax0902>.
- (8) Bosch, B. J.; van der Zee, R.; de Haan, C. A. M.; Rottier, P. J. M. The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the

- Fusion Core Complex. *J. Virol.* **2003**, 77 (16), 8801–8811.  
<https://doi.org/10.1128/jvi.77.16.8801-8811.2003>.
- (9) Batool, M.; Shah, M.; Patra, M. C.; Yesudhas, D.; Choi, S. Structural Insights into the Middle East Respiratory Syndrome Coronavirus 4a Protein and Its DsRNA Binding Mechanism. *Sci. Rep.* **2017**, 7 (1), 1–14. <https://doi.org/10.1038/s41598-017-11736-6>.
- (10) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Velesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, 1–12. <https://doi.org/10.1016/j.cell.2020.02.058>.
- (11) Tian, X.; Li, C.; Huang, A.; Xia, S.; Lu, S.; Shi, Z.; Lu, L.; Jiang, S.; Yang, Z.; Wu, Y.; Ying, T. Potent Binding of 2019 Novel Coronavirus Spike Protein by a SARS Coronavirus-Specific Human Monoclonal Antibody. *Emerg. Microbes Infect.* **2020**, 9 (1), 382–385. <https://doi.org/10.1080/22221751.2020.1729069>.
- (12) Li, F. Receptor Recognition Mechanisms of Coronaviruses: A Decade of Structural Studies. *J. Virol.* **2015**, 89 (4), 1954–1964. <https://doi.org/10.1128/jvi.02615-14>.
- (13) Li, W.; Zhang, C.; Sui, J.; Kuhn, J. H.; Moore, M. J.; Luo, S.; Wong, S. K.; Huang, I. C.; Xu, K.; Vasilieva, N.; Murakami, A.; He, Y.; Marasco, W. A.; Guan, Y.; Choe, H.; Farzan, M. Receptor and Viral Determinants of SARS-Coronavirus Adaptation to Human ACE2. *EMBO J.* **2005**, 24 (8), 1634–1643. <https://doi.org/10.1038/sj.emboj.7600640>.
- (14) Li, F. Structural Analysis of Major Species Barriers between Humans and Palm Civets for Severe Acute Respiratory Syndrome Coronavirus Infections. *J. Virol.* **2008**, 82 (14), 6984–6991. <https://doi.org/10.1128/JVI.00442-08>.
- (15) Li, F. Receptor Recognition and Cross-Species Infections of SARS Coronavirus. *Antiviral Research*. October 2013, pp 246–254. <https://doi.org/10.1016/j.antiviral.2013.08.014>.
- (16) Wu, K.; Peng, G.; Wilken, M.; Geraghty, R. J.; Li, F. Mechanisms of Host Receptor Adaptation by Severe Acute Respiratory Syndrome Coronavirus. *J. Biol. Chem.* **2012**, 287 (12), 8904–8911. <https://doi.org/10.1074/jbc.M111.325803>.
- (17) Li, F. Structural Analysis of Major Species Barriers between Humans and Palm Civets for Severe Acute Respiratory Syndrome Coronavirus Infections. *J. Virol.* **2008**, 82 (14), 6984–6991. <https://doi.org/10.1128/jvi.00442-08>.
- (18) Chakraborti, S.; Prabakaran, P.; Xiao, X.; Dimitrov, D. S. The SARS Coronavirus S Glycoprotein Receptor Binding Domain: Fine Mapping and Functional Characterization.



- Virol. J.* **2005**, 2, 1–10. <https://doi.org/10.1186/1743-422X-2-73>.
- (19) Greenough, T. C.; Babcock, G. J.; Roberts, A.; Hernandez, H. J.; Thomas, Jr., W. D.; Coccia, J. A.; Graziano, R. F.; Srinivasan, M.; Lowy, I.; Finberg, R. W.; Subbarao, K.; Vogel, L.; Somasundaran, M.; Luzuriaga, K.; Sullivan, J. L.; Ambrosino, D. M. Development and Characterization of a Severe Acute Respiratory Syndrome–Associated Coronavirus–Neutralizing Human Monoclonal Antibody That Provides Effective Immunoprophylaxis in Mice. *J. Infect. Dis.* **2005**, 191 (4), 507–514. <https://doi.org/10.1086/427242>.
- (20) Coughlin, M. M.; Prabhakar, B. S. Neutralizing Human Monoclonal Antibodies to Severe Acute Respiratory Syndrome Coronavirus: Target, Mechanism of Action, and Therapeutic Potential. *Rev. Med. Virol.* **2012**, 22 (1), 2–17. <https://doi.org/10.1002/rmv.706>.
- (21) Ter Meulen, J.; Van Den Brink, E. N.; Poon, L. L. M.; Marissen, W. E.; Leung, C. S. W.; Cox, F.; Cheung, C. Y.; Bakker, A. Q.; Bogaards, J. A.; Van Deventer, E.; Preiser, W.; Doerr, H. W.; Chow, V. T.; De Kruif, J.; Peiris, J. S. M.; Goudsmit, J. Human Monoclonal Antibody Combination against SARS Coronavirus: Synergy and Coverage of Escape Mutants. *PLoS Med.* **2006**, 3 (7), 1071–1079. <https://doi.org/10.1371/journal.pmed.0030237>.
- (22) Deming, D.; Sheahan, T.; Heise, M.; Yount, B.; Davis, N.; Sims, A.; Suthar, M.; Harkema, J.; Whitmore, A.; Pickles, R.; West, A.; Donaldson, E.; Curtis, K.; Johnston, R.; Baric, R. Vaccine Efficacy in Senescent Mice Challenged with Recombinant SARS-CoV Bearing Epidemic and Zoonotic Spike Variants. *PLoS Med.* **2006**, 3 (12), 2359–2375. <https://doi.org/10.1371/journal.pmed.0030525>.
- (23) Rockx, B.; Corti, D.; Donaldson, E.; Sheahan, T.; Stadler, K.; Lanzavecchia, A.; Baric, R. Structural Basis for Potent Cross-Neutralizing Human Monoclonal Antibody Protection against Lethal Human and Zoonotic Severe Acute Respiratory Syndrome Coronavirus Challenge. *J. Virol.* **2008**, 82 (7), 3220–3235. <https://doi.org/10.1128/jvi.02377-07>.
- (24) Rockx, B.; Donaldson, E.; Frieman, M.; Sheahan, T.; Corti, D.; Lanzavecchia, A.; Baric, R. S. Escape from Human Monoclonal Antibody Neutralization Affects In Vitro and In Vivo Fitness of Severe Acute Respiratory Syndrome Coronavirus. *J. Infect. Dis.* **2010**, 201 (6), 946–955. <https://doi.org/10.1086/651022>.
- (25) Elbe, S.; Buckland-Merrett, G. Data, Disease and Diplomacy: GISAID’s Innovative

- Contribution to Global Health. *Glob. Challenges* **2017**, *1* (1), 33–46.  
<https://doi.org/10.1002/gch2.1018>.
- (26) Structural Basis for Receptor Recognition by the Novel Coronavirus from Wuhan. **2020**.  
<https://doi.org/10.21203/RS.2.24749/V1>.
- (27) Rodrigues, C. H. M.; Myung, Y.; Pires, D. E. V.; Ascher, D. B. MCSM-PPI2: predicting the effects of mutations on protein-protein interactions  
<https://pubmed.ncbi.nlm.nih.gov/31114883/> (accessed Jun 27, 2020).  
<https://doi.org/10.1093/nar/gkz383>.
- (28) Chakraborti, S.; Prabakaran, P.; Xiao, X.; Dimitrov, D. S. The SARS Coronavirus S Glycoprotein Receptor Binding Domain: Fine Mapping and Functional Characterization. *Viol. J.* **2005**, *2* (1), 73. <https://doi.org/10.1186/1743-422X-2-73>.
- (29) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581* (7807), 215–220. <https://doi.org/10.1038/s41586-020-2180-5>.
- (30) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242. <https://doi.org/10.1093/nar/28.1.235>.
- (31) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.  
<https://doi.org/10.1016/j.softx.2015.06.001>.
- (32) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (8), 1950–1958.  
<https://doi.org/10.1002/prot.22711>.
- (33) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone  $\phi$ ,  $\psi$  and Side-Chain X1 and X2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257–3273. <https://doi.org/10.1021/ct300400x>.
- (34) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J.*

- Chem. Phys.* **2007**, *126* (1), 014101. <https://doi.org/10.1063/1.2408420>.
- (35) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *Journal of Computational Chemistry*. J Comput Chem December 2005, pp 1701–1718. <https://doi.org/10.1002/jcc.20291>.
- (36) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. <https://doi.org/10.1063/1.328693>.
- (37) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method ARTICLES YOU MAY BE INTERESTED IN. *J. Chem. Phys.* **1995**, *103*, 4613. <https://doi.org/10.1063/1.470648>.
- (38) Spoel, D. van der; Buuren, A. R. van; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Hess, B.; Feenstra, K. A.; Lindahl, E.; Drunen, R. van; Berendsen, H. J. C. GROMACS Reference Manual Contributions From. *GROMACS User Man.* **2005**, No. Version 3.1.1.
- (39) Brooks, B. R.; Janežič, D.; Karplus, M. Harmonic Analysis of Large Systems. I. Methodology. *J. Comput. Chem.* **1995**, *16* (12), 1522–1542. <https://doi.org/10.1002/jcc.540161209>.
- (40) David, C. C.; Jacobs, D. J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. *Methods Mol. Biol.* **2014**, *1084*, 193–226. [https://doi.org/10.1007/978-1-62703-658-0\\_11](https://doi.org/10.1007/978-1-62703-658-0_11).
- (41) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- (42) Brown, D. K.; Penkler, D. L.; Amamuddy, O. S.; Ross, C.; Atilgan, A. R.; Atilgan, C.; Bishop, Ö. T. MD-TASK: A software suite for analyzing molecular dynamics trajectories <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5860072/> (accessed Jun 27, 2020). <https://doi.org/10.1093/bioinformatics/btx349>.
- (43) Ren, J.; Yuan, X.; Li, J.; Lin, S.; Yang, B.; Chen, C.; Zhao, J.; Zheng, W.; Liao, H.; Yang, Z.; Qu, Z. Assessing the Performance of the G\_mmpbsa Tools to Simulate the Inhibition of Oseltamivir to Influenza Virus Neuraminidase by Molecular Mechanics Poisson–

- Boltzmann Surface Area Methods. *J. Chinese Chem. Soc.* **2020**, *67* (1), 46–53.  
<https://doi.org/10.1002/jccs.201900148>.
- (44) Wang, C.; Greene, D.; Xiao, L.; Qi, R.; Luo, R. Recent Developments and Applications of the MMPBSA Method. *Frontiers in Molecular Biosciences*. Frontiers Media S.A. January 10, 2018. <https://doi.org/10.3389/fmolb.2017.00087>.
- (45) Yang, T.; Wu, J. C.; Yan, C.; Wang, Y.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Virtual Screening Using Molecular Simulations. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (6), 1940–1951. <https://doi.org/10.1002/prot.23018>.
- (46) Guan, S. S.; Han, W. W.; Zhang, H.; Wang, S.; Shan, Y. M. Insight into the Interactive Residues between Two Domains of Human Somatic Angiotensin-Converting Enzyme and Angiotensin II by MM-PBSA Calculation and Steered Molecular Dynamics Simulation. *J. Biomol. Struct. Dyn.* **2016**, *34* (1), 15–28.  
<https://doi.org/10.1080/07391102.2015.1007167>.
- (47) Wang, Y.; Liu, M.; Gao, J. Enhanced Receptor Binding of SARS-CoV-2 through Networks of Hydrogen-Bonding and Hydrophobic Interactions. *Proc. Natl. Acad. Sci.* **2020**, *117* (25), 202008209. <https://doi.org/10.1073/pnas.2008209117>.
- (48) Spinello, A.; Saltalamacchia, A.; Magistrato, A. Is the Rigidity of SARS-CoV-2 Spike Receptor-Binding Motif the Hallmark for Its Enhanced Infectivity? Insights from All-Atom Simulations. *J. Phys. Chem. Lett.* **2020**, 4785–4790.  
<https://doi.org/10.1021/acs.jpcclett.0c01148>.
- (49) Tai, W.; He, L.; Zhang, X.; Pu, J.; Voronin, D.; Jiang, S.; Zhou, Y.; Du, L. Characterization of the Receptor-Binding Domain (RBD) of 2019 Novel Coronavirus: Implication for Development of RBD Protein as a Viral Attachment Inhibitor and Vaccine. *Cell. Mol. Immunol.* **2020**, No. March. <https://doi.org/10.1038/s41423-020-0400-4>.
- (50) Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2 | Elsevier Enhanced Reader  
<https://reader.elsevier.com/reader/sd/pii/S009286742030338X?token=4DAB72F5F0105372394079C969F6071C34C32021B4B193D5D826E098B07430AB93DD687CFF4B1DB20E853962D960A9A8> (accessed Jun 11, 2020).
- (51) Forster, P.; Forster, L.; Renfrew, C.; Forster, M. Phylogenetic Network Analysis of

- SARS-CoV-2 Genomes. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (17), 9241–9243.  
<https://doi.org/10.1073/pnas.2004999117>.
- (52) Spinello, A.; Saltalamacchia, A.; Magistrato, A. Is the Rigidity of SARS-CoV-2 Spike Receptor-Binding Motif the Hallmark for Its Enhanced Infectivity? An Answer from All-Atoms Simulations. **2020**. <https://doi.org/10.26434/CHEMRXIV.12091260.V2>.
- (53) Brielle, E. S.; Schneidman, D.; Linial, M. The SARS-CoV-2 Exerts a Distinctive Strategy for Interacting with the ACE2 Human Receptor. *bioRxiv* **2020**, 2020.03.10.986398.  
<https://doi.org/10.1101/2020.03.10.986398>.
- (54) Lupala, C. S.; Li, X.; Lei, J.; Chen, H.; Qi, J.; Liu, H.; Su, X. Computational Simulations Reveal the Binding Dynamics between Human ACE2 and the Receptor Binding Domain of SARS-CoV-2 Spike Protein. *bioRxiv* **2020**, 2020.03.24.005561.  
<https://doi.org/10.1101/2020.03.24.005561>.
- (55) Wan, Y.; Shang, J.; Graham, R.; Baric, R. S.; Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: An Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **2020**, *94* (7). <https://doi.org/10.1128/jvi.00127-20>.
- (56) Li, W.; Zhang, C.; Sui, J.; Kuhn, J. H.; Moore, M. J.; Luo, S.; Wong, S. K.; Huang, I. C.; Xu, K.; Vasilieva, N.; Murakami, A.; He, Y.; Marasco, W. A.; Guan, Y.; Choe, H.; Farzan, M. Receptor and Viral Determinants of SARS-Coronavirus Adaptation to Human ACE2. *EMBO J.* **2005**, *24* (8), 1634–1643. <https://doi.org/10.1038/sj.emboj.7600640>.
- (57) Li, W.; Greenough, T. C.; Moore, M. J.; Vasilieva, N.; Somasundaran, M.; Sullivan, J. L.; Farzan, M.; Choe, H. Efficient Replication of Severe Acute Respiratory Syndrome Coronavirus in Mouse Cells Is Limited by Murine Angiotensin-Converting Enzyme 2. *J. Virol.* **2004**, *78* (20), 11429–11433. <https://doi.org/10.1128/jvi.78.20.11429-11433.2004>.
- (58) Li, Y.; Wang, Y.; Yaping Qiu, #; Gong, Z.; Deng, L.; Pan, M.; Yang, H.; Xu, J.; Yang, L.; Li, J. SARS-CoV-2 Spike Glycoprotein Receptor Binding Domain Is Subject to Negative Selection with Predicted Positive Selection Mutations.  
<https://doi.org/10.1101/2020.05.04.077842>.