

Critical Sequence Hotspots for Binding of Novel Coronavirus to Angiotensin Converter Enzyme as Evaluated by Molecular Simulations

Mahdi Ghorbani, Bernard R. Brooks, and Jeffery B. Klauda*

Cite This: <https://dx.doi.org/10.1021/acs.jpcc.0c05994>

Read Online

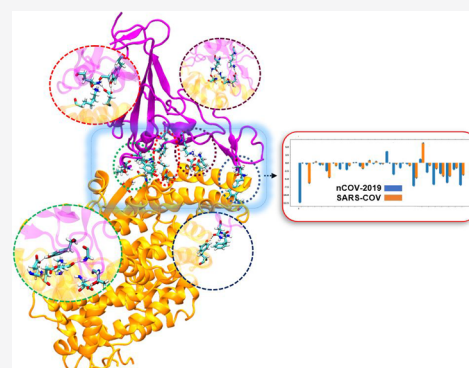
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

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 severe acute respiratory syndrome 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 the RBD/ACE2 complex. It is found that most of the naturally occurring mutations to the RBD either slightly strengthen or have the same binding affinity to ACE2 as the wild-type nCOV-2019. This means that the virus had sufficient binding affinity to its receptor at the beginning of the crisis. This also has implications for 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 even 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 the World Health Organization as of August 28th 2020, there has been about 25 million confirmed cases and approaching 900,000 deaths because of coronavirus in the world.^{1,2} 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 Middle East respiratory syndrome coronavirus (MERS-COV), severe acute respiratory syndrome coronavirus (SARS-COV), and bat-SARS-related coronaviruses.^{3,4} It is worth mentioning that SARS-COV, which was widespread in 2002 caused more than 8000 cases and about 800 deaths and MERS-COV in 2012 also spread in more than 25 countries, causing about 2500 cases and more than 850 deaths. (www.who.int/health-topics/coronavirus).⁵

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.^{3,6} 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.^{3,7} 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.^{7,8} 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 sequence.² The S1 subunit in the spike protein

Received: June 30, 2020

Revised: October 7, 2020

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 1.2Å, which shows the high structural similarity.^{4,8,9} 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 ACE2.⁷ Based on the sequence similarity between RBD of nCoV-2019 and SARS-COV and also the tight binding between the 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.^{3,6,7,10}

The spike protein and specifically the RBD 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.^{7,11} The core of nCoV-2019 RBD is a five-stranded antiparallel β -sheet with connected short α -helices and loops (Figure 1). The binding

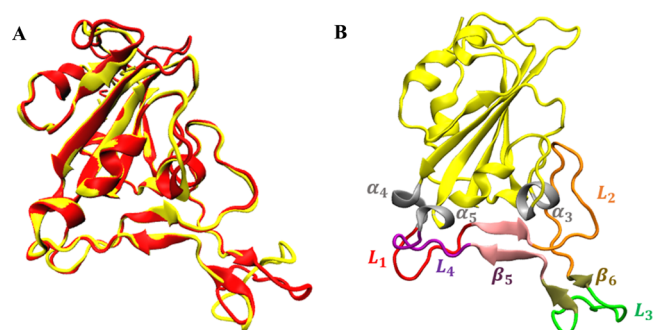


Figure 1. (A) Superposition of the RBD of SARS-COV (yellow) and nCoV-2019 (red). (B) Different regions in the binding domain of nCoV-2019 defining the extended loop (nonyellow).

interface of nCoV-2019 and SARS-COV with ACE2 is very similar with less than 1.3Å RMSD. An extended insertion inside the core containing short strands, α -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 is shown in Figure 1A. The binding interface in nCoV-2019 contains loops L_1 to L_4 and short β -strands β_5 and β_6 and a short helix α_5 . The location of RBM in

nCoV-2019 RBD as well as different helices, strands, and loops is shown in Figure 1B.

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.^{3,12} 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 are 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.^{3,13–15} 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 has been investigated by SPR.^{14,16} 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.^{14,16} 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.¹⁶ Further investigation of RBD residues in binding of SARS-COV and ACE2 was performed through ala-scanning mutagenesis, which resulted in identification of residues that reduce binding affinity to ACE2 upon mutation to alanine.¹⁷ RBM mutations have also been identified in MERS-COV, which affected their affinity to receptor (DPP4) on human cells.¹⁴ Multiple monoclonal antibodies have been developed for SARS since 2002 that neutralized the spike glycoprotein on the SARS-COV surface.^{18–22} 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 a robust treatment.²³

Full genome analysis of nCoV-2019 in different countries and the receptor binding surveillance have shown multiple mutations in the RBD of glycosylated spike. The GISAID database²⁴ (www.gisaid.org/) contains genomes on nCoV-2019 from researchers across the world since December 2019. The latest report by the GISAID database on June 2020 has

SARS-2002	424	NTRNIDATSTGNYNYKYRYLHRGKLRPFERDISNVPFSPDGKPTP-PALNCYWPLNDY	482
SARS-civet	424	NTRNIDATSTGNYNYKYRYLHRGKLRPFERDISNVPFSPDGKPTP-PALNCYWPLKDY	482
Bat RaTG13	446	NSKHIDAKEGGNENYLYRLEFRKANLKPFFERDISTEIYQAGSKPCNGQGLNCYYPYRYG	505
nCoV-2019	437	NSNNLDSKVGGNYNLYRLEFRKSNLKPFFERDISTEIYQAGSTPCNGVEGFNCYFPLOSYG	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 RBM in SARS-2002, SARS-civet, Bat RaTG13, and nCoV-2019. Mutations from SARS-2002 to nCoV-2019 are marked with blue. Important mutations in RBM are marked with yellow. Red color shows the three-residue motif in SARS and civet and four-residue motif in RaTG13 and nCoV-2019.

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

It is not known whether these mutations are linked to the severity of coronavirus in these regions. Starr and co-workers²⁵ performed a deep mutational scanning of nCoV-2019 RBD and used flow cytometry to measure the effect of single mutations on the expression of the folded protein as well as its binding affinity to ACE2. They showed that RBD is very tolerant to these mutations to maintain its expression level as well as binding affinity to ACE2 in most cases. According to their results, most natural mutations exert similar binding affinities to ACE2 as wild-type nCoV-2019. Furthermore, they showed that mutations at critical positions at the RBD-ACE2 interface at nCoV-2019 such as residues Q493 and Q498 do not reduce the binding affinity to ACE2 which shows the substantial plasticity of the interface.²⁵

Different groups have computationally studied the binding of nCoV-2019 RBD with ACE2.^{25–29} All these studies point to the higher binding affinity of nCoV-2019 RBD than SARS-COV RBD to ACE2. Interestingly, the role of water-mediated interactions has been pointed out to be a driving force which is shown to be similar for both SARS-COV and nCoV-2019 RBD.²⁷ Spinello and co-workers³⁰ studied the binding of nCoV-2019 and SARS-COV RBD to ACE2 and found that the former binds its receptor with 30 kcal/mol higher affinity than SARS-COV RBD. Gao et al.³¹ used free energy perturbation (FEP) and showed that most amino acid mutations at the RBM from SARS-COV to nCoV-2019 increase the affinity of RBD to ACE2. 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 the 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.¹ Bat coronavirus RaTG13 seems to be the closest relative of nCoV-2019 sharing about 93% sequence identity in the spike protein.⁶ 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) is shown in Figure 2.⁶ To investigate the roles

of critical mutations on the complex stability of nCoV-2019 with ACE2, mCSM-PPI2 webserver³² was used to find 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, and 21 different residues were identified to be in contact with ACE2 and were chosen for further MD simulation. On the other hand, mutations are also observed in the RBD domain from full genome analysis of different nCoV-2019 variants collected from different countries compiled in the GISAID database.²⁴ The mutations selected are listed in Table S1 along with their location in the RBD.

MD Simulations. The crystal structure of nCoV-2019 in the complex with hACE2 (pdb id:6M0J)¹⁷ as well as the SARS-COV complex with human ACE2 (pdb id: 6ACJ)³³ were obtained from RCSB (www.rcsb.org).³⁴ The RBD domain of nCoV-2019 comprises 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 in GROMACS.³⁵ A TIP3P water model was used for the solvent and Param99SB-ILDN AMBER force field^{36,37} was used for all the complexes. A few counter ions were added to each system to neutralize the charges on the RBD and ACE2 as the PBSA method for binding energy calculation is known to be problematic with high ionic strength. Each system contained about 260,000 atoms. It is important to note that, none of the RBD/ACE2 complexes studied here were glycosylated. The glycosylation sites on RBD are far from the binding interface and do not interfere with the binding of RBD to ACE2. Moreover, there are nine Cys residues at the RBD of nCoV-2019 and eight of them form four pairs of disulfide bonds (Cys336-Cys361, Cys379-Cys432, Cys391-Cys525, and Cys480-Cys488).

In total, 5000 steps of energy minimizations were done using the steepest descent algorithm. In all steps, the LINCS algorithm was used to constrain all bonds containing hydrogen atoms and a time step of 2 fs was used as the integration time step. Equilibration of all systems was 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 310 K with a 0.1 ps coupling constant in an NVT ensemble under periodic boundary conditions and harmonic restraints on the backbone and side-chain atoms of the complex.³⁸ A 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 a temperature of 310 K and pressure of 1 bar using a Berendsen barostat.³⁹ This was done through decreasing the force constant of the restraint on the backbone and side-chain atoms of the complex from 1000 to 100 and finally to 10 $\frac{\text{kJ}}{\text{mol nm}^2}$. The Berendsen barostat was only used for the equilibration step because of its 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 the 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 1 bar using the Parrinello-Rahman barostat⁴⁰ with a compressibility of $4.5 \times$

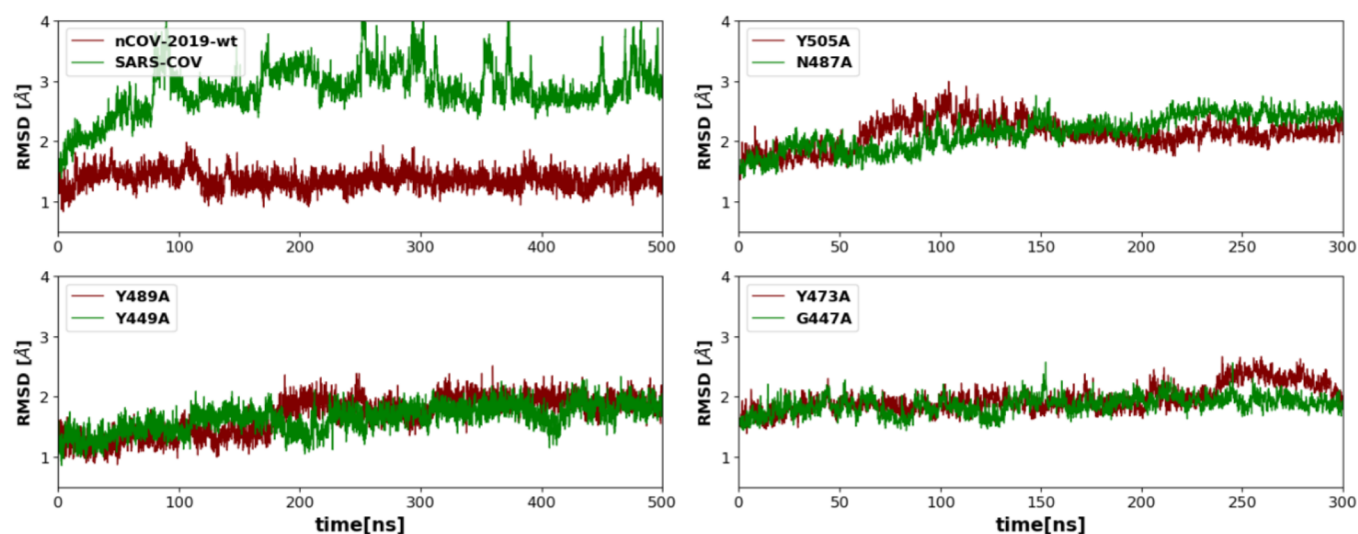


Figure 3. C_{α} RMSD plots for nCOV-2019 and SARS-COV and a few nCOV-2019 mutations.

10^{-5}bar^{-1} and a coupling constant of 0.5 ps. It is important to note that the Berendsen barostat was only used for the equilibration step as it was shown that this barostat can cause unrealistic temperature gradients.⁴¹ The production run lasted for 500 ns for SARS-COV and nCOV-2019 complexes and 300 ns for all the mutants with a 2 fs timestep and the particle-mesh Ewald⁴² for long range electrostatic interactions using the GROMACS 2018.3 package.⁴³ All mutant systems were constructed as described before and ran for 300 ns of production run. In addition, the simulation time for a few mutants (Y449A, T478I, Y489A, and S494P) was extended to 500 ns.

Gibbs Free Energy and Correlated Motions. The last 400 ns of simulation was used to explore the dominant motions in SARS-COV, nCOV-2019 and the mutations with extended simulation, and last 200 ns for all other mutants using principal component analysis (PCA) as part of the quasiharmonic analysis method.⁴⁴ For this method the rotational and translational motions of RBD of all systems were eliminated by fitting to a reference (crystal) structure. Next, 4000 snapshots from the last 400 ns of SARS-COV, nCOV-2019 and mutations with extended simulation time, and 2000 snapshots from last 200 ns of all other mutant systems were taken to generate the covariance matrix between C_{α} atoms of RBD. In the mutant systems with production run, the last 400 ns was used for this analysis. Diagonalization of this matrix resulted in a diagonal matrix of eigenvalues and their corresponding eigenvectors.^{43,45} 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 visual MD (VMD).

The principal components were used to calculate and plot the approximate free energy landscape (aFEL). We refer to the FEL produced by this approach to be approximate in that the ensemble with respect to the first few PC's (lowest frequency quasiharmonic 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 GROMACS³⁵ were used to obtain principal components and aFEL.

The dynamic cross-correlation maps (DCCM) were obtained using the MD_TASK package to identify the

correlated motions of RBD residues.⁴⁶ In DCCM, the cross-correlation matrix C_{ij} is obtained from displacement of backbone C_{α} atoms at a time interval Δt . The DCCM was constructed using the last 400 ns of SARS-COV and nCOV-2019 and the extended mutant systems and last 200 ns of all other mutant systems with a 100 ps time interval. Hydrogen bonds were analyzed in VMD where the distance cutoff was 3.2\AA and the angle cutoff between the donor and acceptor was 30° .

Binding Free Energy from Molecular Mechanics Poisson-Boltzmann Surface Area Method. The molecular mechanics Poisson-Boltzmann surface area (MMPBSA) method was applied to calculate the binding energy between RBD and ACE2 in all complexes.^{47,48} For SARS-COV and nCOV-2019, 200 snapshots of the last 400 ns and for the mutant systems and 100 snapshots of the last 200 ns simulation were used for the calculation of binding free energies with an interval of 2 ns. Simulation for a few mutant systems (Y449A, T478I, Y489A, and S494P) was extended to 400 ns, and the binding energies were calculated for the last 400 ns to assess the convergence of free energies. The binding free energy of a ligand–receptor complex can be calculated as:

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

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

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

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

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

In these equations, ΔE_{MM} , $\Delta G_{\text{bind,solv}}$ and $-T\Delta S$ are calculated in the gas phase. ΔE_{MM} is the gas phase molecular mechanical energy changes which includes covalent, electrostatic, and van der Waals energies. Based on previous studies, the entropy change during binding is small and neglected in these calculations.^{48–50} $\Delta G_{\text{bind,solv}}$ is the solvation free energy which comprises the polar and nonpolar components. The polar solvation is calculated using the MMPBSA method by setting a value of 80 and 2 for solvent and solute dielectric

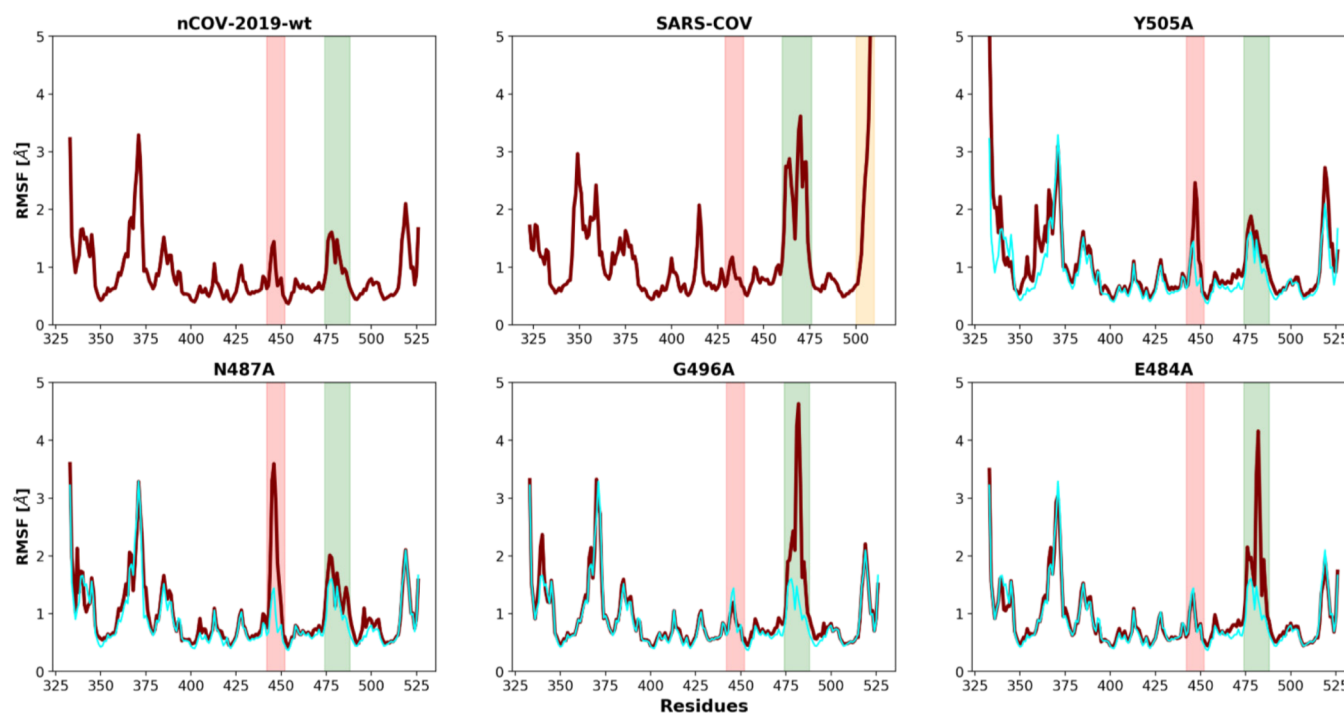


Figure 4. RMSF plots for nCOV-2019-wt, SARS-COV, Y505A, N487A, G496A, and E484A. The red shaded region shows the fluctuation in L_1 and the green shaded region shows the fluctuation in L_3 . The orange shaded region in SARS-COV shows the fluctuation in the C-terminal. For comparison, the RMSF of nCOV-2019-wt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

constants. The nonpolar free energy is simply estimated from the solvent accessible surface area (SASA) of the solute from the eq 5.

RESULTS

Structural Dynamics. To compute the RMSD of systems, the rotational and translational movements were removed by first fitting the C_α atoms of the RBD to the crystal structure and then computing the RMSD with respect to the C_α atoms of RBD in each system.

Figure 3 shows the RMSD plot in the RBD of SARS-COV, nCOV-2019, and some of its variants. Comparison of the RMSD of SARS and nCOV-2019 RBD shows that SARS-COV has a larger RMSD throughout the 500 ns simulation. In nCOV-2019, the 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 very stable during the 300 ns simulation which shows the great tolerance of the interface for mutations. However, a few mutations showed some RMSD variance. In mutation Y489A, the RMSD increases from 1.37 ± 0.21 Å to 1.88 ± 0.16 Å after 2000 ns. Mutation Y505A resulted in an increase in RMSD up to 100 ns to a value to 1.98 ± 0.20 Å and decreased afterward. The RMSD for mutation N487A shows an increasing behavior with a value of 2.10 ± 0.23 Å. Mutations N439K, V483A, and V483F showed a stable RMSD of ~1.5 Å. For mutations T478I, G476S, S494P, and A475V, the RMSD increases up to ~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 α_4 to α_5) of the RBD makes all contacts with ACE2, the RMSD was computed by also fitting to the C_α atoms of this region (Figure S2). The extended loop in nCOV-2019 is very stable with less deviation (RMSD = 0.86 ± 0.017 Å) from the crystal structure compared to SARS-COV having an RMSD of 2.79 ± 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 1.95 ± 0.12 Å and 1.94 ± 0.24 Å, respectively, after about 200 ns of simulation. Mutants G447A and E484A show a loop RMSD values of 2.22 ± 0.03 Å and 1.96 ± 0.02 Å during the last 100 ns of simulation. Other mutants showed a stable extended loop (observed in loop RMSD) during simulation. The stability of extended loop for mutant systems confirms the high tolerance of this region of RBD.

To characterize 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 show 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 had a RMSF of about 4 Å in the L_1 . Mutation Y449A has a higher RMSF of about 3 Å in the L_3 . Mutants G496A, E484A, and G447A show

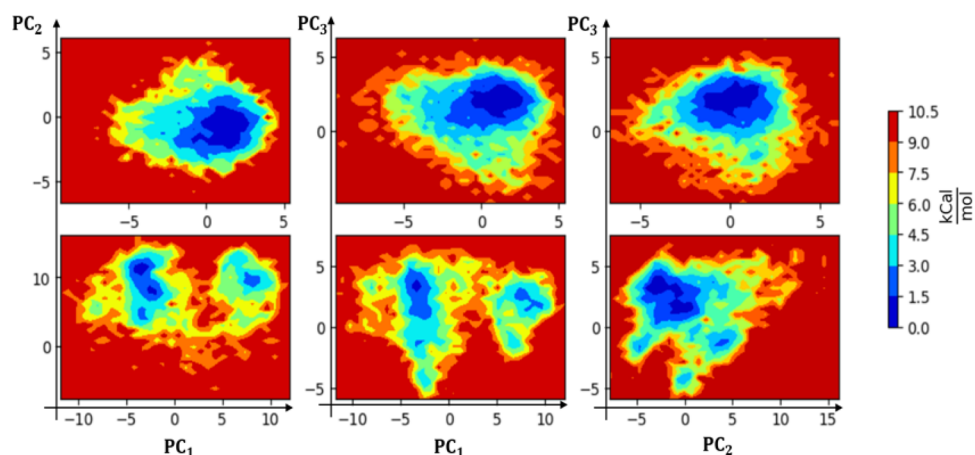


Figure 5. Mapping of the principal components of the RBD for the aFEL from the last 400 ns of simulations for SARS-COV (top row) and nCOV-2019-wt (bottom row). The color bar is relative to the lowest free energy state.

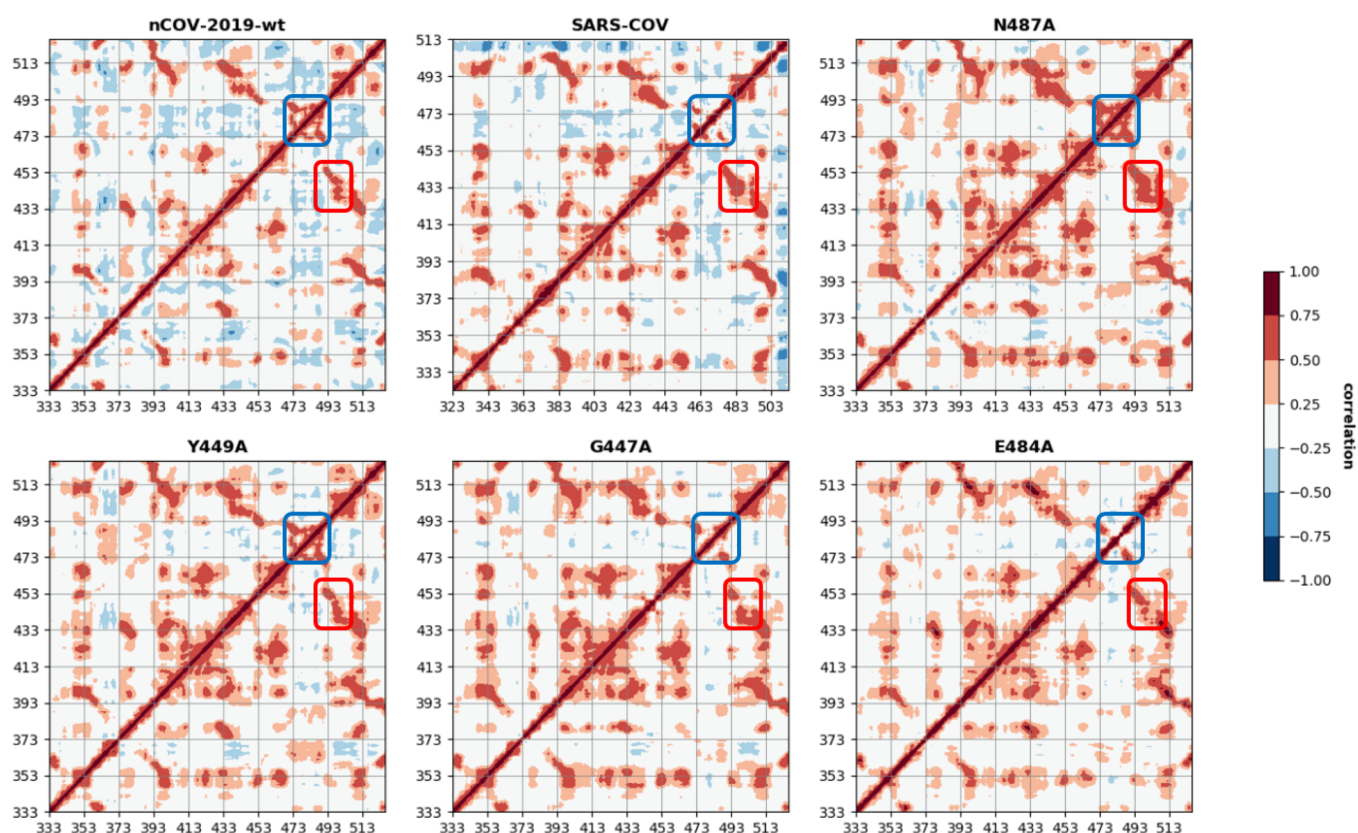


Figure 6. DCCM for nCOV-2019, SARS-COV, and mutants with residue numbers of the RBD domain. Red boxes show the correlation between α_5 , L_4 , and β_5 regions and α_4 , L_1 , and β_5 regions. Blue boxes show the correlation in L_3 and β_6 regions.

a high fluctuation of about 4.5 Å in the L_3 . The RMSF of other variants is shown in Figure S3.

PCA and aFEL. 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 400 ns for SARS-COV, nCOV-2019, and extended mutant systems and the last 200 ns for other nCOV-2019 mutants. 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 the receptor. The first three eigenvectors were used to calculate the aFEL using the last 400 ns of simulation for nCOV-2019 and SARS-COV as 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 the H-bond between N487 on

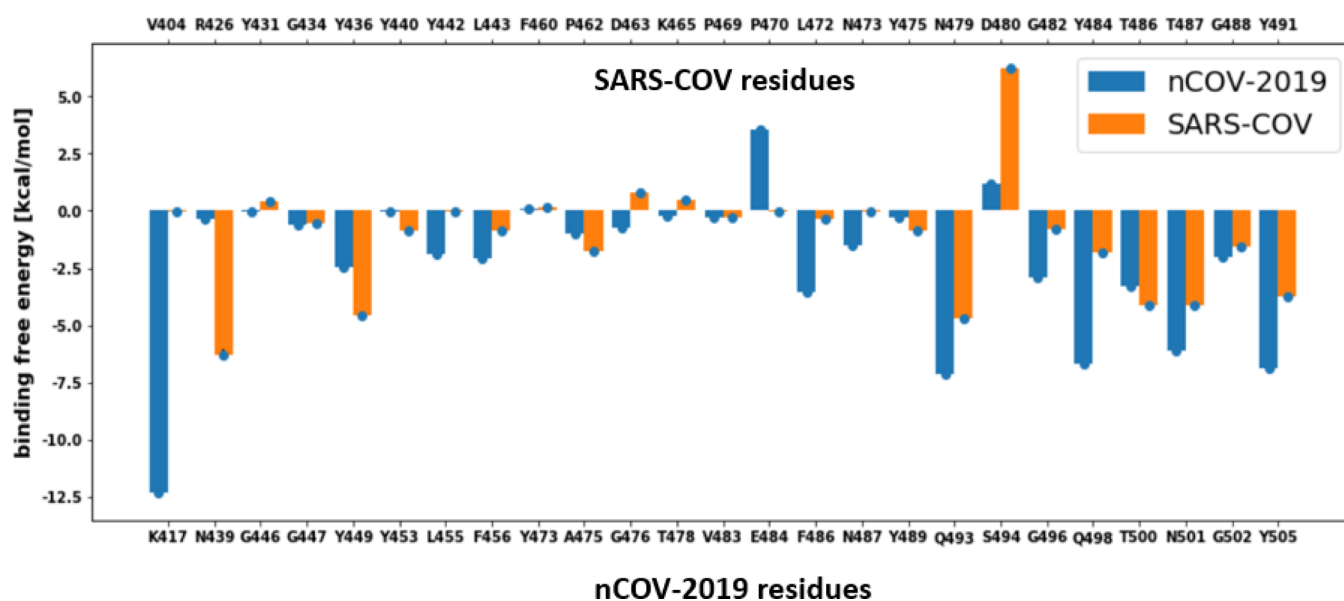


Figure 7. Binding energy decomposition per residue for the RBM of nCOV-2019 and SARS-COV.

RBD and Y83 on ACE2 as well as a π -stacking interaction between F486 and Y83. It is evident that the nCOV-2019 RBD is more stable than SARS-COV RBD and exists in one conformation whereas the SARS-COV interface fluctuates and the aFEL is separated into two different regions. The first two eigenvectors were used to calculate and plot the aFEL as a function of first two principal components using the last 200 ns of the simulation for mutant systems. aFEL for other systems is shown in Figure S4.

In each system, the first eigenvector was used to construct the porcupine plots to visualize the most dominant movements (Figure S5). nCOV-2019 showed a small motion in L_3 and the core region and the extended loop region are 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 (Figure S5-A). In SARS-COV, the C-terminal region shows large motions (Figure S5-B). Mutation N487A showed a large motion in L_1 (Figure S5-C). Mutations Y449A, G447A, and E484A demonstrate large motions in L_3 (Figure S5). Overall, these plots show the involvement of these residues in the dynamic stability of the RBD/ACE2 complex.

DCCM. The correlated motions of RBD atoms were also analyzed with the DCCM based on the C_α atoms of RBD from the last 400 ns of simulation for nCOV-2019, SARS-COV, and extended mutant systems and the last 200 ns for the other mutant systems (Figure 6). The DCCM for nCOV-2019 showed a correlation between residues 490–505 (containing α_5 , L_4 and β_5 regions) and residues 440–455 (containing α_4 , L_1 and β_3 regions) shown in the red rectangle in Figure 6. This correlation showed the coordination of these regions for binding ACE2 effectively. Another important correlation that appears in the DCCM of nCOV-2019 is between residues 473–481 with residues 482–491. These residues are in L_3 and β_6 regions and their correlation in nCOV-2019 is stronger than SARS-COV. This is due to the presence of the β_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 a new secondary structure, which causes strong correlation and makes the loop act as a recognition region for binding. The correlation in L_3 is shown as a blue

rectangle in Figure 6. Some of the mutations disrupted the patterns of correlation and anticorrelation in nCOV-2019-wt. Mutation N487A showed a stronger correlation in L_3 and β_6 strand than the wild-type RBD. In mutation E484A, correlation in L_3 is stronger than the nCOV-2019-wt. DCCM for other mutants are shown in Figure S6. 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 F486 resides in L_3 and plays a crucial role in stabilizing the recognition loop by making a π -stacking interaction with residue Y83A on ACE2.

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.⁴⁸ The binding energy was partitioned into its individual components including: electrostatic, van der Waals, polar solvation, and 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 ± 1.93 kcal/mol, whereas SARS-COV has a much higher binding energy of -18.79 ± 1.53 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 ± 2.66 kcal/mol for nCOV-2019 and -600.14 ± 7.65 kcal/mol for SARS-COV. This high electrostatic contribution is compensated by a large polar solvation free energy which is 797.30 ± 3.12 kcal/mol for nCOV-2019 and 659.61 ± 8.98 kcal/mol for SARS-COV. nCOV-2019 also possess a higher van der Waals (vdw) contribution (-89.93 ± 0.46 kcal/mol than SARS-COV (-70.07 ± 1.22 kcal/mol. Furthermore, the SASA contribution to binding for SARS-COV was -8.30 ± 0.15 kcal/mol and -10.58 kcal/mol for nCOV-2019. Both hydrophobic and electrostatic interactions play major roles in the higher affinity of nCOV-2019 RBD than SARS-COV RBD to ACE2.

The binding free energies for nCOV-2019 and SARS-COV were decomposed into a per-residue based binding energy to find the residues that contribute strongly to the binding and are responsible for higher binding affinity of nCOV-2019 than SARS-COV (Figure 7). Most of the residues in the RBM of

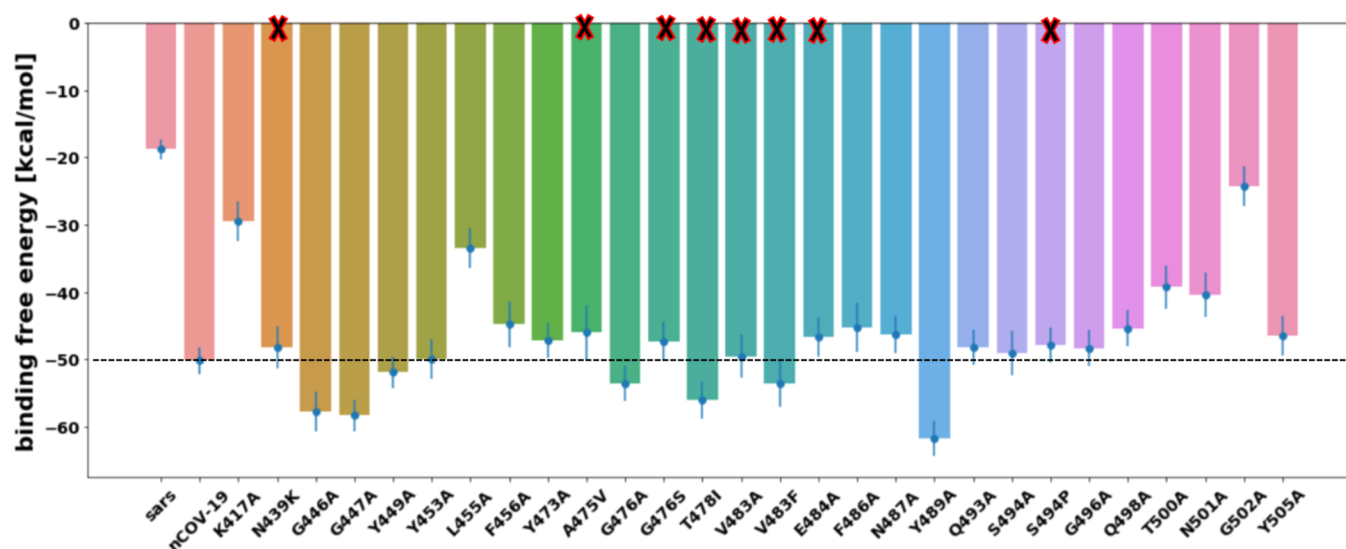


Figure 8. Total free binding energy of SARS-COV, nCOV-2019, and mutants. Natural mutants are shown with X at the bar base.

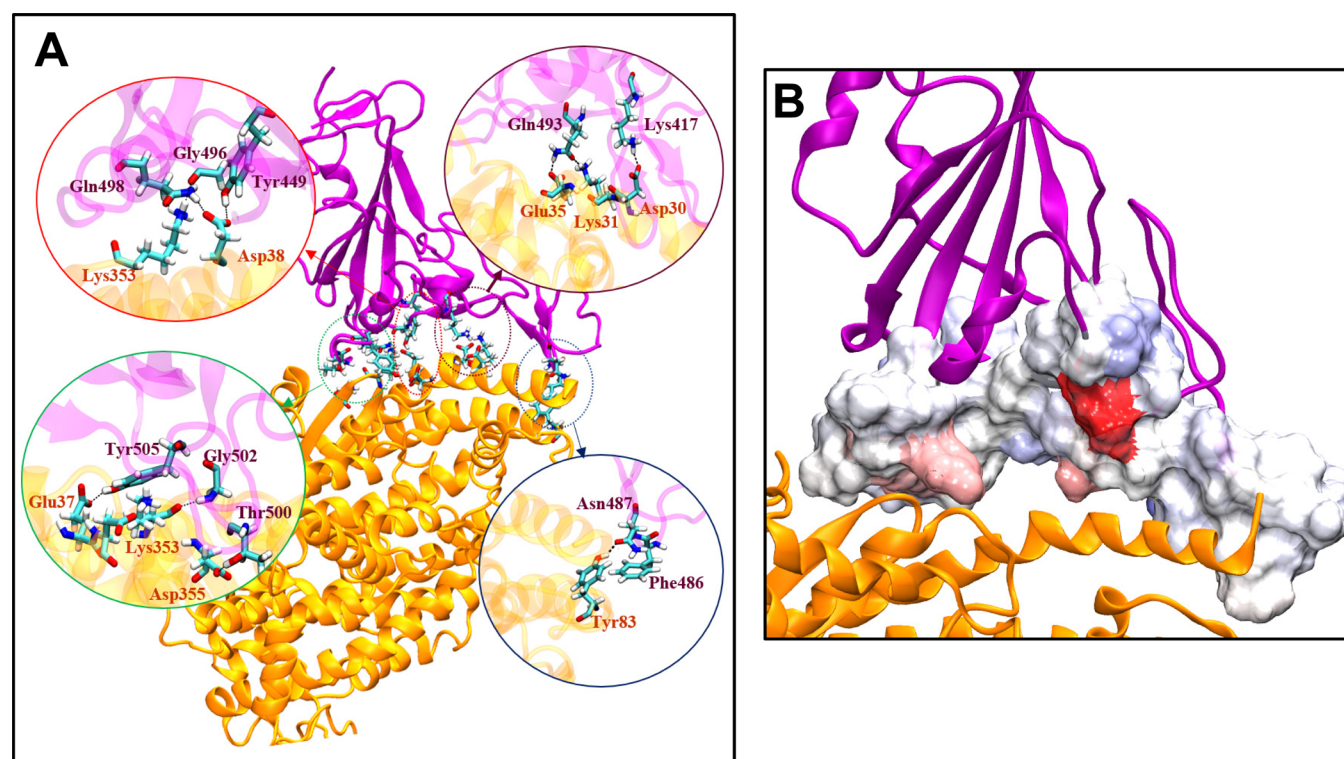


Figure 9. (A) H bonds between RBD of nCOV-2019 and SARS-COV. (B) Mapping contribution of interface residues to structure in the RBD of nCOV-2019. The RBD is purple and the ACE2 is yellow. The RBD in contact with ACE2 is rendered in a surface format with more red being a favorable contribution to binding (more negative) and blue unfavorable contribution (positive).

nCOV-2019 had more favorable contribution to the total binding energy than SARS-COV. Residues Q498, Y505, N501, Q493, and K417 in nCOV-2019 RBM contributed more than 5 kcal/mol to binding affinity and are crucial for complex formation. A few residues such as E484 and S494 contributed unfavorably to the total binding energy. Among all the interface residues K417 had the highest contribution to the total binding energy (-12.34 ± 0.23 kcal/mol). The corresponding residue in SARS-COV is V404 only had a -0.02 ± 0.01 kcal/mol contribution, which points to the importance of this residue for nCOV-2019 binding to ACE2.

Residue Q498 contributed -6.72 ± 0.18 kcal/mol and its corresponding residue in SARS-COV is a Y484 that contributed to total binding by -1.83 ± 0.06 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 (Figure 7). Residue D480 in SARS-COV contributed positively to binding energy by 6.2 ± 0.15 kcal/mol and the corresponding residue in nCOV-2019 which is a S494 residue lowered this positive contribution to only 1.17 ± 0.06 kcal/mol. Mutation D480A/G appeared to be a dominant mutation in SARS-COV in

2002–2003.⁵¹ This mutation was reported to escape neutralization by antibody 80R.⁵² To investigate the effect of this point mutation on binding of SARS-COV RBD to ACE2 we performed an additional simulation and calculated the binding affinity for this mutant in SARS-COV RBD with the same approach for other mutation in this study (Figure 8). D480A mutation showed a binding affinity of 23.46 ± 3.07 kcal/mol which is about 5 kcal/mol higher than the wild-type SARS-COV RBD. In SARS-COV, residue R426 had the highest contribution to the total binding energy (-6.27 ± 0.22 kcal/mol) although the corresponding residue in nCOV-2019 is N439 with a contribution of -0.32 ± 0.02 kcal/mol. These important mutations on RBM of nCOV-2019 from SARS-COV caused RBD of nCOV-2019 to bind ACE2 with much stronger (about 30 kcal/mol) affinity.

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 SASA energies. The vdW free energy of binding ranged from -84.68 ± 0.68 kcal/mol for Q493A to -103.85 ± 0.66 kcal/mol for Y489A. Mutant K417A had the lowest electrostatic contribution to binding -415.67 ± 5.07 kcal/mol and mutants N439K and E484A had the highest electrostatic binding contribution of -989.80 ± 5.6 kcal/mol and -941.20 ± 3.95 kcal/mol, respectively.

Most alanine substitutions exhibited similar or lower total binding affinities to nCOV-2019, however a few mutants had higher binding affinity than the wild type. Mutant Y489A had a total binding energy of -61.78 ± 2.59 kcal/mol which was about 11 kcal/mol lower than wild type binding energy. Mutants G446A, G447A, and T478I also demonstrated higher total binding affinities than nCOV-2019. 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 ± 2.98 kcal/mol. Mutant systems K417A, L455A, T500A, and N501A are the other mutants with total binding affinities significantly lower than the wild type complex. The electrostatic component of binding contributes the most to the low binding affinities for these mutants. The contribution of RBM residues to binding with ACE2 for nCOV-2019 was mapped to the RBD structure and is shown in Figure 9B.

Most natural mutants exhibited similar binding affinities compared to wild-type nCOV-2019 with a few exceptions. Mutation T478I which is one of the most frequent mutations based on the GISAID database has a binding affinity which is about 6 kcal/mol higher than that of the wild-type. S494P and A475V showed a slightly lower binding affinity than the wild-type complex. Other natural mutants showed binding affinities similar to wild-type RBD. 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 ± 3.07 kcal/mol which is similar to nCOV-2019.

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 400 ns 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.

Table 1. H-Bonds and Salt-Bridges between nCOV-2019 and ACE2 and SARS-COV and ACE2 that Persist for >30%^a

#	nCOV-2019	ACE2	% occupancy	SARS-COV	ACE2	% occupancy
1	G502	K353	89	Y436	D38	96
2	Q493	E35	83	R426	E329	87
3	N487	Y83	80	T486	D355	83
4	Q498	D38	73	G488	K353	80
5	K417	D30	55	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			

^aSalt 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 a persistent H-bond with residue K353 on ACE2. This residue was G488 in SARS-COV, which also makes the H-bond with K353 on ACE2. Q493 in nCOV-2019 makes H-bond with E35 and another H-bond with K31 on ACE2. This residue was an N479 in SARS-COV, which only makes one H-bond with K31 on ACE2. 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 the nCOV-2019/ACE2 complex contributes to the total binding energy by -12.34 ± 0.23 kcal/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. Gao et al.²⁷ used a FEP approach and showed that mutation V404 to K417 lowers the binding energy of nCOV-2019 RBD to ACE2 by -2.2 ± 0.9 kcal/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 which is unable to make salt-bridge with ACE2 residue E329. One of the most observed mutations in nCOV-2019 according to the GISAID database is N439K which recovers some of the electrostatic interactions with ACE2 at this position. Y436 in SARS-COV and Y449 in nCOV-2019 both make H-bonds with D38 on ACE2. 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 the RBD/ACE2 complex in nCOV-2019. An important interaction between nCOV-2019 RBD and ACE2 is the π -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 Gao et al.²⁶ that mutation L472 to F486 in nCOV-2019 results in a net change in the binding free energy of -1.2 ± 0.2 kcal/mol. Other interfacial residues in nCOV-2019 RBD that participate in the 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-workers³⁰ performed long-timescale (1 μ 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 β_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 enables 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 affinity by about 5 kcal/mol. The H-bonds between RBD of nCOV-2019 and SARS-COV are shown in Figure 9A. The structural details discussed here are in agreement with other structural studies of the nCOV-2019 RBD/ACE2 complex.^{4,53}

H-bond analysis was also performed for the mutant systems and the results for H-bonds with more than 40% consistency are shown 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 and Q493A, stunningly maintain the number of H-bonds in the wild-type complex. This indicates that the plasticity in the network of H-bonds in RBM of nCOV-2019 can reshape the network and strengthen other H-bonds upon mutation in these locations. However, few mutations 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 K417A decreases the number of H-bonds to only 5 where the H-bond at residue Q498 is broken. This indicates the delicate nature of the 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.

DISCUSSION

In this work, we performed MD simulations to unveil the detailed molecular mechanism for the receptor binding of nCOV-2019 and compared it with SARS-COV. The role of key residues at the interface of nCOV-2019 with ACE2 was investigated by computational ala-scanning. A rigorous 500 ns MD simulation was performed for nCOV-2019, SARS-COV, and few mutants (Y449, T478I, Y489A, and S494P) as well as 300 ns MD simulation on each mutant. These simulations aid in understanding 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.⁵⁴ The effect of these mutations on the stability of the RBD/ACE2 complex was investigated to shed light on their role in the viral infection of coronavirus.

Changes in the 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, a few nCOV-2019 mutations resulted in larger deviations (>2 Å), i.e., Y489A, F456A, Y505A, N487A, K417A,

Y473A, and Y449A. We further investigated the structure of the extended loop domain (Figure 1B) and discovered that nCOV-2019 is stable with an RMSD of less than 1 Å, whereas the extended loop in SARS-COV shows an RMSD of about 3 Å 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 four-residue motif (GQTQ) in nCOV-2019 L_3 , which forces the loop to adopt a compact structure by making two 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 π -stacking and H-bond interactions with residue Y83 on ACE2. Alanine substitution at N487 introduced a large RMSF to L_1 . Mutation L472 to F486 in SARS-COV was shown to favor binding by -1.2 ± 0.2 kcal/mol using FEP.²⁶ In addition, this mutation was shown to be among the five mutations that produce a super affinity ACE2 binder based on SARS-COV RBD.⁶ Alanine mutations at residues Y449, G447, and E484 increased the motion in L_3 characterized by a large RMSF in this region.

Using PCA, the aFEL 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 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 caused the FEL to degenerate into separate multiple low energy regions (Figure S4). Dominant motions in the RBD are visualized in Figure S5 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 the L_1 region and Y449A, G447A, and E484A showed large motions in L_3 (Figure S5).

To better characterize the functional motions of RBD, DCCM for all systems are constructed and shown in Figure 6 and Figure S6. nCOV-2019 showed a large correlation between the α_4 - L_1 - β_5 and α_5 - L_4 - β_5 regions. 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 β_6 . This correlation is stronger in nCOV-2019 than SARS-COV due to the presence of β_6 which makes the loop to adopt correlated motions. Few mutants impact the correlation in this region such as N487A. Interestingly, mutant F486A which is in L_3 and participates in binding by π -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 the 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 ~ 30 kcal/mol difference in the total binding free energy of nCOV-2019 and SARS-COV. The binding energies calculated here for nCOV-2019 and SARS-COV (-50.22 ± 1.93 and -18.79 ± 1.53 kcal/mol,

respectively) are in good agreement with the binding energies calculated using the generalized Born method by Spinello et al.³⁰ Moreover, the patterns of H-bonds between nCOV-2019 and ACE2 were also already characterized by other groups^{26,30} 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 the 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 ± 0.04 kcal/mol and the Ala-substitution at G502 has the highest effect on the binding energy among all the residues by lowering the total binding affinity to 24.31 ± 2.98 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 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. This caused Q493 to have more contribution to total binding than its counterpart N479. However, alanine substitution at Q493 did not affect the total binding energy and this mutant had a total binding energy similar to the wild-type complex as it maintains the number 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 the SARS-COV/ACE2 complex. Q498 contributes to binding by -6.72 ± 0.18 kcal/mol which is more than the contribution of Y484 in SARS-COV (-1.83 ± 0.06 kcal/mol). Ala-substitution at Q498 did not show large impact on the 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 affinity by about 10 kcal/mol. Residue N487 in nCOV-2019 locates in L_3 and plays a crucial role in stabilizing the recognition loop by making a H-bond with Y83 on ACE2. This residue contributes to the total binding energy of nCOV-2019 by -1.52 ± 0.06 kcal/mol, whereas its corresponding residue in SARS-COV does not show any contribution to the binding energy (-0.02 ± 0.05 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 the total binding energy (-12.34 ± 0.23 kcal/mol by making a salt-bridge with residue D30 on ACE2. This residue is crucial for the binding of RBD and ACE2 and alanine substitution lowers the total binding energy to -29.56 ± 2.95 kcal/mol. This salt-bridge is found to be important for the stability of the crystal structure of the RBD/ACE2 complex in nCOV-2019.⁴ K417 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 ± 0.03 kcal/mol. This residue is important for hydrophobic interaction with ACE2 and mutating it to alanine lowers the total binding affinity by about 17 kcal/mol. The hydrophobic residue F456 in nCOV-2019 also has a favorable contribution to the binding energy and F456A lowers the binding affinity by 5 kcal/mol. These results are in fair agreement with experimental binding measurements with deep mutational scanning of RBD in nCOV-2019 where they used flow cytometry for different ACE2 concentrations to

measure the dissociation constant K_D .²⁵ It was shown that mutations at K417, N487, T500, and G502 are detrimental for binding to ACE2, which agrees with the results here. These experiments showed that mutations at Q493 and Q498 do not impact the binding affinity of RBD to ACE2 which demonstrates the high plasticity of the network of H-bonds at the interface where upon mutation at these residues the network can reshape to form new H-bonds. Mutations at hydrophobic residues L455 and F456 are shown to reduce the binding affinity in these experiments.

The total binding energy calculation of all the variants showed that mutation Y489A has the highest binding affinity among all systems which is about 11 kcal/mol stronger than that of the nCOV-2019 complex. This residue is located in β_6 , which is part of the recognition region of RBD for binding to ACE2. Removal of this bulky hydrophobic residue at the interface with ACE2 caused the extended loop to move closer to the ACE2 interface and make more H-bonds with ACE2 (Table S3). A high electrostatic interaction energy is the reason for the higher binding energy of mutant Y489A than the wild-type complex. It is interesting to note that among the five residues L455, F456, Y473, A475, and Y489 that make hydrophobic interactions with ACE2, Y489 is the only residue that is conserved from SARS-COV. However, the experimental binding affinity measurements using deep mutational scanning showed that mutations at this position lower the binding affinity to ACE2. Other alanine substitutions that increase the binding energy are G446A and G447A. Residues G446 and G447 reside in L_1 and mutation to alanine can make L_1 take a more rigid form as shown in the RMSF plot in Figure S3. However, experiment showed that these mutations have similar or lower binding affinities to ACE2 than the wild-type RBD and care must be taken when interpreting these results.²⁵ This discrepancy could be due to force-field inaccuracy and the deficiencies in the PBSA method for the treatment of solvent in binding energy calculation. Further studies are needed to investigate whether these mutations will increase the binding affinity to ACE2. Deep mutational scanning using flow cytometry is a qualitative method to measure the impact of a large number of mutations of protein–protein interactions and further experiments such as SPR or isothermal titration calorimetry which are conventional methods for measuring binding affinities needed to study the effect of these mutations in detail.

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 most frequent mutations according to GISAID database, increases the binding affinity of nCOV-2019 to ACE2 by about 6 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; however, it exerts a binding affinity similar to that of wild-type RBD. Contribution of important interface residues to binding affinity was compared for mutations T478I, N439K, and wild-type nCOV-2019 (Figure S7). The most striking differences between wild-type RBD and mutation T478I are residues Y449 and Q498 which have significantly higher contribution to binding than the wild type residue. Most other residues at the interface have similar binding affinities to the nCOV-2019. A

higher H-bond persistence is also seen for these two residues Y449 and Q498 compared to the wild type RBD which is the reason for their higher contribution to the total binding energy. Mutation N439K has a slightly lower binding affinity to ACE2 than the wild type RBD. Per residue binding energy decomposition showed that K439 in this system has a favorable contribution of -1.80 ± 0.15 kcal/mol to the total binding energy which is balanced by a lower contribution of K417 which resulted in a binding affinity similar to wild-type RBD. Mutant E484A, which is also one of the observed mutations based on GISAID database, demonstrates a high electrostatic interaction with ACE2. E484 contributes to binding by 3.56 ± 0.15 kcal/mol whereas the corresponding residue in SARS-COV, P469 contributes to binding of SARS-COV to ACE2 by -0.27 ± 0.01 kcal/mol. This residue is close to D30 on ACE2 and has electrostatic repulsion with this residue. Most natural mutants including N439K, A475V, G476S, V483A, V483F, E484A, and S494P showed similar or slightly lower binding affinities to ACE2 compared to wild-type complex which agrees with experimental binding measurements.²⁵ However, the experimental binding affinity for T478I also showed similar binding affinity to the wild-type complex. This difference could be due to the use of MMPBSA approach for calculation of polar solvation and further studies are needed to study the effect of this mutation 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 ± 0.14 kcal/mol) and mutating this residue to S494 in nCOV-2019 lowers this negative contribution to 1.17 ± 0.06 kcal/mol. D480 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. Gao and co-workers²⁶ computed the relative free energies of binding because of mutations from the RBD-ACE2 of SARS-COV to the corresponding residues in nCOV-2019. They used a FEP approach and showed that mutation D480S in SARS-COV changed the binding free energy by -1.9 ± 0.8 kcal/mol which is consistent with our study. Furthermore, we performed an additional simulation on D480A mutant in SARS-COV and found that this mutation has a binding affinity of 23.46 ± 3.07 kcal/mol which is about 5 kcal/mol higher than the wild-type SARS-COV RBD. In addition, experimental binding affinity measurements showed that mutations of S494 to an acidic residue highly reduce the binding affinity to ACE2 which confirms the hypothesis here.

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 kcal/mol stronger than SARS-COV and is in fair agreement with the results here.⁵⁶ The critical role of interface residues is 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 hotspot residues for the interaction with receptor ACE2.^{26,30,35,56} 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 have converted an 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 π -stacking interactions with ACE2, which makes this region to act as a recognition loop. Previous studies on SARS-COV have shown that there is a correlation between the higher binding affinity to the receptor and higher infection rate by coronavirus.^{6,13,57} The higher binding affinity of nCOV-2019 for ACE2 than SARS-COV to ACE2 is suggested to be the reason for its high infection rate. Most natural mutations showed similar binding affinities to wild-type ACE2 which indicates that the virus was already effective at the beginning of the crisis for binding ACE2. A few mutations such as N489A and T478I are shown to increase the binding affinity to ACE2. However, more studies are needed to investigate the effect of these mutations in detail. 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 as antibody escape mutants. Escape from monoclonal antibodies is observed for mutations of SARS-COV in 2002 and these mutations should be considered for any antibody design endeavors against these escape mutations.

CONCLUSIONS

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 aid in the understanding of how these mutations can play an important role 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 the dynamic stability of the complex by a π -stacking interaction with ACE2. Per-residue free energy decomposition pinpoints the critical role of residues K417, Y505, Q498, and 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 bind ACE2 with about 7 kcal/mol higher affinity than wild-type. 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 strong RBD attachment to ACE2 and influence the infection virulence. However, details of interaction between these mutants and ACE2 should be carefully studied using experimental techniques. 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 into key structural features of interface residues for the advancement of

effective therapeutic strategies to stop the coronavirus pandemic.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpbc.0c05994>.

Location of residues picked for mutation; binding free energies of mutants; details of H-bonds for all mutations; RMSD of mutants; loop RMSD of all systems; RMSF plots for all mutants; FELs using first two principal components; porcupine plots for structural fluctuations; DCCM plots for mutants; and binding energy decomposition for T478I and N439K mutants (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Jeffery B. Klauda – Department of Chemical and Biomolecular Engineering and Biophysics Graduate Program, University of Maryland, College Park, Maryland 20742, United States;

orcid.org/0000-0001-8725-1870; Email: jbklauda@umd.edu

Authors

Mahdi Ghorbani – Department of Chemical and Biomolecular Engineering, University of Maryland, College Park, Maryland 20742, United States; Laboratory of Computational Biology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20824, United States

Bernard R. Brooks – Laboratory of Computational Biology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20824, United States;

orcid.org/0000-0002-3586-2730

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpbc.0c05994>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the National Heart, Lung, and Blood Institute (NHLBI) at the National Institutes of Health (NIH). The authors acknowledge the Biowulf high performance Linux cluster at the NIH for providing computational resources for this project. 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. J.B.K. would also like to dedicate this work to family friend Joe Kaplan (Silver Spring, MD) who passed away because of COVID-19 on April 22, 2020.

■ REFERENCES

- (1) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733.
- (2) Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273.
- (3) Li, F. Structure, function, and evolution of coronavirus spike proteins. *Ann. rev. of virol.* **2016**, *3*, 237–261.

- (4) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K.-Y. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* **2020**, *181*, 894–904.e9.

- (5) Chan, P. K.; To, W.-K.; Ng, K.-C.; Lam, R. K.; Ng, T.-K.; Chan, R. C.; Wu, A.; Yu, W.-C.; Lee, N.; Hui, D. S. Laboratory diagnosis of SARS. *Emerg. infect. diseases* **2004**, *10*, 825.

- (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*, 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* **2020**, *367*, 1260–1263.

- (8) Bosch, B. J.; Van der Zee, R.; De Haan, C. A.; Rottier, P. J. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* **2003**, *77*, 8801–8811.

- (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*, 1–14.

- (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**, 281.

- (11) Tian, X.; Li, C.; Huang, A.; Xia, S.; Lu, S.; Shi, Z.; Lu, L.; Jiang, S.; Yang, Z.; Wu, Y. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg. micro. & infect.* **2020**, *9*, 382–385.

- (12) Li, F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J. Virol.* **2015**, *89*, 1954–1964.

- (13) 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*, 11429–11433.

- (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*, 6984–6991.

- (15) Li, F. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Res.* **2013**, *100*, 246–254.

- (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*, 8904–8911.

- (17) Chakraborti, S.; Prabakaran, P.; Xiao, X.; Dimitrov, D. S. The SARS coronavirus S glycoprotein receptor binding domain: fine mapping and functional characterization. *Virology journal* **2005**, *2*, 73.

- (18) Greenough, T. C.; Babcock, G. J.; Roberts, A.; Hernandez, H. J.; Thomas, W. D., Jr.; Coccia, J. A.; Graziano, R. F.; Srinivasan, M.; Lowy, I.; Finberg, R. W. Development and characterization of a severe acute respiratory syndrome-associated coronavirus—neutralizing human monoclonal antibody that provides effective immunoprophylaxis in mice. *J. infect. diseases* **2005**, *191*, 507–514.

- (19) Coughlin, M. M.; Prabhakar, B. S. Neutralizing human monoclonal antibodies to severe acute respiratory syndrome coronavirus: target, mechanism of action, and therapeutic potential. *Rev. in med. virology* **2012**, *22*, 2–17.

- (20) Ter Meulen, J.; Van Den Brink, E. N.; Poon, L. L.; Marissen, W. E.; Leung, C. S.; Cox, F.; Cheung, C. Y.; Bakker, A. Q.; Bogaards, J. A.; Van Deventer, E. Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. *PLoS Med* **2006**, *3*, e237.

- (21) Deming, D.; Sheahan, T.; Heise, M.; Yount, B.; Davis, N.; Sims, A.; Suthar, M.; Harkema, J.; Whitmore, A.; Pickles, R. Vaccine efficacy in senescent mice challenged with recombinant SARS-CoV bearing epidemic and zoonotic spike variants. *PLoS med* **2006**, *3*, No. e525.

- (22) 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*, 3220–3235.

(23) 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. diseases* **2010**, *201*, 946–955.

(24) Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* **2017**, *1*, 33–46.

(25) Starr, T. N.; Greaney, A. J.; Hilton, S. K.; Crawford, K. H.; Navarro, M. J.; Bowen, J. E.; Tortorici, M. A.; Walls, A. C.; Velesler, D.; Bloom, J. D. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *BioRxiv* **2020**, 157982.

(26) Wang, Y.; Liu, M.; Gao, J., Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. Proceedings of the National Academy of Sciences 2020.

(27) Malik, A.; Prahlad, D.; Kulkarni, N.; Kayal, A., Interfacial Water Molecules Make RBD of SPIKE Protein and Human ACE2 to Stick Together. *bioRxiv* 2020, 152892 DOI: 10.1101/2020.06.15.152892

(28) Boorla, V. S.; Chowdhury, R.; Maranas, C. D. De novo design of high-affinity antibody variable regions (scFv) against the SARS-CoV-2 spike protein. *BioRxiv* **2020**, No. 034866.

(29) Sharma, S.; Singh, I.; Haider, S.; Malik, M. Z.; Ponnusamy, K.; Rai, E. ACE2 Homo-dimerization, Human Genomic variants and Interaction of Host Proteins Explain High Population Specific Differences in Outcomes of COVID19. *BioRxiv* **2020**, No. 050534.

(30) 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-Atoms Simulations. *J. Phys. Chem. Lett.* **2020**, *11*, 4785–4790.

(31) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F., *Structural basis for receptor recognition by the novel coronavirus from Wuhan.* 2020.

(32) Rodrigues, C. H.; Myung, Y.; Pires, D. E.; Ascher, D. B. mCSM-PPI2: predicting the effects of mutations on protein–protein interactions. *Nucleic Acids Res.* **2019**, *47*, W338–W344.

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

(34) 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*, 235–242.

(35) 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.

(36) 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., Bioinf.* **2010**, *78*, 1950–1958.

(37) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

(38) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J.chem. phys.* **2007**, *126*, No. 014101.

(39) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(40) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(41) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: the Langevin piston method. *J.chem. phys.* **1995**, *103*, 4613–4621.

(42) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. chem. l. phys.* **1993**, *98*, 10089–10092.

(43) Lindahl, E.; Hess, B.; Van Der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Mol. modeling ann.* **2001**, *7*, 306–317.

(44) Brooks, B. R.; Janežič, D.; Karplus, M. Harmonic analysis of large systems. I. Methodology. *J. Comput. Chem.* **1995**, *16*, 1522–1542.

(45) David, C. C.; Jacobs, D. J., Principal component analysis: a method for determining the essential dynamics of proteins. In *Protein dynamics*, Springer: 2014; 193–226.

(46) Brown, D. K.; Penkler, D. L.; Sheik Amamuddy, O.; Ross, C.; Atilgan, A. R.; Atilgan, C.; Tastan Bishop, Ö. MD-TASK: a software suite for analyzing molecular dynamics trajectories. *Bioinformatics* **2017**, *33*, 2768–2771.

(47) Ren, J.; Yuan, X.; Li, J.; Lin, S.; Yang, B.; Chen, C.; Zhao, J.; Zheng, W.; Liao, H.; Yang, 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. Chin. Chem. Soc.* **2019**, *67*, 46–53.

(48) Wang, C.; Greene, D. A.; Xiao, L.; Qi, R.; Luo, R. Recent developments and applications of the MMPBSA method. *Front. mol. biosci.* **2018**, *4*, 87.

(49) 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., Bioinf.* **2011**, *79*, 1940–1951.

(50) 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*, 15–28.

(51) Korber, B.; Fischer, W. M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **2020**, *182*, 812–827.

(52) Sui, J.; Aird, D. R.; Tamin, A.; Murakami, A.; Yan, M.; Yammanuru, A.; Jing, H.; Kan, B.; Liu, X.; Zhu, Q. Broadening of neutralization activity to directly block a dominant antibody-driven SARS-coronavirus evolution pathway. *PLoS Pathog.* **2008**, *4*, No. e1000197.

(53) 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. immuno.* **2020**, *17*, 613–620.

(54) Forster, P.; Forster, L.; Renfrew, C.; Forster, M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Pro. Nat. Aca. of Sci.* **2020**, *117*, 9241–9243.

(55) Brielle, E. S.; Schneidman-Duhovny, D.; Linial, M. The SARS-CoV-2 exerts a distinctive strategy for interacting with the ACE2 human receptor. *Viruses* **2020**, *12*, 497.

(56) Liu, H.; Lupala, C.; Li, X.; Lei, J.; Chen, H.; Qi, J.; Su, X., Computational simulations reveal the binding dynamics between human ACE2 and the receptor binding domain of SARS-CoV-2 spike protein. *bioRxiv* 2020, 005561 DOI: 10.1101/2020.03.24.005561.

(57) Li, Y.; Wang, Y.; Gong, Y.; Qiu, Y.; 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. *bioRxiv* 2020, 077842 DOI: 10.1101/2020.05.04.077842.