Computational Prediction of Mutational Effects on SARS-CoV-2 Binding by Relative Free Energy Calculations

Junjie Zou,[||] Jian Yin,[||] Lei Fang, Mingjun Yang, Tianyuan Wang, Weikun Wu, Michael A. Bellucci, and Peiyu Zhang*

Cite This: https://dx.doi.org/10.1021/acs.jcim.0c00679

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The ability of coronaviruses to infect humans is invariably associated with their binding strengths to human receptor proteins. Both SARS-CoV-2, initially named 2019-nCoV, and SARS-CoV were reported to utilize angiotensin-converting enzyme 2 (ACE2) as an entry receptor in human cells. To better understand the interplay between SARS-CoV-2 and ACE2, we performed computational alanine scanning mutagenesis on the "hotspot" residues at protein–protein interfaces using relative free energy calculations. Our data suggest that the mutations in SARS-CoV-2 lead to a greater binding affinity relative to SARS-CoV. In addition, our free energy calculations provide insight into the infectious ability of viruses on a physical basis and also provide useful information for the design of antiviral drugs.
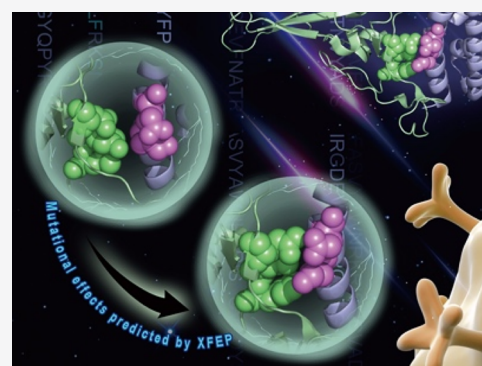
INTRODUCTION

The novel coronavirus strain, SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), that was initially named 2019-nCoV and identified as the causative agent of the ongoing respiratory illness outbreak (coronavirus disease 2019, COVID-19) worldwide, has caused severe damages to global public health and the world economy.[1−4] Genome sequence analysis has shown that SARS-CoV-2 shares an ~80% nucleotide identity with the original virus strains of SARS-CoV (severe acute respiratory syndrome coronavirus),[3] which is another β coronavirus that infects humans and was responsible for the 2002−2003 epidemic of atypical pneumonia.[5,6] As of June 1, 2020, there have been over 6 million confirmed cases of COVID-19 globally and over 300,000 deaths reported to the WHO.[7] In comparison, there was a total of 8098 confirmed cases and 774 deaths during the SARS outbreak.[8]

Both molecular and cellular level studies have revealed that the spike glycoprotein (S protein) on the surface of the coronavirus plays a critical role in facilitating viral entry into host cells.[9,10] The S protein consists of two subunits, the S1 subunit which is responsible for receptor binding and the S2 which mediates membrane fusion of the viral and host membranes. A receptor-binding domain (RBD) of the S protein is located within the central region of the S1 subunit (Figure 1a). More specifically, the RBD consists of a core region and some loop residues that are in direct contact with the receptor. In addition, the receptor binding motif (RBM) is a subdomain of the RBD and is located between residues 437

and 508 in SARS-CoV-2 (Figure 1a and b) and between residues 424 and 494 in SARS-CoV.

It was also reported that both SARS-CoV-2 and SARS-CoV utilize a common host receptor angiotensin-converting enzyme 2 (ACE2) to infect human cells via the S protein RBD.[11−13] The sequence identity between the two RBDs was computed as 72.78%.[14] Aside from variations in the loop regions of the RBDs, mutations of several "hot spot" residues were also identified at the protein–protein interfaces (Figure 1c). Independent studies have suggested that the receptor-binding ability of coronaviruses is crucial for the susceptibility of the host to infection.[15−19] As a result, accurately calculating the change in binding strength of the RBD toward host receptors due to mutations can be useful for assessing the likelihood of coronavirus transmission between different hosts, especially the transmission to humans.

Here we present our results of alanine scanning mutagenesis of SARS-CoV and SARS-CoV-2 S RBD complexed with ACE2 using thermodynamic integration (TI).[21] The hot spot residues at the RBM were transmuted to alanine using an alchemical transformation to generate the relative binding free energy change ($\Delta\Delta G_{binding}$). These relative binding free energies represent the contribution of the perturbed residues
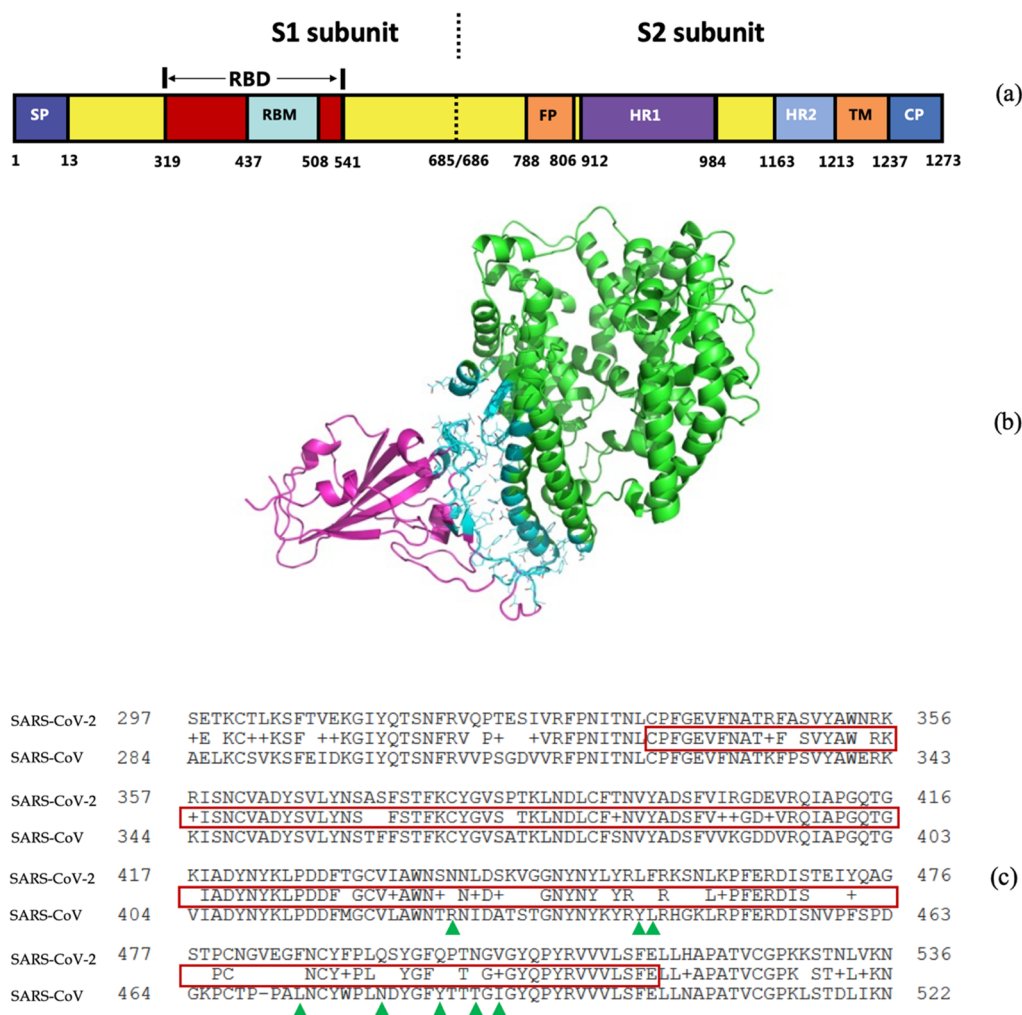
Figure 1. (a) Functional domains in the SARS-CoV-2 S protein: signal peptide (SP), receptor-binding domain (RBD), receptor-binding motif (RBM), fusion peptide (FP), heptad repeat (HR), transmembrane domain (TM), and cytoplasm domain (CP). (b) Crystal structure of the SARS-CoV-2 S RBD complexed with ACE2 (PDB ID: 6LZG[20]): ACE2 (green), SARS-CoV-2 RBD (pink), SARS-CoV-2 S RBM (cyan), and ACE2 residues at the interface (cyan). (c) Sequence alignment of the S protein RBD regions between SARS-CoV-2 and SARS-CoV. The red box indicates the conserved residues in the RBD region between the two viruses. The green triangles designate the "hot spot" residues involved in alanine scanning.

to the molecular recognition of ACE2 by the virus. The results of this study demonstrate how rigorous free energy calculations can be applied to forecast the transmission ability of novel coronaviruses. For SARS-CoV-2, long time scale MD simulations were also performed on the homology-based structure to determine the effectiveness of molecular modeling and assess its predictive power in the early stages of outbreaks when the experimental structure of the virus is not yet available. The limitations of the current computational method are also discussed.
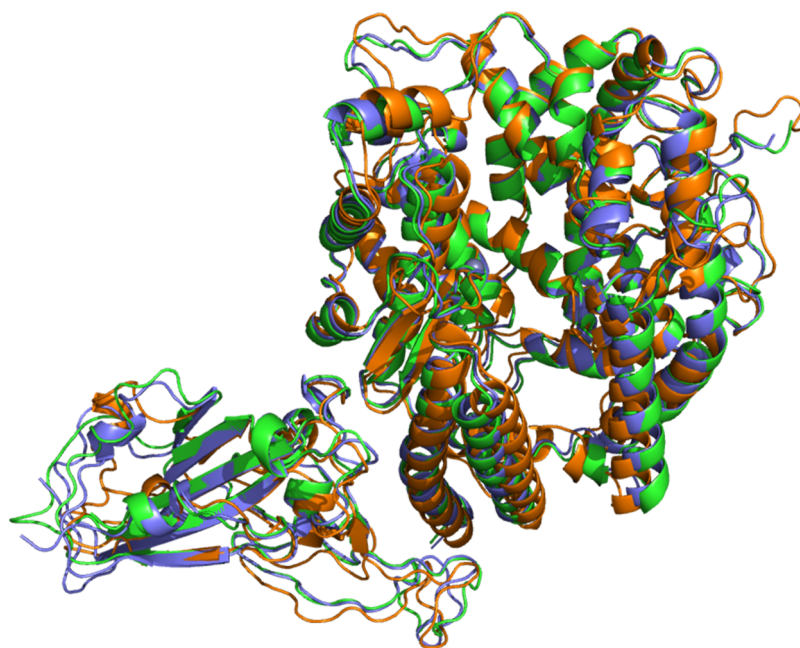
MODELING AND SIMULATION DETAILS

Homology Modeling. Given the close homology of the S protein RBD regions between SARS-CoV-2 and SARS-CoV, a model of the SARS-CoV-2 S RBD/ACE2 complex was constructed based on the X-ray structure of SARS-CoV RBD/ACE2 (PDB ID: 2AJF) using SWISS-MODEL.[14] This homology modeling was performed prior to the release of the SARS-CoV-2 RBD/ACE2 crystal structure. The QMEAN Z-score reported from SWISS-MODEL was −5.59 (Cβ: −1.13; all atom: −1.95; solvation: −2.18; torsion: −4.68). The

sequence of the SARS-CoV-2 S protein was obtained from GenBank (MN908947.3),[22] and the missing loop in the crystal structure was reconstructed via the kinematic closure (KIC) algorithm implemented in Rosetta[23−25] using default parameters. The H++ Web server was used to determine the protonation states of the titratable residues.[26] The internal and external dielectric constant was set to be 10.0 and 80.0 respectively. Salinity was set to be 0.15. Given the pK_a value of 7.4, the Lys and Arg side chains and the N-terminal residue in the complexes were set to be protonated, while the Asp and Glu side chains and the C-terminal residue were set to be unprotonated. Then we used Molprobity to add missing hydrogen atoms and optimize the rotamer states of the Asn, Gln, and His residues.[27]

Long Time Scale Molecular Dynamics Simulations. All molecular dynamics (MD) simulations were performed with the Amber18[28] suite of programs. The Amber force field ff14SB[29] and the TIP3P water model[30] were used. The homology model of SARS-CoV-2 complexed with ACE2 was solvated with an 8 Å buffer of water in a truncated octahedron box. The particle mesh Ewald (PME) scheme[31] was employed
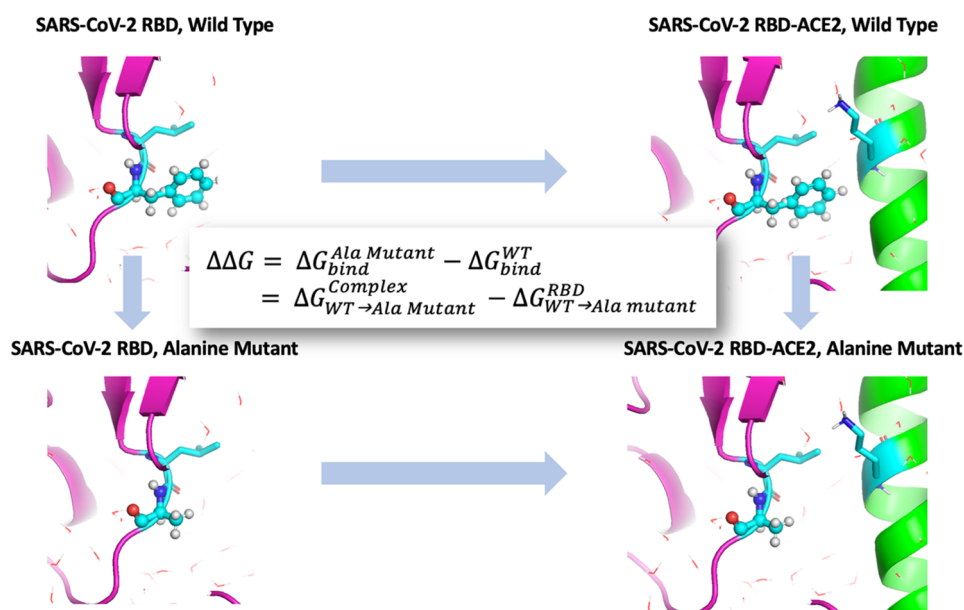
Figure 2. Thermodynamic cycle used for alanine scanning of SARS-CoV and SARS-CoV-2 RBD. Performing TI calculations in accord with the thermodynamic cycle, the relative binding affinity arising from mutating wild-type to alanine mutant can be computed as the difference between the free energy changes of the bound (right: RBD-ACE2 complex) and the unbound state (left: RBD only).

$$\Delta\Delta G = \Delta G_{bind}^{Ala\ Mutant} - \Delta G_{bind}^{WT}$$
$$= \Delta G_{WT \rightarrow Ala\ Mutant}^{Complex} - \Delta G_{WT \rightarrow Ala\ mutant}^{RBD}$$

Figure 3. Superposition of the crystal structures of SARS-CoV-2 (purple, PDB ID: 6LZG) and SARS-CoV (green, PDB ID: 2AJF) complexed with ACE2, and the last frame of the 100 ns MD simulation of the homology model (orange).

to calculate electrostatic energies. The cutoff of nonbonded interactions was set to 8 Å and the long-range dispersion correction was used for van der Waals interactions (VDW).[32]

The whole system was first minimized using a steepest descent algorithm, with the maximal number of cycles set as 10 000. Subsequently, the system was heated from 150 to 298 K over a period of 100 ps, and this was followed by an NPT equilibration for 100 ps. Position restraints of 100 kcal/mol were imposed on all heavy atoms of proteins during these two steps.

After that, the strength and scale of the positional restrains were gradually reduced to gently relax the system. An NPT

ensemble and step size of 1 fs were used in all the following equilibration steps: first, the system was equilibrated with 10 kcal/mol positional restraints on all heavy atoms for 250 ps. To relax the side chains, the system was further equilibrated over a series of simulations with 10, 1, and 0.1 kcal/mol positional restraints on all CA, C, and N atoms. Each equilibration step lasted for 100 ps. Finally, a 250 ps NPT equilibration was conducted with no restraints.

MD was then performed to relax and verify the structure obtained from homology modeling. The production phase was simulated in an NPT ensemble with the Berendsen barostat[33] for 100 ns. Temperature was controlled with a Langevin

thermostat[34] at 298 K, and the collision frequency was 1.0 ps^{-1}. Pressure was controlled by isotropic position scaling, and the pressure relaxation time was set to 0.1 ps. Hydrogen atoms were constrained using the SHAKE[35,36] algorithm. A time step of 2 fs was used for the production runs.

Relative Binding Free Energy Calculations. XFEP[37] coupled with Amber-TI[38−42] was used to calculate the relative binding free energies. The crystal structures of both SARS-CoV and SARS-CoV-2 RBD complexed with ACE2 were obtained as the starting structures of FEP calculations. An identical protein preparation and equilibration protocol for the 100 ns MD simulations described above was used for these structures. The perturbation for the alanine scanning was carried out in two steps, which consists of decoupling the coulomb and VDW interactions. To decouple the coulomb interactions, the partial charges were turned off for all the side chain atoms on the residues of interest except for the beta o carbon atom (CB). The charges on all backbone atoms N, H, C, O, CA, HA, as well as CB were unchanged. Then, the residues of interest were transformed into alanine using a dual-topology scheme with soft-core potentials[43,44] implemented in Amber-TI. For both of the perturbed residue and alanine, scmask was set to exclude all the backbone atoms and atom CB. A total of 11 λ windows were used for both the coulomb and VDW steps, with the λ values for both steps evenly spaced between 0.0 and 1.0 with an interval of 0.1, including 0.0 and 1.0. Each λ window was simulated for 4 ns and trapezoidal integration was used to obtain the ΔG. Free energy changes were calculated for both of the bound and the unbound state of the RBD, and the difference between the bound (RBD-ACE2 complex) and unbound (RBD only) calculations gives the relative binding affinity, ΔΔG (Figure 2).

■ RESULTS

The structure of the SARS-CoV-2 RBD/ACE2 complex constructed using homology modeling remained stable during the simulation. The configuration of the complex from the last frame of the MD simulation aligned reasonably well with the crystal structure (see Figure 3), having an RMSD of 1.74 Å for the heavy atoms, as shown in Figure 3. The largest structural deviation came from the RBD loop regions that were distant from the binding interface. Surprisingly, despite having only ~50% sequence similarity between SARS-CoV and SARS-CoV-2 in the RBM region, the ACE2 contacting loop did not undergo significant conformational change (Figure 3).

N479 and T487 in SARS-CoV S RBD were reported as hot spot residues sensitive to host receptor recognition in previous experimental studies.[45] Therefore, TI calculations were conducted at these two positions in order to validate our relative free energy calculation protocols. Our results showed that the N479K and T487S single point mutations reduced the binding affinity of SARS-CoV RBD by 2.1 ± 0.3 and 0.8 ± 0.3 kcal/mol, respectively. In comparison, the experimental values were reported as 2.0 and 1.8 kcal/mol.[16] The good agreements between the computed and the measured binding affinity shifts suggest that our procedure is able to yield reliable estimates of ΔΔG$_{binding}$ caused by point mutations on the coronavirus strains.

In addition, based on a cutoff distance of 4 Å, a total of 16 residues (including N479 and T487) in the SARS-CoV RBM were considered to be in close contact with ACE2.[46] Among these residues, eight residues are mutated in SARS-CoV-2 (Table 1), and six out of the eight mutated residues in SARS-

Table 1. Relative Binding Free Energy Changes of the Coronavirus S RBD/ACE2 Complex Caused by Single-Point Alanine Mutations, Calculated Using XFEP[37],[a]
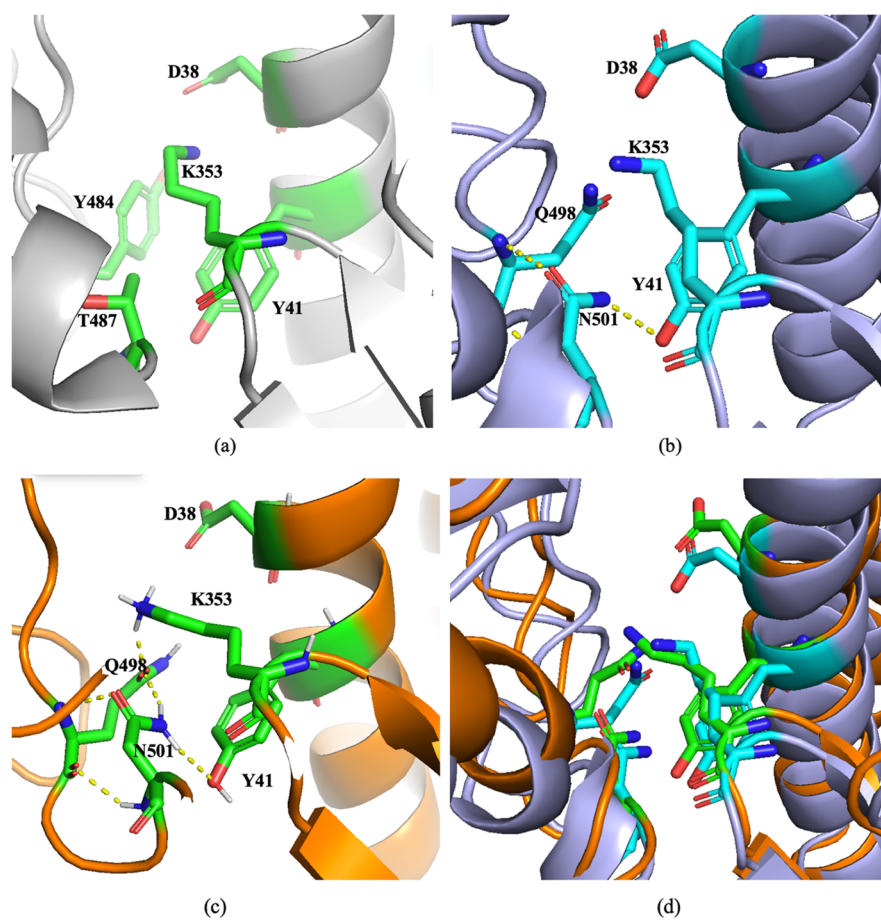
SARS-CoV		SARS-CoV-2	
mutated residue	ΔΔG$_{binding}$[b]	mutated residue	ΔΔG$_{binding}$[b]
R426	1.1 ± 0.3	N439[c]	0.6 ± 1.1
Y442	0.5 ± 1.1	L455[d]	3.3 ± 1.4
L443	1.5 ± 1.0	F456	2.2 ± 1.3
L472	0.6 ± 0.4	F486	1.9 ± 0.4
N479	0.4 ± 0.4	Q493	2.9 ± 1.4
Y484[d]	1.0 ± 0.3	Q498	0.2 ± 0.2
T487	0.7 ± 0.3	N501[d]	0.6 ± 0.9
I489	0.1 ± 0.4	V503[c]	0.2 ± 0.2

^{a}The amino acids at the same row reside at equivalent positions of the aligned sequences. ^{b}All in kcal/mol. Uncertainties were estimated from the standard deviations of three replicated runs. ^{c}These two residues N439 and V503 are further away from ACE2 than the other six mutated ones in SARS-CoV-2 RBD. ^{d}ΔΔG$_{binding}$ values of Y484A, L455A and N501A computed from the data obtained from the last 6 ns of the extended simulations are listed here.

CoV-2 remain in close contact with ACE2. These close contact residues consist of the L455, F456, F486, Q493, Q498, and N501 residues. Taking this into consideration, we performed computational point mutations on the hot spot residues that were primarily involved in the interactions at the interface. We then compared the resulting free energy changes calculated from the SARS-CoV and SARS-CoV-2 systems.

Our results of alanine scanning showed that all native to alanine mutations studied here disfavored binding. All mutations on SARS-CoV showed a mild to moderate effect, whereas three mutations, L455A, F456A, and Q493A on SARS-CoV-2 weakened the binding by over 2 kcal/mol. Furthermore, L455A and Q493A in the SARS-CoV-2 S RBD showed significantly higher (>2.0 kcal/mol) contributions to binding than their counterparts in SARS-CoV-2 RBD. To ensure that the simulation protocols were sufficient for generating well-converged relative binding affinities, we recomputed ΔΔG using a reduced number of λ windows for all mutations, with the spacing of windows enlarged from 0.1 to 0.2. It was found that the removal of λ windows led to a change of ΔΔG of no more than 0.5 kcal/mol for all point mutations except for the Y442A in SARS-CoV S RBD (Table S1). Yet increasing the total number of λ windows for Y442A from 11 to 22 using a spacing of 0.05 showed little effect on its ΔΔG estimate (Table S1). Therefore, we concluded that 11 λ windows for both the coulomb and VDW steps are sufficient for the alanine scanning of the current coronavirus systems, and adding more windows does not significantly improve the convergence.

Aside from altering the spacing of the λ windows, we also examined the ΔΔG estimates as a function of the simulation time, in both the forward and reverse λ directions (Figure S1 and S2). The difference between the cumulative ΔΔG$_{forward}$ and the ΔΔG$_{reverse}$ estimates obtained from 50% of the simulation data, referred to as the middle error, has been proposed as an analysis tool of convergence.[47] In most cases, the time forward and reverse estimates quickly approached the same final values, with a middle error of less than 0.5 kcal/mol (Figure S1 and S2), which suggests good convergence for these calculations. There were only three residues that showed relatively high convergence risks, including L455A and N501A

Figure 4. Comparison of the local interactions around (a) T487 in SARS-CoV RBD, (b) N501 in the crystal structure of SARS-CoV-2 RBD, and (c) N501 in the MD-relaxed homologous structure of SARS-CoV-2 RBD. In d, we show the local view of the crystal-MD structure alignment.

of SARS-CoV-2 and Y484A of SARS-CoV, with middle errors of 0.6, 0.7, and 1.8 kcal/mol, respectively. Therefore, for these three residues, the simulation time of each λ window was extended from 4 to 8 ns, and only the data from the last 6 ns was used to recompute $\Delta\Delta G$ (Table 1). The new middle errors were reduced to 0.1, 0.3, and 0.2 kcal/mol, respectively (Figure S3).

We then examined in more detail how each mutation affected local interactions. It was observed that the methyl group on the side chain of T487 in SARS-CoV RBD was in close contact with K353 on ACE2, which facilitated the interaction formed between K353 and D38 on ACE2 at the interface (Figure 4a).[45] Our simulation showed that the amide group on the side chain of N501 formed a hydrogen bond with Y41 on ACE2 and thus stabilized the contact between the SARS-CoV-2 RBD and ACE2. This finding was confirmed by the crystal structure (Figure 4b and c). In the homology-based simulations, Q498 disrupted the K353-D38 salt bridge through a hydrogen bond with K353, whereas the hydrogen bonding interaction was not observed in the crystal structure (Figure 4b and c). The multiple rotamer states and interacting modes of residues near N501 may explain the convergence difficulties of the calculated $\Delta\Delta G_{\text{binding}}$ of N501.

Another interesting observation was the hydrogen bonding interactions formed by Q493 in SARS-CoV-2 RBD with E35 and K31 across the interface. In the crystal structure of SARS-CoV RBD/ACE2, N479 at the corresponding position was protected within a local hydrophobic pocket formed by Y440

and Y442 (Figure 5a). In the crystal structure of SARS-CoV-2 RBD/ACE2, the longer side chain of Q493 made it capable of extending out of the pocket and reaching out to ACE2 residues to form hydrogen bonding interaction (Figure 5b). L455 and F456 on SARS-CoV-2 facilitated those interactions by making a better fit to the hydrophobic environment than Y442 and L443 in SARS-CoV. The MD simulations accurately predicted the local interactions and the orientations of these residues (Figure 5c and d).

The more favorable binding free energy contribution of F486 in SARS-CoV-2 RBD computed by FEP than L472 in SARS-CoV RBD can also be rationalized from the structures. In SARS-CoV-2, the F486 residue is more tightly packed with two other hydrophobic residues across the interface, M82 and Y83 on ACE2, simply because of the larger VDW volume of the aromatic ring than the methylene side chain of L472 (Figure 6a and b). The same local hydrophobic interactions were also captured by the homology-based MD simulations (Figure 6c and d). Residues N439 and V503 in SARS-CoV-2 RBD have shorter side chains compared with R426 and I489 in SARS-CoV RBD, respectively, which lead to little contacts with ACE2 and this was well reflected in their corresponding binding contributions.

◼ DISCUSSION

In the present work, we demonstrate that the relative binding affinity calculation obtained from alanine scanning mutagenesis is a useful tool to study the transmission ability of the
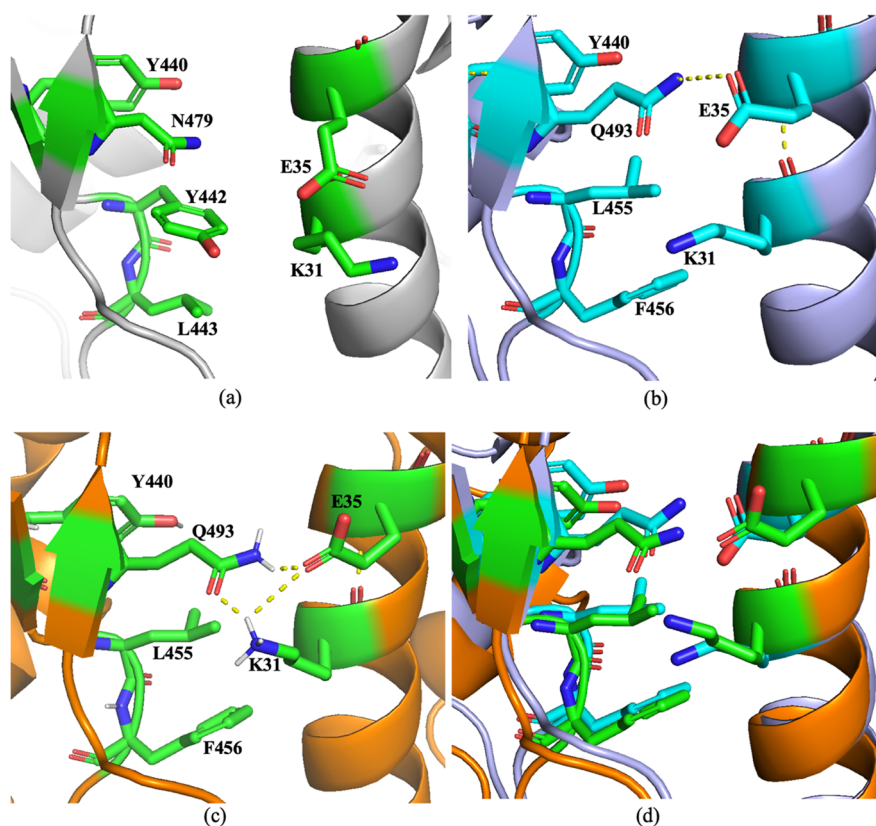
Figure 5. Comparison of the local interactions around (a) N479 in SARS-CoV RBD, (b) Q493 in the crystal structure of SARS-CoV-2 RBD, and (c) Q493 in the MD-relaxed homologous structure of SARS-CoV-2 RBD. In d we show the local view of the crystal-MD structure alignment.
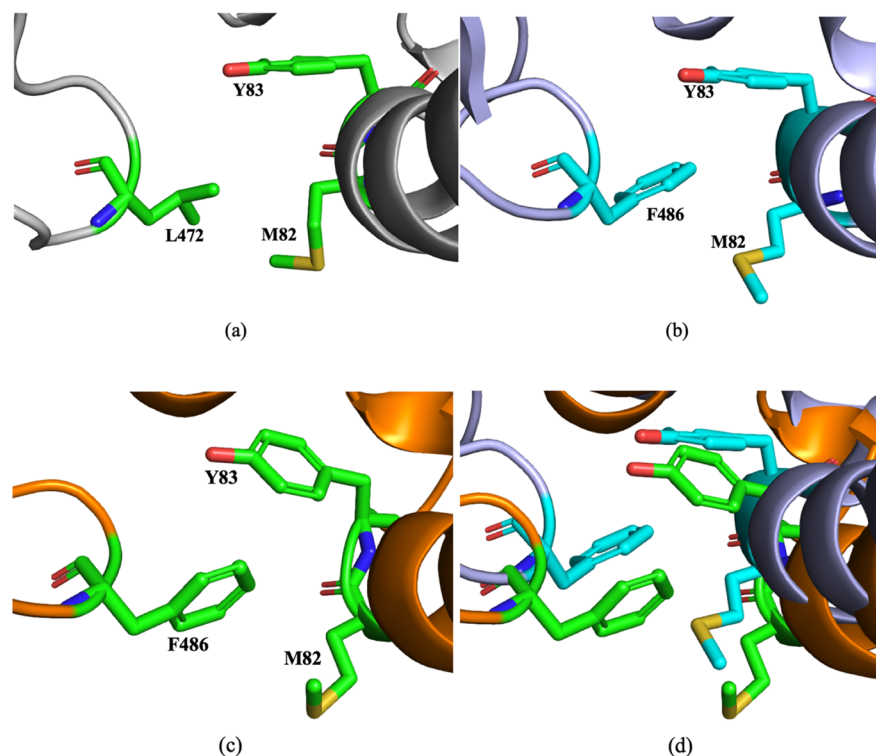
Figure 6. Comparison of the local interactions around (a) L472 in SARS-CoV RBD, (b) F486 in the crystal structure of SARS-CoV-2 RBD, and (c) F486 in the MD-relaxed homologous structure of SARS-CoV-2 RBD. In d we show the local view of the crystal-MD structure alignment.

coronavirus. The recently emerged novel coronavirus SARS-CoV-2, which is responsible for COVID-19, is genetically related to the well-studied coronavirus SARS-CoV. Our free energy results showed that, the point mutations of most hot

spot residues in SARS-CoV RBD maintained or strengthened the binding of SARS-CoV-2 to ACE2, yet further investigation is required to validate our results for the following reasons. First, MD simulations were necessary to relax the undesired contacts both in the homology model and in the crystal structure, and thus the quality of the protein force field was heavily relied on to produce conformations close to the native states. A more accurate general force field and water model could be developed and used to improve predicting binding properties.[48−50] Second, our single-point alanine-scanning calculations only yielded the contribution of the individual residue without taking the coupling effect caused by double or multiple mutations into account. For example, L455 and F456 both had their side chains pointing toward the interface between SARS-CoV-2 RBD and ACE2, and since the side chains of L455 and F456 were in proximity, it is reasonable to speculate that a significant part of the $\Delta\Delta G_{binding}$ is due to collective contributions. Furthermore, we only scanned the mutated hot spot between SARS-CoV and SARS-CoV-2 RBDs, whereas new mutations may occur at other positions that are sensitive to receptor binding. Finally, simulating the protein−protein contacts is more challenging than protein−ligand interactions, due to the larger size of the system. Therefore, enhanced sampling methods can be leveraged to improve the rotation of the protein side chains and may lower uncertainties and generate more accurate results when computational cost is not a concern.

It is rather encouraging that the protein conformations and local interactions observed based on the homology model agreed well with those produced based on the crystal structure of SARS-CoV-2. More importantly, our long time scale MD simulations were completed 3 weeks before the crystal structure of SARS-CoV-2 RBD/ACE2 was released. Information obtained from reliable modeling and simulations can be critical when a novel virus emerges and the experimental structure is not yet known. More importantly, alanine-scanning data obtained using the protein mutation workflow in XFEP could reveal the new hot spots on the novel virus and the host receptor and thus support the rational design of peptide and antibody therapeutics, as well as small-molecule drugs. In our ongoing work, we aim to carry out a thorough scanning of the interface residues and improve our prediction protocols by bringing together more accurate force field parameters, enhanced sampling, group mutations, and experimental validation.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c00679.

Decomposed pairwise interactions for the studied residues (XLSX)

Convergence plots, hydrogen bond analysis, and RMSF of SARS-CoV/SARS-CoV-2 RBDs based on the TI trajectories (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Peiyu Zhang − Shenzhen Jingtai Technology Co., Ltd. (XtalPi), Shenzhen, China 518000; Email: peiyu.zhang@xtalpi.com

Authors

Junjie Zou − Shenzhen Jingtai Technology Co., Ltd. (XtalPi), Shenzhen, China 518000

Jian Yin − Shenzhen Jingtai Technology Co., Ltd. (XtalPi), Shenzhen, China 518000; ⓞ orcid.org/0000-0001-9668-3704

Lei Fang − Shenzhen Jingtai Technology Co., Ltd. (XtalPi), Shenzhen, China 518000

Mingjun Yang − Shenzhen Jingtai Technology Co., Ltd. (XtalPi), Shenzhen, China 518000

Tianyuan Wang − XtalPi−AI Research Center (XARC), Beijing, China 100083

Weikun Wu − XtalPi−AI Research Center (XARC), Beijing, China 100083

Michael A. Bellucci − XtalPi, Cambridge, Massachusetts 02142, United States

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.0c00679

Author Contributions

∥J.Z. and J.Y. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Hui, D. S.; I Azhar, E.; Madani, T. A.; Ntoumi, F.; Kock, R.; Dar, O.; Ippolito, G.; Mchugh, T. D.; Memish, Z. A.; Drosten, C.; Zumla, A.; Petersen, E. The Continuing 2019-nCoV Epidemic Threat of Novel Coronaviruses to Global Health - The Latest 2019 Novel Coronavirus Outbreak in Wuhan, China. *Int. J. Infect. Dis.* **2020**, *91*, 264−266.

(2) Wu, J. T.; Leung, K.; Leung, G. M. Nowcasting and Forecasting the Potential Domestic and International Spread of the 2019-nCoV Outbreak Originating in Wuhan, China: A Modelling Study. *Lancet* **2020**, *395* (10225), 689−697.

(3) Gralinski, L. E.; Menachery, V. D. Return of the Coronavirus: 2019-nCoV. *Viruses* **2020**, *12* (2), 135.

(4) Gorbalenya, A. E.; Baker, S. C.; Baric, R. S.; de Groot, R. J.; Drosten, C.; Gulyaeva, A. A.; Haagmans, B. L.; Lauber, C.; Leontovich, A. M.; Neuman, B. W.; Penzar, D.; Perlman, S.; Poon, L. L. M.; Samborskiy, D.; Sidorov, I. A.; Sola, I.; Ziebuhr, J. Severe Acute Respiratory Syndrome-Related Coronavirus: The Species and its Viruses−A Statement of the Coronavirus Study Group. *BioRxiv.org* **2020**, DOI: 10.1101/2020.02.07.937862.

(5) Holmes, K. V.; Enjuanes, L. Virology. The SARS Coronavirus: A Postgenomic Era. *Science* **2003**, *300* (5624), 1377−1378.

(6) Holmes, K. V. SARS Coronavirus: a New Challenge for Prevention and Therapy. *J. Clin. Invest.* **2003**, *111* (11), 1605−1609.

(7) World Health Organization. Coronavirus Disease (COVID-19) Situation Report—133. https://www.who.int/docs/default-source/coronavirus/situation-reports/20200601-covid-19-sitrep-133.pdf?sfvrsn=9a56f2ac_4 (accessed Jun 15).

(8) World Health Organization. Summary of Probable SARS Cases with Onset of Illness from 1 November 2002 to 31 July 2003. https://www.who.int/csr/sars/country/table2004_04_21/en/ (accessed Jun 15).

(9) Li, F.; Li, W.; Farzan, M.; Harrison, S. C. Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science* **2005**, *309* (5742), 1864−1868.

(10) Simmons, G.; Reeves, J. D.; Rennekamp, A. J.; Amberg, S. M.; Piefer, A. J.; Bates, P. Characterization of Severe Acute Respiratory Syndrome-Associated Coronavirus (SARS-CoV) Spike Glycoprotein-Mediated Viral Entry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (12), 4240−4245.

(11) Kuba, K.; Imai, Y.; Rao, S.; Gao, H.; Guo, F.; Guan, B.; Huan, Y.; Yang, P.; Zhang, Y.; Deng, W.; Bao, L.; Zhang, B.; Liu, G.; Wang, Z.; Chappell, M.; Liu, Y.; Zheng, D.; Leibbrandt, A.; Wada, T.; Slutsky, A. S.; Liu, D.; Qin, C.; Jiang, C.; Penninger, J. M. A Crucial Role of Angiotensin Converting Enzyme 2 (ACE2) in SARS Coronavirus-Induced Lung Injury. *Nat. Med.* **2005**, *11* (8), 875−879.

(12) Li, W.; Moore, M. J.; Vasilieva, N.; Sui, J.; Wong, S. K.; Berne, M. A.; Somasundaran, M.; Sullivan, J. L.; Luzuriaga, K.; Greenough, T. C.; Choe, H.; Farzan, M. Angiotensin-Converting Enzyme 2 is a Functional Receptor for the SARS Coronavirus. *Nature* **2003**, *426* (6965), 450−454.

(13) Xu, X.; Chen, P.; Wang, J.; Feng, J.; Zhou, H.; Li, X.; Zhong, W.; Hao, P. Evolution of the Novel Coronavirus from the Ongoing Wuhan Outbreak and Modeling of its Spike Protein for Risk of Human Transmission. *Sci. China: Life Sci.* **2020**, *63* (3), 457−460.

(14) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296−W303.

(15) 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.

(16) 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.

(17) McCray, P. B., Jr.; Pewe, L.; Wohlford-Lenane, C.; Hickey, M.; Manzel, L.; Shi, L.; Netland, J.; Jia, H. P.; Halabi, C.; Sigmund, C. D.; Meyerholz, D. K.; Kirby, P.; Look, D. C.; Perlman, S. Lethal Infection of K18-hACE2 Mice Infected with Severe Acute Respiratory Syndrome Coronavirus. *J. Virol.* **2007**, *81* (2), 813−821.

(18) Moore, M. J.; Dorfman, T.; Li, W.; Wong, S. K.; Li, Y.; Kuhn, J. H.; Coderre, J.; Vasilieva, N.; Han, Z.; Greenough, T. C.; Farzan, M.; Choe, H. Retroviruses Pseudotyped with the Severe Acute Respiratory Syndrome Coronavirus Spike Protein Efficiently Infect Cells Expressing Angiotensin-Converting Enzyme 2. *J. Virol.* **2004**, *78* (19), 10628−10635.

(19) Qu, X. X.; Hao, P.; Song, X. J.; Jiang, S. M.; Liu, Y. X.; Wang, P. G.; Rao, X.; Song, H. D.; Wang, S. Y.; Zuo, Y.; Zheng, A. H.; Luo, M.; Wang, H. L.; Deng, F.; Wang, H. Z.; Hu, Z. H.; Ding, M. X.; Zhao, G. P.; Deng, H. K. Identification of Two Critical Amino Acid Residues of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for its Variation in Zoonotic Tropism Transition via a Double Substitution Strategy. *J. Biol. Chem.* **2005**, *280* (33), 29588−29595.

(20) Qi, J. X-ray Structure of 2019 Novel Coronavirus Spike Protein RBD and ACE2 Receptor Complex at 2.5 Å Resolution. http://nmdc.cn/?from=groupmessage#/resource/detail?no=NMDCS0000001 (accessed Feb 26).

(21) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3* (5), 300−313.

(22) Severe Acute Respiratory Syndrome Coronavirus 2 Isolate Wuhan-Hu-1, Complete Genome. https://www.ncbi.nlm.nih.gov/nuccore/MN908947 (accessed Mar 25).

(23) Mandell, D. J.; Coutsias, E. A.; Kortemme, T. Sub-Angstrom Accuracy in Protein Loop Reconstruction by Robotics-Inspired Conformational Sampling. *Nat. Methods* **2009**, *6* (8), 551−552.

(24) Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R. F.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P. Macromolecular Modeling and Design in Rosetta: Recent Methods and Frameworks. *Nat. Methods* **2020**, *17*, 665.

(25) Das, R.; Baker, D. Macromolecular Modeling with Rosetta. *Annu. Rev. Biochem.* **2008**, *77*, 363−382.

(26) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A Server for Estimating pKas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33* (Web Server issue), W368−371.

(27) Headd, J. J.; Immormino, R. M.; Keedy, D. A.; Emsley, P.; Richardson, D. C.; Richardson, J. S. Autofix for Backward-Fit Sidechains: Using MolProbity and Real-Space Refinement to Put Misfits in Their Place. *J. Struct. Funct. Genomics* **2009**, *10* (1), 83−93.

(28) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, I. V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H. *AMBER 2018*; University of California, San Francisco, 2018.

(29) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696−3713.

(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.

(31) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.

(32) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, Second ed.; Oxford University Press: 1991.

(33) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684−3690.

(34) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: the Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers* **1992**, *32* (5), 523−535.

(35) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327−341.

(36) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952−962.

(37) XFEP is the binding calculation module on XtalPi's Intelligent Digital Drug Discovery and Development (ID4) platform (to be released soon. For more information, please refer to the company website: https://www.xtalpi.com/en/core-tech/xfep/ or contact us directly.)

(38) Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542−1555.

(39) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Grand, S. L.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878−3888.

(40) Lee, T. S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58* (10), 2043−2050.

(41) Lee, T. S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D. M. Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *J. Chem. Theory Comput.* **2017**, *13* (7), 3077−3084.

(42) Mermelstein, D. J.; Lin, C.; Nelson, G.; Kretsch, R.; McCammon, J. A.; Walker, R. C. Fast and Flexible GPU Accelerated Binding Free Energy Calculations within the Amber Molecular Dynamics Package. *J. Comput. Chem.* **2018**, *39* (19), 1354−1358.

(43) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear Scaling Schemes for Lennard-Jones Interactions in Free Energy Calculations. *J. Chem. Phys.* **2007**, *127* (21), 214108.

(44) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. *J. Comput. Chem.* **2011**, *32* (15), 3253−3263.

(45) 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.

(46) 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.

(47) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the Analysis of Free Energy Calculations. *J. Comput.-Aided Mol. Des.* **2015**, *29* (5), 397−411.

(48) Yin, J.; Henriksen, N. M.; Muddana, H. S.; Gilson, M. K. Bind3P: Optimization of a Water Model Based on Host-Guest Binding Data. *J. Chem. Theory Comput.* **2018**, *14* (7), 3621−3632.

(49) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528−552.

(50) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5* (21), 3863−3871.