

# MMGB/SA Consensus Estimate of the Binding Free Energy Between the Novel Coronavirus Spike Protein to the Human ACE2 Receptor.

Negin Forouzesh<sup>†</sup> and Alexey V. Onufriev<sup>\*,‡,¶,§</sup>

<sup>†</sup>*Department of Computer Science, California State University, Los Angeles, Los Angeles, CA 90032, USA*

<sup>‡</sup>*Department of Computer Science, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061, USA*

<sup>¶</sup>*Department of Physics, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061, USA*

<sup>§</sup>*Center for Soft Matter and Biological Physics, Virginia Polytechnic Institute & State University, Blacksburg, VA 24061, USA*

E-mail: alexey@cs.vt.edu

## Abstract

The ability to estimate protein-protein binding free energy in a computationally efficient via a physics-based approach is beneficial to research focused on the mechanism of viruses binding to their target proteins. Implicit solvation methodology may be particularly useful in the early stages of such research, as it can offer valuable insights into the binding process, quickly. Here we evaluate the potential of the related molecular mechanics generalized Born surface area (MMGB/SA) approach to estimate the binding free energy  $\Delta G_{bind}$  between the SARS-CoV-2 spike receptor-binding domain

and the human ACE2 receptor. The calculations are based on a recent flavor of the generalized Born model, GBNSR6. Two estimates of  $\Delta G_{bind}$  are performed: one based on standard bondi radii, and the other based on a newly developed set of atomic radii (OPT1), optimized specifically for protein-ligand binding. We take the average of the resulting two  $\Delta G_{bind}$  values as the consensus estimate. For the well-studied Ras-Raf protein-protein complex, which has similar binding free energy to that of the SARS-CoV-2/ACE2 complex, the consensus  $\Delta G_{bind} = -11.8 \pm 1$  kcal/mol, vs. experimental  $-9.7 \pm 0.2$  kcal/mol.

The consensus estimates for the SARS-CoV-2/ACE2 complex is  $\Delta G_{bind} = -9.4 \pm 1.5$  kcal/mol, which is in near quantitative agreement with experiment ( $-10.6$  kcal/mol). The availability of a conceptually simple MMGB/SA-based protocol for analysis of the SARS-CoV-2 /ACE2 binding may be beneficial in light of the need to move forward fast.

Keywords: binding free energy, implicit solvent, coronavirus, COVID-19

## Introduction

Emerged as a global threat to human health, the SARS-CoV-2 virus that causes COVID-19 disease has been studied widely since the start of 2020.<sup>1</sup> Despite sequence and structure similarities with other viruses,<sup>2</sup> no highly effective treatment option for the novel coronavirus is available. As of today, about 14 million people across the globe have tested positively for the virus, and around 600,000 have died of COVID-19.<sup>3</sup> This fast-growing pandemic highlights the role of computational structural biology and computer-aided drug design (CADD), which have the ability to accelerate the slow and expensive process of drug discovery.<sup>4</sup> In structure-based drug discovery, accuracy and speed of the binding free energy prediction of drug-like compounds (ligands) to target biomolecules plays a key role in virtual screening of drug candidates.<sup>5-7</sup> Despite decades of research, efficient and accurate computational prediction of binding free energies is still a challenge.<sup>8-12</sup>

In theory, the binding free energy of a molecular system can be estimated directly from thermodynamic first principles.<sup>13</sup> However, for any realistic molecular systems, *e.g.*, the complex of interest in this work that is made of more than 12,000 atoms,<sup>14</sup> approximations must be made to make the estimate computationally feasible. For example, alchemical methods,<sup>15,16</sup> simulate changes in the free energy along a pathway that sometimes reflects non-physical properties or literally “alchemy”. The required sample points along the pathway are generated via Monte Carlo or Molecular Dynamics (MD) simulations. Some of the popular methods in this class are thermodynamic integration (TI) and free energy perturbations (FEP).<sup>17</sup> However, these simulations are still computationally expensive, specifically when it comes to high throughput virtual screening of thousands of potential drugs.

Remarkably more efficient, end-point free energy methods ignore details of the complex to unbound state pathway and estimate free energy on an ensemble of snapshots representing the initial (complex) and the final (unbound) states only. These snapshots can be generated by an MD simulation. Molecular mechanics Poisson-Boltzmann surface area (MMPB/SA) and molecular mechanics generalized Born surface area (MMGB/SA)<sup>18–20</sup> are arguably among the most popular methods in this category. They are often used in docking projects where a quick estimate of binding affinities is required.<sup>21</sup> Leading docking software, for instance, AutoDock Vina<sup>22</sup> and DOCK,<sup>23</sup> rank the feasible poses of a ligand in a binding pocket based on a scoring function in which binding affinity plays an important role. End-point free energies can improve the accuracy of these scoring functions on-the-fly. Recently, MMGB/SA was employed to improve the accuracy of AutoDock Vina and Dock in the Drug Design Data Resource (D3R) Grand Challenge 4 (GC4).<sup>24</sup>

While calculations based on practical implicit solvation models such as generalized Born (GB) are arguably not as accurate as corresponding estimates based on the best available explicit solvent models, the use of implicit solvent has an undeniable appeal. And not only of computational efficiency: reasoning about physical origins of observed effects is often much more transparent in an implicit than explicit solvent.<sup>25–27</sup> That last advantage may be

particularly valuable now, when so much about COVID-19 structure-infectivity relationship remains unknown.

In this work, we employ MMGB/SA implemented in AmberTools18<sup>28</sup> for binding free energy calculation of the SARS-CoV-2 spike receptor-binding domain (SARS-CoV-2 S RBD) and the human ACE2 receptor (PDB ID:6m0j), see Fig. 1. Through the MMGB/SA approach, the absolute binding free energy of a complex is calculated as the sum of gas-phase energy, solvation free energy, and entropic contributions averaged over several snapshots extracted from the main MD trajectory. A grid-based surface GB model is used for estimating the polar component of solvation free energy, coupled with water and atomic radii introduced earlier.<sup>29</sup> Human H-Ras and the Ras-binding domain of C-Raf1, so-called Ras-Raf complex,<sup>6,30</sup> is chosen as the reference for the initial evaluation of the MMGB/SA model. Final results are compared with those from a few available relevant studies. The main goal of the work is a quick assessment of the potential of the simple and efficient MMGB/SA method to future studies of SARS-CoV-2 to ACE2 binding.

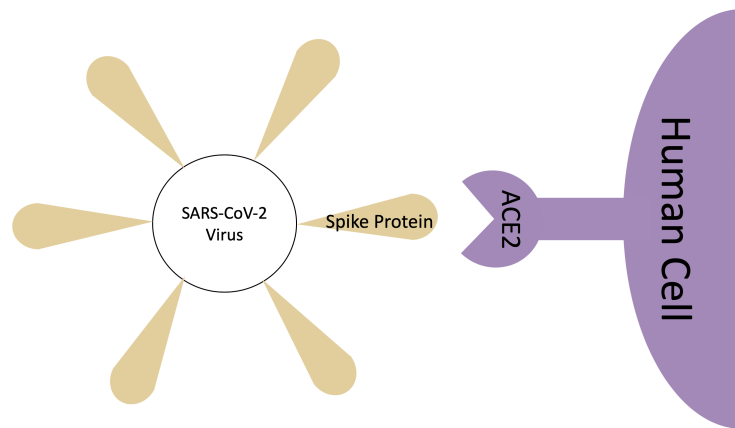


Figure 1: Binding scheme of the SARS-CoV-2 spike protein to the ACE2 human receptor.

## Methods and Materials

### Binding Free Energy Decomposition

Binding free energy,  $\Delta G_{bind}$ , of a molecular system is calculated as follows:

$$\Delta G_{bind} = \Delta H - T\Delta S, \quad (1)$$

where  $\Delta H$  is the enthalpy change of the system,  $T$  is the absolute temperature in K, and  $\Delta S$  is the entropy change of the system. A high-level illustration of  $\Delta G_{bind}$  between bound and unbound states of a solvated complex is shown in Fig. 2.

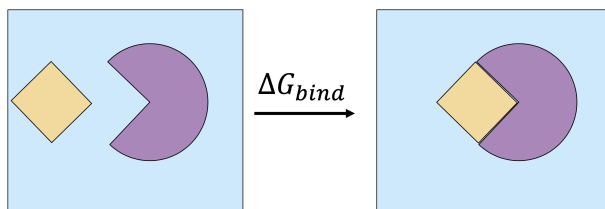


Figure 2: Binding a ligand (shown in yellow) to a protein receptor (shown in purple) in a box of solvent (shown in blue) releases binding free energy of  $\Delta G_{bind}$ . A negative sign of  $\Delta G_{bind}$  indicates that spontaneous binding occurs, the magnitude of  $|\Delta G_{bind}|$  characterizes the binding strength (affinity).

In theoretical/computational studies, a useful way of calculating  $\Delta G_{bind}$  is through a thermodynamic cycle shown in Fig. 3.

With this approach,  $\Delta G_{bind, solv}$  is calculated as follows:

$$\Delta G_{bind, solv} = \Delta G_{bind, vacuum} + \Delta G_{solv, complex} - (\Delta G_{solv, ligand} + \Delta G_{solv, receptor}). \quad (2)$$

The solvation free energy,  $\Delta G_{solv}$ , is broken into the polar and non-polar components:

$$\Delta G_{solv} = \Delta G_{pol} + \Delta G_{nonpol}. \quad (3)$$

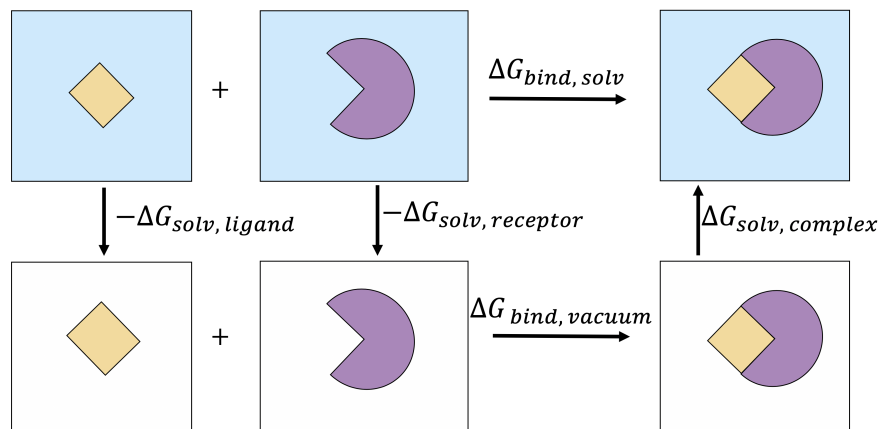


Figure 3: The thermodynamic cycle used here to estimate the binding free energy of a protein-ligand complex in the solvent.

The free energy in vacuum,  $\Delta G_{vacuum}$ , is decomposed into the gas-phase energy ( $\Delta E_{MM}$ ) and the configurational entropy of the solute ( $T \Delta S$ ):

$$\Delta G_{vacuum} = \Delta E_{MM} - T \Delta S. \quad (4)$$

Note that the  $T\Delta S$  above does not exactly correspond to  $T\Delta S$  in Eq. 1; specifically, the entropy of solvent re-arrangement<sup>25,27</sup> is subsumed into  $\Delta G_{solv}$ , see below, which is then considered a part of  $\Delta H$ . Combining the free energy components defined above, we obtain:  $\Delta H = \Delta E_{MM} + \Delta G_{pol} + \Delta G_{nonpol}$ . Our approaches for calculating  $\Delta G_{solv}$ ,  $\Delta E_{MM}$  and  $T \Delta S$  are explained below.

## MMPB/SA Free Energy Methodology

MMPB/SA is a popular end-point free energy method which estimates  $\Delta G_{solv}$  by Poisson-Boltzmann implicit solvent model,<sup>31</sup> while components of  $\Delta E_{MM}$  are estimated based on a classical Molecular Mechanics force-field. (In MMGB/SA, discussed below and used here, the role of the PB is played by the faster GB). Significantly faster than the conventional Alchemical methods, MMPB/SA can be very useful, particularly in the early stages of structure-based

virtual screening. As another important advantage, is that through MMPB/SA it is possible to decompose the total free energy into sub-components and measure their contributions separately.<sup>6,30</sup> This feature is certainly useful when it comes to comparing several different free energy methods. Finally, MMPB/SA is applicable to a wide range of structures,<sup>32</sup> from small host-guest systems to large protein-protein complexes with thousands of atoms.<sup>6</sup>

Through the MMPB/SA approach, the average of  $\Delta G_{solv}$  is calculated on a collection of snapshots extracted from an MD simulation. Several decisions have to be made in applying the approach in practice. First, the computational protocol must be selected between the “single-trajectory” (one trajectory of the complex), or “separate-trajectory” (three separate trajectories of the complex, receptor and ligand). In this study, we choose the former protocol as it was shown<sup>33</sup> to be not only much faster than the alternative, but also less “noisy” due to the cancellation of inter-molecular energy contributions. This protocol applies to cases where significant structural changes upon binding are not expected. Shown in Fig. 4, the single-trajectory MMPB/SA starts with the initial structure of the complex in vacuum. After solvating the structure in a solvent model, an MD simulation is performed to generate the snapshots for further analysis. Then, a relatively large number (typically  $N > 100$ ) of uncorrelated snapshots are extracted to represent the structural ensemble. Next, binding free energies of these structures are calculated in the implicit solvent after removing the explicit solvent molecules. The average binding free energy over these snapshots is reported as the final  $\Delta G_{bind}$ .

With the single-trajectory protocol, the binding free energy of a protein-protein complex is formally calculated as follows:

$$\Delta G_{bind} = \langle G^{complex}(i) - G^{protein1}(i) - G^{protein2}(i) \rangle_i, \quad (5)$$

where  $\langle \dots \rangle_i$  denotes an average over  $i$  snapshots extracted from the main MD trajectory. The implementation of this protocol is available in AmberTools18 in Perl<sup>33</sup> and Python.<sup>34</sup>

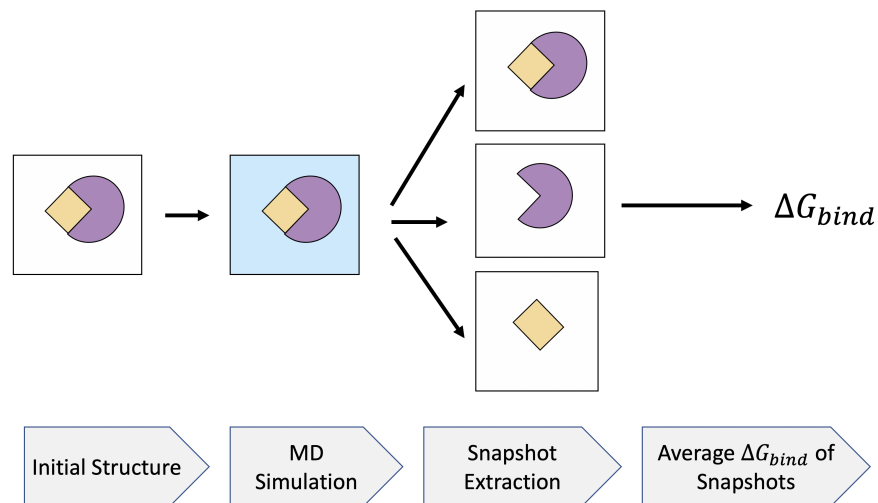


Figure 4: MMPB/SA flowchart. The initial structure of the complex is solvated using a water model. An MD simulation is run from which a relatively large number of snapshots are extracted. The average binding free energy of the snapshots is assigned as the binding free energy of the system.

In this work the former is used to maintain consistency with the reference study<sup>30</sup> opted for tuning the MMGB/SA model.

## Solvation Free Energy

**Polar Component.** A computationally efficient alternative to the PB, the GB implicit solvent model<sup>35,36</sup> can be used for computing  $\Delta G_{solv}$ . Generally speaking, GB models have shown to be computationally less expensive than the PB models, although the deterioration of the accuracy has always been a concern. Here, we employed a grid-based surface GB model called GBNSR6,<sup>37</sup> which, in a recent study,<sup>38</sup> was shown to be the most accurate among several GB models in terms of the ability to approximate  $\Delta G_{pol}$  relative to the numerical PB. In this work,  $\Delta G_{pol}$  is calculated with the ALPB modification<sup>39,40</sup> of the generalized Born<sup>41</sup> model:

$$\Delta G_{pol} = \sum_{ij} \Delta G_{ij}^{pol} \approx -\frac{1}{2} \left( \frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}} \right) \frac{1}{1 + \beta\alpha} \sum_{ij} q_i q_j \left( \frac{1}{f_{ij}^{GB}} + \frac{\alpha\beta}{A} \right), \quad (6)$$



where  $\epsilon_{\text{in}} = 1$  and  $\epsilon_{\text{out}} = 80$  are the dielectric constants of the solute and the solvent, respectively,  $\beta = \epsilon_{\text{in}}/\epsilon_{\text{out}}$ ,  $\alpha = 0.571412$ , and  $A$  is the electrostatic size of the molecule, which is essentially the overall size of the structure that can be computed analytically. Here we employ the most widely used functional form  $f_{ij}^{\text{GB}} = \left[ r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j) \right]^{\frac{1}{2}}$ , where  $r_{ij}$  is the distance between atomic charges  $q_i$  and  $q_j$ , and  $R_i$ ,  $R_j$  are the so-called *effective Born radii* of atoms  $i$  and  $j$ , which represent each atom's degree of burial within the solute. The effective Born radii,  $R$ , are calculated by the “ $R^6$ ” equation:<sup>42,43</sup>

$$R_i^{-3} = \left( -\frac{1}{4\pi} \oint_{\partial V} \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^6} \cdot d\mathbf{S} \right), \quad (7)$$

where  $\partial V$  represents the chosen representation of the dielectric boundary of the molecule,  $d\mathbf{S}$  is the infinitesimal surface element vector,  $\mathbf{r}_i$  is the position of atom  $i$ , and  $\mathbf{r}$  represents the position of the infinitesimal surface element. Uniform offset to the inverse effective radii is set to the default (optimal) value that is  $0.028 \text{ \AA}^{-1}$ .<sup>44</sup> The screening effect of monovalent salt is introduced into Eq. 6 as is standard for the GB model;<sup>35</sup> in our MMGB/SA calculations the salt concentration was set to 0.1 M.

**Non-polar Component.** A common method to estimate the non-polar contribution to the solvation free energy in Eq. 3 is to assume that it is proportional to the solvent accessible surface area (SASA) of the molecule:

$$G_{\text{nonpol}} = \gamma * \text{SASA}. \quad (8)$$

While there are more accurate methods to estimate the non-polar<sup>45</sup> contribution, here we use the simple Eq. 8 for the sake of simplicity and consistency with ref.<sup>30</sup> Also for consistency with the same work, here we use  $\gamma = 0.0072 \text{ kcal/mol/\AA}^2$ . Atomic radii that form SASA not only play an important role in the non-polar component, but also enter the polar component through the dielectric boundary. Therefore, the right choice of atomic radii is crucial to the accuracy of binding free energy estimation. Three sets of atomic radii are used here:

OPT1,<sup>29,46</sup> bondi, and mbondi2. The first two are listed in Tab. 1. Mbonid2 is indeed bondi whose hydrogen atoms bound to a nitrogen are expanded from 1.2 Å to 1.3 Å, see .<sup>47</sup> Carbon (C), hydrogen (H), oxygen (O), nitrogen (N) and sulfur (S) are the main atomic types in this study. The water probe radius is fixed to 1.4 Å.

Table 1: Two sets of atomic radii in Å used in this study.

	$\rho_C$	$\rho_H$	$\rho_N$	$\rho_O$	$\rho_S$
bondi	1.70	1.20	1.55	1.52	1.80
OPT1	1.40	1.55	2.35	1.28	1.80

## Gas-Phase Energy

Gas-phase energy of the solute,  $\Delta E_{MM}$ , is the summation of internal energies, electrostatic energies, and van der Waals energies. In all of the MMGB/SA calculations reported here,  $\Delta E_{MM}$  is calculated using the ff99 AMBER force field. The choice of this old force field is deliberate, and was initially motivated to ensure maximum consistency with ,<sup>6</sup> which provides a very detailed analysis of MMGB/SA performance on Ras-Raf. Good agreement with experiment, Table 2, motivated us to use the same ff99 force field for all the subsequent MMGB/SA calculations reported here. All of the enthalpy calculations in this study are averages over 500 snapshots extracted from the main MD trajectory.

## Configurational Entropy

Normal-mode analysis (NMA) and quasi-harmonic analysis are of the two common methods for calculating configurational entropy of the solute.<sup>33</sup> Since the latter has shown poor convergence in several cases, NMA is selected for entropy calculations.<sup>30</sup> The main drawback of this method is the computational cost that becomes intractable for large systems, e.g., systems with more than 8,000 atoms in MMGB/SA (Perl version) of AMBER18 are not supported for NMA. To tackle this problem, one approach is to truncate the complex so that the binding interface is preserved in its original shape.<sup>48</sup> In this study, the SARS-CoV-2 S RBD

and ACE2 complex is truncated for NMA feasible calculations. An offset of 1.92 *kcal/mol* has been subtracted from the  $-T\Delta S$  component of GBNSR6 calculations to address the concentration- dependency of the translational entropy at 1M, see <sup>30</sup> for details. NMA entropy calculations are done over 150 snapshots extracted from the main MD trajectory.

## Structure Preparation

**Ras-Raf Complex.** This well-studied complex was selected as the reference for testing the parameters of the MM/GBSA model. We used `t1eap` module in AMBER18 to set up the input coordinate and topology files. The structure was solvated in a box of TIP3P<sup>49</sup> water model (10 Å buffer). This choice of old water model and ff99 AMBER force field was deliberate, to ensure full consistency with <sup>6</sup>. The GTP molecule and the magnesium ion ( $Mg^{2+}$ ) were eliminated for the sake of simplification. Since the net charge was 0, no counterion was added.

**SARS-CoV-2 S RBD and ACE2 Complex: Full Structure.** H++ server<sup>50</sup> was employed to protonate the complex at pH=7.5. The server automatically generates the solvated structure in a box of OPC<sup>51</sup> explicit water model (10 Å buffer), with AMBER ff14SB force field. This full structure is used only for enthalpy calculations that are compared with those of the truncated complex structure for justifying the truncation approach, see below.

**SARS-CoV-2 S RBD and ACE2 Complex: Truncated Structure.** To execute NMA entropy calculation the original structure (PDB ID:6m0j) was truncated from 12,515 atoms (791 residues) to 7,286 atoms (463 residues) by removing residues, one by one, starting from the N-terminus of the spike protein, and the C-terminus of the ACE2 protein. The goal was to have fewer than 8,000 atoms remaining, while preserving sequence continuity of the resulting structure to facilitate the set up of MD simulations, see Fig. 5. The remaining atoms are still within 8 Å from the binding interface. The same protocol used for the full structure was employed for parameterization and solvation.

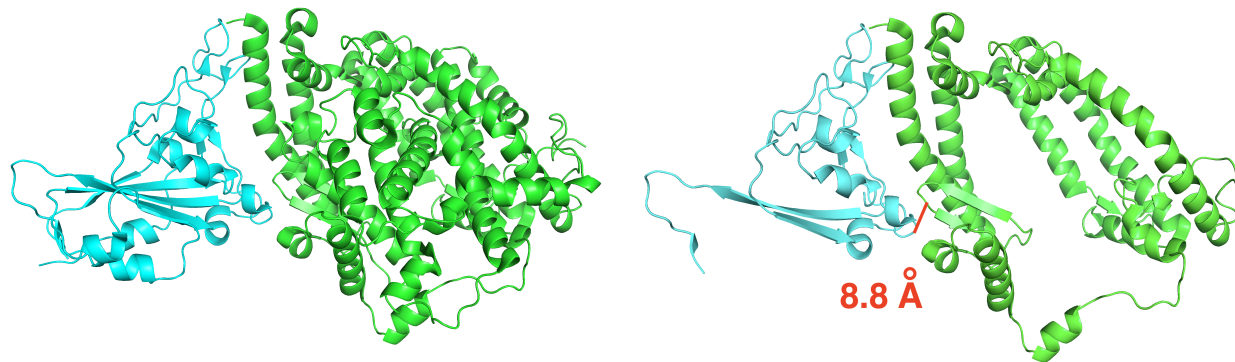


Figure 5: Truncation of SARS-CoV-2 S RBD used in the entropy estimate. The spike protein is in cyan, and the ACE2 receptor is in green. Left: original complex. Right: truncated complex. A pair of atoms on the binding interface that are 8.8 Å apart is shown in a solid red segment to illustrate the length scale.

## Trajectory Generation

All of the MMGB/SA estimates are based on snapshots extracted from MD trajectories generated as described below. The structures were prepared and equilibrated as follows. The solvated complexes were first energy minimized (max. minimization cycle of 1000), followed by 50 ps of heating (from 1 K to 300 K) at constant volume, followed by 50 ps of density equilibration at 300 K at constant 1 bar pressure, followed by another 2 ns of constant (1 bar) pressure equilibration at 300K. In these stages, atomic coordinates were restrained to their initial positions with 2  $kcal/mol/\text{\AA}^2$ . All simulations, including the production runs described below, were executed with the GPU-enabled pmemd.cuda MD engine in AMBER18, using Langevin dynamics with a collision frequency of  $2ps^{-1}$  and an integration time step of 2 fs while the bonds involving hydrogen atoms were constrained by the SHAKE algorithm. Electrostatic interactions were approximated via the Particle Mesh Ewald (PME) method, with a non-bond cutoff set to 9 Å. Coordinates were recorded every 10 ps.

**Ras-Raf Complex.** A production of 10 ns was performed using the Ras-Raf structure prepared with the protocol described in Sec. .

**SARS-CoV-2 S RBD and ACE2 Complex: Full Structure.** A production of 50 ns was carried out using the full structure described in Sec. .

**SARS-CoV-2 S RBD and ACE2 Complex: Truncated Structure.** The same protocol used for the full structure was employed. A weak restraint of  $0.01 \text{ kcal/mol/\AA}^2$  was applied to the atoms of the truncated complex, relative to the X-ray positions, during the 50 ns production to prevent the truncated complex from falling apart. This restraint diminishes the discrepancy between the force field and water model in the structure used for MD simulation (OPC, ff14SB) and the one for  $\Delta G_{bind}$  calculations (TIP3P, ff99).

## Results and Discussion

### MM/GBSA on Ras-Raf

Here, we study the accuracy of  $\Delta G_{bind}$  calculation using a different GB model (GBNSR6) coupled with two sets of atomic radii. According to Tab. 2, it is observed that  $\Delta G_{bind}$  calculated by GBNSR6 with OPT1 radii underestimates the binding affinity whereas GBNSR6 coupled with bondi radii overestimates it. Yet, both of these results have better agreement with the experiment<sup>52</sup> compared to the reference MGB model in .<sup>30</sup> We have also noticed that a *consensus* estimate  $\Delta G_{bind} = (\Delta G_{bind}(bondi) + \Delta G_{bind}(OPT1))/2$  is only within 2 kcal/mol off the experimental reference. Encouraged by this reasonable agreement with experiment, we have decided to utilize GBNSR6 with both bondi and OPT1 radii to produce a consensus estimate of  $\Delta G_{bind}$  for the SARS-CoV-2 S RBD and ACE2 complex.

### MM/GBSA on the Truncated SARS-CoV-2 S RBD and ACE2

The RMSD of the truncated SARS-CoV-2 S RBD and ACE2 backbone compared to the crystal structure of the full complex is shown in Fig. 6. The trajectory is stable after 50 ns of production, with the RMSD from the X-ray reference of around  $3.15 \text{ \AA}$ . Comparing the estimated  $\Delta H$  between the truncated and original structures (results not shown) demonstrates that the relatively small difference between the two (about 1 kcal/mol using mbondi2 radii and 8 kcal/mol using OPT1 radii) indirectly validates the use of the truncation procedure.

Table 2: MMGB/SA results on RAS-RAF. Means and the standard errors of the mean are listed. All the components are in *kcal/mol*. Consensus  $\Delta G_{bind} = (\Delta G_{bind}(bondi) + \Delta G_{bind}(OPT1))/2$  For comparison, an earlier MMGB/SA estimate is also listed (it does not contribute to the consensus estimate). The experimental value is from isothermal titration calorimetry.<sup>52</sup> An offset of 1.79 *kcal/mol* has been subtracted from the  $\Delta H$  component of MGB-based estimate for consistency with the author’s recommendation in .<sup>30</sup>

	GBNSR6 (bondi)	GBNSR6 (OPT1)	Consensus	MGB (bondi) ref <sup>30</sup>	Exp.
$E_{MM}$	$-937.42 \pm 2.07$	$-937.42 \pm 2.07$		$-1308.6 \pm 2.5$	
$\Delta G_{nonpol}$	$-9.63 \pm 0.02$	$-11.02 \pm 0.02$		$-9.5 \pm 0$	
$\Delta G_{pol}$	$887.39 \pm 2.06$	$897.59 \pm 2.15$		$1275.3 \pm 2.4$	
$\Delta H$	$-59.66 \pm 0.27$	$-50.85 \pm 0.43$		$-42.7 \pm 0.3$	
$-T\Delta S$	$43.39 \pm 0.98$	$43.39 \pm 0.98$		$41.4 \pm 1.6$	
$\Delta G_{bind}$	$-16.27 \pm 1.01$	$-7.46 \pm 1.07$	<b><math>-11.8 \pm 1</math></b>	$-3.1 \pm 1.6$	$-9.7 \pm 0.2$

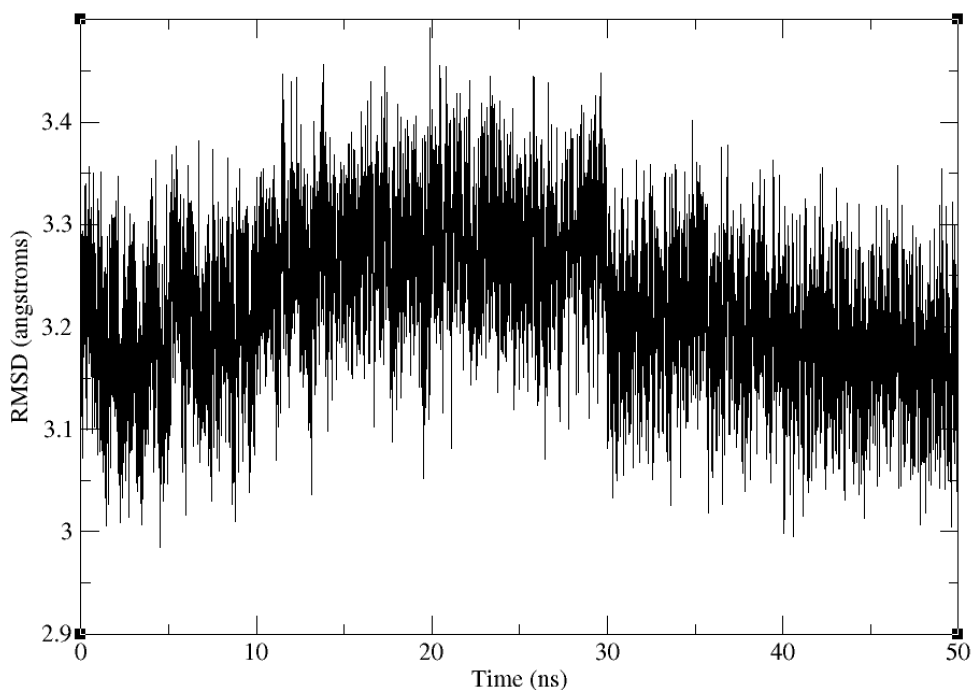


Figure 6: Backbone RMSD of the truncated SARS-CoV-2 S RBD and ACE2 complex, relative to the truncated part of the experimental crystal structure of the full complex, along the 50 ns production trajectory.

Shown in Tab. 3, our final estimates are presented, and compared to a previous computational study<sup>53</sup> and experiment. These findings are briefly discussed in Conclusion.

Table 3: MMGB/SA results on the truncated SARS-CoV-2 S RBD and ACE2 complex. Means and the standard errors of the mean are listed. All the components are in *kcal/mol*. Consensus  $\Delta G_{bind} = (\Delta G_{bind}(bondi) + \Delta G_{bind}(OPT1))/2$  For comparison, a recently published MMGB/SA estimate is also listed (it does not contribute to the consensus estimate). Experimental value derived from a fit to surface plasmon resonance sensogram.<sup>54</sup>

	GBNSR6 (mbondi2)	GBNSR6 (OPT1)	Consensus	$GB^{OBC}$ (mbondi2) ref <sup>53</sup>	Exp.
$E_{MM}$	$-453.76 \pm 0.87$	$-453.76 \pm 0.87$		$-761.06 \pm 3.96$	
$\Delta G_{nonpol}$	$-14.71 \pm 0.02$	$-16.35 \pm 0.02$		$-12.21 \pm 0.06$	
$\Delta G_{pol}$	$401.55 \pm 0.81$	$413.72 \pm 0.87$		$737.98 \pm 3.86$	
$\Delta H$	$-66.93 \pm 0.29$	$-56.39 \pm 0.37$		$-35.30 \pm 0.60$	
$-T\Delta S$	$52.28 \pm 1.49$	$52.28 \pm 1.49$		$13.56 \pm 0.70$	
$\Delta G_{bind}$	$-14.65 \pm 1.52$	$-4.11 \pm 1.54$	<b><math>-9.4 \pm 1.5</math></b>	$-21.74 \pm 0.65$	<b>-10.6</b>

## Entropy Convergence of the Truncated SARS-CoV-2 S RBD and ACE2 Complex

Entropy calculation is one of the most challenging and time-consuming parts of  $\Delta G_{bind}$  estimation. In order to maintain the consistent protocol, 150 snapshots were selected for  $-T\Delta S$  calculations. In addition, we conducted an investigation to examine whether a fewer number of snapshots would suffice to lead to a similar  $-T\Delta S$ . A subset of 15 and 50 equidistant snapshots were collected from the set of 150 snapshots. According to Tab. 4 it is observed that  $-T\Delta S$  calculated on 15 and 50 sample snapshots leads to a similar  $-T\Delta S$  calculated on the whole set. Naturally, the standard error of the mean decreases as the sample size increases. However, this increase doesn't affect the stability of the mean around  $52 \text{ kcal/mol}$ . Given entropy calculation as the bottleneck of  $\Delta G_{bind}$  estimation, this observation suggests that with a relatively small set of snapshots it is possible to compute  $-T\Delta S$  quite accurately.

Table 4: Entropy convergence of the truncated SARS-CoV-2 S RBD and ACE2 complex. Means and standard error of the means are listed. Increasing the number of equidistant snapshots  $n$  from 15 to 150 shows the stability of the entropy around  $52 \text{ kcal/mol}$ .

	$n = 15$	$n = 50$	$n = 150$
$-T\Delta S$	$52.69 \pm 5.05$	$52.30 \pm 2.83$	$52.28 \pm 1.49$

## Conclusion

In this study, we have evaluated the performance of a relatively new GB model, GBNSR6, and the newly introduced intrinsic atomic radii, in calculations of binding free energy of a well-studied protein-protein complex, Ras-Raf and of SARS-CoV-2 S RBD and ACE2 complex, critical in the mechanism of the novel coronavirus infection. Unlike many previous efforts, the new radii were specifically optimized to best reproduce the explicit solvent results particularly in the implicit solvent-based binding MMGB/SA estimates. We also employed the common bondi radii for the same calculations.

A better agreement with experiment of the absolute binding free energy for Ras-Raf, compared to previous work, was achieved for both radii sets; however each individual estimate either under- or over-estimated the experimental binding energy. We have therefore proposed a consensus estimate, which is the average of the two: for Ras-Raf the consensus based on OPT1 and bondi is in near quantitative agreement with experiment.

We applied the same approach to estimate  $\Delta G_{bind}$  of the SARS-CoV-2 S RBD and ACE2 complex. As in the Ras-Raf case, the under- and over- estimation by each radii sets nearly cancelled, resulting in a consensus estimate essentially within 1 kcal/mol from the experimental reference. That essentially quantitative agreement paves the way for further exploration of SARS-CoV-2 S RBD and ACE2 complex with MMGB/SA, which has a number of advantages. Specifically, the MMGB/SA approach could be reasonably accurate for future analysis of *relative* binding free energies in this system, including the effects of mutations, relative contributions from various residues to  $\Delta G_{bind}$ , congeneric series of ligands, etc. Equally importantly, MMGB/SA is well-suited for reasoning for physical reasoning. Needless to say, additional thorough investigation is needed to see if the proposed approach can be extended to other complexes.

Both of the two separate estimates of  $\Delta G_{bind}$  based on either bondi or OPT1 radii result in the expected near cancellation of the relatively large  $\Delta H$  and  $-T\Delta S$  terms, suggesting that each of these estimates “makes sense” on its own. Moreover, note that for both of



the protein-protein complexes studied here, the two radii sets provide a fairly narrow range within which the experimental value lies. Specifically,  $\Delta G_{bind}$  calculated with bondi radii is underestimated (too negative), whereas  $\Delta G_{bind}$  calculated by OPT1 radii is over-estimated. One rationale for this behavior, which is the basis of the proposed consensus approach, could be that bondi and OPT1 radii sets have very different physical foundations behind them (geometry for the former and global optimization of the electrostatics for the latter), so the resulting errors in the corresponding electrostatic estimates are not as strongly correlated as for radii derived on the same principle.

All of the MD trajectories generated in this work are available from the authors upon request.

## Acknowledgment

The authors thank Dr. Holger Gohlke and Dr. Nadine Homeyer for their assistance with developing GBNSR6 in AMBER and their consultation. This work was supported by the NIH R21 GM131228.

## References

- (1) Amaro, R. E.; Mulholland, A. J. A Community Letter Regarding Sharing Biomolecular Simulation Data for COVID-19. *Journal of Chemical Information and Modeling* **2020**,
- (2) Devaux, C. A.; Rolain, J.-M.; Colson, P.; Raoult, D. New insights on the antiviral effects of chloroquine against coronavirus: what to expect for COVID-19? *International journal of antimicrobial agents* **2020**, 105938.
- (3) Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* **2020**,

- (4) Harrer, S.; Shah, P.; Antony, B.; Hu, J. Artificial intelligence for clinical trial design. *Trends in pharmacological sciences* **2019**,
- (5) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813–1818.
- (6) Gohlke, H.; Kiel, C.; Case, D. A. Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes. *Journal of molecular biology* **2003**, *330*, 891–913.
- (7) Zhou, H.-X.; Gilson, M. K. Theory of Free Energy and Entropy in Noncovalent Binding. *Chemical reviews* **2009**, *109*, 4092–4107.
- (8) Mobley, D. L.; Dill, K. A. Binding of Small-Molecule Ligands to Proteins: ” What You See” Is Not Always ” What You Get”. *Structure* **2009**, *17*, 489–498.
- (9) Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L., et al. Overview of the SAMPL6 Host–Guest Binding Affinity Prediction Challenge. *Journal of computer-aided molecular design* **2018**, *32*, 937–963.
- (10) Zou, J.; Tian, C.; Simmerling, C. Blinded prediction of protein–ligand binding affinity using Amber thermodynamic integration for the 2018 D3R grand challenge 4. *Journal of computer-aided molecular design* **2019**, *33*, 1021–1029.
- (11) Wang, J.; Alekseenko, A.; Kozakov, D.; Miao, Y. Improved Modeling of Peptide–Protein Binding through Global Docking and Accelerated Molecular Dynamics Simulations. *Frontiers in molecular biosciences* **2019**, *6*, 112.
- (12) Siebenmorgen, T.; Zacharias, M. Computational prediction of protein–protein binding affinities. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2020**, *10*, e1448.

- (13) Straatsma, T.; Berendsen, H.; Postma, J. Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water. *The Journal of chemical physics* **1986**, *85*, 6720–6727.
- (14) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L., et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, 1–6.
- (15) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Current opinion in structural biology* **2011**, *21*, 150–160.
- (16) Abel, R.; Wang, L.; Mobley, D. L.; Friesner, R. A. A critical review of validation, blind testing, and real-world use of alchemical protein-ligand binding free energy calculations. *Current topics in medicinal chemistry* **2017**, *17*, 2577–2585.
- (17) Michel, J.; Foloppe, N.; Essex, J. W. Rigorous free energy calculations in structure-based drug design. *Molecular informatics* **2010**, *29*, 570–578.
- (18) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z.; Hou, T. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chemical reviews* **2019**, *119*, 9478–9508.
- (19) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery* **2015**, *10*, 449–461.
- (20) Wang, C.; Greene, D.; Xiao, L.; Qi, R.; Luo, R. Recent developments and applications of the MMPBSA method. *Frontiers in molecular biosciences* **2018**, *4*, 87.
- (21) Homeyer, N.; Stoll, F.; Hillisch, A.; Gohlke, H. Binding free energy calculations for lead optimization: assessment of their accuracy in an industrial drug design context. *Journal of chemical theory and computation* **2014**, *10*, 3331–3344.

- (22) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **2010**, *31*, 455–461.
- (23) Allen, W. J.; Balias, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of new features and current docking performance. *Journal of computational chemistry* **2015**, *36*, 1132–1156.
- (24) Sasmal, S.; El Khoury, L.; Mobley, D. L. D3R Grand Challenge 4: ligand similarity and MM-GBSA-based pose prediction and affinity ranking for BACE-1 inhibitors. *Journal of Computer-Aided Molecular Design* **2020**, *34*, 163–177.
- (25) Onufriev, A. V.; Izadi, S. Water models for biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1347.
- (26) Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118*, 1691–1741.
- (27) Onufriev, A. Continuum electrostatics solvent modeling with the generalized Born model. *Modeling Solvent Environments* **2010**, *1*.
- (28) IYB-S, D. C.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; DG, R. D.; Gilson, M.; Gohlke, H.; Goetz, A., et al. AMBER 2018. *University of California, San Francisco* **2018**,
- (29) Forouzesh, N.; Mukhopadhyay, A.; Watson, L. T.; Onufriev, A. V. Multidimensional Global Optimization and Robustness Analysis in the Context of Protein-Ligand Binding. *Journal of Chemical Theory and Computation* **2020**,
- (30) Gohlke, H.; Case, D. A. Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf. *Journal of computational chemistry* **2004**, *25*, 238–250.

- (31) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W., et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research* **2000**, *33*, 889–897.
- (32) Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. The SAMPL4 host–guest blind prediction challenge: an overview. *Journal of computer-aided molecular design* **2014**, *28*, 305–317.
- (33) Homeyer, N.; Gohlke, H. Free energy calculations by the molecular mechanics Poisson–Boltzmann surface area method. *Molecular Informatics* **2012**, *31*, 114–122.
- (34) 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.
- (35) Onufriev, A. V.; Case, D. A. Generalized Born implicit solvent models for biomolecules. *Annual review of biophysics* **2019**, *48*, 275–296.
- (36) 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.
- (37) Forouzesh, N.; Izadi, S.; Onufriev, A. V. Grid-based Surface Generalized Born Model for Calculation of Electrostatic Binding Free Energies. *Journal of Chemical Information and Modeling* **2017**, *57*, 2505–2513.
- (38) Izadi, S.; Harris, R. C.; Fenley, M. O.; Onufriev, A. V. Accuracy Comparison of Generalized Born Models in the Calculation of Electrostatic Binding Free Energies. *Journal of chemical theory and computation* **2018**, *14*, 1656–1670.

- (39) Sigalov, G.; Scheffel, P.; Onufriev, A. Incorporating Variable Dielectric Environments Into the Generalized Born Model. *The Journal of chemical physics* **2005**, *122*, 094511.
- (40) Sigalov, G.; Fenley, A.; Onufriev, A. Analytical Electrostatics for Biomolecules: Beyond the Generalized Born Approximation. *The Journal of chemical physics* **2006**, *124*, 124902.
- (41) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *Journal of the American Chemical Society* **1990**, *112*, 6127–6129.
- (42) Aguilar, B.; Shadrach, R.; Onufriev, A. V. Reducing the Secondary Structure Bias in the Generalized Born Model via R6 Effective radii. *Journal of Chemical Theory and Computation* **2010**, *6*, 3613–3630.
- (43) Aguilar, B.; Onufriev, A. V. Efficient Computation of the Total Solvation Energy of Small Molecules via the R6 Generalized Born Model. *Journal of chemical theory and computation* **2012**, *8*, 2404–2411.
- (44) Mongan, J.; Svrcek-Seiler, W. A.; Onufriev, A. Analysis of Integral Expressions for Effective Born Radii. *The Journal of chemical physics* **2007**, *127*, 11B608.
- (45) Gallicchio, E.; Levy, R. M. AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *Journal of computational chemistry* **2004**, *25*, 479–499.
- (46) Forouzesh, N.; Watson, L. T.; Onufriev, A. V. Robustness of multidimensional optimization outcomes: a general approach and a case study. Proceedings of the 2020 Spring Simulation Conference. 2020; pp 1–12.
- (47) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale

- conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **2004**, *55*, 383–394.
- (48) Genheden, S.; Kuhn, O.; Mikulskis, P.; Hoffmann, D.; Ryde, U. The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant. *Journal of chemical information and modeling* **2012**, *52*, 2079–2088.
- (49) 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.
- (50) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating p K prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic acids research* **2012**, *40*, W537–W541.
- (51) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building water models: a different approach. *The journal of physical chemistry letters* **2014**, *5*, 3863–3871.
- (52) Rudolph, M. G.; Linnemann, T.; Grünewald, P.; Wittinghofer, A.; Vetter, I. R.; Herrmann, C. Thermodynamics of Ras/effecter and Cdc42/effecter interactions probed by isothermal titration calorimetry. *Journal of Biological Chemistry* **2001**, *276*, 23914–23921.
- (53) Peng, C.; Zhu, Z.; Shi, Y.; Wang, X.; Mu, K.; Yang, Y.; Zhang, X.; Xu, Z.; Zhu, W. Exploring the binding mechanism and accessible angle of SARS-CoV-2 spike and ACE2 by molecular dynamics simulation and free energy calculation. *Chemrxiv. Org* **2020**,
- (54) 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.