Article

Key Interacting Residues between RBD of SARS-CoV-2 and ACE2 Receptor: Combination of Molecular Dynamics Simulation and Density Functional Calculation

Bahaa Jawad, Puja Adhikari, Rudolf Podgornik,* and Wai-Yim Ching*

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The spike protein of SARS-CoV-2 binds to the ACE2 receptor *via* its receptor-binding domain (RBD), with the RBD−ACE2 complex presenting an essential molecular target for vaccine development to stall the virus infection proliferation. The computational analyses at molecular, amino acid (AA), and atomic levels have been performed systematically to identify the key interacting AAs in the formation of the RBD−ACE2 complex for SARS-CoV and SARS-CoV-2 with its Alpha and Beta variants. Our study uses the molecular dynamics (MD) simulations with the molecular mechanics generalized Born surface area (MM-GBSA) method to predict the binding free energy (BFE) and to determine the actual interacting AAs, as well as two *ab initio* quantum chemical protocols based on the density functional theory (DFT)

implementation. Based on MD results, Q^{493}, Y^{505}, Q^{498}, N^{501}, T^{500}, N^{487}, Y^{449}, F^{486}, K^{417}, Y^{489}, F^{456}, Y^{495}, and L^{455} have been identified as hotspots in SARS-CoV-2 RBD, while those in ACE2 are K^{353}, K^{31}, D^{30}, D^{355}, H^{34}, D^{38}, Q^{24}, T^{27}, Y^{83}, Y^{41}, and E^{35}. RBD with Alpha and Beta variants has slightly different interacting AAs due to N501Y mutation. Both the electrostatic and hydrophobic interactions are the main driving force to form the AA−AA binding pairs. We confirm that Q^{493}, Q^{498}, N^{501}, F^{486}, K^{417}, and F^{456} in RBD are the key residues responsible for the tight binding of SARS-CoV-2 with ACE2 compared to SARS-CoV. RBD with the Alpha variant binds with ACE2 stronger than the wild-type RBD or Beta. In the Beta variant, K417N reduces the binding, E484K slightly enhances it, and N501Y significantly increases it as in Alpha. The DFT results reveal that N^{487}, Q^{493}, Y^{449}, T^{500}, G^{496}, G^{446}, and G^{502} in RBD of SARS2 form pairs *via* specific hydrogen bonding with Q^{24}, H^{34}, E^{35}, D^{38}, Y^{41}, Q^{42}, and K^{353} in ACE2.
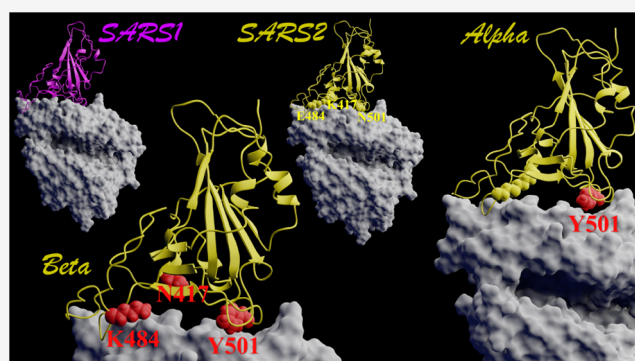
1. INTRODUCTION

The COVID-19 pandemic, caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has infected millions of people worldwide due to its rapid global spread, long incubation period, proclivity to evolve new variants, and lack of safe and effective therapies or vaccines.[1] This pandemic caused unprecedented life disruption, serious threat to public health, and enormous economic damage. In response, intensive efforts have been undertaken by scientific and medical communities to combat this pandemic. Many SARS-CoV-2 vaccine candidates have been developed and tested at various stages of clinical trials.[2−5] Three of these vaccines, the Pfizer-BioNTech, Moderna, and Johnson and Johnson, are now authorized for emergency use by the Food and Drug Administration in the United States.[6] The key elements in developing these vaccines are the rapid determination of the SARS-CoV-2 genomic sequence,[7] the three-dimensional (3D) structure of its spike protein,[8,9] and its mechanism in penetrating the human cell receptor.[9,10]

SARS-CoV-2 is an enveloped single-stranded RNA virus with spikelike glycoproteins protruding from its exterior membrane

surface forming a "corona".[1] It has four main structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N).[11] The S-protein is responsible for viral entry into the human host cell during the infection, making it a primary target for vaccine development,[12−14] repurposing antiviral drugs or discovery of new ones,[15,16] and development of therapeutic antibodies.[17,18] S-protein exists in the trimeric form, with each protomer having two functional subunits, S1 and S2. S1 contains the receptor-binding domain (RBD) responsible for direct binding to angiotensin-converting enzyme 2 (ACE2), while S2 functions in the host cell membrane fusion.[19] To engage with ACE2, RBD undergoes hingelike movements to change from down to up states for easier accessibility.[8] The nature of interaction between RBD and ACE2 not only is important for

Table 1. Summary of Models for MD Simulation and DFT Calculation

model	name of model	# of atoms in RBD	# of atoms in ACE2	# of water molecules	# of ions			total # of atoms
					Zn^{+2}	Cl^-	Na^+	
Interface-MD	SARS1	2827	9523	27 000	1	1	25	93 377
	SARS2	3001	9523	27 000	1	1	25	93 551
	Alpha	3008	9523	27 000	1	1	25	93 558
	Beta	3007	9523	27 000	1	1	24	93 556
Interface-DFT	SARS1	1116	1822				4	2942
	SARS2	1102	1822				6	2930
	Beta	1116	1822				4	2942
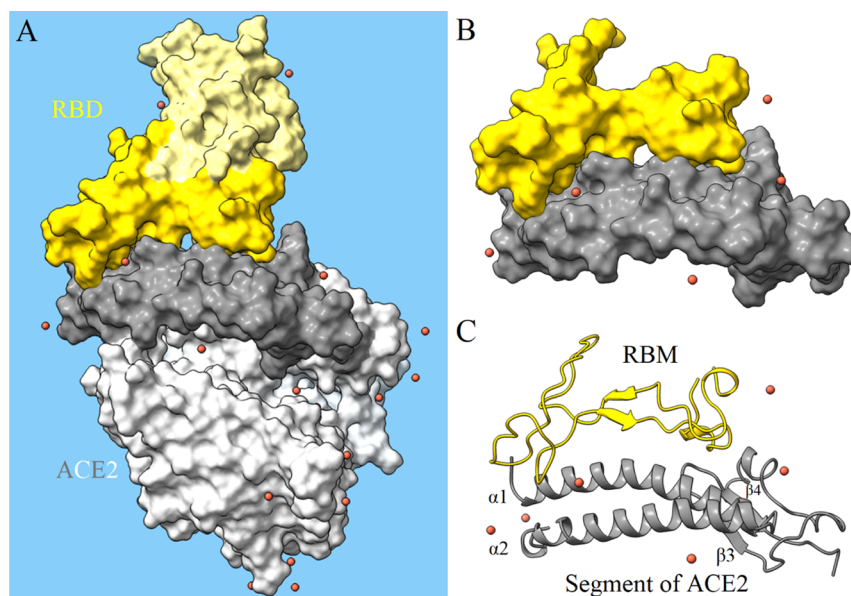
Figure 1. Interface complex models of bound SARS-CoV-2 RBD to ACE2 receptor. (A) Interface-MD model of the whole RBD−ACE2 complex (in surface representation) solvated in a water box (light blue) and neutralized with Na^+ ions (orange). Light yellow and yellow, RBD; yellow, RBM; gray, ACE2; and dark gray, segment of ACE2 used for DFT calculation. (B) Interface-DFT model of RBM and a segment of ACE2 in surface representation. (C) Ribbon structure of RBM and segment of ACE2 with $\alpha 1$, $\alpha 2$, $\beta 3$, and $\beta 4$ motifs.

understanding the initial step of virus infection but also provides necessary information to guide the effective development of drugs or vaccines.

The crucial role of the binding mechanism between RBD and ACE2 is in the infection initiation process, triggering a cascade of events and directing the focus of the research to the RBD−ACE2 interface complex. Its structure has been determined by X-ray diffraction (PDB ID: 6M0J, 6VW1),[20,21] showing that the overall ACE2-binding mode of the SARS-CoV-2 RBD, apart from its more compact form, is similar to that of the SARS-CoV RBD. However, there are no quantitative assessments of the energetics of the actual interacting residues between RBD and ACE2. While numerous biophysical and simulation studies on this complex have been conducted,[22−40] more research effort is necessary to resolve still unanswered questions. First, it is not clear how the SARS-CoV-2 RBD recognizes and binds to ACE2 and what is the pattern of the binding interaction. Second, how does the novel SARS-CoV-2 variant influence its binding pattern and explain the increased infectivity, especially of the Alpha and Beta variants? Alpha and Beta variants have been detected in the United Kingdom (B.1.1.7)[41] and South Africa (B.1.351).[42] The RBD with the Alpha variant contains a mutation of asparagine to tyrosine at position 501 (N501Y), while the Beta variant has two new mutations, K417N and E484K, in addition to N501Y. Third, it remains to be established what is the main driving force

responsible for the RBD−ACE2 complex formation. There is insufficient information on the physical factors that govern the complexation process and their functions in the receptor recognition process. Fourth, the role of the aqueous environment and its impact on the RBD−ACE2 interactions remains to be highlighted. An aqueous solvent is present at all stages of viral infection, starting with the role of droplets and aerosols in the spreading and transmission of the virus[43] and then continuing throughout the viral life-cycle processes such as replication, transcription, and genome packaging occurring in the aqueous bathing environment. Finally, quantitative information on the role of key amino acids is important for unraveling the source of the differences between the binding pattern of the coronaviruses and the ACE2 receptor. The precise identification of these residues requires more rigorous approaches using efficient molecular dynamics (MD) simulations as well as *ab initio* calculations based on the density functional theory (DFT). Recently, our group has succeeded in the largest *ab initio* quantum chemical computation involving seven structural domains of the S-protein,[44] as well as in formulating a *de novo* method to evaluate the inter-amino acid interactions in 3D of these subdomains.[45,46]

To address the questions raised above, we undertook a comprehensive and systematic study focusing on the understanding of the molecular mechanisms and specific interactions

at the interface. We used the MD simulation methodology for assessment of the binding free energy (BFE) and the *ab initio*-based DFT methodology to elucidate the binding phenomenology. Two replicate MD simulations for all-atom RBD−ACE2 complexes of SARS-CoV, SARS-CoV-2, Alpha, and Beta variants are performed with explicit water molecules over a time scale of 100 ns followed by the molecular mechanics generalized Born surface area (MM-GBSA) method to compute the BFE of the RBD−ACE2. At the same time, the per-residue and pairwise BFE decompositions are used to capture the actual interacting AAs between the individual residue in RBD and ACE2. Finally, DFT calculations are applied to a specially designed restricted interface model for detailed structural elucidation of the binding mechanism, based on the structure relaxation using the Vienna *ab initio* simulation package (VASP) and followed by applying the orthogonalized linear combination of atomic orbital (OLCAO) method to investigate the atomic-scale resolved interaction, partial charge distribution, and intra- and intermolecular binding of the RBD−ACE2 interface model similar to the one used for the S-protein.[44]

2. METHODS AND MODEL CONSTRUCTION

2.1. Construction of the Molecular Model for MD Simulations. We explicitly designed and simulated the interface complex models in a bathing solution environment, composed of water and ions. These models are labeled as the Interface-MD models, which include SARS-CoV RBD−ACE2, SARS-CoV-2 RBD(WT)−ACE2, SARS-CoV-2 RBD(Alpha)−ACE2, and SARS-CoV-2 RBD(Beta)−ACE2 complexes, and are referred to as SARS1, SARS2, Alpha, and Beta, respectively. Here, we use SARS2 as an example to explain details involved in construction of the Interface-MD models. SARS2 contains residues from T^{333} to G^{526} of RBD (194 AAs) and the residues from S^{19} to D^{615} of the ACE2 N-terminal peptidase domain (PD) (597 AAs). Also, it contains one zinc ion (Zn^{+2}), one chloride ion ($Cl^{−1}$), 25 sodium ions (Na^+), and 27 000 water molecules (Table 1 and Figure 1A). The details of creating this model are summarized as follows. The initial structure of the interface SARS2 was obtained from the crystal structure PDB (ID: 6M0J).[20] This initial structure is solvated in a periodic water box using the TIP3P explicit water model implemented in the Assisted Model Building with Energy Refinement (AMBER) package, after adding H atoms using the LEaP module.[47,48] In particular, 27 000 water molecules are added to the initial complex with 25 Na^+ ions to neutralize the total charge of the model because the ACE2 carries a total charge of $28e^−$, zinc with $2e^+$, chloride ion with $1e^−$, and RBD has a net charge of $2e^+$. The placing of the Na^+ ions was performed *via* a Coulomb potential on a grid using the LEaP program in the AMBER package. The structure of the SARS2 complex (6M0J) is used as a template to generate the Alpha and Beta models. In the Alpha model, we changed the N^{501} to Y^{501} by adopting the Dunbrack backbone-dependent rotamer library[49] implemented by UCSF Chimera.[50] Similarly, the K^{417}, E^{484}, and N^{501} are changed to N^{417}, K^{484}, and Y^{501} in the Beta model. The SARS1 complex model was built based on its crystal structure from PDB ID: 2AJF.[51] This initial structure contains similar residues of ACE2 (S^{19} to D^{615}) as SARS2 and residues of C^{323} to E^{502} of RBD, except for the six missing residues (D^{376} to N^{381}). We used the program Modeller *via* Chimera software to model these missing residues.[50,52] This structure has two copies of the complex, and we only selected Chains A and E of ACE2 and RBD, respectively. We also solvated this complex in 27 000 water molecules and 25 Na^+ ions

(see Table 1). The most recent AMBER force field ff14SB is used to represent the parameterizations of the inter- and intramolecular interactions of these Interface-MD models.[53]

2.2. Construction of the Molecular Model for DFT Calculations. For gaining deeper insights into the interface interactions, the use of *ab initio* quantum chemical methods for accurate atomic positions and interatomic interactions is necessary. However, exceptionally large-scale *ab initio* all-atom calculations are obviously impossible at present since MD simulation could involve in excess of hundreds of thousands of atoms. To this end, we have reduced the size of the interface model to a much smaller and manageable size containing only the most relevant AAs at the interface. These models are labeled as the Interface-DFT models derived from three Interface-MD models: SARS1, SARS2, and Beta (with the removal of water and some ions). Here, the SARS2 model is used as an example to show the details of its DFT model. It contains all residues of the receptor-binding motif (RBM) of SARS-CoV-2 (71 AAs from S^{438} to Y^{508}) and 117 residues from ACE2 (70 residues of $\alpha1$ and $\alpha2$ motifs from the S^{19} to I^{88} plus 47 residues from G^{319} to T^{365} of $\beta3$ and $\beta4$ motifs and with some other residues of ACE2) as well as six Na^+ ions to neutralize the system (Figure 1B,C). The SARS-CoV-2 RBM is the main functional motif in RBD that forms the interface between the S-protein and ACE2. According to the high-resolution crystal structure information,[20,21] the ACE2 segments selected for the Interface-DFT model include all of the interacting AAs from ACE2. Therefore, this model captures almost all binding properties between SARS-CoV-2 RBD and ACE2. It has a total of 2930 atoms, as summarized in Table 1. The details of SARS1 and Beta DFT models are also listed in Table 1. They are fully optimized using VASP and used as the input for the calculation of electronic structure and properties using OLCAO methods.

2.3. Molecular Dynamics (MD) Simulation. All-atom MD simulations for the Interface-MD model are carried out using the AMBER 18 simulation package in an explicit solvent with periodic boundaries.[54] Prior to MD simulations, the model underwent six stages of energy minimization using 5000 cycles of steepest descent followed by an additional 5000 cycles of a conjugate gradient to eliminate possible steric overlaps and allow the system to adapt to the chosen force field. In the first five stages, five different restraint force constants of 500, 250, 100, 10, and 1 kcal/mol-Å^2 are applied to hold the solute (RBD−ACE2 complex) in a fixed position and optimize the positions of water molecules and ions. In the final stage, the whole system is minimized without any constraints. Subsequently, a heating phase is introduced by gradually raising the temperature from 0 to 310 K for 310 picoseconds (ps) using the NVT ensemble with a weak restraint of 10 kcal/mol-Å^2 on the solute, followed by system equilibration for 5 ns without restraint at a constant pressure of 1 bar and temperature at 310 K (NPT ensemble) to reach the required density. Finally, two replicate NPT production MD simulations over 100 ns (200 ns in total) are initiated at constant pressure (1 bar) and temperature (310 K) using the same equilibrated starting coordinates. During the production runs, the atomic coordinates from the trajectory are saved every 2 ps for subsequent BFE analysis. The following settings are used in all equilibration and production protocols: Langevin dynamics for temperature scaling, 2 ps as the pressure relaxation time, and long-range electrostatic interactions with the particle mesh Ewald (PME) method.[55] Both the direct space PME and the Lennard-Jones cutoffs are set at 10 Å. The SHAKE algorithm is applied for

constraining all bonds involving hydrogen atoms with a 2 fs time step.[56] All of these protocols are conducted using the PMEMD.CUDA module in AMBER.[57,58]

2.4. Binding Free Energy (BFE) Calculation. We have used the most well-known end-point free energy method, the MM-GBSA method, to calculate the BFE between RBD and the ACE2 receptor. It combines the molecular MM energies with the GBSA continuum solvation approach. The BFEs are calculated utilizing the single-trajectory protocol (STP) of the MM-GBSA method. The common procedure for performing BFE calculations using MM-GBSA begins with running MD simulations with an explicit solvent to generate the conformational ensemble. In this step, the gas-phase energies (MM) are calculated from the interactions consistent with the force field and based on the structures that are extracted from the generated MD trajectories. Subsequently, explicit water molecules and counterions are removed, and the implicit GBSA solvent model is utilized to compute the solvation energy at 0.15 M ionic concentration. Finally, the solute conformational entropy change is estimated using the quasi-harmonic approximation. In the STP, all ensembles can be extracted from a single MD simulation of the bounded RBD−ACE2 complex to obtain the average ensemble of the native ACE2 receptor and RBD. BFE is determined as the difference between the free energies of the bound RBD−ACE2 complex ($G_{COM, sol}$) and the unbound states of ACE2 ($G_{ACE2, sol}$) and RBD ($G_{RBD, sol}$)[59−61]

$$\Delta G_{bind} = G_{(COM,sol)} - G_{(ACE2,sol)} - G_{(RBD,sol)} \quad (1)$$

Each term in eq 1 can be computed from contributions of interactions and expressed as

$$G = E_{MM} + G_{sol} - TS \quad (2)$$

Thus, eq 1 can be written as

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S = \Delta G_{vac} + \Delta G_{sol} \quad (3)$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW} \quad (4)$$

$$\Delta G_{sol} = \Delta G_{GB} + \Delta G_{SA} \quad (5)$$

$$\Delta G_{SA} = \gamma \cdot SASA + b \quad (6)$$

$$\Delta G_{vac} = \Delta E_{MM} - T\Delta S \quad (7)$$

Here, ΔE_{MM}, ΔG_{sol}, and $-T\Delta S$ represent the changes in the gas-phase MM energy, solvent-free energy, and conformational entropy upon binding, respectively. ΔE_{MM} is the sum of the changes in the bonded energy (ΔE_{int}), the nonbonded electrostatic energy (ΔE_{ele}), and the van der Waals energy (ΔE_{vdW}). ΔG_{sol} is divided into an electrostatic or polar solvation energy part (ΔG_{GB}) and a nonelectrostatic or nonpolar part (ΔG_{SA}) between the solute and the continuum solvent. $\Delta G_{GB/PB}$ is typically computed using either the generalized Born (GB) model, as in this study, or the Poisson−Boltzmann (PB) model, whereas ΔG_{SA} follows from a linear dependence on the solvent-accessible surface area (SASA).[62] The total electrostatic contribution to free energy is the sum of ΔE_{ele} and ΔG_{GB} ($\Delta G_{ele} = \Delta E_{ele} + \Delta G_{GB}$). The entropy contribution ($-T\Delta S$) is the sum of the change in translational, rotational, and vibrational entropies.[61] The first two terms are determined from the standard statistical mechanical formula, while the vibrational term can be approximated using a normal-mode analysis or a quasi-harmonic approximation, as in this study.[61] Once the MD

simulations are completed, snapshots are taken for every 10 ps over the whole 100 ns. In total, 5000 snapshots are extracted for the BFE postprocess analysis. This analysis has been carried out using the MM-GBSA method *via* the MM-PBSA.py module of AMBER.[61] The model developed by Onufriev et al. (GB^OBC, where GB = 2) is employed as the GB model, and a set of radii mbondi2 is prepared.[63] A surface tension coefficient (γ) of 0.0072 kcal/mol-Å² and a zero correction constant ($b = 0$) are employed to calculate ΔG_{SA}. The value of the exterior dielectric constant of water was set to 78.3, whereas three values of protein dielectric constants were tested ($\varepsilon_i = 1$,[64] 5, and 10) to investigate their impact on the binding process, especially the electrostatic interactions. Only five snapshots along simulations are used to evaluate $-T\Delta S$ at 310 K.

Two types of BFE decomposition have been undertaken to dissect the calculated BFE of the RBD−ACE2 complex in terms of an individual residue coined as "per-residue BEF decomposition" and residue−residue contributions as "pairwise BFE decomposition".[61] They can provide critical information about the local dominant interactions between the AAs in RBD and ACE2, especially those located at the interface. These decomposition analyses are carried out using the MM-PBSA.py module. A key aspect of these analyses is the ability to further break down the decomposition energy for each AA or AA−AA pair into backbone, side-chain, and total contributions and/or into interaction components (ΔE_{vdW}, ΔE_{ele}, *etc.*).

2.5. Structural Relaxation for DFT Calculation Using VASP. The Interface-DFT models are fully relaxed using the Vienna *ab initio* simulation package (VASP), known for its efficiency in structure optimization.[65] We use the projector augmented wave (PAW) method with the Perdew−Burke−Ernzerhof (PBE) exchange correlation functional[66] within the generalized gradient approximation (GGA). The input parameters used in VASP are as follows: energy cutoff 500 eV, electronic convergence of 10^{-4} eV, force convergence criterion for ionic steps at -10^{-2} eV/Å, and a single k-point sampling. For the optimization, there is complete freedom for ionic position but not for cell volume and cell shape. All VASP relaxations were carried out at the National Energy Research Scientific Computing (NERSC) facility at the Lawrence Berkeley Laboratory with special allocations and at the Research Computing Support Services (RCSS) of the University of Missouri System. The computational resources used for the structural relaxation are quite substantial because of the high accuracy required in the final structure and the slow convergence for the large complex biomolecular systems.

2.6. Electronic Structure and Interatomic Bonding Using OLCAO. The electronic structure and interatomic interactions of the Interface-DFT model are calculated using the all-electron orthogonalized linear combination of atomic orbital (OLCAO) method,[67] developed in-house. The efficacy of using these two different DFT codes is well documented[68−70] and is especially beneficial for large complex biomolecular systems such as the SARS-CoV-2 virus. The key feature of the OLCAO method is the provision for the effective charge (Q^*) on each atom and the bond order (BO) values $\rho_{\alpha\beta}$ between any pair of atoms. They are obtained from the *ab initio* wave functions with atomic basis expansion

$$Q_\alpha^* = \sum_i \sum_{m,occ} \sum_{j,\beta} C_{i\alpha}^{*m} C_{j\beta}^m S_{i\alpha,j\beta} \quad (8)$$

Table 2. Predicted BFE (kcal/mol) and Energetic Components at 0.15 M Salt and $\varepsilon_i = 1$ for RBD−ACE2 Complexes[a]

energy	SARS1		SARS2		Alpha		Beta	
	ave.	SE (SEM)	ave.	SE (SEM)	ave.	SE (SEM)	ave.	SE (SEM)
ΔE_{vdW}	−89.52	0.1 (0.1)	−90.10	0 (0.1)	−102.26	2.6 (0.2)	−101.57	1.9 (0.1)
ΔE_{ele}	−688.31	1.7 (1.1)	−700.92	0.6 (0.6)	−696.94	3.3 (0.7)	−844.02	5.4 (0.7)
ΔE_{MM}	−777.84	1.6 (1.1)	−791.03	0.6 (0.7)	−799.20	0.7 (0.7)	−945.59	3.4 (0.7)
ΔG_{GB}	737.51	1.7 (1)	748.49	0.5 (0.6)	755.39	1.2 (0.7)	903.33	4.1 (0.7)
ΔG_{SA}	−13.00	0 (0)	−13.21	0 (0)	−14.35	0.1 (0)	−14.45	0 (0)
ΔG_{sol}	724.51	1.7 (1)	735.28	0.5 (0.6)	741.04	1.3 (0.7)	888.88	4.1 (0.7)
ΔG_{ele}	49.19	0.1 (0.1)	47.57	0.1 (0.1)	58.45	2.1 (0.2)	59.31	1.3 (0.2)
$-T\Delta S$	42.52	0.2	42.89	0	43.50	0.2	43.2	0.5
ΔG_{vac}	−735.32	1.7 (1.1)	−748.14	0.6 (0.7)	−755.70	0.9 (0.7)	−902.39	3.9 (0.7)
ΔG_{bind}	−10.81	0.1 (0.2)	−12.86	0.1 (0.1)	−14.66	0.4 (0.1)	−13.52	0.2 (0.1)

[a]Ave. is the averaged value for both MD simulations. From each MD simulation, the energy is estimated based on the 5000 snapshots. SE is a standard error from both MD runs. SEM is a standard error of the mean that is computed based on 5000 snapshots of each run.

$$\rho_{\alpha\beta} = \sum_{m,occ} \sum_{i,j} C_{i\alpha}^{*m} C_{j\beta}^{m} S_{i\alpha,j\beta} \tag{9}$$

In the above equations, $S_{i\alpha,j\beta}$ are the overlap integrals between the ith orbital in the αth and the jth orbital in the βth atom. $C_{j\beta}^{m}$ are the eigenvector coefficients of the mth occupied molecular orbital level. The partial charge (PC) or ($\Delta Q_\alpha = Q_\alpha^0 - Q_\alpha^*$) is the deviation of the effective charge Q_α^* from the neutral atomic charge Q_α^0 on the same atom α. The BO quantifies the strength of the bond between two atoms and usually scales with the bond length (BL). The BL should be more accurately interpreted as the distance of separation of the two atoms since the BO value is influenced by the surrounding atoms. The calculations of PC and BO are based on the Mulliken scheme.[71,72]

The BO $\rho_{\alpha\beta}$ in eq 9 is further extended to quantify the bonding strength between a pair of amino acids (u, ν), called the amino acid−amino acid bond pair (AABP).[46] In many cases, the use of AABP is more useful than interatomic bonding between a pair of atoms for biomolecular systems.

$$AABP(u, \nu) = \sum_{\alpha \in u} \sum_{\beta \in \nu} \rho_{\alpha i, \beta j} \tag{10}$$

AABP considers all possible bondings between two amino acids, including both covalent and hydrogen bonding (HB).

3. RESULTS AND DISCUSSION

We investigate the interface interaction between RBD and ACE2 for SARS1, SARS2, Alpha, and Beta complexes using both MD simulations and DFT calculations. For MD simulations, all-atom RBD−ACE2 complexes in an explicit solvent are created (see Section 2.1) and labeled as the Interface-MD models with the following names: SARS1, SARS2, Alpha, and Beta. For DFT calculations, the smaller interface models are designed based only on the relevant interacting AAs at the interface (see Section 2.2) and labeled as the Interface-DFT models, consistent with the MD designations.

3.1. MD Results and Discussion. *3.1.1. Binding Free Energy between RBD and ACE2.* We performed two replicate MD runs for the Interface-MD model over the time range of 100 ns (see Sections 2.3 and 2.4). The root-mean-square deviation (RMSD) and the root-mean-square fluctuation (RMSF) of these two simulations are presented in the Supporting Information (SI, Figures S1 and S2). The STP of the MM-GBSA method[59] was adopted to calculate the BFE at 310 K (37 °C) under a neutral pH and 0.15 M univalent NaCl salt

concentration. Table 2 lists the net BFE (ΔG_{bind}) with their energetic components at the protein dielectric constant (ε_i) of 1. Our predicted values of ΔG_{bind} for SARS1 and SARS2 are −10.81 ± 0.2 and −12.86 ± 0.1 kcal/mol, respectively, which are in excellent agreement with the experimentally determined binding affinities of −10.7 and −11.83 kcal/mol using the surface plasmon resonance (SPR) assay.[20] More specifically, the measured values of the equilibrium dissociation constant (K_D) of the interface complex are 31 and 4.7 nM for SARS1 and SARS2, respectively,[20] which are equivalent to ΔG_{bind} values of −10.7 and −11.83 kcal/mol at 37 °C (i.e., $\Delta G = RT \ln(K_D)$, where R is the gas constant). Our calculated value is thus in line with the general trend that the binding affinity of RBD−ACE2 in SARS-CoV-2 is higher than that of SARS-CoV. This is the main reason why SARS-CoV-2 is much more contagious than SARS-CoV. The BFEs for Alpha and Beta are −14.66 ± 0.4 and −13.52 ± 0.2 kcal/mol, respectively. Our results conclude that the RBD with the Alpha variant binds strongly with ACE2 than the wild-type RBD or the one with the Beta variant. This is in line with the experimental observations.[73,74] This demonstrates increased transmissivity and infectivity of the Alpha variant compared to the wild-type SARS-CoV-2 or Beta variant.

Table 2 lists the various energy components contributing to ΔG_{bind} (see Section 2.4). Since the STP approach of the MM-GBSA method was used in our calculations, there is no contribution from internal energy, or $\Delta E_{int} = 0$. The sign of the energetic components that reflects favorability is uniform across all models. The nonbonded interactions of ΔE_{vdW} and ΔE_{ele} both favorably contribute to form a stable RBD−ACE2 complex for all MD models. Such a highly favored ΔE_{ele} behavior can be explained by the strong electrostatic attractions between the positively charged AAs of RBD and the negatively charged AAs of ACE2. Thus, strong interaction at the interface complex is electrostatic in origin, which agrees with the recent study by Nguyen et al.[35] However, ΔE_{ele} is completely overshadowed by ΔG_{GB}, leading to an overall unfavorable contribution of the total ΔG_{ele} (Table 2). Upon binding in the complex, the solvent−solvent and solvent−solute interactions are reorganized as are certain polar and/or charged AAs at the interface that become buried in the protein core and thus hidden away from the bathing solution, resulting in a largely unfavorable ΔG_{GB} and favorable ΔG_{SA}. The ΔG_{ele} is quite sensitive to the choice of the internal dielectric constant of the protein, as shown in Figure S3. Larger ε_i causes a significant drop in the value of ΔG_{ele}, making it favorable, but it still opposes the binding even at $\varepsilon_i = 10$. This enhanced favorability in ΔG_{ele} results in a tighter binding
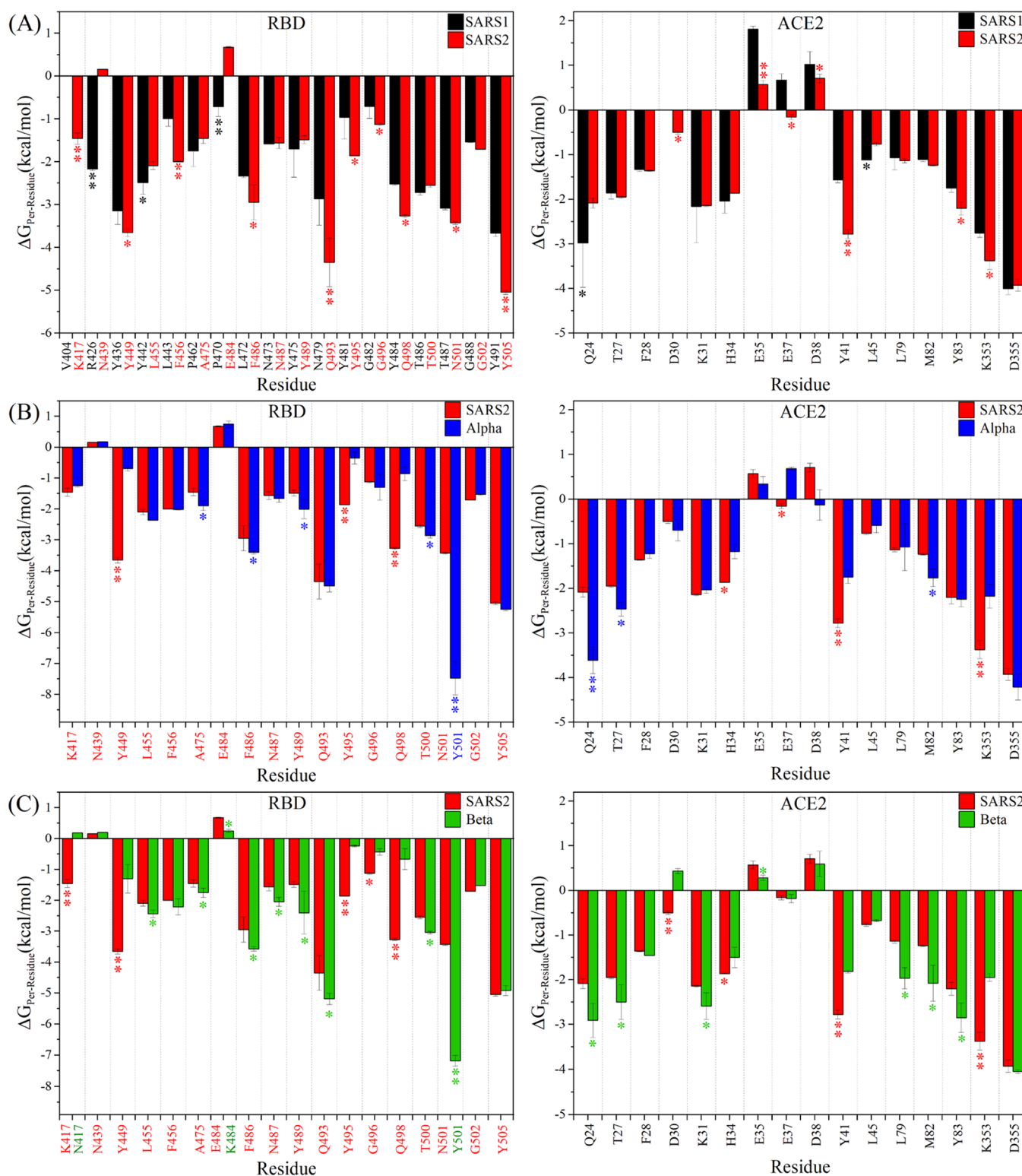
Figure 2. Per-residue energy decomposition comparison of the key interacting AAs of the RBD−ACE2 complex of Group A (with some important AAs of Groups B and C) at a protein dielectric constant of 1. The left panel shows that the interacting AAs belong to RBD, and the right panel shows that the interacting AAs in ACE2. (A) SARS1 vs SARS2, (B) SARS2 vs Alpha, and (C) SARS2 vs Beta. The significant difference larger than −0.3 kcal/mol but less than −1 kcal/mol is labeled with one asterisk, and the more significant differences that are larger than −1 kcal/mol are labeled with double asterisks.

between RBD and ACE2 (Figure S3), an observation well consistent with general trends regarding the impact of the internal dielectric constant of proteins on the electrostatic interaction.[75] Additionally, our results predict that increasing

the ε_i beyond 10 could result in having a ΔG_{ele} favoring the binding. The total entropic contribution $(-T\Delta S)$ is important and in the same range in models, reflecting the significant entropic cost of confining the RBD with little movement when

associated with ACE2. Finally, the ΔG_{bind} can be further partitioned in terms of the nature of the interactions in a vacuum (ΔG_{vac}) and/or a solvent (ΔG_{sol}). This pattern indicates that ΔG_{vac} dominates and stabilizes the dimer association between RBD and ACE2, while the aqueous solution opposes it. However, the unavoidable solvent effect cannot be ignored. When the energetic components of SARS1 and SARS2 are compared, it becomes clear that the source of tighter binding in SARS2 originates from its more favorable ΔG_{ele} of 47.57 kcal/mol compared to 49.2 kcal/mol of SARS1. This result fully agrees with that of other studies.[33−35] On the other hand, the van der Waals interaction (ΔE_{vdW}) plays a crucial role in the tighter binding of both Alpha and Beta as compared to SARS2 or SARS1 (Table 2).

From the above analysis, we can conclude that the long-range Coulombic electrostatic interaction (ΔE_{ele}) is responsible for initiating the recognition process between the RBD and ACE2, while the attractive component of the van der Waals interaction (ΔE_{vdW}) is the main factor to stabilize the RBD−ACE2 complex. Interestingly, the ΔE_{vdW} is also the main force to stabilize other complexes such as drug−DNA,[59,60] ligand−protein,[76] and protein−protein.[77] The salient feature of this trend can provide guidance on what should be targeted when a new drug is designed.

3.1.2. BFE Decomposition. Two BFE decompositions have been used to further explore the binding mechanism of the RBD−ACE2 complex and to identify the essential interacting AAs between them. First, per-residue BFE decomposition is performed to capture the energetic contribution from every individual residue. This is followed by pairwise BFE decomposition to further investigate the key AAs capable of forming AA−AA pairs. Unlike other analyses from structural studies,[20,21] these decompositions, to be discussed separately, can quantify the energetic interaction for the key AAs and AA−AA pairs in the complex. Note that the entropic contributions are not involved in these analyses.[60]

3.1.3. Per-Residue BFE Decomposition. We divide the interacting AAs in the complex into three groups: "group A" has highly favored *per-residue* binding ($\Delta G \leq -1$ kcal/mol); "group B" has ΔG of -1 to -0.15 kcal/mol; and "group C" has $\Delta G \geq 0.15$ kcal/mol. We ignore the interactions stemming from thermal fluctuation. These three groups are summarized in detail as follows.

3.1.3.1. Group A. There are 15 (12), 16 (11), 13 (11), and 12 (11) AAs in RBD (ACE2) for SARS1, SARS2, Alpha, and Beta, respectively, which contribute significantly to the total BFE. When comparing the interaction spectra of SARS1 and SARS2, we observe that six mutations from SARS1 to SARS2 (V^{404} to K^{417}, L^{443} to F^{456}, L^{472} to F^{486}, N^{479} to Q^{493}, Y^{484} to Q^{498}, and T^{487} to N^{501}) enhance the binding, while two mutations (R^{426} to N^{439} and Y^{442} to L^{455}) reduce it (labeled with asterisks in Figure 2A). Even unmutated AAs at RBD such as Y^{449}, Y^{495}, G^{496}, and Y^{505} of SARS2 increase the binding. All of these AAs in RBD are located in RBM, except for V^{404} (SARS1) or K^{417} (SARS2). Unlike V^{404} (SARS1), K^{417} (SARS2) forms a salt-bridge with D^{30} of ACE2. ΔG of $K^{417} = -1.46 \pm 0.13$ kcal/mol with a favored ΔG_{ele} of -1.18 ± 0.13 kcal/mol (Table S1). Residues Y^{505} (Y^{491}), Q^{493} (N^{479}), Y^{449} (Y^{434}), N^{501} (T^{487}), Q^{498} (Y^{484}), F^{486} (L^{472}), and T^{500} (T^{486}) of SARS2 (SARS1) have the most favored binding with $\Delta G < -2.55$ (-2.4) kcal/mol. They form interface hotspots with ACE2 in both coronaviruses, but they have different strengths. For the AAs in ACE2 in both SARS1 and SARS2, six are common in $\alpha 1$ (Q^{24}, T^{27}, F^{28}, K^{31}, H^{34}, and Y^{41}) with one

addition in SARS1 (L^{45}), and three are common in $\alpha 2$ (L^{79}, M^{82}, and Y^{83}) motifs. The loop connecting $\beta 3$ and $\beta 4$ motifs contributes *via* two highly interacting AAs (D^{355} and K^{353}). The obvious difference between contributing ACE2 in SARS1 vs SASR2 is mainly stemming from Y^{41}, Y^{83}, and K^{353} (Figure 2A). Most of these interacting AAs in RBD or ACE2 are also reported by crystal structure data.[20,51] As an example, Table S1 lists the per-residue BFE analysis in terms of the energetic components of SARS2.

Figure 2B compares the contribution from RBD and ACE2 in both SARS2 and Alpha models. Y^{501} of Alpha contributes twice as much as N^{501} of SARS2 (-7.5 vs -3.5 kcal/mol). In addition, the N501Y mutation promotes the binding of some RBD AAs, such as T^{500}, Y^{489}, F^{486}, and L^{475}, but it interrupts the binding of others, such as Q^{498}, Y^{495}, and Y^{449}. In ACE2, this mutation also enhances the binding of M^{82}, T^{27}, and Q^{24}; however, it influences the others like K^{353}, Y^{41}, and H^{34} (marked with asterisks in Figure 2B). At first glance, the effect of this mutation on reducing the favorability of K^{353} and Y^{41} as compared to SARS2 appears to contradict previous findings.[74,78] As we will see in the next Section, this influence does not arise from the actual pair of Y^{501} with its partners in ACE2, but it comes from the disruption in the interaction of its neighboring AAs, specifically Q^{498}. These observations of N501Y are also true when comparing the Beta vs SARS2 model (Figure 2C). The K417N mutation in Beta reduces the binding due to loss of the salt-bridge between K^{417} (SARS2) and D^{30} of ACE2, while the E484K mutation (i.e., Group C) provides a slight increase in the binding *via* forming an ion pair with E^{75} of ACE2. The other consequences of these mutations will be discussed in Groups B and C.

3.1.3.2. Group B. Figure S4A reveals that the mutation from T^{433} (SARS1) to G^{446} (SARS2) increases the binding of SARS2, which is not observed in the Alpha or Beta models. On the other hand, the mutation from P^{470}, W^{476}, and L^{478} of SARS1 to E^{484}, F^{490}, and L^{492} of SARS2 does the opposite, as shown in Figures 2A and S4A, respectively. Formation of an ionic pair between E^{329} from ACE2 and R^{426} at RBD in SARS1 is unique. D^{30} in SARS2 and Alpha is essential to form the salt-bridge with K^{417}, as shown in Figure 2. The per-residue BFE for SARS2 is shown in Table S2. There are notable influences in Alpha and Beta models, particularly at Y^{453} and T^{478} of RBD and A^{386} and A^{387} of ACE2, as shown in Figure S4A. Interestingly, the T^{478} changes to K^{478} in the B.1.1.222 variant.[79]

3.1.3.3. Group C. Even though E^{35}, E^{37}, and D^{38} in ACE2 of SARS2 have a positive sign of per-residue BFE (Figure 2), they contribute in a more favorable way than in SARS1. In Alpha and Beta models, E^{393} in ACE2 favorably contributes to the BFE. In addition, some AAs with charged side chains located outside RBM could contribute to the binding in all models (Figure S4B). Our analysis of energetic components of per-residue BFE in SARS2 (Table S3) reveals that they are driven by unfavorable ΔG_{ele} and ΔG_{GB}. None of them were reported in the X-ray studies, and only the E^{35}, D^{38}, R^{357}, and R^{393} of ACE2 were mentioned.[20,21] This can be attributed to the cutoff distance of 4 Å used in the experiment, while our calculation has no such constraints. Additionally, our findings show that the long-range electrostatic interactions, which are not considered in X-ray studies, are the main driving force for the interaction in group C. Some of these AAs have also been identified in other studies.[29,30,35]

3.1.4. Pairwise BFE Decomposition. To gain more details on the interacting AA−AA pairs between RBD and ACE2, we apply the pairwise BFE decomposition scheme. The pair interaction
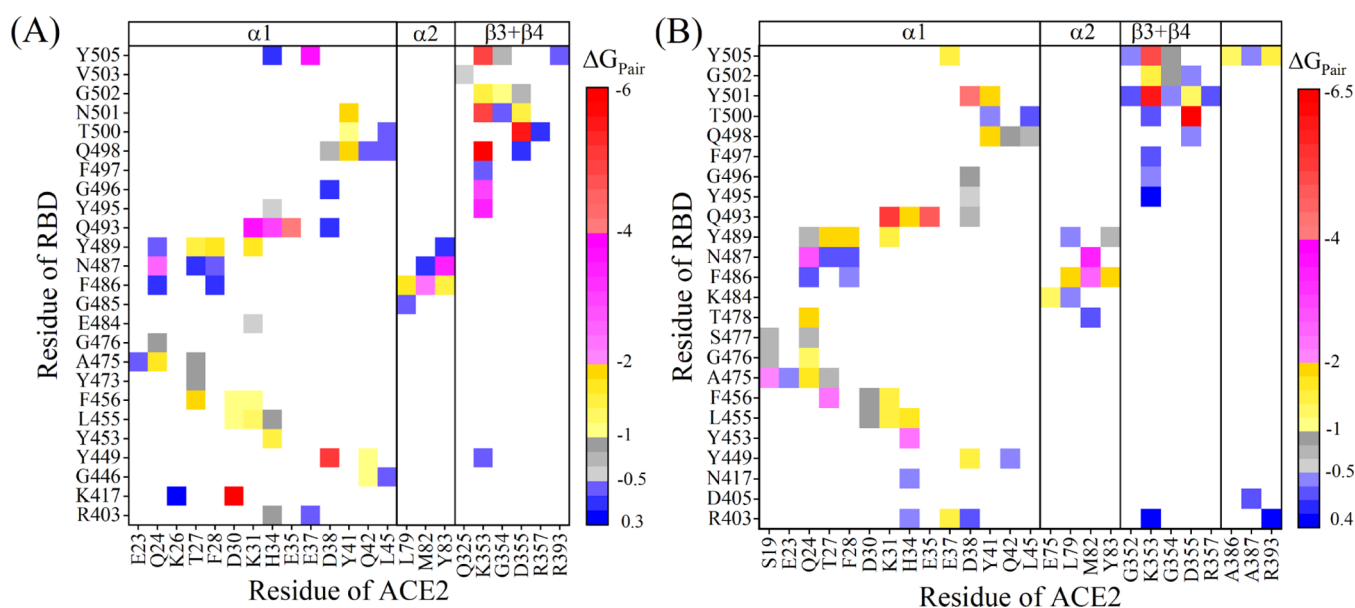
Figure 3. AA−AA interaction pair map between RBD and ACE2 of SARS2 (A) and Beta (B). Each square cell represents one pair for the intersection AA from RBD on the vertical axis and AA from ACE2 on the horizontal axis. These pairs have different strengths based on ΔG_{Pair}.

maps of the RBD−ACE2 for SARS2 and Beta are shown in Figure 3A and B, respectively. Similar pair interaction maps for SARS1 and Alpha are shown in Figure S5A,B.

The energetic components for interaction pairing in SARS1, SARS2, Alpha, and Beta are listed in Tables S4−S7, respectively. Some AAs that have been assigned in the previous Section are not engaged in AA−AA pairs, particularly those from Groups B and C. Quantitatively, the number of pairings for SARS1, SARS2, Alpha, and Beta are 65, 70, 80, and 77, respectively. In SARS2, seven pairs ($K^{417}:D^{30}$, $Y^{449}:D^{38}$, $Q^{493}:E^{35}$, $Q^{498}:K^{353}$, $T^{500}:D^{355}$, $N^{501}:K^{353}$, and $Y^{505}:K^{353}$) shown as red square cells have very strong $\Delta G \leq -4$ kcal/mol (Figure 3A). The first five pairs have favorable ΔG_{ele} (Table S5) due to formation of HBs (Table S8). The other two pairs ($N^{501}:K^{353}$ and $Y^{505}:K^{353}$) have similar favorable contributions from ΔE_{vdW}, ΔG_{ele}, and ΔG_{SA} (Table S5). The $N^{501}:K^{353}$ pair is stabilized by HBs but has a low HB occupancy. X-ray study[20] did not report any of these HBs, and only one has been mentioned elsewhere.[36] The $Y^{505}:K^{353}$ pair is driven by the π−alkyl interaction between the π-orbital in the phenol ring on Y^{505} and the CG carbon atom on K^{353}. There are also nine other strong pairs ($F^{486}:M^{82}$, $N^{487}:Q^{24}$, $N^{487}:Y^{83}$, $Q^{493}:K^{31}$, $Q^{493}:H^{34}$, $Y^{495}:K^{353}$, $G^{496}:K^{353}$, $N^{501}:Y^{41}$, and $Y^{505}:E^{37}$, purple cells in Figure 3A) with $\Delta G > -4$ but ≤ -2 kcal/mol. Except for $F^{486}:M^{82}$, all pairs are held together by HBs (Table S8). The attractive components of ΔG_{ele} are the main force for their interactions, especially for $N^{487}:Y^{83}$, $Q^{493}:K^{31}$, $Y^{495}:K^{353}$, $G^{496}:K^{353}$, and $Y^{505}:E^{37}$ pairs (Table S5). On the other hand, the $F^{486}:M^{82}$ pair is purely driven by favorable ΔE_{vdW} and ΔG_{SA} arising from π−alkyl and π−sulfur interactions between the π-orbital in the benzene ring in the F^{486} side chain and the CG carbon atom and the SD sulfur atom of M^{82}. The sequence alignment between RBD of SARS1 and SARS2 is shown in Figure S6. On comparing the pairs between SARS1 (shown in Figure S5A or Table S4) and SARS2 (shown in Figure 3A) or Table S5, we observed the following points. First, K^{417} (SARS2) forms the most attractive electrostatic pair with D^{30} of ACE2, but a hydrophobic residue of V^{404} in SARS1 does not pair with D^{30}.[20,36] Second, $F^{456}:K^{31}$, $F^{456}:T^{27}$, $F^{486}:Y^{83}$, $Q^{493}:E^{35}$, $Q^{498}:K^{353}$, $N^{501}:Y^{41}$, $N^{501}:K^{353}$, and $Y^{505}:E^{37}$ of SARS2 are stronger than the

corresponding pairs in SARS1. Third, the $F^{456}:D^{30}$ pair is formed in SARS2 but not in SARS1. We conclude that the Q^{493}, Q^{498}, N^{501}, F^{486}, K^{417}, and F^{456} are the most critical AAs in SARS2 in comparison to SARS1.

The Y^{501} in Alpha forms more pairs in comparison to the N^{501} of SARS2 (Figures S5B vs 3A). Both Y^{501} and N^{501} form pairs with Y^{41}, K^{353}, G^{354}, and D^{355}. The $Y^{501}:K^{353}$ (Alpha) pair is stronger by −1.3 kcal/mol than the $N^{501}:K^{353}$ (SARS2). This results in no pair formation between Q^{498} and K^{353} of Alpha. The interaction in $Y^{501}:Y^{41}$ (Alpha) is mostly due to hydrophobic and van der Waals interactions, whereas $N^{501}:Y^{41}$ (SARS2) is from electrostatic, hydrophobic, and van der Waals interactions, which surprisingly leads to a similar strength, as shown in Tables S5 and S6. Y^{501} forms additional pairs with D^{38}, G^{352}, and R^{357} in Alpha. The pair $Y^{501}:D^{38}$ forms significant HBs but interferes in the formation of HBs in Y^{449} and D^{38}. However, the $Y^{449}:D^{38}$ pair is observed in SARS2. The N501Y facilitates Y^{505} to create three pairings with A^{386}, A^{387}, and E^{393} in Alpha that were not seen in SARS2. The pair $Y^{505}:E^{37}$ in SARS2 has a strong binding, which is disrupted due to N501Y mutation in Alpha. This mutation causes formation of $T^{478}:Q^{24}$ and $T^{500}:N^{333}$ in addition to other weak pairs. These implications caused by the N501Y mutation are also present in the Beta model. The main differences in Beta are the loss of the salt-bridge between K^{417} and D^{30} when K^{417} changes to N^{417} and forms a pair between K^{484} and E^{75}.

Our result for all models shows that RBD has more AAs in forming pairs with the $\alpha 1$ than the $\alpha 2$ motif of ACE2. This suggests that focusing on the RBD blocker with $\alpha 1$ in ACE2 may be a promising therapeutic strategy for inhibiting the S-protein SARS-CoV-2.[80−82]

3.2. DFT Results and Discussion. *3.2.1. Structural Optimization.* The structural optimization of the Interface-DFT model with high accuracy in atomic coordinates is critical for realistic quantum chemical calculations.[44] Such a high-precision simulation of the interface model is expected to complement experimental studies to provide a deeper understanding at atomic levels of the various aspects of the initial viral infection and at a much lower cost. In the present study, the
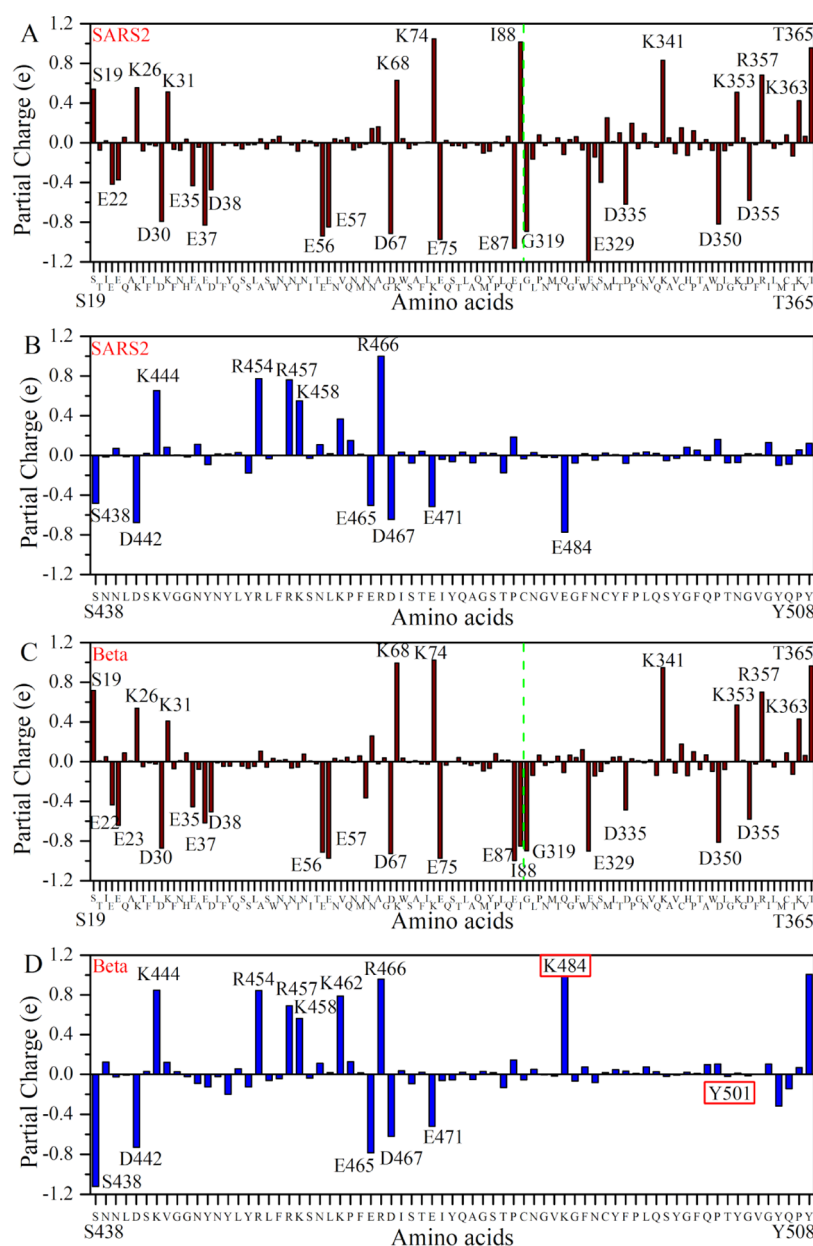
Figure 4. Partial charge (PC) distribution in the Interface-DFT SARS2 model of the segment of ACE2 in sequences 19−88 and 319−365 (A) and RBM in the sequence 438−508 (B). Interface-DFT Beta model of the segment of ACE2 in sequences 19−88 and 319−365 (C) and RBM in the sequence 438−508 (D). Amino acids with positive and negative PCs higher than 0.4e and lower than −0.4e, respectively, are marked. The green dashed line in (A) and (C) shows the break in the sequence of ACE2. Mutated AAs are marked with red boxes in (D).

geometrical optimization was carried out using the VASP package based on the DFT methodology (see Section 2.5).

The Interface-DFT SARS2 model consists of 2930 atoms. The initial unrelaxed structure has a total energy of −18052.021 eV, and the final relaxed structure has a total energy of −18039.025 eV. The reduction of energy is −12.996 eV or −0.0044354 eV/atom (−0.102 kcal/mol). Such a small energy difference accentuates the importance of exact atomic positions in quantum chemical DFT calculations, which control the details of the interatomic bonding, including HB and the partial charge (PC) distributions. It is also completely off the energy scale used in the Interface-MD model, which focuses on the statistical fluctuations of atomic positions in the MD steps over the long-time scale but never at the real atomic scale. The importance of the accurate structural optimization will be

illustrated in the following sections when we present the results for PC distribution and interatomic bonding in the Interface-DFT model. It should also be emphasized that the DFT calculation is a single-point calculation on a specific model of fixed atomic positions. The fully optimized structure for the Interface-DFT model requires a large support on top-notch supercomputers, which was provided by the NERSC of the Lawrence Berkeley National Laboratory managed by US DOE. We have added Figure S7 to show the differences between the crystal and our relaxed structure of SARS2 model with the calculated RMSD value.

3.2.2. Electronic Structure and Partial Charge Distribution. The electronic structure in the form of density of states (DOS) and partial DOS (PDOS) for the Interface-DFT SARS2 model is calculated and presented in Figure S8. It is noted that the total

I
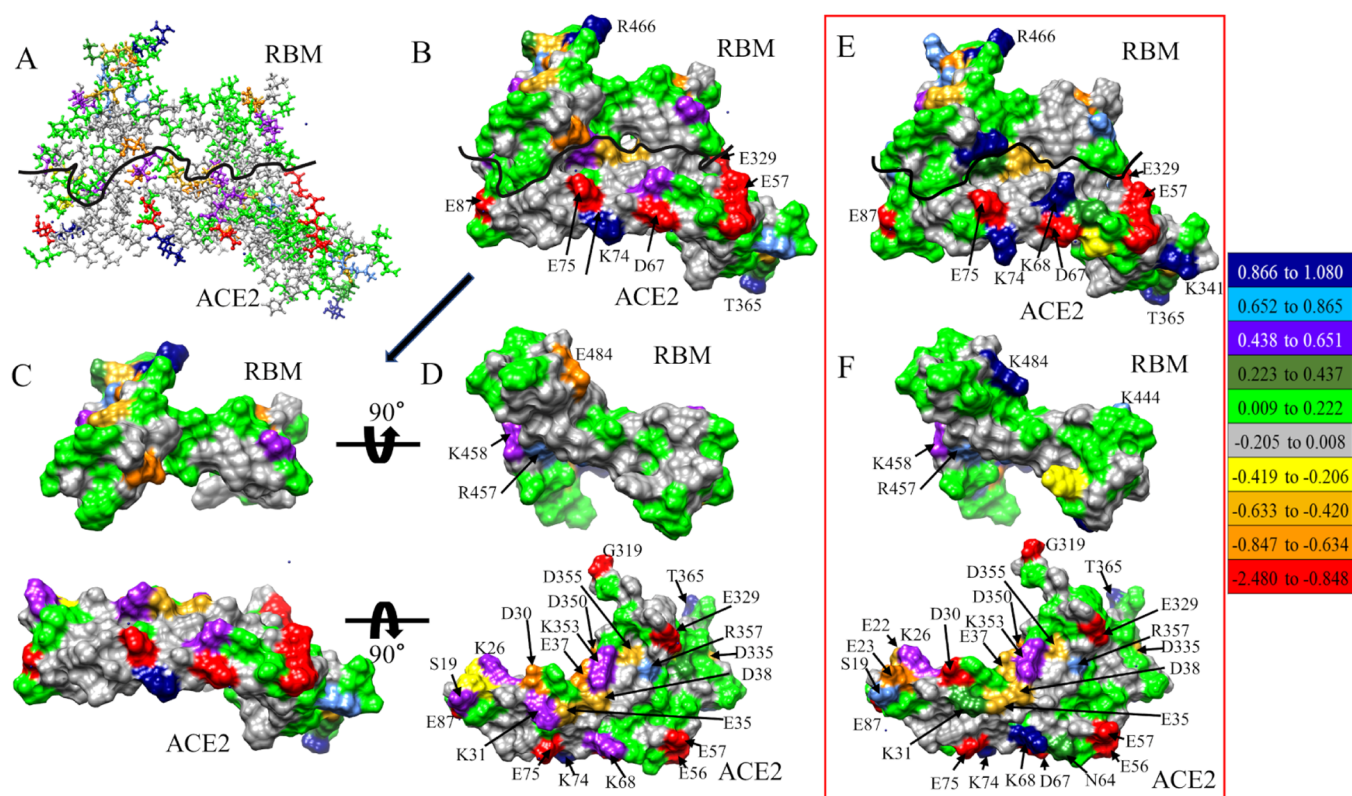
Figure 5. PC distribution in the Interface-DFT SARS2 model in ball and stick (A) and surface (B, C) shows separate RBM and ACE2, and (D) shows the rotated surface of RBM and ACE2 to show the PC distribution in amino acids. Surface PC distribution in the Interface-DFT Beta model (E) and its rotated surface of RBM and ACE2 (F). The curved black lines in (A, B, and E) roughly show the partition between RBM and ACE2 in the interface. The color bar shows the total PC for different amino acids from red (very negative) to blue (very positive). The navy-blue, light-blue, and red amino acids are identified explicitly in (B) and (E). In (D) and (F), we have explicitly marked amino acids lower and higher than −0.4e and 0.4e, respectively.

DOS and the PDOS for RBM and ACE2 parts are almost identical in shape and the structures they contain.

Partial charge (PC) distribution in biomolecules is crucial in understanding the long-range electrostatic interactions that play a significant role in many biological processes and are often implicated in drug design.[83] For the Interface-DFT models, we have determined the PC on every atom and on every residue using the OLCAO method (see Section 2.6). For the SARS2 model, the calculated PC in units of electron charge (e) on each of the 2930 atoms is grouped into 194 AAs and six Na$^+$ ions. Figure 4A,B shows the PC of the 194 AAs in the segments of ACE2 (19−88 and 319−365) and RBM (438−508). The PC can be positive or negative, with some AAs actually exhibiting a very large PC. They are E^{329}, E^{87}, E^{75}, E^{56}, D^{67}, G^{319}, E^{57}, E^{37}, and D^{350} with PC values of −2.480e, −1.061e, −0.976e, −0.939e, −0.915e, −0.892e, −0.849e, −0.831e, and −0.818e and K^{341}, T^{365}, R^{466}, I^{88}, and K^{74} with PC values of 0.832e, 0.956e, 0.999e, 1.014e, and 1.047e, respectively. All Na ions have a positive PC ranging from 0.880e to 1.080e. The complete data for the PC in every AA in the Interface-DFT mode are listed in Table S9. Similar to Figure 4A,B, we show the PC for the Interface-DFT Beta model in Figure 4C,D. Among the two mutated amino acids, K^{484} has a higher positive PC, which contrasts with the high negative PC of E^{484} in SARS2. I^{88} in Beta has a higher negative PC, which contrasts with the higher positive PC in SARS2. Besides these two AAs, K^{462} and Y^{508} in Beta have a higher positive PC. A similar figure for SARS1 is shown in Figure S9. The complete data for the PC shown in Figures S9 and 4C,D are listed in Tables S10 and S11. We have also provided atom-

based PC data for the three models SARS1, SARS2, and Beta in the Supporting Information excel sheet. The atom-based PC for these three models is also shown in Figure S11.

The PC distribution on the ball and stick sketch for all AAs is displayed in Figure 5A, together with the solvent excluded surfaces in Figure 5B. The black curve in Figure 5A,B roughly shows the interface boundary separating the RBM and ACE2. Figure 5C shows the separated RBM and ACE2, which are further rotated in Figure 5D to show the interacting surfaces in the RBM and ACE2. Figures similar to Figure 5B,D for Beta are shown in Figure 5E,F, respectively. In Figure 5F, we can see mutated and positively charged K^{484} in the surface of RBM, which might be the reason for it being more infectious. In Figure 5D, we have marked residues with PCs lower and higher than −0.4e and 0.4e, respectively. Note that the AAs at the interface boundary are neither highly positively charged nor highly negatively charged, which could have implications for the stability of the interface. PC distribution for SARS1 similar to Figure 5 is shown in Figure S10.

3.2.3. Interatomic Bonding. The bond order (BO) vs bond length (BL) for Interface-DFT SARS2 is shown in Figure 6A. There are 17 types of different bonds including O⋯H and N⋯H hydrogen bonding (HB). The bonds with stronger BO are C−O and C−C. The C−C bonds can be partitioned into two groups, of which one exhibits a higher BO, which may be due to double bonds. Some other strong bonds are N−C, C−H, N−H, and O−H. Cys and Met are the sulfur-containing residues and the C−S bonds are associated with them. However, there are no disulfide bonds that are usually common between Cys residues
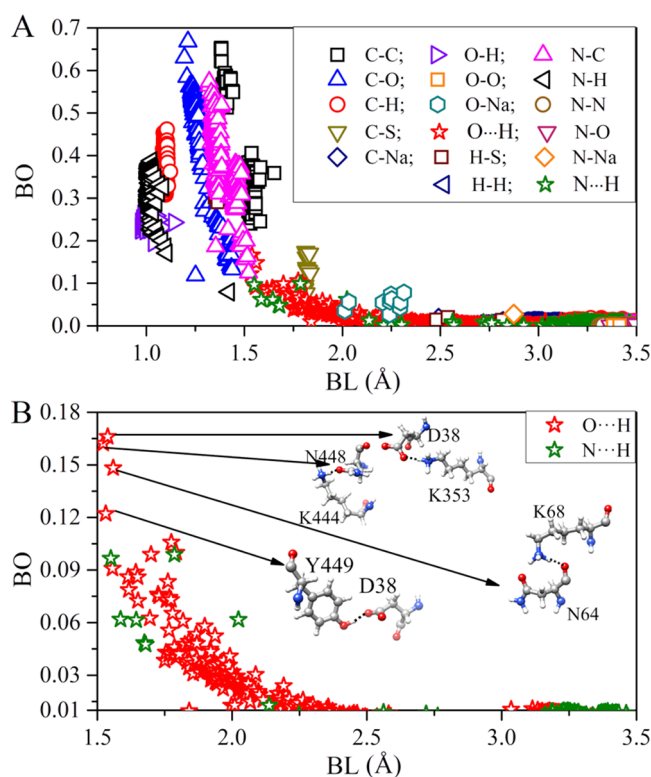
Figure 6. Interatomic interactions in the Interface-DFT SARS2 model. (A) Bond order (BO) vs bond length (BL) for every pair of atoms. (B) Hydrogen bonding (HB). HBs are represented by the star sign.

in the protein. This may be due to the presence of only RBM, a section of RBD, in the Interface-DFT model. O−Na BL ranges from 2.0 to 2.3 Å. C−Na and N−Na belong to weaker bonds. There is only one C−Na bond with BL 2.5 Å (0.018e) and one N−Na bond with BL of 2.9 Å (0.028e). Figure 6B separately shows the HB (O⋯H and N⋯H) distributions in the Interface-DFT SARS2 model. It can be seen that most of the HBs are from O⋯H and their BL ranges from 1.5 Å and becomes very weak after 2.5 Å and certainly cannot be classified as HBs for distance of separation greater than 3.0 Å. The upper limit of BO values for the HB calculated using the OLCAO method is around 0.1e but can be larger in special cases. Figure 6B shows that some so-called HBs in O⋯H can have BO values close to 0.16e. More details on the different ways the HBs are analyzed in MD and in OLCAO will be discussed later.

3.3. Combining MD and DFT Results. The main motivation for this paper is to combine the detailed classical MD simulation with the highly accurate DFT calculations, applying this methodology to the interface between SARS-CoV-2 RBD and ACE2 receptor. Each of the two methods involved has its own distinct advantages but also obvious limitations. By skillfully combining them, most of the drawbacks can be mitigated and new, previously absent insights can be explicitly revealed.

3.4. Nature of Interfacial Cohesion and Interaction between RBD and ACE2. Detailed MD simulation on the Interface-MD model using the generalized Born surface area (MM-GBSA) method provided a wealth of information on the BFE and some generalized conclusions on the specific AA pairs at the much larger and realistic scale for the RBD−ACE2 interface complex. Several pairs are governed by attractive ΔG_{ele} since the HB networks or ionic interactions can stabilize them

(see Section 3.1.4). Other pairs are identified by the favorably ΔE_{vdW} and ΔG_{SA}. For instance, F[486]:L[79], F[486]:M[82], and F[486]:Y[83] pairs are driven by ΔE_{vdW} and ΔG_{SA} (Table S5) and form hydrophobic pockets between SARS-CoV-2 RBD and ACE2, important for anchoring the interface complex.

By constructing a smaller Interface-DFT model between RBM and α1, α2, β3, and β4 motifs in the ACE2 membrane receptor, large-scale *ab initio* DFT calculations could be carried out similar to what has been accomplished in our unprecedented large-scale calculation of the structure and properties of the S-protein in COVID-SARS-2 virus[44−46] and related electronic dielectric constants of small proteins.[84] The Interface-DFT SARS2 model has a total of 2930 atoms, and the highly accurate DFT calculations for the interface model in this work again qualify as an unprecedented computational accomplishment. The detailed electronic structure and interatomic bonding and the partial charge distribution on each atom and residue clearly complement what is missing in the MD simulations. For example, specific bonding across the interface boundary and interactions between residues on two parts of the interface becomes possible, as will be described in the next Section.

**3.5. Interfacial Amino Acid−Amino Acid Interaction.** There are 194 AAs in the Interface-DFT model, 71 in RBM, and 117 in ACE2. Figure 7 (Table S12) shows the calculated AABP
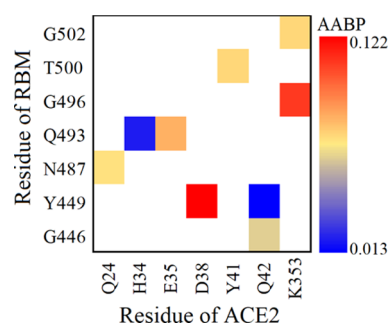
Figure 7. AA to AA bond pair (AABP) between RBM and ACE2 in the Interface-DFT SARS2 model.

values for each interacting pair in RBM and ACE2 with the strength of the interactions (see Section 2.6). The strongest binding is between two pairs: G[496]:K[353] and Y[449]:D[38] in RBM and ACE2 with AABP values of 0.100e and 0.122e, respectively.

Figure 8 further shows the intricate bonding configurations between RBM and ACE2, with HBs playing a key role in the AA−AA interaction across the interfacial boundary for SARS2 and Beta models. The AABPs for the interfacial interaction between a pair of amino acids are characterized by having only one hydrogen bonding. However, it is interesting to see the relation between these nine pairs. Figure 8B shows two pairs with one common AA. Here, both E[35] and H[34] have one HB with Q[493]. Similarly, Figure 8C shows the interaction of Y[449] with D[38] and Q[42] as well as G[446] with Q[42]. The pairs between T[500]:Y[41] and G[496]:K[353] or K[353]:G[502] are shown in Figure 8D,E, respectively. This intertwined bonding in the interface region shows a complex topology that would be difficult to discern with less advanced computational methodology. Comparing Interface-DFT SARS2 and Beta models, we identified some same AA pair interactions; they are shown in Figure 8F,G,J, which are similar to those in Figure 8A−C, respectively. However, there are some new pairs in the Beta model; they are A[475]:S[19], Y[505]:E[37], N[330]:T[500], and Y[501]:K[353]. We would like to point out that among these new pairs Y501 is the mutated AA, which interacts with
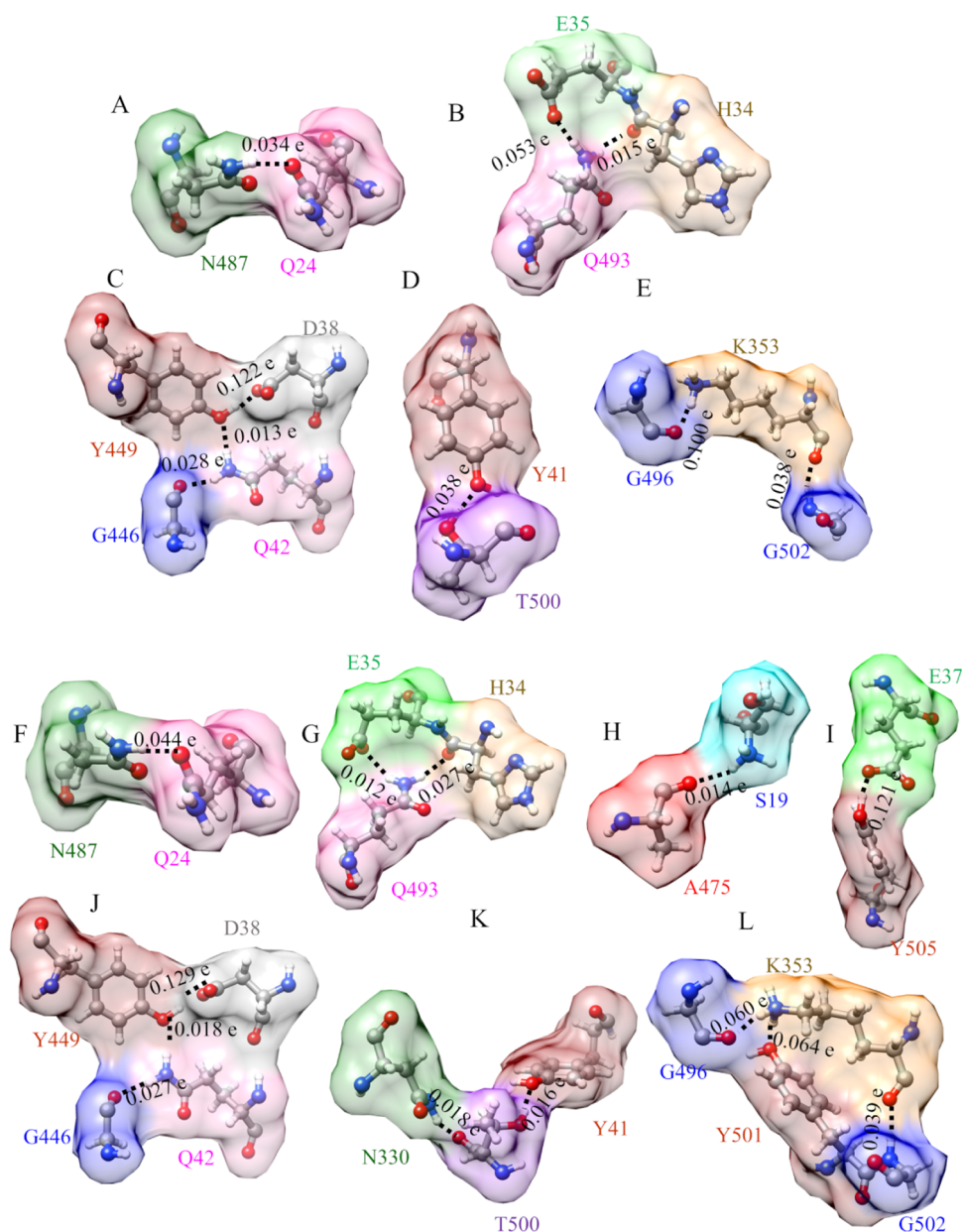
Figure 8. Amino acid−amino acid bond pair (AABP) between RBM and ACE2 in ball and stick for participating pairs of Interface-DFT SARS2 (blue box) (A) N[487]:Q[24], (B) Q[493]:E[35] and Q[493]:H[34], (C) Y[449]:D[38], Y[449]:Q[42], and G[446]:Q[42], (D) T[500]:Y[41], and (E) G[496]:K[353], K[353]:G[502]. Participating pairs for Interface-DFT Beta (red box) (F) N[487]:Q[24], (G) E[35]:Q[493], Q[493]:H[34], (H) A[475]:S[19], (I) Y[505]:E[37], (J) D[38]:Y[449], Y[449]:Q[42], and Q[42]:G[446], (K) N[330]:T[500], T[500]:Y[41], and (L) G[496]:K[353], K[353]:G[502], Y[501]:K[353]. It is interesting that there is only hydrogen bonding in the interface of SARS2 and Beta. There is more interaction in the interface of Beta in comparison to SARS2. One of the new hydrogen bondings is between mutated Y[501] with K[353].

K[353]. The Y[501]:K[353] interaction in the interface might be the reason for the Beta being more infectious. We have shown a similar AABP interaction for SARS1 in Figure S12.

3.6. Hydrogen-Bonding Analysis. Analysis of the HB network can provide insights into the nature of binding between RBD and ACE2. This section emphasizes the prominent role of intermolecular HBs in the RBD−ACE2 interface complex of SARS2, using MD simulations and DFT calculations, as well as comparing them to available experimental data.

It must be pointed out that the ways HBs are calculated in these two models are fundamentally different. In the DFT, they are calculated quantum mechanically, as demonstrated in the case of our previous work on a large water box,[85] extending now the same bonding criteria to the HBs in more complex biomolecules. These entail the use of BO value and BL (*aka*,

the distance of separation between H in one H_2O and O in another H_2O). In the MD simulation, following the guidelines of IUPAC in 2011 and applying them to more complex biomolecular systems, the HB identification is based solely on the geometric consideration, with no quantitative definition for the strength of the HB.[86] Both approaches have limitations. The DFT calculation is a single-point calculation with all its atomic positions fixed, while MD simulations entail dynamic data but cannot offer the quantitative strength of HBs. This leads to some confusion or apparent differences in the HB analysis from the two calculations using very different approaches. We believe that both approaches provide insights into HBs at the interface, and we interpret the results based on both methodologies. They should refer to the same HBs in the biomolecular systems and
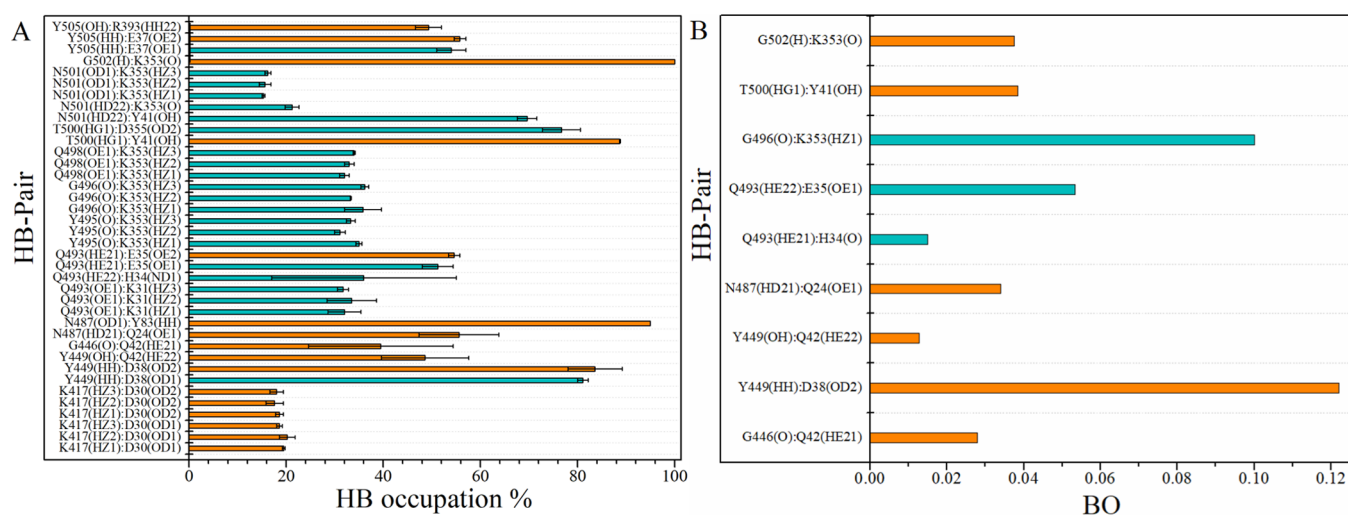
Figure 9. Histogram diagram of the HB network of the SARS2 interface complex. (A) HBs with their percentage occupancies from MD simulations. (B) HBs from the single Interface-DFT SARS2 model with their bond orders (BOs) representing their strength. The orange bars indicate that the HB pairs have been experimentally observed, while the green bar represents the one predicted in the present study.

complete each other for a deeper understanding of their impact on the stability and conformation of the biomolecules involved.

Figure 9A,B shows the possible HBs in the SARS2 model from MD simulations and DFT calculations, respectively. In MD simulations, the CPPTRAJ program in AMBER is utilized to calculate HBs based on geometric criteria using a cutoff acceptor and donor (A...D) distance < 4.0 Å and a cutoff angle > 130° among the acceptor, hydrogen, and donor atoms.[87] The percentage of HB occupancy gives an account of how many times a certain HB forms during the MD simulations. The higher the frequency, the higher the stability of this HB. In DFT calculations, the HB is determined by its BL and BO values, as shown in Figure 9B. We use a BO cutoff of 0.01e to characterize only the actual HBs. DFT calculation is a single-point calculation providing the exact geometry as well as strength of each HB. Figure 9A shows that in MD simulation seven HBs have a very high frequency of occupancy of more than 70%, with four of them having been also reported in the X-ray study:[20] $Y^{449}(HH):D^{38}(OD2)$, $N^{487}(OD1):Y^{83}(HH)$, $T^{500}(HG1):Y^{41}(OH)$, and $G^{502}(H):K^{353}(O)$. The other three HBs predicted are $Y^{449}(HH):D^{38}(OD1)$, $T^{500}(HG1):D^{355}(OD2)$, and $N^{501}(HD22):Y^{41}(OH)$. The remaining possible HBs have occupancy of less than 70% (Figure 9A). On the other hand, the HB analysis based on DFT calculation shows that two HBs, $Y^{449}(HH):D^{38}(OD2)$ and $G^{496}(O):K^{353}(HZ1)$, have a higher BO of ≥0.1e and are also observed in MD simulations. Five other HBs have medium BO values in the range of 0.053−0.027e (Figure 9B), with four of them observed also experimentally as well as in MD simulations. The remaining two HBs are relatively weaker with BO less than 0.02e and only one of them is experimentally detected.[20]

Table S8 summarizes the comparison and reveals that both MD simulations and DFT calculations on a single model can predict a large number of HBs, far more than that from the experimental X-ray study.[20] Unlike the static single structure of X-ray or the relaxed Interface-DFT model, each AA in the MD simulation moves dynamically around and thus increases the chance of forming a large number of HBs even though the criterion of what qualifies as a *bona fide* HB is not clear. On the other hand, the static DFT calculation on the fully relaxed structure has the benefits of providing a precise definition for HB

in the presence of real H atoms that are missing in the X-ray experiments. Additionally, the separation distances between the acceptor and H atom attached to the donor in DFT calculations are less ambiguous compared to the experimental X-ray crystal structure and certain HBs such as $N^{501}(N):Y^{41}(OH)$ and $Y^{489}(OH):Y^{83}(OH)$ that have been claimed experimentally as HBs[20] but were not found in MD simulations or DFT calculations. Those may indeed not be actual HBs.

3.7. Implications for Electrostatic Interactions of Different PCs. The electrostatic interactions in the protein−protein binding process are well known.[83] In general, the electrostatic interactions depend on the distribution of atomic PC, the dielectric constant, and the presence of a bathing solution (water and ions). In the current study, the PC distribution at the atomic and AA levels for the Interface-DFT model has been quantified in detail using *ab initio* quantum calculations (see Section 3.2). Additionally, the electrostatic interactions and solvent effects of the Interface-MD model have been investigated based on fixed PCs (see Section 3.1). These electrostatic interactions play a significant role in explaining the tight binding of SARS-CoV-2 RBD with ACE2 than in SARS-CoV, as shown earlier in MD results and in other existing studies.[33−35,88,89] However, these interactions should be carefully reexamined not only based on the fixed PCs from the force field of MD since we now have accurate PCs from *ab initio* calculations.

Calculating the free energy of the total electrostatic contribution (ΔG_{ele}) is quite sensitive to the value of the assumed dielectric constant. Very recently, we have forcefully argued that the dielectric constants should be much higher than the values of 1−4 that have been standardly invoked in many continuum approaches used in MD simulations.[84] To this end, we separately calculated ΔG_{ele} of the Interface-DFT SARS2 model based on the fixed PCs from the AMBER force field ff14SB[53] and on the PCs from the *ab initio* OLCAO method at three different interior dielectric constants (1, 5, and 10). To carry out such an investigation, two different atomic PCs are implemented in the DelPhi v.8.4.5 program.[90,91] DelPhi calculates the polar solvation energy implicitly by solving the nonlinear and/or linear variants of the Poisson−Boltzmann (PB) equation numerically using the finite difference method-

Table 3. Predicted Electrostatic Interactions and Their Energetic Components Using Two Different Sets of PC Values at a 0.15 M Salt Concentration and Three Assumed Interior Dielectric Constants (ε_i) of the Interface-DFT SARS2 model

PCs	$\varepsilon_i = 1$			$\varepsilon_i = 5$			$\varepsilon_i = 10$		
	ΔE_{ele}	ΔG_{PB}	ΔG_{ele}	ΔE_{ele}	ΔG_{PB}	ΔG_{ele}	ΔE_{ele}	ΔG_{PB}	ΔG_{ele}
AMBER	−200.30	249.10	48.80	−40.06	47.34	7.28	−20.03	22.11	2.08
OLCAO	−132.49	302.45	169.96	−26.50	57.36	30.86	−13.24	26.52	13.27

ology.[90,91] It can provide the Coulombic electrostatic interaction (ΔE_{ele}) and requires four input files: the atomic structure, the atomic PCs, the atomic radii, and parameter files. These files are all available from the current study. The following input parameters are adopted in DelPhi: linear PB, scale 2 grids/A, 0.0001 kT/e of the convergence criterion root-mean-square change of potential (RMSC) and potential maximum change threshold of potential (MAXC), external dielectric constant 78.3, and ionic concentration 0.15 M. The three values of internal dielectric constants chosen as a test were 1, 5, and 10. The energy in units of kT is converted to kcal/mol at temperature T = 310 K.

Table 3 shows the results of ΔG_{ele} and its components using two different sets of atomic PCs with different assumed internal dielectric constants, which are unknown for the protein.[84] The favorability pattern of ΔG_{ele} and its components is similar for both sets of PCs and exhibits the same behavior in MD simulations (Table 2). The computed values of ΔG_{ele}, ΔE_{ele}, and ΔG_{PB} are highly sensitive to atomic PCs and are higher, i.e., less favorable, when using PCs from DFT. Importantly, regardless of which atomic PCs are used, our analysis reveals that increasing the interior dielectric constant results in less favored ΔE_{ele} and more favored ΔG_{PB} and ΔG_{ele}. Again, the computed values of ΔG_{ele} and its components greatly depend on the internal dielectric constant, which is unknown.[84] Many MD simulations chose them *a posteriori* to justify their result to be in line with experiments.

4. CONCLUSIONS

The methodology and the ensuing analysis of biomolecular interaction presented above have broader repercussions in the general context of biomolecular computational physics and should not be seen as specifically delimited only to the cases of the SARS-CoV-2 RBD−ACE2 complex and its Alpha and Beta variants that were analyzed in detail. Its main advantages can be summarized as follows.

1. We have combined the MD simulations with DFT calculations to significantly enhance the power and utility of comprehensive computational modeling applied to large complex biomolecular systems.
2. From detailed MD simulations, we have extracted the binding free energy of the RBD−ACE2 complex of SARS1, SARS2, Alpha and Beta with a complete energetic profile including the main driving force for the interaction in the interfacial region of this biomolecular complex. The key interacting AAs between RBD and ACE2 have been identified, along with their quantitative energies.
3. With detailed DFT calculations, we have further quantified the interaction between amino acids involved in the interface interaction between RBD and ACE2 at the atomistic level, including the contribution from hydrogen-bonding topology.
4. The two combined and complementary methodologies provide the missing link that has been bothering

computational scientists for decades, partly due to the different quantification of the HBs in molecular interactions.
5. The combined approach presented in this contribution puts forth a new road map for critical evaluation of problems related to drug design and mutation analysis in the SARS-CoV-2 virus and in complex biomolecular systems in general.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00560.

Tables of key interacting AAs from BFE decompositions with their energetic components and of HB comparisons; tables for PC distributions or AABP values of the interface-DFT model; RMSD and RMSF figures for RBD and ACE2 into the interface-MD model; sequence alignment figure of RBD in SARS-CoV and SARS-CoV-2; and TDOS and PDOS figures of the interface-DFT model (PDF)

Atom-based PC data for the three models SARS1, SARS2, and Beta (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Rudolf Podgornik − *Wenzhou Institute of the University of Chinese Academy of Sciences, Wenzhou, Zhejiang 325000, China; School of Physical Sciences and Kavli Institute of Theoretical Science, University of Chinese Academy of Sciences, Beijing 100049, China; CAS Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100090, China; Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia*; Email: rudipod@gmail.com

Wai-Yim Ching − *Department of Physics and Astronomy, University of Missouri-Kansas City, Kansas City, Missouri 64110, United States;* ⓞ orcid.org/0000-0001-7738-8822; Email: Chingw@umkc.edu

Authors

Bahaa Jawad − *Department of Physics and Astronomy, University of Missouri-Kansas City, Kansas City, Missouri 64110, United States;* ⓞ orcid.org/0000-0001-6252-3999

Puja Adhikari − *Department of Physics and Astronomy, University of Missouri-Kansas City, Kansas City, Missouri 64110, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.1c00560

participated in the discussion and interpretation of the results. All authors edited and proofread the final manuscript.

Notes

The authors declare no competing financial interest.

All data are listed in tables in the Supporting Information (SI) including the per-residue and pairwise BFE decompositions, HB comparisons, partial charge distributions of each amino acid in the Interface-DFT model as well as amino acid−amino acid bond pair (AABP) values. Additionally, the RMSD and RMSF from MD simulations of the Interface-MD model are shown in the SI. Density of states (DOS) and partial DOS (PDOS) are also shown in the SI. Specific data are available by contacting the authors.The AMBER 18 package and DelPhi v.8.4.5 open-source programs were used, and all are cited in the manuscript.VASP (Vienna *Ab initio* Simulation Package) is a commercially available code https://www.vasp.at/. Users must be licensed. We used the older version 5.6 available at the National Energy Research Scientific Computing Center (NERSC).The OLCAO method is an in-house-developed package at the University of Missouri-Kansas City. It is not publicly available yet and used by students and collaborators. The method used is fully described in the book by Wai-Yim Ching and Paul Rulis, *Electronic Structure Methods for Complex Materials: The Orthogonalized Linear Combination of Atomic Orbitals*, Oxford University Press, 2012.

■ REFERENCES

(1) Shin, M. D.; Shukla, S.; Chung, Y. H.; Beiss, V.; Chan, S. K.; Ortega-Rivera, O. A.; Wirth, D. M.; Chen, A.; Sack, M.; Pokorski, J. K.; Steinmetz, N. F. COVID-19 vaccine development and a potential nanomaterial path forward. *Nat. Nanotechnol.* **2020**, *15*, 646−655.

(2) Lurie, N.; Saville, M.; Hatchett, R.; Halton, J. Developing Covid-19 vaccines at pandemic speed. *N. Engl. J. Med.* **2020**, *382*, 1969−1973.

(3) Krammer, F. SARS-CoV-2 vaccines in development. *Nature* **2020**, *586*, 516−527.

(4) Deming, M. E.; Michael, N. L.; Robb, M.; Cohen, M. S.; Neuzil, K. M. Accelerating development of SARS-CoV-2 vaccines—the role for controlled human infection models. *N. Engl. J. Med.* **2020**, *383*, No. e63.

(5) Jackson, L. A.; Anderson, E. J.; Rouphael, N. G.; Roberts, P. C.; Makhene, M.; Coler, R. N.; McCullough, M. P.; Chappell, J. D.; Denison, M. R.; Stevens, L. J.; et al. An mRNA vaccine against SARS-CoV-2—preliminary report. *N. Engl. J. Med.* **2020**, *383*, 1920−1931.

(6) U.S. Food and Drug Administration. COVID-19 Vaccines. https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines.

(7) Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265−269.

(8) 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.

(9) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**, *181*, 281−292.e6.

(10) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **2020**, *181*, 271−280.e8.

(11) Schoeman, D.; Fielding, B. C. Coronavirus envelope protein: current knowledge. *Virol. J.* **2019**, *16*, No. 69.

(12) Tian, J.-H.; Patel, N.; Haupt, R.; Zhou, H.; Weston, S.; Hammond, H.; Logue, J.; Portnoff, A. D.; Norton, J.; Guebre-Xabier, M.; et al. SARS-CoV-2 spike glycoprotein vaccine candidate NVX-CoV2373 immunogenicity in baboons and protection in mice. *Nat. Commun.* **2021**, *12*, No. 372.

(13) Kang, Y.-F.; Sun, C.; Zhuang, Z.; Yuan, R.-Y.; Zheng, Q.; Li, J.-P.; Zhou, P.-P.; Chen, X.-C.; Liu, Z.; Zhang, X.; et al. Rapid development of SARS-CoV-2 spike protein receptor-binding domain self-assembled nanoparticle vaccine candidates. *ACS Nano* **2021**, *15*, 2738−2752.

(14) Vogel, A. B.; Kanevsky, I.; Che, Y.; Swanson, K. A.; Muik, A.; Vormehr, M.; Kranz, L. M.; Walzer, K. C.; Hein, S.; Güler, A. BNT162b vaccines protect rhesus macaques from SARS-CoV-2. *Nature* **2021**, *592*, 283−289.

(15) Panda, P. K.; Arul, M. N.; Patel, P.; Verma, S. K.; Luo, W.; Rubahn, H.-G.; Mishra, Y. K.; Suar, M.; Ahuja, R. Structure-based drug designing and immunoinformatics approach for SARS-CoV-2. *Sci. Adv.* **2020**, *6*, No. eabb8097.

(16) Braga, L.; Ali, H.; Secco, I.; Chiavacci, E.; Neves, G.; Goldhill, D.; Penn, R.; Jimenez-Guardeño, J. M.; Ortega-Prieto, A. M.; Bussani, R. Drugs that inhibit TMEM16 proteins block SARS-CoV-2 Spike-induced syncytia. *Nature* **2021**, *594*, 88−93.

(17) Kim, C.; Ryu, D.-K.; Lee, J.; Kim, Y.-I.; Seo, J.-M.; Kim, Y.-G.; Jeong, J.-H.; Kim, M.; Kim, J.-I.; Kim, P. A therapeutic neutralizing antibody targeting receptor binding domain of SARS-CoV-2 spike protein. *Nat. Commun.* **2021**, *12*, No. 288.

(18) Chen, R. E.; Zhang, X.; Case, J. B.; Winkler, E. S.; Liu, Y.; VanBlargan, L. A.; Liu, J.; Errico, J. M.; Xie, X.; Suryadevara, N.; et al. Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat. Med.* **2021**, *27*, 717−726.

(19) Xia, S.; Zhu, Y.; Liu, M.; Lan, Q.; Xu, W.; Wu, Y.; Ying, T.; Liu, S.; Shi, Z.; Jiang, S.; Lu, L. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell. Mol. Immunol.* **2020**, *17*, 765−767.

(20) 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*, 215−220.

(21) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221−224.

(22) Letko, M.; Marzi, A.; Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **2020**, *5*, 562−569.

(23) Xu, C.; Wang, Y.; Liu, C.; Zhang, C.; Han, W.; Hong, X.; Wang, Y.; Hong, Q.; Wang, S.; Zhao, Q.; et al. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci. Adv.* **2021**, *7*, No. eabe5575.

(24) 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, *947* e00127-20 DOI: 10.1128/JVI.00127-20.

(25) Yan, R.; Zhang, Y.; Li, Y.; Ye, F.; Guo, Y.; Xia, L.; Zhong, X.; Chi, X.; Zhou, Q. Structural basis for the different states of the spike protein of SARS-CoV-2 in complex with ACE2. *Cell Res.* **2021**, *31*, 717−719.

(26) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K.-Y.; et al. Structural and functional basis of

SARS-CoV-2 entry by using human ACE2. *Cell* **2020**, *181*, 894–904.e9.

(27) He, J.; Tao, H.; Yan, Y.; Huang, S.-Y.; Xiao, Y. Molecular mechanism of evolution and human infection with SARS-CoV-2. *Viruses* **2020**, *12*, 428.

(28) Starr, T. N.; Greaney, A. J.; Hilton, S. K.; Ellis, D.; Crawford, K. H.; Dingens, A. S.; Navarro, M. J.; Bowen, J. E.; Tortorici, M. A.; Walls, A. C.; et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **2020**, *182*, 1295–1310.e20.

(29) Laurini, E.; Marson, D.; Aulic, S.; Fermeglia, A.; Pricl, S. Computational Mutagenesis at the SARS-CoV-2 Spike Protein/Angiotensin-Converting Enzyme 2 Binding Interface: Comparison with Experimental Evidence. *ACS Nano* **2021**, *15*, 6929–6948.

(30) Laurini, E.; Marson, D.; Aulic, S.; Fermeglia, M.; Pricl, S. Computational alanine scanning and structural analysis of the SARS-CoV-2 spike protein/angiotensin-converting enzyme 2 complex. *ACS Nano* **2020**, *14*, 11821–11830.

(31) Mercurio, I.; Tragni, V.; Busto, F.; De Grassi, A.; Pierri, C. L. Protein structure analysis of the interactions between SARS-CoV-2 spike protein and the human ACE2 receptor: from conformational changes to novel neutralizing antibodies. *Cell. Mol. Life Sci.* **2021**, *78*, 1501–1522.

(32) Yan, F.-F.; Gao, F. Comparison of the binding characteristics of SARS-CoV and SARS-CoV-2 RBDs to ACE2 at different temperatures by MD simulations. *Briefings Bioinf.* **2021**, *22*, 1122–1136.

(33) Spinello, A.; Saltalamacchia, A.; Magistrato, A. Is the rigidity of SARS-CoV-2 spike receptor-binding motif the hallmark for its enhanced infectivity? Insights from all-atom simulations. *J. Phys. Chem. Lett.* **2020**, *11*, 4785–4790.

(34) Amin, M.; Sorour, M. K.; Kasry, A. Comparing the binding interactions in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *J. Phys. Chem. Lett.* **2020**, *11*, 4897–4900.

(35) Nguyen, H. L.; Lan, P. D.; Thai, N. Q.; Nissley, D. A.; O'Brien, E. P.; Li, M. S. Does SARS-CoV-2 bind to human ACE2 more strongly than does SARS-CoV? *J. Phys. Chem. B* **2020**, *124*, 7336–7347.

(36) Wang, Y.; Liu, M.; Gao, J. Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 13967–13974.

(37) Peng, C.; Zhu, Z.; Shi, Y.; Wang, X.; Mu, K.; Yang, Y.; Zhang, X.; Xu, Z.; Zhu, W. Computational Insights into the Conformational Accessibility and Binding Strength of SARS-CoV-2 Spike Protein to Human Angiotensin-Converting Enzyme 2. *J. Phys. Chem. Lett.* **2020**, *11*, 10482–10488.

(38) Chowdhury, R.; Boorla, V. S.; Maranas, C. D. Computational biophysical characterization of the SARS-CoV-2 spike protein binding with the ACE2 receptor and implications for infectivity. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 2573–2582.

(39) Qiao, B.; de la Cruz, Olvera. M., Enhanced binding of SARS-CoV-2 spike protein to receptor by distal polybasic cleavage sites. *ACS Nano* **2020**, *14*, 10616–10623.

(40) Zou, J.; Yin, J.; Fang, L.; Yang, M.; Wang, T.; Wu, W.; Bellucci, M. A.; Zhang, P. Computational Prediction of Mutational Effects on SARS-CoV-2 Binding by Relative Free Energy Calculations. *J. Chem. Inf. Model.* **2020**, *60*, 5794–5802.

(41) Rambaut, A.; Loman, N.; Pybus, O.; Barclay, W.; Barrett, J.; Carabelli, A.; Connor, T.; Peacock, T.; Robertson, D. L.; Volz, E. *Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations*; On behalf of COVID-19 Genomics Consortium UK (CoG-UK), 2020.

(42) Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E. J.; Msomi, N., Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *MedRxiv* **2020**.

(43) Klompas, M.; Baker, M. A.; Rhee, C. Airborne transmission of SARS-CoV-2: theoretical considerations and available evidence. *JAMA* **2020**, *324*, 441.

(44) Ching, W.-Y.; Adhikari, P.; Jawad, B.; Podgornik, R. Ultra-large-scale ab initio quantum chemical computation of bio-molecular systems: The case of spike protein of SARS-CoV-2 virus. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1288–1301.

(45) Adhikari, P.; Li, N.; Shin, M.; Steinmetz, N. F.; Twarock, R.; Podgornik, R.; Ching, W.-Y. Intra-and intermolecular atomic-scale interactions in the receptor binding domain of SARS-CoV-2 spike protein: implication for ACE2 receptor binding. *Phys. Chem. Chem. Phys.* **2020**, *22*, 18272–18283.

(46) Adhikari, P.; Ching, W.-Y. Amino acid interacting network in the receptor-binding domain of SARS-CoV-2 spike protein. *RSC Adv.* **2020**, *10*, 39831–39841.

(47) 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*, 926–935.

(48) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.

(49) Shapovalov, M. V.; Dunbrack, R. L., Jr A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* **2011**, *19*, 844–858.

(50) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(51) Li, F.; Li, W.; Farzan, M.; Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **2005**, *309*, 1864–1868.

(52) Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.

(53) 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*, 3696–3713.

(54) Case, D.; Cerutti, D.; Cheatham, T., III; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Greene, D.; Homeyer, N. *AMBER Reference Manual*; University of California, 2018.

(55) 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*, 10089–10092.

(56) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(57) Götz, 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*, 1542–1555.

(58) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.

(59) Jawad, B.; Poudel, L.; Podgornik, R.; Steinmetz, N. F.; Ching, W.-Y. Molecular mechanism and binding free energy of doxorubicin intercalation in DNA. *Phys. Chem. Chem. Phys.* **2019**, *21*, 3877–3893.

(60) Jawad, B.; Poudel, L.; Podgornik, R.; Ching, W.-Y. Thermodynamic Dissection of the Intercalation Binding Process of Doxorubicin to dsDNA with Implications of Ionic and Solvent Effects. *J. Phys. Chem. B* **2020**, *124*, 7803–7818.

(61) Miller, B. R., III; McGee, T. D., Jr; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA. py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.

(62) Wang, J.; Hou, T.; Xu, X. Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. *Curr. Comput.-Aided Drug Des.* **2006**, *2*, 287–306.

(63) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712−3720.

(64) Homeyer, N.; Gohlke, H. Free energy calculations by the molecular mechanics Poisson− Boltzmann surface area method. *Mol. Inf.* **2012**, *31*, 114−122.

(65) Vienna Ab initio Simulation Package. https://www.vasp.at/.

(66) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(67) Ching, W.-Y.; Rulis, P., *Electronic Structure Methods for Complex Materials: The Orthogonalized Linear Combination of Atomic Orbitals*. Oxford University Press, 2012.

(68) Poudel, L.; Steinmetz, N. F.; French, R. H.; Parsegian, V. A.; Podgornik, R.; Ching, W.-Y. Implication of the solvent effect, metal ions and topology in the electronic structure and hydrogen bonding of human telomeric G-quadruplex DNA. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21573−21585.

(69) Adhikari, P.; Wen, A. M.; French, R. H.; Parsegian, V. A.; Steinmetz, N. F.; Podgornik, R.; Ching, W.-Y. Electronic Structure, Dielectric Response and Surface Charge Distribution of RGD (1FUV) Peptide. *Sci. Rep.* **2014**, *4*, No. 5605.

(70) Poudel, L.; Twarock, R.; Steinmetz, N. F.; Podgornik, R.; Ching, W.-Y. Impact of hydrogen bonding in the binding site between capsid protein and MS2 bacteriophage ssRNA. *J. Phys. Chem. B* **2017**, *121*, 6321−6330.

(71) Mulliken, R. S. Electronic population analysis on LCAO−MO molecular wave functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.

(72) Mulliken, R. Electronic population analysis on LCAO−MO molecular wave functions. II. Overlap populations, bond orders, and covalent bond energies. *J. Chem. Phys.* **1955**, *23*, 1841−1846.

(73) Laffeber, C.; de Koning, K.; Kanaar, R.; Lebbink, J. H. Experimental evidence for enhanced receptor binding by rapidly spreading SARS-CoV-2 variants. *J. Mol. Biol.* **2021**, *433*, 167058.

(74) Supasa, P.; Zhou, D.; Dejnirattisai, W.; Liu, C.; Mentzer, A. J.; Ginn, H. M.; Zhao, Y.; Duyvesteyn, H. M.; Nutalai, R.; Tuekprakhon, A.; et al. Reduced neutralization of SARS-CoV-2 B. 1.1. 7 variant by convalescent and vaccine sera. *Cell* **2021**, *184*, 2201−2211.e7.

(75) Talley, K.; Ng, C.; Shoppell, M.; Kundrotas, P.; Alexov, E. On the electrostatic component of protein-protein binding free energy. *PMC Biophys.* **2008**, *1*, 2.

(76) Barratt, E.; Bingham, R. J.; Warner, D. J.; Laughton, C. A.; Phillips, S. E.; Homans, S. W. Van der Waals interactions dominate ligand− protein association in a protein binding site occluded from solvent water. *J. Am. Chem. Soc.* **2005**, *127*, 11827−11834.

(77) Nilofer, C.; Sukhwal, A.; Mohanapriya, A.; Kangueane, P. Protein-protein interfaces are vdW dominant with selective H-bonds and (or) electrostatics towards broad functional specificity. *Bioinformation* **2017**, *13*, 164.

(78) Zhu, X.; Mannar, D.; Srivastava, S. S.; Berezuk, A. M.; Demers, J.-P.; Saville, J. W.; Leopold, K.; Li, W.; Dimitrov, D. S.; Tuttle, K. S.; Zhou, S.; Chittori, S.; Subramaniam, S. Cryo-electron microscopy structures of the N501Y SARS-CoV-2 spike protein in complex with ACE2 and 2 potent neutralizing antibodies. *PLoS Biol.* **2021**, *19*, e3001237.

(79) Di Giacomo, S.; Mercatelli, D.; Rakhimov, A.; Giorgi, F. M. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K. *J. Med. Virol.* **2021**, *93*, 5638−5643.

(80) Han, Y.; Král, P. Computational design of ACE2-based peptide inhibitors of SARS-CoV-2. *ACS Nano* **2020**, *14*, 5143−5147.

(81) Karoyan, P.; Vieillard, V.; Gómez-Morales, L.; Odile, E.; Guihot, A.; Luyt, C.-E.; Denis, A.; Grondin, P.; Lequin, O. Human ACE2 peptide-mimics block SARS-CoV-2 pulmonary cells infection. *Commun. Biol.* **2021**, *4*, No. 197.

(82) Chan, K. K.; Dorosky, D.; Sharma, P.; Abbasi, S. A.; Dye, J. M.; Kranz, D. M.; Herbert, A. S.; Procko, E. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* **2020**, *369*, 1261−1265.

(83) Zhou, H.-X.; Pang, X. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem. Rev.* **2018**, *118*, 1691−1741.

(84) Adhikari, P.; Podgornik, R.; Jawad, B.; Ching, W.-Y., Ab Initio Electronic Dielectric "Constant" of Proteins: A Baseline for Electrostatic Interaction in Biomolecular Systems. *chemrxiv* **2021**.

(85) Liang, L.; Rulis, P.; Ouyang, L.; Ching, W. Ab initio investigation of hydrogen bonding and network structure in a supercooled model of water. *Phys. Rev. B* **2011**, *83*, No. 024201.

(86) Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; Alkorta, I.; Clary, D. C.; Crabtree, R. H.; Dannenberg, J. J.; Hobza, P.; et al. Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl. Chem.* **2011**, *83*, 1637−1641.

(87) Roe, D. R.; Cheatham, T. E., III PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.

(88) Bai, C.; Warshel, A. Critical differences between the binding features of the spike proteins of SARS-CoV-2 and SARS-CoV. *J. Phys. Chem. B* **2020**, *124*, 5907−5912.

(89) Xie, Y.; Karki, C. B.; Du, D.; Li, H.; Wang, J.; Sobitan, A.; Teng, S.; Tang, Q.; Li, L. Spike proteins of SARS-CoV and SARS-CoV-2 utilize different mechanisms to bind with human ACE2. *Front. Mol. Biosci.* **2020**, *7*, No. 591873.

(90) Li, L.; Li, C.; Sarkar, S.; Zhang, J.; Witham, S.; Zhang, Z.; Wang, L.; Smith, N.; Petukh, M.; Alexov, E. DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys.* **2012**, *5*, No. 9.

(91) Li, C.; Jia, Z.; Chakravorty, A.; Pahari, S.; Peng, Y.; Basu, S.; Koirala, M.; Panday, S. K.; Petukh, M.; Li, L.; Alexov, E. DelPhi suite: New developments and review of functionalities. *J. Comput. Chem.* **2019**, *40*, 2502−2508.