

# Key interacting residues between RBD of SARS-CoV-2 and ACE2 receptor: Combination of molecular dynamic simulation and density functional calculation

Bahaa Jawad<sup>§</sup>, Puja Adhikari<sup>§</sup>, Rudolf Podgornik<sup>▽#†¶</sup>, Wai-Yim Ching<sup>§\*</sup>

<sup>§</sup>Department of Physics and Astronomy, University of Missouri-Kansas City, Kansas City Missouri, USA

<sup>▽</sup>Wenzhou Institute of the University of Chinese Academy of Sciences, Wenzhou, Zhejiang 325000, China

<sup>#</sup>School of Physical Sciences and Kavli Institute of Theoretical Science, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>†</sup>CAS Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100090, China

<sup>¶</sup>Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia

\*Corresponding Author: Wai-Yim Ching, Email: [Chingw@umkc.edu](mailto:Chingw@umkc.edu)

## Abstract

The spike protein of SARS-CoV-2 binds to 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 analysis 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, including the MD simulations with molecular mechanics generalized Born surface area (MM-GBSA) method to predict 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<sup>493</sup>, Y<sup>505</sup>, Q<sup>498</sup>, N<sup>501</sup>, T<sup>500</sup>, N<sup>487</sup>, Y<sup>449</sup>, F<sup>486</sup>, K<sup>417</sup>, Y<sup>489</sup>, F<sup>456</sup>, Y<sup>495</sup>, and L<sup>455</sup> have been identified as hotspots in RBD, while those in ACE2 are K<sup>353</sup>, K<sup>31</sup>, D<sup>30</sup>, D<sup>355</sup>, H<sup>34</sup>, D<sup>38</sup>, Q<sup>24</sup>, T<sup>27</sup>, Y<sup>83</sup>, Y<sup>41</sup>, E<sup>35</sup>, and E<sup>37</sup>. Both the electrostatic and hydrophobic interactions are the main driving force to form the AA-AA binding pairs. We confirm that Q<sup>493</sup>, N<sup>501</sup>, F<sup>486</sup>, K<sup>417</sup>, and F<sup>456</sup> in RBD are the key residues responsible for the tight binding of SARS-CoV-2 with ACE2 compared to SARS-CoV. The DFT results reveal that N<sup>487</sup>, Q<sup>493</sup>, Y<sup>449</sup>, T<sup>500</sup>, G<sup>496</sup>, G<sup>446</sup> and G<sup>502</sup> in RBD form pairs *via* specific hydrogen bonding with Q<sup>24</sup>, H<sup>34</sup>, E<sup>35</sup>, D<sup>38</sup>, Y<sup>41</sup>, Q<sup>42</sup> and K<sup>353</sup> in ACE2.

KEYWORDS: SARS-CoV-2 RBD-ACE2 interface complex, Binding free energy, Key amino acids, Interatomic bonding, Density functional calculation.

## 1. Introduction

The COVID-19 pandemic, caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has infected millions of people world-wide due to its rapid global spread, long incubation period, proclivity to evolve new variants, and lack of safe and effective therapies or vaccines.<sup>1</sup> 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.<sup>2-5</sup> Three of these vaccines, the Pfizer-BioNTech, Moderna, Johnson and Johnson, are now authorized for emergency use by Food and Drug Administration in the United States.<sup>6</sup> The key elements in developing these vaccines are the rapid determination of the SARS-CoV-2 genomic sequence,<sup>7</sup> the 3D structure of its spike protein,<sup>8,9</sup> and its mechanism in penetrating human cell receptor.<sup>9,10</sup>

SARS-CoV-2 is an enveloped single-stranded RNA virus with spike-like glycoproteins protruding from its exterior membrane surface forming a ‘corona’.<sup>1</sup> It has four main structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N).<sup>11</sup> The S-protein is responsible for viral entry into the human host cell during the infection, making it a primary target for vaccine development,<sup>12-14</sup> repurposing antiviral drugs or discovery of new ones,<sup>15,16</sup> and development of therapeutic antibodies.<sup>17,18</sup> S-protein exists in 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.<sup>19</sup> To engage with ACE2, RBD undergoes hinge-like movements to change from down to up states for easier accessibility.<sup>8</sup> The nature of interaction between RBD and ACE2 is not only important for understanding the initial step of virus infection, but also provides necessary information to guide effective development of drugs or vaccine.

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),<sup>20,21</sup> 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,<sup>22-40</sup> 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 binding interaction. *Second*, 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. *Third*, the role of the aqueous environment and its impact on the RBD-ACE2 interactions remains to be highlighted. 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,<sup>41</sup> 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 by using efficient 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 S-protein,<sup>42</sup> as well as in formulating a *de novo* method to evaluate the inter-amino acid interactions in 3-D of these subdomains.<sup>43,44</sup>

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 well-known MD simulation methodology for assessment of binding free energy

(BFE), and *ab initio* based DFT methodology to elucidate the binding phenomenology. Two replicate MD simulations for all-atom SARS-CoV-2 RBD-ACE2 complex are performed with explicit water molecules over a time scale of 100 ns followed by 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 is 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 orthogonalized linear combination of atomic orbitals (OLCAO) method to investigate the atomic-scale resolved interaction, partial charge distribution, and intra- and inter-molecular binding of the RBD-ACE2 interface model similar to the one used for the spike protein.<sup>42</sup>

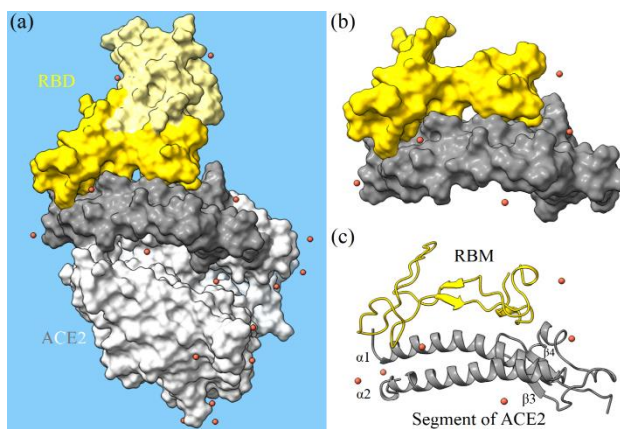
## 2. Model Construction

### 2.1 Molecular Model for MD simulations

We explicitly designed and simulated the interface complex model in a bathing solution environment, composed of water and ions. This model is labeled as the *Interface-MD model* which contains residues from T<sup>333</sup> to G<sup>526</sup> of SARS-CoV-2 RBD (194 AAs), and the residues from S<sup>19</sup> to D<sup>615</sup> of the ACE2 N-terminal peptidase domain (PD) (597 AAs). Also, it contains one zinc ion (Zn<sup>+2</sup>), one chloride ion (Cl<sup>-1</sup>), 25 sodium ions (Na<sup>+</sup>) and 27000 water molecules (**Table 1** and **Figure 1(a)**). The details of creating this model are summarized as follows. The initial structure of the interface complex was obtained from the crystal structure PDB (ID: 6M0J).<sup>20</sup> This initial structure is solvated in a periodic water box using the TIP3P explicit water model implemented in AMBER (Assisted Model Building with Energy Refinement) package, after adding H atoms using the LEaP module.<sup>45, 46</sup> In particular, 27,000 water molecules are added to the initial complex with 25 Na<sup>+</sup> ions to neutralize the total charge of the model because the ACE2 carries a total charge of 28 e<sup>-</sup>, zinc with 2 e<sup>+</sup>, chloride ion with 1 e<sup>-</sup> and RBD has a net charge of 2 e<sup>+</sup>. The placing of the Na<sup>+</sup> ions was performed *via* a Coulomb potential on a grid using the LEaP program in the AMBER package. The most recent AMBER force field ff14SB is used to represent the parameterizations of the inter- and intra-molecular interactions of this Interface-MD model.<sup>47</sup>

### 2.2 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-atoms calculations are obviously impossible at present since MD simulation could involve in excess of hundred thousand atoms. To this end, we have reduced the size of the interface model



**Figure 1.** The interface complex models of bound SARS-CoV-2 RBD to ACE2 receptor. (a) The Interface-MD model of the whole RBD-ACE2 complex (in surface representation) solvated in a water box (light blue) and neutralized with Na<sup>+</sup> ions (orange). Light yellow and yellow: RBD, yellow: RBM, gray: ACE2, and dark gray: segment of ACE2 used for DFT calculation. (b) The 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.

to a much smaller and manageable size containing only the most relevant AAs at the interface. This model is labelled as the Interface-DFT model derived from the Interface-MD model (with the removal of water and some ions). The Interface-DFT model contains all residues of the receptor-binding motif (RBM) of SARS-CoV-2 (71 AAs from S<sup>438</sup> to Y<sup>508</sup>) and 117 residues from ACE2 (70 residues of  $\alpha$ 1 and  $\alpha$ 2 motifs from the S<sup>19</sup> to I<sup>88</sup> plus 47 residues from G<sup>319</sup> to T<sup>365</sup> of  $\beta$ 3 and  $\beta$ 4 motifs and with some other residues of ACE2) as well as 6 Na<sup>+</sup> ions to neutralize the system (**Figure 1 (b)**). 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,<sup>20, 21</sup> the ACE2 segments selected for the Interface-DFT model include all 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 Interface-DFT model is fully optimized using VASP and used as the input for the electronic structure and properties calculation using OLCAO methods.

**Table 1.** Summary of models for MD simulation and DFT calculation.

Model	# of Atoms in RBD	# of Atoms in ACE2	# of water molecules	# of Zn <sup>+2</sup> ion	# of Cl <sup>-</sup> ion	# of Na <sup>+</sup> ions	Total # of atoms
Interface-MD	3001	9523	27000	1	1	25	93551
Interface-DFT	1102	1822	---	---	---	6	2930

### 3. Methods

#### 3.1 Molecular dynamic (MD) simulation

All-atom MD simulations for the Interface-MD model are carried out using AMBER 18 simulation package in explicit solvent with periodic boundaries.<sup>48</sup> Prior to MD simulations, the model underwent six stages of energy minimization using 5000 cycles of steepest descent followed by 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, 1 kcal/mol-Å<sup>2</sup> are applied to hold the solute (RBD-ACE2 complex) in 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 K to 310 K for 310 picoseconds (ps) using the NVT ensemble with a weak restraint of 10 kcal/mol-Å<sup>2</sup> on the solute, followed by the 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, long-range electrostatic interactions with the Particle Mesh Ewald (PME) method.<sup>49</sup> Both the direct space PME and the Lennard-Jones cut-offs are set at 10 Å. The SHAKE algorithm is applied for constraining all bonds involving hydrogen atoms with a 2 fs time step.<sup>50</sup> All these protocols are conducted using the PMEMD.CUDA module in AMBER.<sup>51, 52</sup>

### 3.2 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 of S-protein and ACE2 receptor. It combines the molecular MM energies with the GBSA continuum solvation approach. The BFEs are calculated by utilizing single-trajectory protocol (STP) of MM-GBSA method based on long-timescale MD simulations in the context of an explicit solvent TIP3P water model, with a neutralized total charge of the system. Subsequently, all solvent molecules and charged ions are removed from each MD snapshot, and the implicit GBSA solvent models are used to estimate the solvation free energies at 0.15 M ionic concentration. In the STP approach, 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 ( $G_{RBD, sol}$ )<sup>53-55</sup>

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

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,  $\Delta E_{MM}$ ,  $\Delta G_{sol}$ , and  $-T\Delta S$  represent the changes in the gas phase MM energy, solvent-free energy, and conformational entropy upon binding, respectively.  $\Delta E_{MM}$  is the sum of the changes in the bonded energy ( $\Delta E_{int}$ ), the non-bonded electrostatic energy ( $\Delta E_{ele}$ ) and the van der Waals energy ( $\Delta E_{vdW}$ ).  $\Delta G_{sol}$  is divided into electrostatic or polar solvation energy part ( $\Delta G_{GB}$ ), and a non-electrostatic or nonpolar part ( $\Delta G_{SA}$ ) 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 (BP) model, whereas  $\Delta G_{SA}$  follows from a linear dependence on the solvent-accessible surface area (SASA).<sup>56</sup> The total electrostatic contribution to free energy is the sum of  $\Delta E_{ele}$  and  $\Delta G_{GB}$  ( $\Delta G_{ele} = \Delta E_{ele} + \Delta G_{GB}$ ). The entropy contribution ( $-T\Delta S$ ) is the sum of the change in translational, rotational, and vibrational entropies.<sup>55</sup> 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.<sup>55</sup> Once the MD simulations are completed, snapshots are taken for every 10 ps over the whole 100 ns. 5000 snapshots are extracted for the BFE post-process analysis. This analysis has been carried out using the MM-GBSA method *via* MMPBSA.py module of AMBER.<sup>55</sup> The model developed by Onufriev et al. (GB<sup>OBC</sup> which GB = 2) is employed as the GB model and a set of radii *mbondi2* is prepared.<sup>57</sup> A surface tension coefficient ( $\gamma$ ) of 0.0072 kcal/mol-Å<sup>2</sup> and zero correction constant ( $b = 0$ ) are employed to calculate  $\Delta G_{SA}$ . The value of the exterior dielectric constant of water was

set to 78.3 while the dielectric constant for the solute is set to the vacuum value of 1.<sup>58</sup> Only 5 snapshots along simulations are used to evaluate -TAS at 310 K.

Two types of BFE decomposition have been undertaken to dissect the calculated BFE of the RBD-ACE2 complex in terms of individual residue coined as “*per-residue BFE decomposition*” and residue-residue contributions as “*pairwise BFE decomposition*”.<sup>55</sup> 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 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, sidechain, and total contributions. It can also decompose into interaction components ( $\Delta E_{vdW}$ ,  $\Delta E_{ele}$ , *etc.*).

### 3.3 Structural relaxation for DFT calculation using VASP

The Interface-DFT model is fully relaxed by using Vienna *ab initio* simulation package (VASP) known for its efficiency in structure optimization.<sup>59</sup> We use the projector augmented wave (PAW) method with Perdew-Burke-Ernzerhof (PBE) exchange correlation functional<sup>60</sup> within the generalized gradient approximation (GGA). The input parameters used in VASP are as follows: energy cut-off 500 eV, electronic convergence of  $10^{-4}$  eV, force convergence criteria for ionic steps at  $-10^{-2}$  eV/Å and a single k-point sampling. 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.

### 3.4 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 orbitals (OLCAO) method,<sup>61</sup> developed in-house. The efficacy of using these two different DFT codes is well documented,<sup>62-64</sup> 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 pairs of atoms. They are obtained from the *ab initio* wave functions with atomic basis expansion:

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

$$\rho_{\alpha\beta} = \sum_{m,occ} \sum_{i,j} C_{i\alpha}^{*m} C_{j\beta}^m S_{i\alpha,j\beta}. \quad (9)$$

In the above equations,  $S_{i\alpha,j\beta}$  are the overlap integrals between the  $i^{th}$  orbital in  $\alpha^{th}$  atom and the  $j^{th}$  orbital in the  $\beta^{th}$  atom.  $C_{j\beta}^m$  are the eigenvector coefficients of the  $m^{th}$  occupied molecular orbital level. The partial charge (PC) or ( $\Delta Q_{\alpha} = Q_{\alpha}^0 - Q_{\alpha}^*$ ) is the deviation of the effective charge  $Q_{\alpha}^*$  from the neutral atomic charge  $Q_{\alpha}^0$  on the same atom  $\alpha$ . 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 calculation of PC and BO are based on the Mulliken scheme.<sup>65, 66</sup>

The BO  $\rho_{\alpha\beta}$  in Eq. (9) is further extended to quantify the bonding strength between a pair of amino acids (u,v) called amino acid -amino acid bond pair (AABP).<sup>44</sup> In many cases, the use of AABP is more useful than interatomic bonding between a pair of atoms for biomolecular systems

$$AABP(u, v) = \sum_{\alpha \in u} \sum_{\beta \in v} \rho_{\alpha i, \beta j}. \quad (10)$$

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

## 4. Results

### 4.1 MD Results

#### 4.1.1 Binding free energy between RBD and ACE2

We performed two replicate MD runs for Interface-MD model over the time range of 100 ns. The RMSD (root mean square deviation) and RMSF (root mean square fluctuation) of these two simulations are presented in Supplementary Information (SI, Figures S1, S2). The STP approach of the MM-GBSA method<sup>53</sup> 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 ( $\Delta G_{\text{bind}}$ ) with their energetic components from both MD runs. Our predicted value of  $\Delta G_{\text{bind}} = -12.86 \pm 0.1$  kcal/mol is in excellent agreement with experimentally determined binding affinity of -11.83 kcal/mol using surface plasmon resonance (SPR) assay.<sup>20</sup> More specifically, the measured value of the equilibrium dissociation constant ( $K_D$ ) of the interface complex is 4.7 nM,<sup>20</sup> which is equivalent to  $\Delta G_{\text{bind}}$  of -11.83 kcal/mol at 37 °C (i.e.  $\Delta G = RT \ln(K_D)$ , where R is the gas constant). A similar experimental study showed that the  $K_D$  of the SARS-CoV RBD and ACE2 is 31 nM (or  $\Delta G$  is -10.7 kcal/mol).<sup>20</sup> 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.

**Table 2** lists the various energy components contributing to  $\Delta G_{\text{bind}}$  (**Method Section**). Since the STP approach of MM-GBSA method was used in our calculations, there is no contribution from internal energy, or  $\Delta E_{\text{int}} = 0$ . On the other hand, the non-bonded interactions of  $\Delta E_{\text{vdw}} = -90.1$  kcal/mol and  $\Delta E_{\text{ele}} = -700.92 \pm 0.59$  kcal/mol both favorably contribute to form a stable RBD-ACE2 complex. Such highly favored  $\Delta E_{\text{ele}}$  behavior can be explained by the strong electrostatic attractions between the positively charged AAs of RBD and the negative charged AAs of ACE2. Thus, strong

**Table 2.** Predicted BFE (kcal.mol<sup>-1</sup>) and energetic components at 0.15 M salt. SE: standard error.

Energy	Run1	Run2	Aver.	SE
$\Delta E_{\text{vdw}}$	-90.10	-90.11	-90.10	0
$\Delta E_{\text{ele}}$	-699.75	-702.10	-700.92	0.59
$\Delta E_{\text{MM}}$	-789.84	-792.21	-791.03	0.59
$\Delta G_{\text{GB}}$	747.47	749.51	748.49	0.51
$\Delta G_{\text{SA}}$	-13.27	-13.16	-13.21	0.03
$\Delta G_{\text{sol}}$	734.20	736.36	735.28	0.54
$\Delta G_{\text{ele}}$	47.72	47.42	47.57	0.08
$-T\Delta S$	42.95	42.83	42.89	0.03
$\Delta G_{\text{vac}}$	-746.89	-749.38	-748.14	0.62
$\Delta G_{\text{bind}}$	-12.69	-13.02	-12.86	0.08

interaction at the interface complex is electrostatic in origin, which agrees with the recent study by

Nguyen et al.<sup>35</sup> However,  $\Delta E_{\text{ele}}$  is completely overshadowed by  $\Delta G_{\text{GB}}$  of  $748.49 \pm 0.51$  kcal/mol, leading to an overall unfavorable contribution of the total  $\Delta G_{\text{ele}}$  of  $47.57 \pm 0.1$  kcal/mol. 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  $\Delta G_{\text{GB}}$  and favorable  $\Delta G_{\text{SA}}$  of  $-13.21 \pm 0.03$  kcal/mol. The total entropic contribution ( $-T\Delta S$ ) is 42.89 kcal/mol reflecting the significant entropic cost of confining the RBD with little movement when associated with ACE2. Finally, the  $\Delta G_{\text{bind}}$  can be further partitioned in terms of the nature of the interactions in vacuum ( $\Delta G_{\text{vac}} = -748.14 \pm 0.62$  kcal/mol) and/or solvent ( $\Delta G_{\text{sol}} = 735.28 \pm 0.54$  kcal/mol). This pattern indicates that  $\Delta G_{\text{vac}}$  dominates and stabilizes the dimer association between RBD and ACE2 while the aqueous solution opposes to it. However, the unavoidable solvent effect cannot be ignored.

From the above analysis, we conclude that the long-range Coulombic electrostatic interaction ( $\Delta E_{\text{ele}}$ ) is responsible for initiating the recognition process between the RBD and ACE2, while the attractive component of the van der Waals interaction ( $\Delta E_{\text{vdw}}$ ) is the main factor to stabilize the RBD-ACE2 complex. Interestingly, the  $\Delta E_{\text{vdw}}$  is also the main force to stabilize other complexes such as drug-DNA,<sup>53, 54</sup> ligand-protein,<sup>67</sup> and protein-protein.<sup>68</sup> The salient feature of this trend can provide the guidance on what should be targeted when a new drug is designed.

#### 4.1.2 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,<sup>20, 21</sup> 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.<sup>54</sup>

##### *i. 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  $\Delta G$  of -1 to -0.15 kcal/mol; “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:

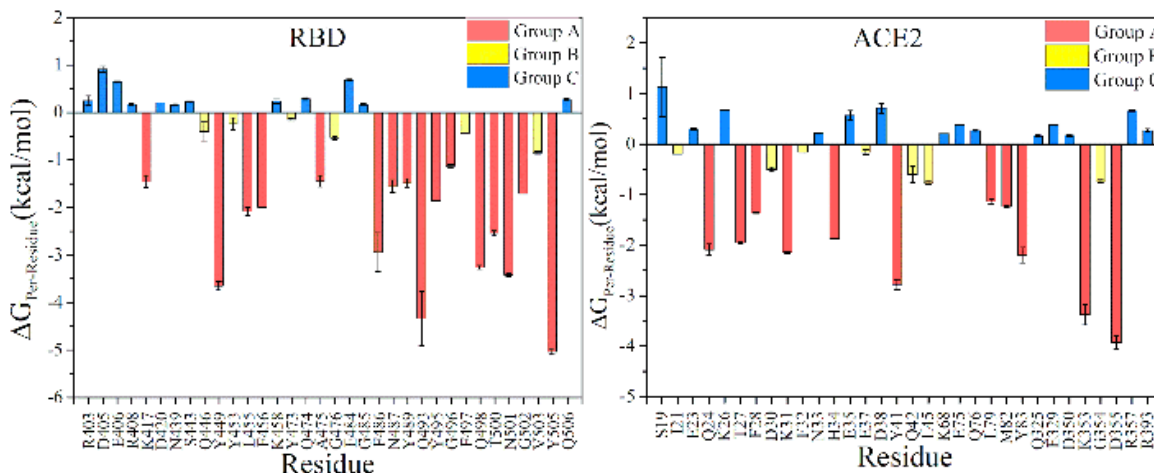
(a) **Group A.** There are 16 (11) AAs in RBD (ACE2) that contribute significantly to the total BFE. They are Y<sup>505</sup>, Q<sup>493</sup>, Y<sup>449</sup>, N<sup>501</sup>, Q<sup>498</sup>, F<sup>486</sup>, T<sup>500</sup>, L<sup>455</sup>, F<sup>456</sup>, Y<sup>495</sup>, G<sup>502</sup>, N<sup>487</sup>, Y<sup>489</sup>, K<sup>417</sup>, A<sup>475</sup>, and G<sup>496</sup> in RBD and D<sup>355</sup>, K<sup>353</sup>, Y<sup>41</sup>, Y<sup>83</sup>, K<sup>31</sup>, Q<sup>24</sup>, T<sup>27</sup>, H<sup>34</sup>, F<sup>28</sup>, M<sup>82</sup>, and L<sup>79</sup> in ACE2 (**Figure 2, Table S1**). All these AAs in RBD are located in RBM except K<sup>417</sup> which forms a salt-bridge with D<sup>30</sup> of ACE2.  $\Delta G$  of K<sup>417</sup> =  $-1.46 \pm 0.13$  kcal/mol with a favored  $\Delta G_{\text{ele}}$  of  $-1.18 \pm 0.13$  kcal/mol. Residues Y<sup>505</sup>, Q<sup>493</sup>, Y<sup>449</sup>, N<sup>501</sup>, Q<sup>498</sup>, F<sup>486</sup>, and T<sup>500</sup> have the most favored binding with  $\Delta G < -2.5$  kcal/mol. They form interface hotspots with ACE2. A close examination of *per-residue* BFE reveals that  $\Delta E_{\text{vdw}}$  are consistently favorable for all 16 AAs but with different strengths. The  $\Delta G_{\text{ele}}$  displays more variance in favorability with only eight AAs (Y<sup>505</sup>, Q<sup>493</sup>, Y<sup>449</sup>, Q<sup>498</sup>, Y<sup>495</sup>, G<sup>502</sup>, K<sup>417</sup>, and G<sup>496</sup>) having a favorable  $\Delta G_{\text{ele}}$ . The  $\Delta G_{\text{SA}}$  are favorable (negative) and similar in all 16 AAs except G<sup>496</sup>. These 16 AAs are also reported by crystal structure data except Y<sup>495</sup>.<sup>20</sup>  $\Delta G$  of Y<sup>495</sup> is



-1.86 kcal/mol, and differs from other tyrosine residues (Y<sup>505</sup>, Y<sup>449</sup> and Y<sup>489</sup>) in that its backbone participates in the interactions. For the 11 AAs in ACE2, six are in  $\alpha 1$  (Q<sup>24</sup>, T<sup>27</sup>, F<sup>28</sup>, K<sup>31</sup>, H<sup>34</sup>, and Y<sup>41</sup>) and three in  $\alpha 2$  (L<sup>79</sup>, M<sup>82</sup> and Y<sup>83</sup>) motifs. The loop connecting  $\beta 3$  and  $\beta 4$  motifs contributes *via* two highly interacting AAs (D<sup>355</sup> and K<sup>353</sup> with  $\Delta G$  of  $-3.93 \pm 0.13$  and  $-3.38 \pm 0.2$  kcal/mol respectively). Both  $\Delta E_{vdW}$  and  $\Delta G_{SA}$  contribute favorably to the total  $\Delta G_{bind}$  in all 11 AAs of ACE2. On other hand, only two AAs (D<sup>355</sup> and Y<sup>83</sup>) have favorable contributions from  $\Delta G_{ele}$ . All these residues in ACE2 were also confirmed by crystal structure.<sup>20</sup>

(b) **Group B.** These are AAs V<sup>503</sup>, G<sup>476</sup>, F<sup>497</sup>, Q<sup>446</sup>, Y<sup>453</sup>, and Y<sup>473</sup> in RBD and L<sup>45</sup>, G<sup>354</sup>, Q<sup>42</sup>, D<sup>30</sup>, I<sup>21</sup>, F<sup>32</sup>, and E<sup>37</sup> in ACE2 (**Figure 2, Table S2**). Interestingly, only three AAs in RBD (Q<sup>446</sup>, Y<sup>453</sup>, and Y<sup>473</sup>) and five in ACE2 residues (D<sup>30</sup>, E<sup>37</sup>, Q<sup>42</sup>, L<sup>45</sup> and G<sup>354</sup>) have been reported.<sup>20, 29, 30, 35, 36</sup> The  $\Delta E_{vdW}$  and  $\Delta G_{SA}$  are favorable, but there are no contributions to  $\Delta G_{SA}$  from F<sup>497</sup>, I<sup>21</sup>, and F<sup>32</sup>. On the other hand,  $\Delta G_{ele}$  is unfavorable in all AAs in Group B.

(c) **Group C.** These are 12 AAs in RBD (N<sup>439</sup>, R<sup>408</sup>, G<sup>485</sup>, D<sup>420</sup>, S<sup>443</sup>, K<sup>458</sup>, R<sup>403</sup>, Q<sup>506</sup>, Q<sup>474</sup>, E<sup>406</sup>, E<sup>484</sup>, and D<sup>405</sup>) and 14 in ACE2 (Q<sup>325</sup>, D<sup>350</sup>, K<sup>68</sup>, N<sup>33</sup>, Q<sup>76</sup>, R<sup>393</sup>, E<sup>23</sup>, E<sup>329</sup>, E<sup>75</sup>, E<sup>35</sup>, R<sup>357</sup>, K<sup>26</sup>, D<sup>38</sup>, and S<sup>19</sup>) (**Figure 2, Table S3**). Five residues are located outside RBM. Their distinctive sidechains are either positively (R<sup>403</sup> and R<sup>408</sup>) or negatively (D<sup>405</sup>, E<sup>406</sup>, and D<sup>420</sup>) charged. They have unfavorable  $\Delta G_{ele}$  and  $\Delta G_{GB}$ . None of these 12 AAs were reported in the X-ray studies and only the E<sup>35</sup>, D<sup>38</sup>, R<sup>357</sup>, and R<sup>393</sup> of ACE2 were mentioned.<sup>20, 21</sup> This can be attributed to the cut-off 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 interaction in group C. Some of these AAs have also been identified in other studies.<sup>29, 30, 35</sup>



**Figure 2.** Per-residue energy decomposition of the key interacting AAs of RBD-ACE2 complex. Left panel is shown the interacting AAs belong to RBD, and right panel for interacting AAs in ACE2.

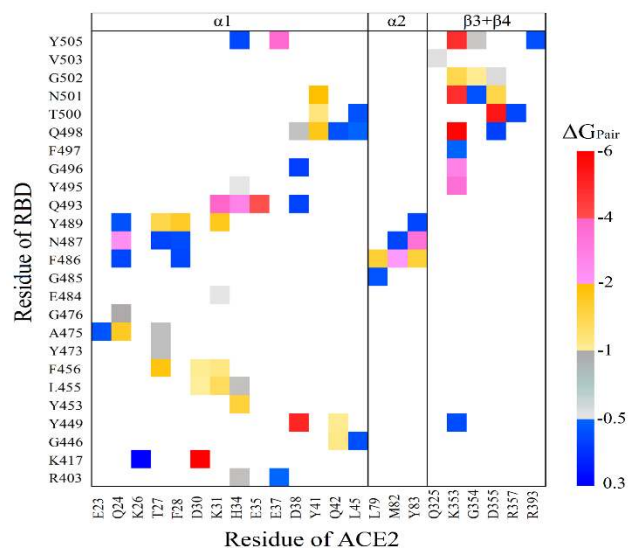
## ii. Pairwise BFE decomposition

To gain more details on the interacting AA-AA pairs between RBD and ACE2, we apply the *pairwise BFE decomposition* scheme. The pair interactions map of the RBD-ACE2 is shown in **Figure 3** and listed in **Table S4**. Remarkably, 9 AAs in RBD (D<sup>405</sup>, E<sup>406</sup>, R<sup>408</sup>, D<sup>420</sup>, N<sup>439</sup>, S<sup>443</sup>, K<sup>458</sup>, Q<sup>474</sup>, and Q<sup>506</sup>) and 9 AAs in ACE2 (S<sup>19</sup>, I<sup>21</sup>, F<sup>32</sup>, N<sup>33</sup>, K<sup>68</sup>, Q<sup>76</sup>, E<sup>75</sup>, E<sup>329</sup>, and D<sup>350</sup>), that have

been assigned in the previous subsection, are not engaged in AA-AA pairs. On other hand, 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. The first five pairs have favorable  $\Delta G_{\text{ele}}$  (Table S4) due to formation of HBs (Table S5). Interestingly,  $K^{417}:D^{30}$  is the most attractive pair arising from the ionic interactions between a cation of NZ nitrogen atom on  $K^{417}$  sidechain and the anions of OD1 and OD2 oxygen atoms on the sidechain of  $D^{30}$  in the presence of HBs (Table S5). In fact, the most obvious difference in the RBD–ACE2 complex between SARS-CoV-2 and SARS-CoV occurs at this  $K^{417}$  residue. This residue is in fact replaced by a hydrophobic residue of  $V^{404}$  in SARS-CoV RBD that does not participate in the ACE2 binding.<sup>20, 36</sup> Significantly, the stronger binding between SARS-CoV-2 RBD and ACE2 is mainly attributed to the electrostatic interaction of  $K^{417}:D^{30}$  pair. Other two pairs ( $N^{501}:K^{353}$  and  $Y^{505}:K^{353}$ ) have similar favorable contributions from  $\Delta E_{\text{vdW}}$ ,  $\Delta G_{\text{ele}}$  and  $\Delta G_{\text{SA}}$  (Table S4). The  $N^{501}:K^{353}$  pair is stabilized by HBs but have low HB occupancy (see Figure 9(a) later). X-ray study<sup>20</sup> did not report any of these HBs and only one has been mentioned elsewhere.<sup>36</sup> The  $Y^{505}:K^{353}$  pair is driven by  $\pi$ -Alkyl interaction between the  $\pi$ -orbital in phenol ring on  $Y^{505}$  and 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 3) with  $\Delta G < -4$  but  $\geq -2$  kcal/mol. Except  $F^{486}:M^{82}$ , all pairs are held together by HBs (Table S5). The attractive components of  $\Delta G_{\text{ele}}$  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 S4). On the other hand, the  $F^{486}:M^{82}$  pair is purely driven by favorable  $\Delta E_{\text{vdW}}$  and  $\Delta G_{\text{SA}}$  arising from  $\pi$ -Alkyl and  $\pi$ -Sulfur interactions between the  $\pi$ -orbital in the benzene ring in  $F^{486}$  sidechain and the CG carbon atom and SD sulfur atom of  $M^{82}$ . Other pairs shown in Figure 3 and Table S4 have less strong interactions. Figure 3 clearly shows that RBD has more AAs in forming pairs with  $\alpha 1$  than  $\alpha 2$  of ACE2 (45 pairs vs 7 pairs). This suggests that focusing on RBD blocker with  $\alpha 1$  in ACE2 may be a promising therapeutic strategy for inhibiting S-protein SARS-CoV-2.<sup>69-71</sup> Additionally, RBD forms 15 pairs with  $K^{353}$  to  $D^{355}$  connecting  $\beta 3$  and  $\beta 4$  of ACE2.

In conclusion, the highest interface hotspots of RBD based on total  $\Delta G$  of pairs with ACE2 are  $Q^{493}$ ,  $Y^{505}$ ,  $Q^{498}$ ,  $N^{501}$ ,  $T^{500}$ ,  $N^{487}$ ,  $Y^{449}$ ,  $F^{486}$ ,  $K^{417}$ ,  $Y^{489}$ ,  $F^{456}$ ,  $Y^{495}$ , and  $L^{455}$ . After the sequence alignment between interacting AAs of RBD as shown in Figure S3, the most interacting AAs enhancing BFE of SARS-CoV-2 compared to SARS-CoV are:  $Q^{493}$ ,  $Q^{498}$ ,  $N^{501}$ ,  $F^{486}$ ,  $K^{417}$ ,  $F^{456}$ ,  $Y^{495}$ , and  $L^{455}$ . Among them, we believe  $Q^{493}$ ,  $N^{501}$ ,  $F^{486}$ ,  $K^{417}$ , and  $F^{456}$  are the most critical.



**Figure 3.** The AA-AA interaction pair map between SARS-CoV-2 RBD and ACE2. 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  $\Delta G_{\text{Pair}}$ .

## 4.2 DFT Results

### 4.2.1 Structural optimization

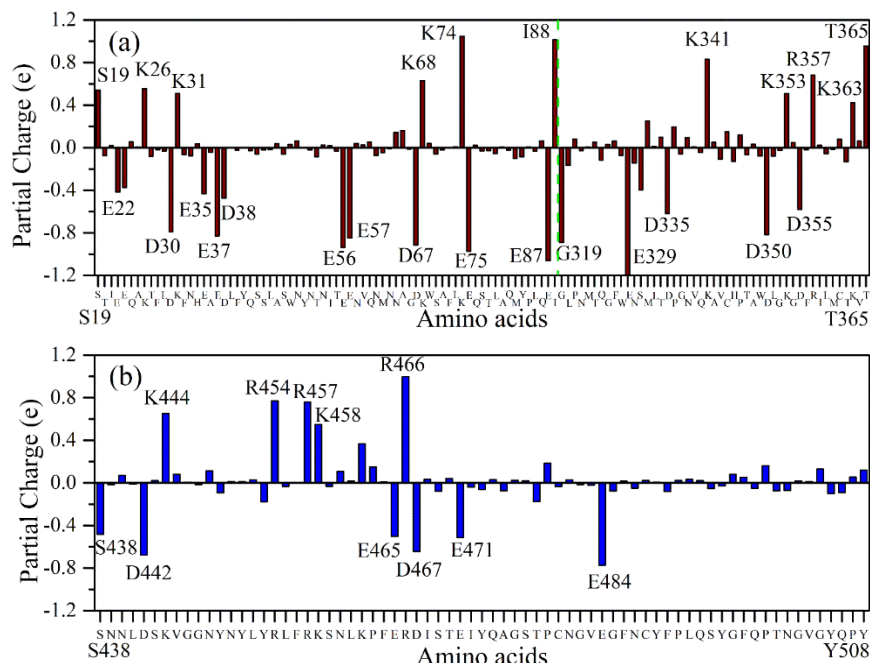
The structural optimization of Interface-DFT model with high accuracy in atomic coordinates is critical for realistic quantum chemical calculations.<sup>42</sup> Such high-precision simulation of interface model is expected to complement experimental studies to provide a deeper understanding at atomic levels of the various aspects of the initial viral infection and at a much lower cost. In the present study, the geometrical optimization has been carried out by using VASP package based on the DFT methodology (see **Method section**).

The Interface-DFT model consists of 2930 atoms. The initial unrelaxed structure has a total energy of -18052.021 eV and the final relaxed structure has -18039.025eV respectively. 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 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 longtime 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 DFT calculation is a single point calculation on a specific model of fixed atomic positions. The fully optimized structure for 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.

### 4.2.2 Electronic structure and partial charge distribution

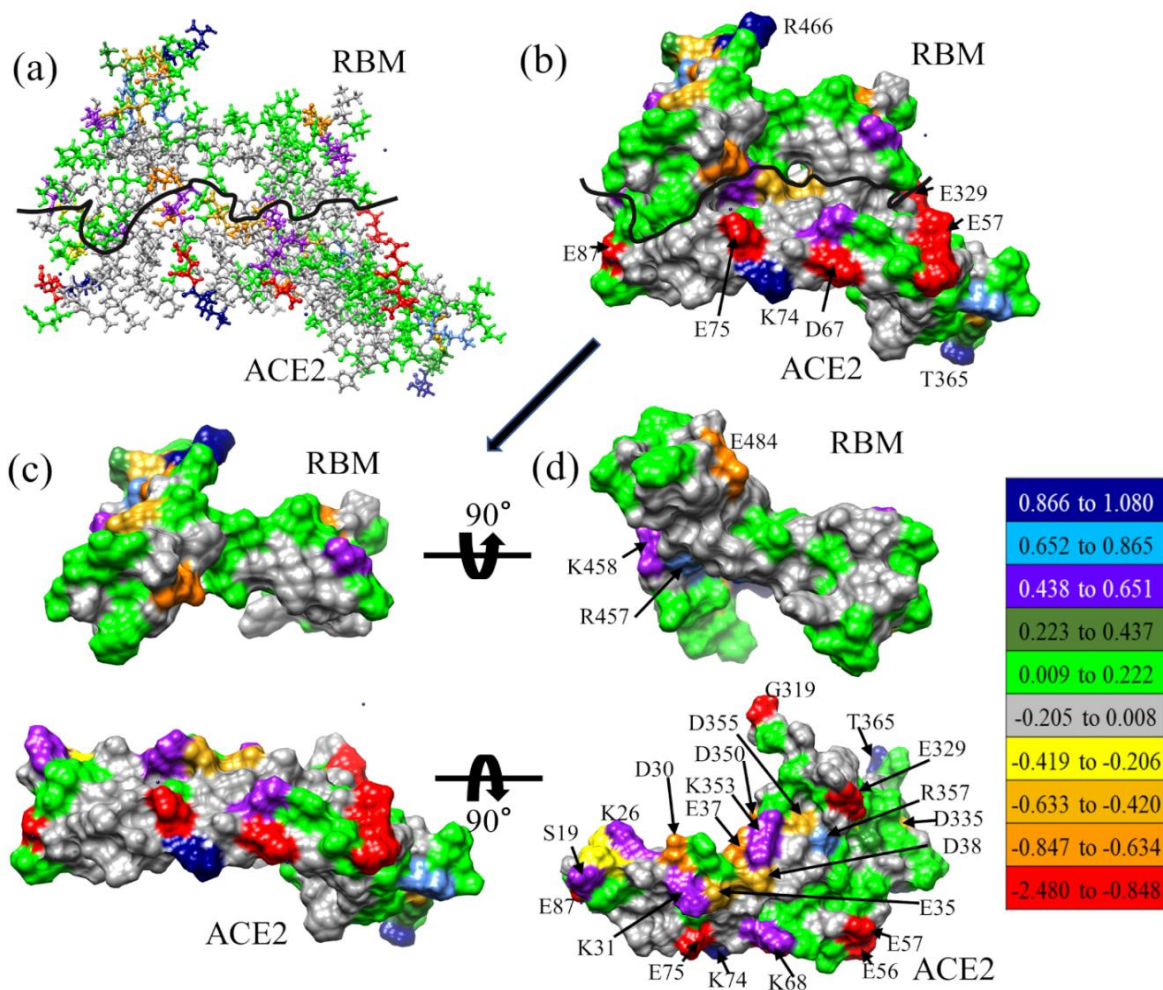
The electronic structure in the form of density of states (DOS) and partial DOS (PDOS) for the Interface-DFT model are calculated and presented in **Figure S4**. It is noted that the total DOS and the PDOS for RBM and ACE2 parts are almost identical in shape and the structures they contain.

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.<sup>72</sup> For the Interface-DFT model, we have determined the PC on every atom and on every residue using the OLCAO method (see **Method section**). The calculated PC in unit of electron charge (e) on each of the 2930 atoms are grouped into 194 AAs and 6 Na<sup>+</sup> ions. **Figure 4 (a) and (b)** shows the PC of the 194 AAs in the segment of ACE2 (19-88 and 319-365) and RBM (438-508). The PC can be positive or negative, with some AAs exhibiting a very large PC. They are: E<sup>329</sup>, E<sup>87</sup>, E<sup>75</sup>, E<sup>56</sup>, D<sup>67</sup>, G<sup>319</sup>, E<sup>57</sup>, E<sup>37</sup>, D<sup>350</sup> with PC values of -2.480 e, -1.061 e, -0.976 e, -0.939 e, -0.915 e, -0.892 e, -0.849 e, -0.831 e, -0.818 e and K<sup>341</sup>, T<sup>365</sup>, R<sup>466</sup>, I<sup>88</sup>, K<sup>74</sup> with PC values of 0.832 e, 0.956 e, 0.999 e, 1.014 e, 1.047 e, respectively. All Na ions have positive PC ranging from 0.880 e<sup>-</sup> to 1.080 e. The complete data for the PC in every AA in the Interface DFT mode is listed in **Table S6**.



**Figure 4.** PC distribution in Interface-DFT model of (a) segment of ACE2 in sequence 19-88 and 319-365 and (b) RBM in sequence 438-508. Amino acids

The PC distribution on the ball and stick sketch for all AAs is displayed in **Figure 5 (a)**, together with the solvent excluded surfaces in **Figure 5 (b)**. The black curve in **Figure 5 (a)** and **(b)** roughly shows the interface boundary separating the RBM and ACE2. **Figure 5 (c)** shows separated RBM and ACE2, which are further rotated in **Figure 5 (d)** to show the interacting surfaces in the RBM and ACE2. In **Figure 5 (d)**, we have marked residues with PC lower and higher than  $-0.4 e$  and  $0.4 e$ , 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.



**Figure 5.** PC distribution in Interface-DFT model in (a) ball and stick and (b) surface, (c) shows separate RBM and ACE2, and (d) shows rotated surface of RBM and ACE2 to show the PC distribution in amino acids. The curved black line in (a) and (b) roughly shows the partition between RBM and ACE2 in the interface. The color bar shows 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). In (d) we have explicitly marked amino acids lower and higher than  $-0.4$  e and  $0.4$  e respectively.



### 4.2.3 Interatomic bonding

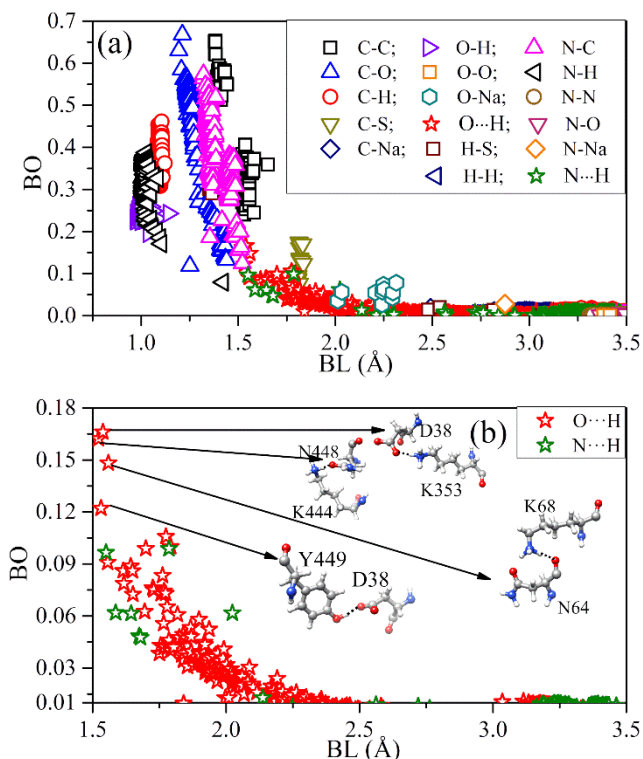
The bond order (BO) vs. bond length (BL) for Interface-DFT is shown in **Figure 6(a)**. There are 17 types of different bonds including O $\cdots$ H and N $\cdots$ H hydrogen bonding (HBs). The bonds with stronger BO are C-O, and C-C. The C-C bonds can be partitioned into two groups, of which one exhibits a higher BO that 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 Sulphur containing residues and the C-S bonds are associated with them. However, there are no disulfide bonds which are usually common between Cys residues in 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 Å 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 (0.018 e) and one N-Na bond with BL of 2.9 Å (0.028 e). **Figure 6(b)** separately shows the HB (O $\cdots$ H and N $\cdots$ H) distributions in the Interface-DFT model. Most of the HBs are from O $\cdots$ H and they BL range from 1.5 Å and become very weak after 2.5 Å and certainly cannot be classified as HBs for distance of separation > 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 6(b)** shows that some so-called HB in O $\cdots$ H can have BO values close to 0.16 e. More details on the different ways the HBs are analyzed in MD and in OLCAO will be discussed in **Section 5.3**.

## 5. 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 their 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.

### 5.1 Nature of interfacial cohesion and interaction between RBD and ACE2

1) Detailed MD simulation on Interface-MD model using 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  $\Delta G_{\text{ele}}$  since the HB networks or ionic interactions



**Figure 6.** Interatomic interactions in the interface-DFT model. (a) BO vs BL for every pair of atoms (b) Hydrogen bonding. HBs are represented by star sign.

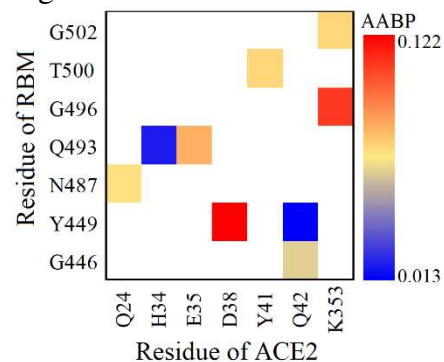
can stabilize them (see subsection 4.1.2.ii). Other pairs are identified by the favorably  $\Delta E_{\text{vdw}}$  and  $\Delta G_{\text{SA}}$ . For instance,  $F^{486}:L^{79}$ ,  $F^{486}:M^{82}$ , and  $F^{486}:Y^{83}$  pairs are driven by  $\Delta E_{\text{vdw}}$  and  $\Delta G_{\text{SA}}$  (**Table S4**) and form the hydrophobic pockets between SARS-CoV-2 RBD and ACE2, important for anchoring the interface complex.

2) By construct a smaller Interface-DFT model between RBM and  $\alpha 1$ ,  $\alpha 2$ ,  $\beta 3$  and  $\beta 4$  motifs in the ACE2 membrane receptor, large-scale *ab initio* DFT calculations are carried out similar to what has been accomplished in our unprecedented large-scale calculation of the structure and properties of S-protein in COVID-SARS-2 virus,<sup>42-44</sup> and related electronic dielectric constants of small proteins.<sup>73</sup> Interface-DFT 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. Detailed electronic structure and interatomic bonding, partial charge distribution on each atom and residues 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 become possible as will be described in **Section 5.2** below.

## 5.2 Interfacial amino acid -amino acid interaction

There are 194 AAs in the Interface-DFT model, 71 in RBM and 117 in ACE2. **Table 3** or **Figure 7** shows the calculated AABP values for each interacting pair in RBM and ACE2 with the strength of the interactions. The strongest binding is between two pairs:  $G^{496}:K^{353}$  and  $Y^{449}:D^{38}$  in RBD and ACE2 with AABP values of (0.100 e, 0.122 e) respectively.

**Figure 8** further shows the intricate bonding configurations between these two AAs, with HBs playing a key role in the AA-AA interaction across the interfacial boundary. The AABP 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 8 (b)** shows two pairs with one common AA. Here both  $E^{35}$  and  $H^{34}$  have one HB with  $Q^{493}$ . Similarly, **Figure 8(c)** shows interaction of  $Y^{449}$  with  $D^{38}$  and  $Q^{42}$ , and  $G^{446}$  with  $Q^{42}$ . The pairs between  $T^{500}:Y^{41}$  and  $G^{496}:K^{353}$  or  $K^{353}:G^{502}$  are shown in **Figure 8(d)** and **(e)** respectively. This intertwined bonding in the interface region shows a complex topology that would be difficult to discern with less advanced computational methodology.



**Figure 7.** AA to AA bond pair between RBM and ACE2.

**Table 3.** Amino acid to amino acid bond pair between RBM and ACE2 in the Interface-DFT model.

Pair in RBM:ACE2	AABP
$N^{487}:Q^{24}$	0.034
$Q^{493}:H^{34}$	0.015
$Q^{493}:E^{35}$	0.053
$Y^{449}:D^{38}$	0.122
$T^{500}:Y^{41}$	0.038
$G^{446}:Q^{42}$	0.028
$Y^{449}:Q^{42}$	0.013
$G^{496}:K^{353}$	0.100
$G^{502}:K^{353}$	0.038

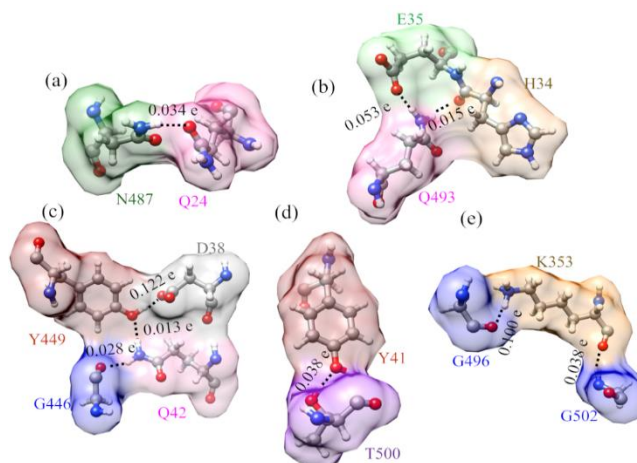
### 5.3 Hydrogen Bonding analysis

Analysis of 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, 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,<sup>74</sup> 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<sub>2</sub>O and O in another H<sub>2</sub>O). In the MD simulation, following the guidelines of IUPAC in 2011 and applying them to more complex biomolecular systems, the HBs identification is based solely on the geometric consideration, with no quantitative definition for the strength of the HB.<sup>75</sup>

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

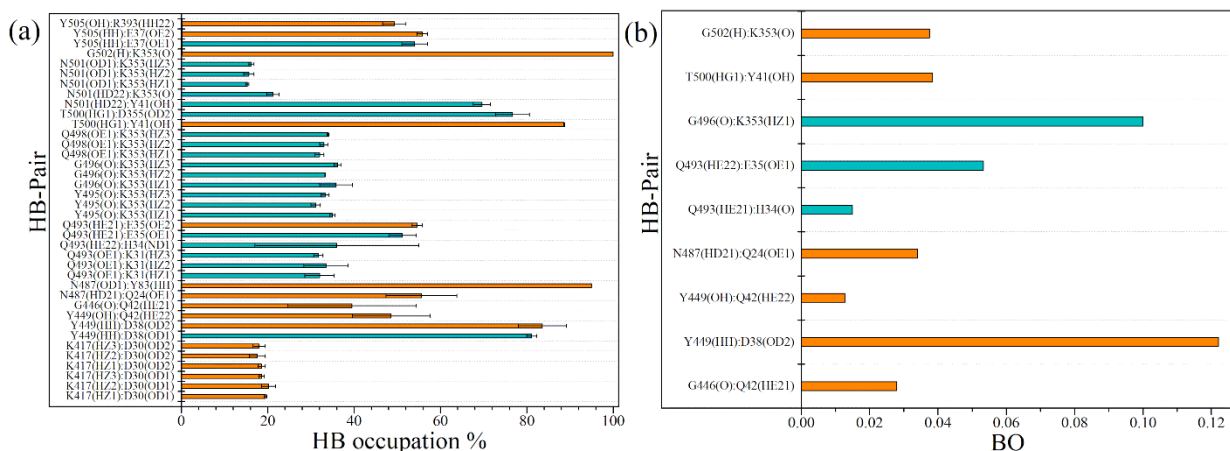
**Figures 9(a)** and **9(b)** show the possible HBs from MD simulations and DFT calculations respectively. In MD simulations, the CPPTRAJ program in AMBER is utilized to calculate HBs based on geometric criteria using a cutoff acceptor and donor (A...D) distance < 4.0 Å and a cutoff angle > 130° between the acceptor, hydrogen, and donor atoms.<sup>76</sup> 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 calculation, the HB is determined by its BL and BO values as shown in **Figure 9(b)**. We use BO cutoff of 0.01 e 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 9(a)** shows that in MD simulation 7 HBs have a very high frequency of occupancy of more than 70%, with 4 of them having been also reported in the X-ray study<sup>20</sup>: Y<sup>449</sup>(HH):D<sup>38</sup>(OD2), N<sup>487</sup>(OD1):Y<sup>83</sup>(HH), T<sup>500</sup>(HG1):Y<sup>41</sup>(OH), and G<sup>502</sup>(H):K<sup>353</sup>(O). The other 3 HBs predicted are: Y<sup>449</sup>(HH):D<sup>38</sup>(OD1), T<sup>500</sup>(HG1):D<sup>355</sup>(OD2), and N<sup>501</sup>(HD22):Y<sup>41</sup>(OH). The remaining possible HBs have occupancy of less than 70% (**Figure 9(a)**). On the other hand, the HB analysis based on DFT calculation shows that two HBs, Y<sup>449</sup>(HH):D<sup>38</sup>(OD2) and G<sup>496</sup>(O):K<sup>353</sup>(HZ1), have a higher BO of ≥ 0.1 e<sup>-</sup>, and are also observed in MD simulations. Five



**Figure 8.** Amino acid-amino acid bond pair (AABP) between RBD and ACE2 in ball and stick for participating pairs. (a) N<sup>487</sup>:Q<sup>24</sup>, (b) Q<sup>493</sup>:E<sup>35</sup> and Q<sup>493</sup>:H<sup>34</sup>, (c) Y<sup>449</sup>:D<sup>38</sup>, Y<sup>449</sup>:Q<sup>42</sup>, and G<sup>446</sup>:Q<sup>42</sup>, (d) T<sup>500</sup>:Y<sup>41</sup>, and (e) G<sup>496</sup>:K<sup>353</sup>, K<sup>353</sup>:G<sup>502</sup>. It is interesting that there is only one hydrogen bonding in each interacting pairs.



other HBs have medium BO values in the range of 0.053 to 0.027  $e^-$  (**Figure 9(b)**), with 4 of them observed also experimentally as well as in MD simulations. The remaining two HBs are relatively weaker with BO less than 0.02  $e^-$  and only one of them is experimentally detected.<sup>20</sup>



**Figure 9.** Histogram diagram of HB network of the interface complex. (a) MD simulation with their percentage HB occupancies. (b) HBs from the single Interface-DFT model with their bond orders (BO) represent their strength. The orange bars indicate the HB pairs have been experimentally observed while the green bar represents the one predicted in the present study.

**Table S5** summarizes the comparison and reveals that both MD simulations and DFT calculation on a single model can predict large number of HBs, far more than that from the experimental X-ray study.<sup>20</sup> Unlike the static single structure of X-ray or the relaxed Interface-DFT model, each AA in 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 acceptor and H atom attached to donor in DFT calculations are less ambiguous compared to the experimental X-ray crystal structure, and certain HBs such as N<sup>501</sup>(N):Y<sup>41</sup>(OH) and Y<sup>489</sup>(OH):Y<sup>83</sup>(OH) that have been claimed experimentally as HBs,<sup>20</sup> but were not found in MD simulations or DFT calculations. Those may indeed not be actual HBs.

#### 5.4 Implications for electrostatic interactions of different PCs

The electrostatic interactions in the protein-protein binding process are well known.<sup>72</sup> 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 (section 4.2.2). Additionally, the electrostatic interactions and solvent effects of Interface-MD model have been investigated based on fixed PCs (section 4.1). 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 Section 4.1 and in other existing studies.<sup>33-35, 77, 78</sup> 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* calculation.

Calculating the free energy of the total electrostatic contribution ( $\Delta G_{\text{ele}}$ ) 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 simulation.<sup>73</sup> To this end, we separately calculated  $\Delta G_{\text{ele}}$  of the Interface-DFT model based on the fixed PCs from AMBER force field ff14SB,<sup>47</sup> 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.<sup>79, 80</sup> DelPhi calculates the polar solvation energy implicitly by solving the non-linear and/or linear variants of the Poisson Boltzmann (PB) equation numerically using finite difference methodology.<sup>79, 80</sup> It can provide the Coulombic electrostatic interaction ( $\Delta E_{\text{ele}}$ ) 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 RMSC (root mean square change of potential) and MAXC (potential maximum change threshold of potential), the external dielectric constant 78.3 and ionic concentration was 0.15 M. Three values of internal dielectric constants chosen as a test are 1, 5 and 10. The energy in unit of kT are converted to kcal/mol at temperature  $T = 310$  K.

**Table 4** shows the results of  $\Delta G_{\text{ele}}$  and its components using two different sets of atomic PCs with different assumed internal dielectric constants which are unknown for the protein.<sup>73</sup> The favorability pattern of  $\Delta G_{\text{ele}}$  and its components are similar for both sets of PCs and exhibits the same behavior in MD simulations (**Table 2**). The computed values of  $\Delta G_{\text{ele}}$ ,  $\Delta E_{\text{ele}}$  and  $\Delta G_{\text{PB}}$  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  $\Delta E_{\text{ele}}$  and more favored  $\Delta G_{\text{PB}}$  and  $\Delta G_{\text{ele}}$ . Again, the computed values of  $\Delta G_{\text{ele}}$  and its components greatly depend on the internal dielectric constant which is unknown.<sup>73</sup> Many MD simulations chose them *a posteriori* to justify their result to be in line with experiments.

**Table 4.** Predicted electrostatic interactions and their energetic components using two different set of PC values at a 0.15 M salt concentration and three assumed interior dielectric constants  $\epsilon_i$ .

PCs	$\epsilon_i = 1$			$\epsilon_i = 5$			$\epsilon_i = 10$		
	$\Delta E_{\text{ele}}$	$\Delta G_{\text{PB}}$	$\Delta G_{\text{ele}}$	$\Delta E_{\text{ele}}$	$\Delta G_{\text{PB}}$	$\Delta G_{\text{ele}}$	$\Delta E_{\text{ele}}$	$\Delta G_{\text{PB}}$	$\Delta G_{\text{ele}}$
<b>AMBER</b>	-200.30	249.10	48.81	-40.06	47.34	7.28	-20.03	22.11	2.08
<b>OLCAO</b>	-132.49	302.45	169.96	-26.50	57.36	30.86	-13.24	26.52	13.27

## 6. Conclusions

The methodology and the ensuing analysis of biomolecular interaction presented above has broader repercussions in the general context of biomolecular computational physics, and should not be seen as specifically delimited only to the case of the SARS-CoV-2 RBD-ACE2 complex that was 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 SARS-CoV-2 RBD-

ACE2 complex with 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 ACEs at atomistic level including the contribution from hydrogen bonding topology.

4. The two combined and complementary methodologies provide the missing link that was 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.

## Reference

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., Covid-19 Vaccine Development and a Potential Nanomaterial Path Forward. *Nature nanotechnology*. **2020**, *15*, 646-655.
2. Lurie, N.; Saville, M.; Hatchett, R.; Halton, J., Developing Covid-19 Vaccines at Pandemic Speed. *New England Journal of Medicine*. **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. *New England Journal of Medicine*. **2020**, *383*, e63.
5. Jackson, L. A.; Anderson, E. J.; Roupheal, N. G.; Roberts, P. C.; Makhene, M.; Coler, R. N.; McCullough, M. P.; Chappell, J. D.; Denison, M. R.; Stevens, L. J., An Mrna Vaccine against Sars-Cov-2—Preliminary Report. *New England Journal of Medicine*. **2020**.
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., 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.; Veessler, 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., 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. *Virology journal*. **2019**, *16*, 1-22.
12. Tian, J.-H.; Patel, N.; Haupt, R.; Zhou, H.; Weston, S.; Hammond, H.; Logue, J.; Portnoff, A. D.; Norton, J.; Guebre-Xabier, M., Sars-Cov-2 Spike Glycoprotein Vaccine Candidate Nvx-

- Cov2373 Immunogenicity in Baboons and Protection in Mice. *Nature communications*. **2021**, *12*, 1-14.
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., 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**, 1-7.
  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. *Science advances*. **2020**, *6*, 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**, 1-8.
  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. *Nature communications*. **2021**, *12*, 1-10.
  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., Resistance of Sars-Cov-2 Variants to Neutralization by Monoclonal and Serum-Derived Polyclonal Antibodies. *Nature medicine*. **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., Fusion Mechanism of 2019-Ncov and Fusion Inhibitors Targeting Hr1 Domain in Spike Protein. *Cellular & molecular immunology*. **2020**, 1-3.
  20. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L., Structure of the Sars-Cov-2 Spike Receptor-Binding Domain Bound to the Ace2 Receptor. *Nature*. **2020**, *581*, 215-220.
  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. *Nature microbiology*. **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., Conformational Dynamics of Sars-Cov-2 Trimeric Spike Glycoprotein in Complex with Receptor Ace2 Revealed by Cryo-Em. *Science advances*. **2021**, *7*, 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. *Journal of virology*. **2020**, *94*.
  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 research*. **2021**, 1-3.
  26. Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K.-Y., Structural and Functional Basis of Sars-Cov-2 Entry by Using Human Ace2. *Cell*. **2020**, *181*, 894-904. e9.
  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., 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**.
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. *Cellular and Molecular Life Sciences*. **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 in bioinformatics*. **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 and Pathogenesis? **2020**.
34. Amin, M.; Sorour, M. K.; Kasry, A., Comparing the Binding Interactions in the Receptor Binding Domains of Sars-Cov-2 and Sars-Cov. *The journal of physical chemistry letters*. **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? *The Journal of Physical Chemistry 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. *Proceedings of the National Academy of Sciences*. **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. *The journal of physical chemistry letters*. **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. *Computational and structural biotechnology journal*. **2020**, *18*, 2573-2582.
39. Qiao, B.; Olvera de la Cruz, 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. *Journal of chemical information and modeling*. **2020**.
41. Klompas, M.; Baker, M. A.; Rhee, C., Airborne Transmission of Sars-Cov-2: Theoretical Considerations and Available Evidence. *Jama*. **2020**.
42. 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. *Computational and Structural Biotechnology Journal*. **2021**, *19*, 1288-1301.

43. 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. *Physical Chemistry Chemical Physics*. **2020**, *22*, 18272-18283.
44. Adhikari, P.; Ching, W.-Y., Amino Acid Interacting Network in the Receptor-Binding Domain of Sars-Cov-2 Spike Protein. *RSC Advances*. **2020**, *10*, 39831-39841.
45. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics*. **1983**, *79*, 926-935.
46. 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. *Computer Physics Communications*. **1995**, *91*, 1-41.
47. 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. *Journal of Chemical Theory and Computation*. **2015**, *11*, 3696-3713.
48. Case, D.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Greene, D.; Homeyer, N., Amber Reference Manual. *University of California*. **2018**.
49. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald: An N·Log(N) Method for Ewald Sums in Large Systems. *The Journal of chemical physics*. **1993**, *98*, 10089-10092.
50. Miyamoto, S.; Kollman, P. A., Settle: An Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *Journal of computational chemistry*. **1992**, *13*, 952-962.
51. Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with Amber on Gpus. 1. Generalized Born. *Journal of chemical theory and computation*. **2012**, *8*, 1542-1555.
52. 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. *Journal of chemical theory and computation*. **2013**, *9*, 3878-3888.
53. Jawad, B.; Poudel, L.; Podgornik, R.; Steinmetz, N. F.; Ching, W.-Y., Molecular Mechanism and Binding Free Energy of Doxorubicin Intercalation in DNA. *Physical Chemistry Chemical Physics*. **2019**, *21*, 3877-3893.
54. 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. *The Journal of Physical Chemistry B*. **2020**, *124*, 7803-7818.
55. Miller III, B. R.; McGee Jr, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., Mmpbsa. Py: An Efficient Program for End-State Free Energy Calculations. *Journal of chemical theory and computation*. **2012**, *8*, 3314-3321.
56. Wang, J.; Hou, T.; Xu, X., Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Current Computer-Aided Drug Design*. **2006**, *2*, 287-306.
57. Onufriev, A.; Bashford, D.; Case, D. A., Modification of the Generalized Born Model Suitable for Macromolecules. *The Journal of Physical Chemistry B*. **2000**, *104*, 3712-3720.
58. Homeyer, N.; Gohlke, H., Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Molecular informatics*. **2012**, *31*, 114-122.
59. Package, V.-V. A. i. S. <https://www.vasp.at/>.



60. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Physical review letters*. **1996**, *77*, 3865.
61. Ching, W.-Y.; Rulis, P., *Electronic Structure Methods for Complex Materials: The Orthogonalized Linear Combination of Atomic Orbitals*. Oxford University Press: 2012.
62. 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. *Physical Chemistry Chemical Physics*. **2016**, *18*, 21573-21585.
63. 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. *Scientific reports*. **2014**, *4*, 1-7.
64. 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. *The Journal of Physical Chemistry B*. **2017**, *121*, 6321-6330.
65. Mulliken, R. S., Electronic Population Analysis on Lcao–Mo Molecular Wave Functions. I. *The Journal of Chemical Physics*. **1955**, *23*, 1833-1840.
66. Mulliken, R., Electronic Population Analysis on Lcao–Mo Molecular Wave Functions. II. Overlap Populations, Bond Orders, and Covalent Bond Energies. *The Journal of Chemical Physics*. **1955**, *23*, 1841-1846.
67. 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. *Journal of the American Chemical Society*. **2005**, *127*, 11827-11834.
68. Nilofer, C.; Sukhwai, A.; Mohanapriya, A.; Kanguane, P., Protein-Protein Interfaces Are Vdw Dominant with Selective H-Bonds and (or) Electrostatics Towards Broad Functional Specificity. *Bioinformatics*. **2017**, *13*, 164.
69. Han, Y.; Král, P., Computational Design of Ace2-Based Peptide Inhibitors of Sars-Cov-2. *ACS nano*. **2020**, *14*, 5143-5147.
70. 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. *Communications biology*. **2021**, *4*, 1-9.
71. 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.
72. Zhou, H.-X.; Pang, X., Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chemical reviews*. **2018**, *118*, 1691-1741.
73. 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*, 10.26434/chemrxiv.14403188.v1, **2021**.
74. Liang, L.; Rulis, P.; Ouyang, L.; Ching, W., Ab Initio Investigation of Hydrogen Bonding and Network Structure in a Supercooled Model of Water. *Physical Review B*. **2011**, *83*, 024201.
75. 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., Definition of the Hydrogen Bond (Iupac Recommendations 2011). *Pure and applied chemistry*. **2011**, *83*, 1637-1641.

76. Roe, D. R.; Cheatham III, T. E., Ptraaj and Cpptraaj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of chemical theory and computation*. **2013**, *9*, 3084-3095.
77. Bai, C.; Warshel, A., Critical Differences between the Binding Features of the Spike Proteins of Sars-Cov-2 and Sars-Cov. *The Journal of Physical Chemistry B*. **2020**, *124*, 5907-5912.
78. 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. *Frontiers in molecular biosciences*. **2020**, *7*.
79. 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 biophysics*. **2012**, *5*, 1-11.
80. Li, C.; Jia, Z.; Chakravorty, A.; Pahari, S.; Peng, Y.; Basu, S.; Koirala, M.; Panday, S. K.; Petukh, M.; Li, L., Delphi Suite: New Developments and Review of Functionalities. *Journal of computational chemistry*. **2019**, *40*, 2502-2508.

## ASSOCIATED CONTENT

### SUPPORTING INFORMATION:

Additional figures and Tables are provided in the Supporting Information.

### AUTHOR INFORMATION

#### Corresponding Author

Wai-Yim Ching, Email: [Chingw@umkc.edu](mailto:Chingw@umkc.edu)

ORCID 0000-0001-7738-8822

Rudolf Podgornik, Email: [rudipod@gmail.com](mailto:rudipod@gmail.com)

ORCID 0000-0002-3855-4637

#### Author Contributions

BJ and WC conceived the project. BJ and PA performed the calculations and made most of the figures. WC, BJ and RP drafted the paper with inputs from RP. All authors participated in the discussion and interpretation of the results. All authors edited and proofread the final manuscript.

#### Funding Sources

This project is funded by the National Science Foundation of USA: RAPID DMR/CMMT-2028803.

#### ACKNOWLEDGMENT

This research used the resources of the National Energy Research Scientific Computing Center supported by DOE under Contract No. DE-AC03-76SF00098 and the Research Computing Support Services (RCSS) of the University of Missouri System. This project is funded by the National Science Foundation of USA: RAPID DMR/CMMT-2028803. RP acknowledges funding from the Key project #12034019 of the National Natural Science Foundation of China.