

The pH Effects on SARS-CoV and SARS-CoV-2 Spike Proteins in the Process of Binding to hACE2

Yixin Xie

The University of Texas at El Paso

Wenhan Guo

The University of Texas at El Paso

Alan Lopez-Hernandez

The University of Texas at El Paso

Shaolei Teng

Howard University

Lin Li (✉ lli5@utep.edu)

The University of Texas at El Paso

Research Article

Keywords: COVID-19, electrostatic features, SARS-CoV/SARS-CoV-2, hACE2

DOI: <https://doi.org/10.21203/rs.3.rs-871118/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

1 **The pH Effects on SARS-CoV and SARS-CoV-2 Spike Proteins in the Process**
2 **of Binding to hACE2**

3
4 **Yixin Xie¹, Wenhan Guo¹, Alan Lopez-Hernandez¹, Shaolei Teng², Lin Li^{1,3*}**

5
6 ¹ Computational Science Program, University of Texas at El Paso, El Paso, TX.

7 ² Department of Biology, Howard University, Washington, D.C.

8 ³ Department of Physics, University of Texas at El Paso, El Paso, TX.

9
10 ***Correspondence:**

11 Lin Li: lli5@utep.edu

12
13
14 **Keywords:** SARS-CoV, SARS-CoV-2, COVID-19, Electrostatic Features, Angiotensin-
15 Converting Enzyme 2, hACE2, spike protein, pH dependence, binding energy, folding energy

16
17 **Abstract**

18 COVID-19 has been threatening human health since the late 2019, which has significant impact
19 on human health and economy. Understanding the SARS-CoV-2 and other coronaviruses is
20 important to develop effective treatments for COVID-19 and other coronaviruses-caused
21 diseases. In this work, we applied multi-scale computational approaches to study the electrostatic
22 features of spike (S) proteins for SARS-CoV and SARS-CoV-2. From our results, we found that
23 SARS-CoV and SARS-CoV-2 have similar charge distributions and electrostatic features when
24 binding with the human angiotensin-converting enzyme 2 (hACE2). The energy pH-dependence
25 calculations revealed that the complex structures of hACE2 and the S proteins of SARS-
26 CoV/SARS-CoV-2 are stable at pH values ranging from 7.5 to 9. Molecular dynamics
27 simulations were performed using NAMD to investigate the hydrogen bonds between S proteins
28 and hACE2. From the MD simulations it was found that SARS-CoV-2 has four pairs of essential
29 hydrogen bonds (high occupancy, >80%), while SARS-CoV has three pairs, which indicates the
30 SARS-CoV-2 S protein has relatively more robust binding strategy than SARS-CoV S protein.
31 Four key residues forming essential hydrogen bonds from SARS-CoV-2 are identified, which are
32 potential drug targets for COVID-19 treatments. The findings in this study shed lights on the
33 current and future treatments for COVID-19 and other coronaviruses-caused diseases.

34

35 1 Introduction

36 The ongoing COVID-19 pandemic is changing human society significantly and causing both
37 economic and social consequences all over the world [1]. Coronaviruses are named for their
38 crown-like spikes on their surface, and they are commonly found in many mammal species [2].
39 Human coronaviruses were firstly identified in the mid-1960s. There are four main sub-
40 groupings of coronaviruses, known as alpha, beta, gamma, and delta [3]. Among all the
41 coronaviruses, there are seven known types of coronaviruses that can infect human beings.
42 People around the world commonly get infected by human coronaviruses 229E, NL63, OC43,
43 and HKU1 [4, 5]. And some coronaviruses that infect animals are able to evolve and infect
44 humans, among which the three recent cases are SARS-CoV-2, SARS-CoV, and MERS-CoV[6].
45 The SARS-CoV-2 virus is the novel coronavirus that causes coronavirus disease 2019, or
46 COVID-19. Other than COVID-19, coronaviruses have caused several pandemics before,
47 including severe acute respiratory syndrome (SARS) which was caused by SARS-CoV and the
48 Middle East respiratory syndrome (MERS) which was caused by MERS-CoV. To end the
49 current pandemic soon and be prepared for the future similar challenges for human society, it is
50 essential to understand the binding mechanisms of SARS-CoV-2 infecting human cells. This is
51 achievable by studying the stabilities of SARS-CoV-2 at different pH conditions, and identify the
52 key residues that play significant roles in the binding processes.

53
54 Coronaviruses contain membrane glycoprotein (M), nucleocapsid protein (N), spike protein (S),
55 envelope protein (E) and an RNA single chain[7]. For all enveloped viruses, one of the most
56 important steps during the binding process is membrane fusion, which allows viruses to get into
57 host cells [8]. For coronaviruses, the fusion protein is the S protein that leads the binding process
58 to attack human cells through the host cell receptor angiotensin-converting enzyme 2 (hACE2)
59 [9]. Human hACE2 (hACE2) is an enzyme located widely in the human body, including the
60 lungs, kidneys, adipose tissue, central nervous system and cardiovascular system [9-11] and it
61 has multiple essential functions such as the regulation of amino acid transport in the kidney
62 controlling the blood pressure, and viral receptors including both SARS-CoV-2 and SARS-CoV
63 [11]. Since it is of extreme importance to human health, there are numerous research groups have
64 been or are currently working on S proteins and hACE2 using various approaches.

65
66 The traditional process of the de novo drug design is a challenging task which consumes
67 resources and time significantly. With the fast developments of computing technology,
68 computational methods have been widely used to drug-related research[12], including protein-
69 protein interactions[13, 14], MD simulations[15], coarse-grained models[16], pH dependence of
70 protein-protein interactions[17-20], etc. Our previous studies have applied multi-scale
71 computational methods to study several pathogens [21-25] including the SARS-CoV-2 viruses
72 [26, 27], which revealed some mechanisms of the SARS-CoV-2 S protein. Besides, many other
73 research groups have made successful progress to understand the SARS-CoV-2 using
74 computational methods [28, 29].

75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94

In this work, we first calculated the electrostatic potentials on the surface of S proteins from both SARS-CoV and SARS-CoV-2, followed by the electric field line comparison between SARS-CoV and SARS-CoV-2 when they bind to hACE2. We found that the two viruses have similar pH responses: The pH-dependence of folding energies for S protein receptor binding domains (RBDs) demonstrated that both the S protein RBDs of these two viruses are at the most stable status when pH values ranging from 6 to 9. Also, the pH-dependence of binding energies for S protein RBDs and hACE2 RBD showed that the complex structures of the two viruses are at the most stable status at pH values ranging from 7.5 to 10.5. Therefore, SARS-CoV and SARS-CoV-2 survive in a similar pH environment. The pH 7.5 to 9 is the best condition for both SARS-CoV and SARS-CoV-2 to best perform their functions to bind with hACE2. Also, we analyzed the trajectories from 100ns MD simulations using NAMD [30] and identified hydrogen bonds with the involved key residues using VMD [31]. It is shown that for the high-frequency (>80%) hydrogen bonds, SARS-CoV-2 has four pairs while SARS-CoV has three pairs, which indicates that the S protein of SARS-CoV-2 uses more residues to form strong hydrogen bonds. The key residues forming essential hydrogen bonds from SARS-CoV-2 are ARG-121, TYR103, THR182 and TYR171, which are potential drug targets for COVID-19 treatments. Using multiple computational approaches, the findings in this work pave the way for the current and future treatment development of COVID-19 and other coronaviruses-caused diseases.

95 **2 Methods**

96 **2.1 Structure Preparation**

97 The complex structures of SARS-CoV/hACE2 and SARS-CoV-2/hACE2 were downloaded
98 from the Protein Data Bank (PDB ID 6ACG [32] and 7AD1 [33], respectively) . Please note that
99 in 7AD1, the mutations that the authors made during their experiments are not on the interface
100 area. Since we only focus on the interface area between S proteins and hACE2, the mutations do
101 not affect our results. In this work, we used the complex structures to study the electrostatic
102 binding interactions and the relative binding energies in different pH environments between S
103 proteins and hACE2 RBDs. For the missing loops in proteins, we used MODELLER [34] to
104 model the structures based on the sequences. To understand the mechanisms of S protein binding
105 to hACE2 at the interface particularly, S protein RBDs were separated from the hACE2 binding
106 domain by a distance of 10Å for the best results and visualization.

107

108 **2.2 Electrostatic Potential Calculation**

109 In order to study the electrostatic features, DelPhi [35, 36] was utilized to calculate the
110 electrostatic potential for the S proteins and hACE2 RBDs. In the framework of continuum
111 electrostatics, DelPhi calculates the electrostatic potential ϕ (in systems comprised of biological

112 macromolecules and water in the presence of mobile ions) by solving the Poisson-Boltzmann
113 equation (PBE):

$$114 \quad \nabla \cdot [\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) + \epsilon(\mathbf{r})\kappa^2(\mathbf{r}) \sinh(\phi(\mathbf{r})/k_B T) \quad (1)$$

115 where $\phi(\mathbf{r})$ is the electrostatic potential, $\epsilon(\mathbf{r})$ is the dielectric distribution, $\rho(\mathbf{r})$ is the charge
116 density based on the atomic structures, κ is the Debye-Huckel parameter, k_B is the Boltzmann
117 constant, and T is the temperature. Due to the irregular shape of macromolecules, DelPhi uses a
118 finite difference (FD) method to solve the PBE.

119
120 Before the DelPhi calculations, the PQR file of each trimer was generated by PDB2PQR [37].
121 We used AMBER [38] force field for PDB2PQR calculation, and removed water molecules. For
122 the better results, we ensured the new atoms are not rebuilt too close to existing atoms and
123 optimized the hydrogen bonding network.

124
125 During DelPhi calculations, the resolution was set as 0.5 grids/Å. The dielectric constants were
126 set as 2.0 for protein and 80.0 for the water environment, respectively. The pH value for the
127 solvent environment was set to be 7.0. The probe radius for generating the molecular surface was
128 1.4 Å. Salt concentration was set as 0.15 M. The boundary condition for the Poisson Boltzmann
129 equation was set as a dipolar boundary condition. The calculated electrostatic potential on the
130 surface was visualized with Chimera (figure 2). VMD was used to illustrate electric field lines
131 between S protein and hACE2 (figure 3). Finally, the color scale range was set to be from -1.0 to
132 1.0 kT/e for the best visual presentation. Besides the calculations of electrostatic potentials, we
133 also used DelphiForce [39] to calculate the electrostatic binding forces between each S protein
134 and hACE2 while separating them in the direction of the mass center connection line (figure S2).
135 Besides the net forces between each S protein and hACE2, the X, Y, Z components of the net
136 forces are also calculated and shown in figure S2.

137

138 **2.3 Relative Folding Energy Calculation**

139 We used DelPhiPKa [40, 41] to calculate pKa values of DNA and UDG, given the pH ranging
140 from 0 to 14 with the pH interval of 0.5. During the calculations, we used AMBER force field,
141 and removed water molecules and HETATM. For the hydrogen of ASP and GLU attached atom,
142 we used OD1 and OE1, respectively. Variance of Gaussian Distribution was set to be 0.7, salt
143 concentration was 0.15, reference dielectric was 8.0, and external dielectric was 80.0.

144

145 The net charges of proteins at the unfolded state were calculated using this equation:

$$146 \quad Q_u(pH) = \sum_{i=1}^N \frac{10^{-2.3y(i)(pH-pKa(i))}}{1 + 10^{-2.3y(i)(pH-pKa(i))}} \quad (2)$$

147 where the summation is of all the titratable groups, $y(i)$ value is -1 for acidic groups and +1 for
148 basic groups, respectively. As for the folding free energy, we used this equation:

$$149 \quad \Delta N(\text{pH}_{\text{folding}}) = 2.3RT \int_{\text{pH}_i}^{\text{pH}_f} (Q_f(\text{pH}) - Q_u(\text{pH}))d(\text{pH}) \quad (3)$$

150 where $Q_f(\text{pH})$ and $Q_u(\text{pH})$ stand for the net charge of folded and unfolded state, respectively.
151 R is the universal gas constant taken as $1.9872 \times 10^{-3} \frac{\text{kcal}}{\text{Mol} \cdot \text{K}}$. T is the temperature with the
152 value of 300 K.

153 Please note that the algorithms we applied to calculate the folding energies are for the relative
154 values, that is, at $\text{pH}=0$ the folding energy is 0 and at any other pH values the folding energies
155 are the relative values to the $\text{pH}=0$ condition.

156

157 **2.4 Relative Binding Energy Calculation**

158 For the binding energy calculation, we involved two methods, which are DelPhiPKa and
159 MM/PBSA[42]. To calculate binding energy using DelPhiPKa, the following equation was used:

$$160 \quad \Delta N(\text{pH}_{\text{binding}}) = 2.3RT \int_{\text{pH}_i}^{\text{pH}_f} (Q_t(\text{pH}) - Q_n(\text{pH}) - Q_r(\text{pH}))d(\text{pH}) \quad (4)$$

161 where $\Delta N(\text{pH}_{\text{binding}})$ is the the binding free energy at different pH values, $Q_t(\text{pH})$, $Q_n(\text{pH})$,
162 and $Q_r(\text{pH})$ are the net charges of complexes of each model. R is the universal gas constant
163 taken as $1.9872 \times 10^{-3} \frac{\text{kcal}}{\text{Mol} \cdot \text{K}}$. T is the temperature with the value of 300 K.

164

165 Please note that the algorithms we applied to calculate the binding energies are for the relative
166 values, that is, at $\text{pH}=0$ the binding energy is 0 and at any other pH values the binding energies
167 are the relative values to the $\text{pH}=0$ condition.

168

169 **2.5 Molecular Dynamic (MD) Simulations**

170 To simulate the dynamic interactions between S proteins RBD and hACE2 protein, MD
171 simulations [15] were carried out using NAMD [30] with the help of GPUs on Lonestar5 clusters
172 at the Texas Advanced Computing Center (TACC <https://www.tacc.utexas.edu/>). A 2000-step
173 minimization was performed for each simulation, followed by a 100 million steps, during which
174 20,000 frames were saved from two 100ns simulations of both SARS-CoV and SARS-CoV-2
175 separately (1.0 fs per step, 1 frame at each 5000 steps, 100 million steps in total). The RMSDs of
176 the SARS-CoV and SARS-CoV-2 trajectories are about 3.4Å and 1.1 Å, respectively (figure S1).
177 During the MD simulations, we used CHARMM [43] force field, the temperature was set to be

178 300 K, and the pressure was set to be standard using the Langevin dynamics. For PME, which is
179 set for full-system periodic electrostatics, with the grid size (86, 88, 132) as (x, y, z) value
180 respectively. In those two simulations, atoms that are not located in binding domains were
181 constrained within a margin of 10.0 Å of their natural movement maximum length values. In
182 order to get a more accurate result of the simulation, data of the last 50 ns of simulations were
183 selected and used for data analysis, since the structure of the first 50 ns is not as stable as the last
184 50 ns of simulations. The simulation processes are visualized in movies 4 and 5, generated by
185 VMD.

186
187 To analyze the interaction between S proteins and hACE2, the hydrogen bonds that formed
188 within the distance of 4 Å were extracted from the last 10,000 frames (50 ns) of simulations. The
189 several top-strongest hydrogen bonds in each binding domain were determined by calculating
190 their formation frequency (the frequency in figure of essential hydrogen bonds is shown in figure
191 7.

192

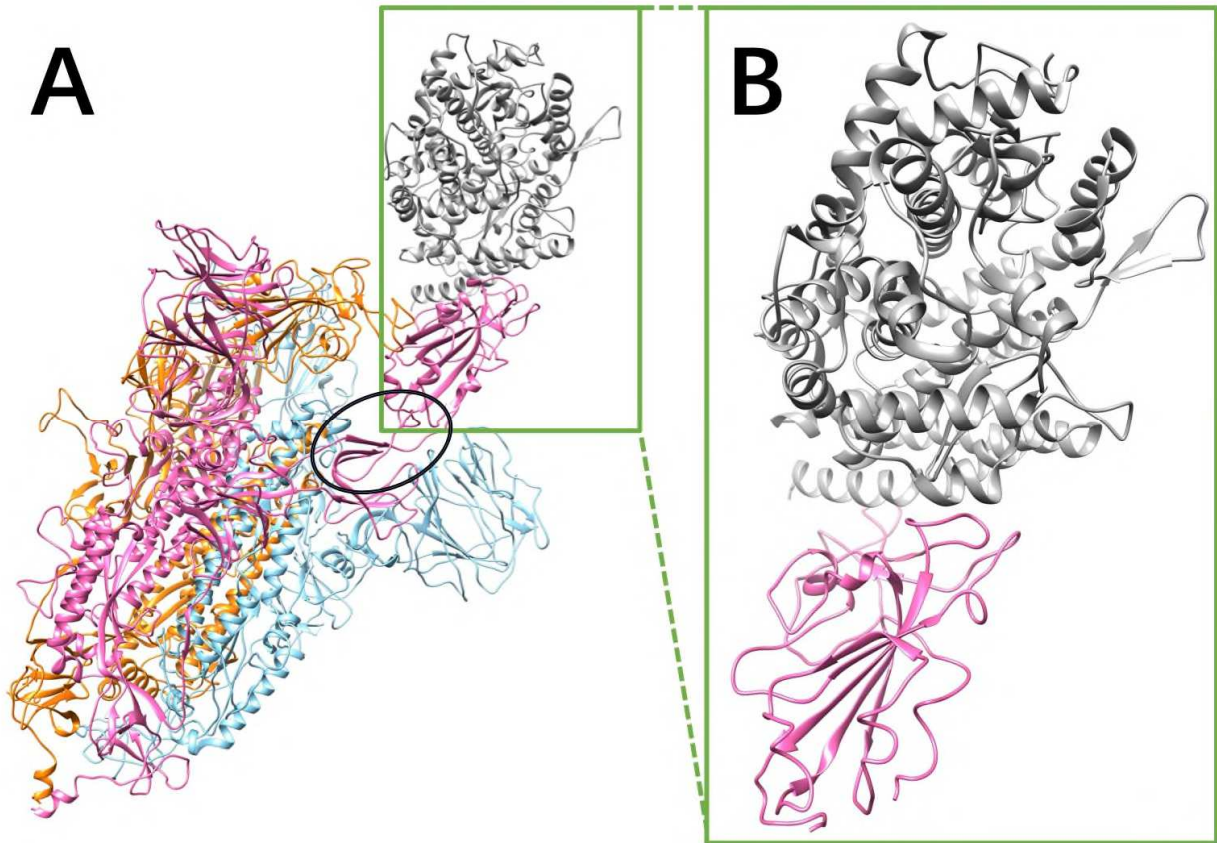
193 **3 Results and Discussions**

194 First of all, the electrostatic features of SARS-CoV and SARS-CoV-2 S proteins were
195 investigated, including electrostatic potential and electric field lines. Secondly, the relative
196 binding energies of complex structures and folding energies of S proteins at different pH values
197 were analyzed. Finally, the hydrogen bonds and related key residues in each complex structure
198 were obtained using MD simulations.

199

200 **3.1 S Protein Trimer Structure**

201



202
203

204 **Figure 1.** SARS-CoV S protein structure. Only the SARS-CoV S protein structure is illustrated
 205 in this figure, because SARS-CoV and SARS-CoV-2 S proteins are very similar (the RMSD
 206 between two S protein RBDs is 0.973 Å). (A) The S protein is a homotrimer (orange, blue, pink),
 207 of which one chain (pink) flips out when it binds to hACE2 (gray). The hinge connecting the
 208 RBD and the other part of S protein is shown in a black circle; (B) The closeup view of binding
 209 domains when S protein RBD (pink) binds to hACE2 RBD (gray).

210

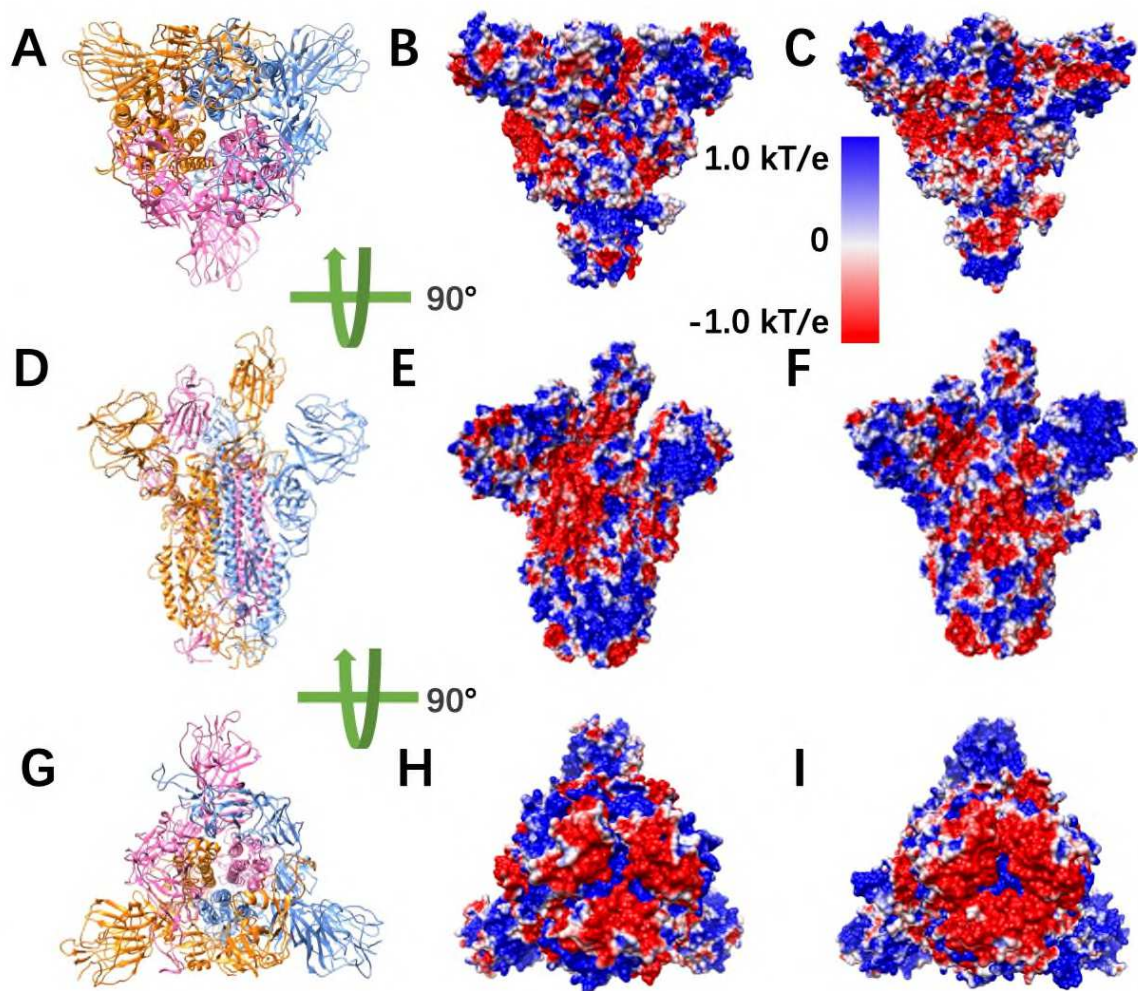
211 The RMSD between the S proteins of SARS-CoV and SARS-CoV-2 is 0.973 Å, showing that
 212 the S proteins of SARS-CoV and SARS-CoV-2 are very similar. The S proteins of SARS-CoV
 213 and SARS-CoV-2 are both homotrimers. Each monomer contains an RBD which connects the
 214 other part of the monomer via a hinge composed by two flexible loops (as shown in the black
 215 circle of figure 1A). The RBD is in closed configuration when there is no hACE2 binds to the S
 216 protein. When binding to hACE2, the RBD of one monomer flips out as open configuration and
 217 it binds to the RBD of hACE2.

218

219 3.2 Electrostatic Potential on Surfaces

220 To study the electrostatic features, DelPhi was utilized to calculate the electrostatic potential on
221 surfaces of the S protein trimer (full structure) and hACE2 RBD. The electrostatic potential
222 distribution on SARS-CoV S protein trimer structure is showed in figure 2BEH and movie 1,
223 which were rendered by Chimera with a color scale from -1.0 to 1.0 kT/e. The charge
224 distribution on SARS-CoV-2 S protein trimer structure is shown in figure 2CFI and movie 2,
225 which were rendered by Chimera with a color scale from -1.0 to 1.0 kT/e as well, for the
226 comparison. Negatively and positively charged areas are colored in red and blue, respectively.
227

228 By comparing the electrostatic potential on surfaces of two trimer structures, it is obvious that
229 the charge distribution of SARS-CoV and SARS-CoV-2 S proteins are different. From the top
230 view (figure 2A-C) and the bottom view (figure 2G-I), we noticed that SARS-CoV has slightly
231 more positively charged area (blue), compared to SARS-CoV-2. It indicates that the SARS-CoV
232 may attract the hACE2 more easily, since the hACE2 binding interface is overall negatively
233 charged (movie 3). Such finding supports the previous studies of our research group [26, 27].
234 The electrostatic distribution differences observed from front views (figure 2D-F) of the S
235 proteins demonstrate that the electrostatic features may have impacts on the stabilities of the
236 trimers. Here it was not investigated several details about the binding stabilities among
237 monomers in an S protein, due to the scope of this work that mainly focusses on the binding
238 between S protein and hACE2. The electrostatic distributions on S protein RBDs show that the
239 SARS-CoV RBD is more positive, which is consistent with the top view (figure 2BC). The
240 bottom of the SARS-CoV (Figure 2EH) has more positive potential than SARS-CoV-2 (figure
241 2FI).
242



243
 244 **Figure 2.** Electrostatic potential on surfaces of SARS-CoV and SARS-CoV-2 S proteins. (A)
 245 Top view of S protein structure; (B-C) Top views of electrostatic potential on surfaces of SARS-
 246 CoV and SARS-CoV-2 S protein, respectively; (D) Front view of S protein structure; (E-F)
 247 Front views of electrostatic potential on surfaces of SARS-CoV and SARS-CoV-2 S protein,
 248 respectively; (G) Bottom view of S protein structure; (H-I) Bottom views of electrostatic
 249 potential on surfaces of SARS-CoV and SARS-CoV-2 S protein, respectively. Negatively and
 250 positively charged areas are colored in red and blue respectively, with the color scale from -1.0
 251 to 1.0 kT/e.

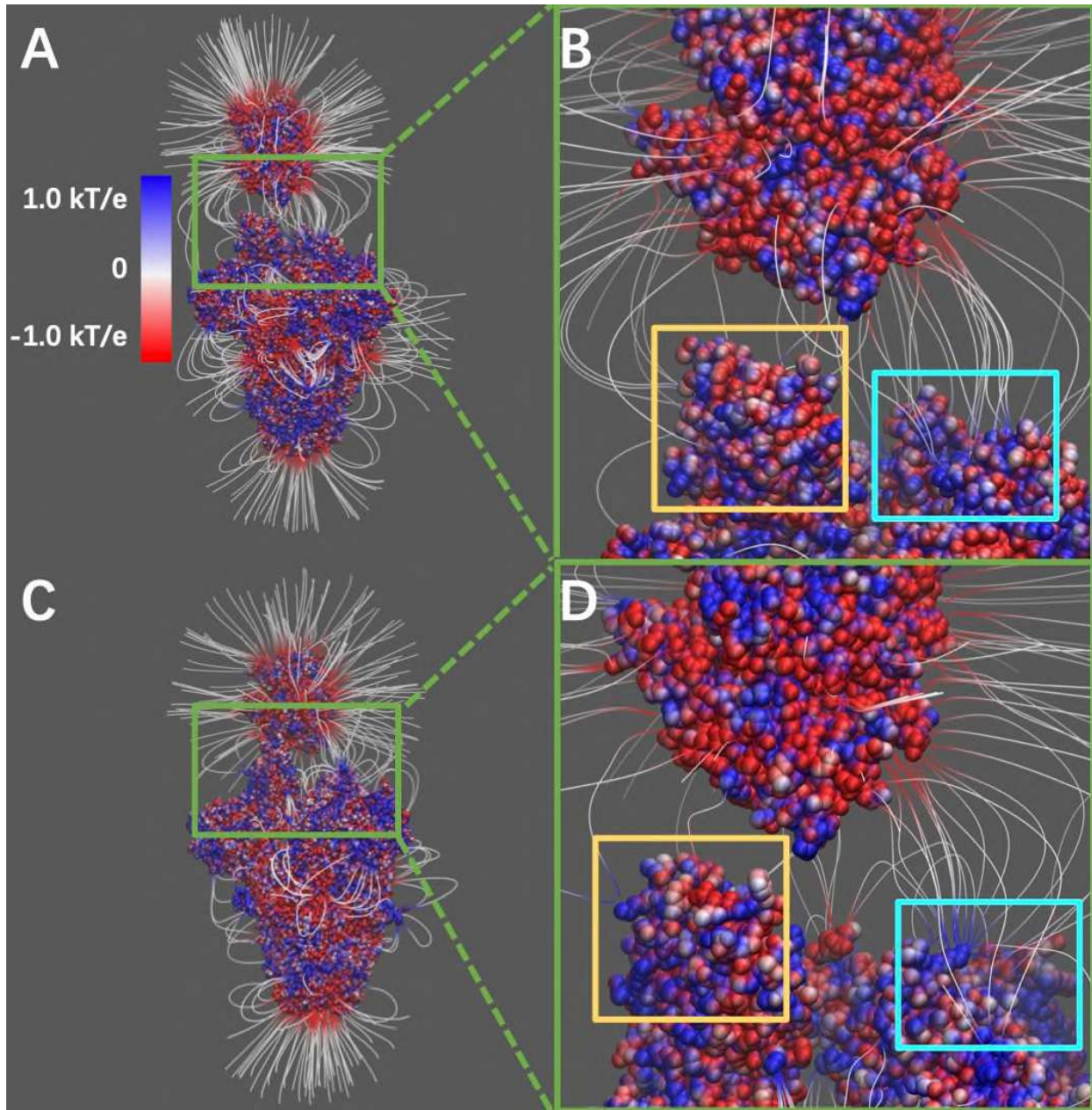
252

253 3.3 Electric Filed Lines

254 Electric field lines surrounding the two complex structures were calculated. To better visualize
 255 the field lines between interfaces, the S protein RBDs are separated from hACE2 RBDs by 10Å
 256 (figure 3). The field lines confirmed that both the SARS-CoV and SARS-CoV-2 S protein RBDs
 257 have attractive forces to hACE2 protein. In the analysis of field lines, the density of field lines

258 indicates the strength of binding force, which means the denser area has the stronger interactions.
259 The electric field lines demonstrate that when hACE2 is away from S protein, all the three S
260 protein monomers provide attractive interactions to the hACE2. This is expected because the S
261 protein RBDs are positively charged while the hACE2 is negatively charged, as shown in figure
262 2 and movie 3, respectively. When hACE2 binds to S proteins (as shown in figure 1), the hACE2
263 only binds with one S protein RBD, which is in open state. Combining the information from
264 figure 1 and 3, it demonstrates that all the three S protein RBDs generate attractive forces to
265 hACE2. However, when hACE2 gets closer to S protein, one S protein RBD flips out and binds
266 to the hACE2 tightly, while the other two S protein RBDs stay in closed state. Even though the
267 monomer with flipped-out S protein RBD is the closest to hACE2 and forms most of the salt
268 bridges and hydrogen bonds, the other two monomers also provide dense field lines and show
269 strong attractive interactions between S proteins and hACE2.

270



271

272

273 **Figure 3.** Electrostatic field lines at the interfaces of S protein and hACE2. (A) Electrostatic
 274 field lines between SARS-CoV S protein and hACE2; (B) A closeup view of binding domain
 275 between SARS-CoV S protein and hACE2 (C) Electrostatic field lines between SARS-CoV-2 S
 276 protein and hACE2; (D) A closeup view of binding domain between SARS-CoV-2 S protein and
 277 hACE2. Negatively and positively charged areas are colored in red and blue, respectively. Color
 278 scale is -1.0 to 1.0 kT/e. Yellow square areas are the RBD of S proteins at open state to reach the
 279 hACE2, cyan square areas are the the RBD of S proteins at closed state.

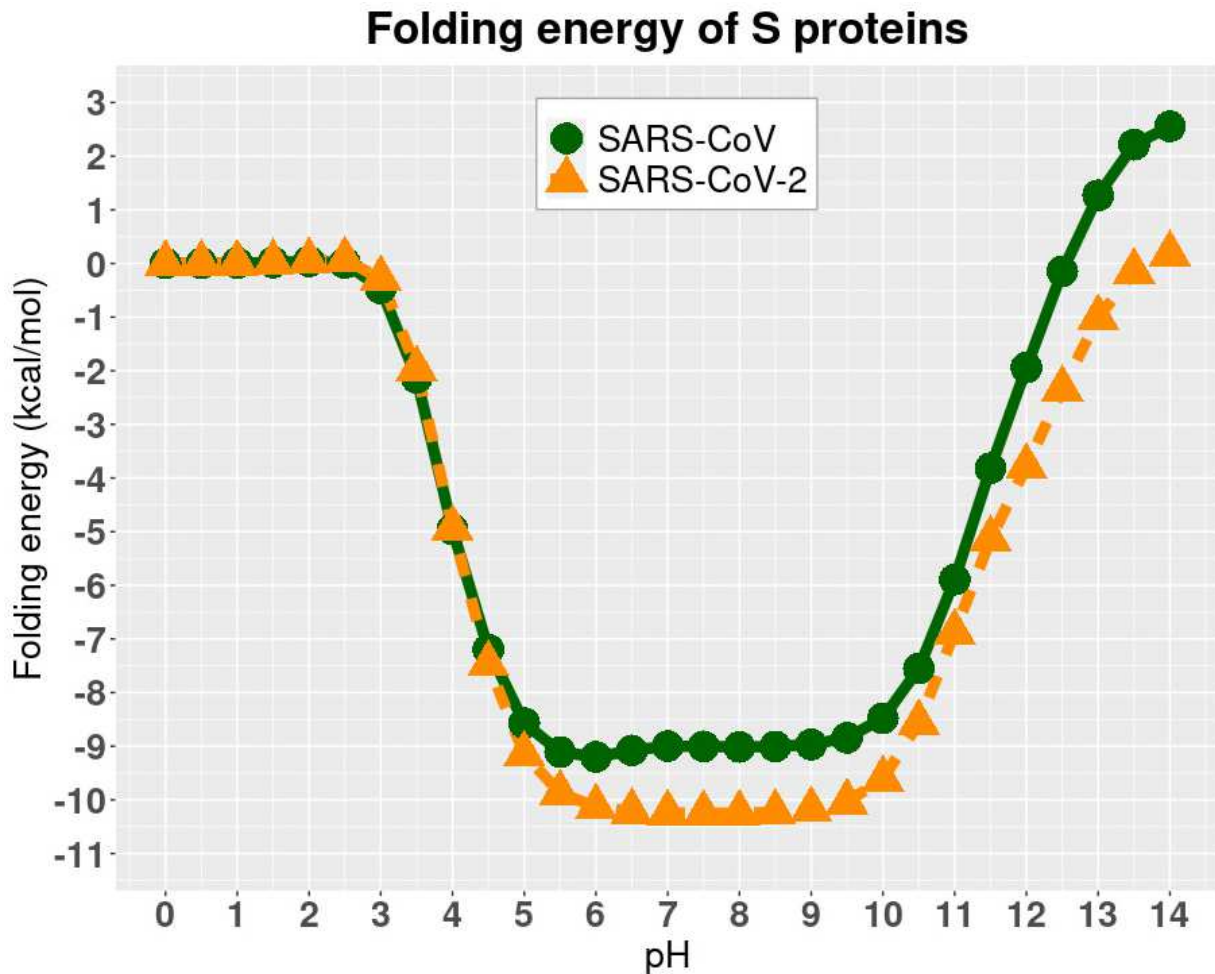
280

281 **3.4 pH-Dependence of Relative Folding Energies**

282 The folding energy of SARS-CoV and SARS-CoV-2 complexes were calculated using
283 DelPhiPKa at different pH values ranging from 0 to 14 with an interval of 0.5 (Figure 4). We
284 observed that SARS-CoV and SARS-CoV-2 have the same trend of folding energy with the
285 change of pH values, which is decreasing from 0 to 6, then becoming stable from 6 to 9, and
286 increasing from 10 to 14. Other than the trend, the optimal values locate between 6 to 9 for both
287 of the viruses.

288 Please note that the folding energies in figure 4 are relative values because we set the reference
289 energy to be 0 kcal/mol when pH is equal to 0. We did not calculate the absolute values of
290 folding energies since we focused on the pH dependency of the folding energies.

291



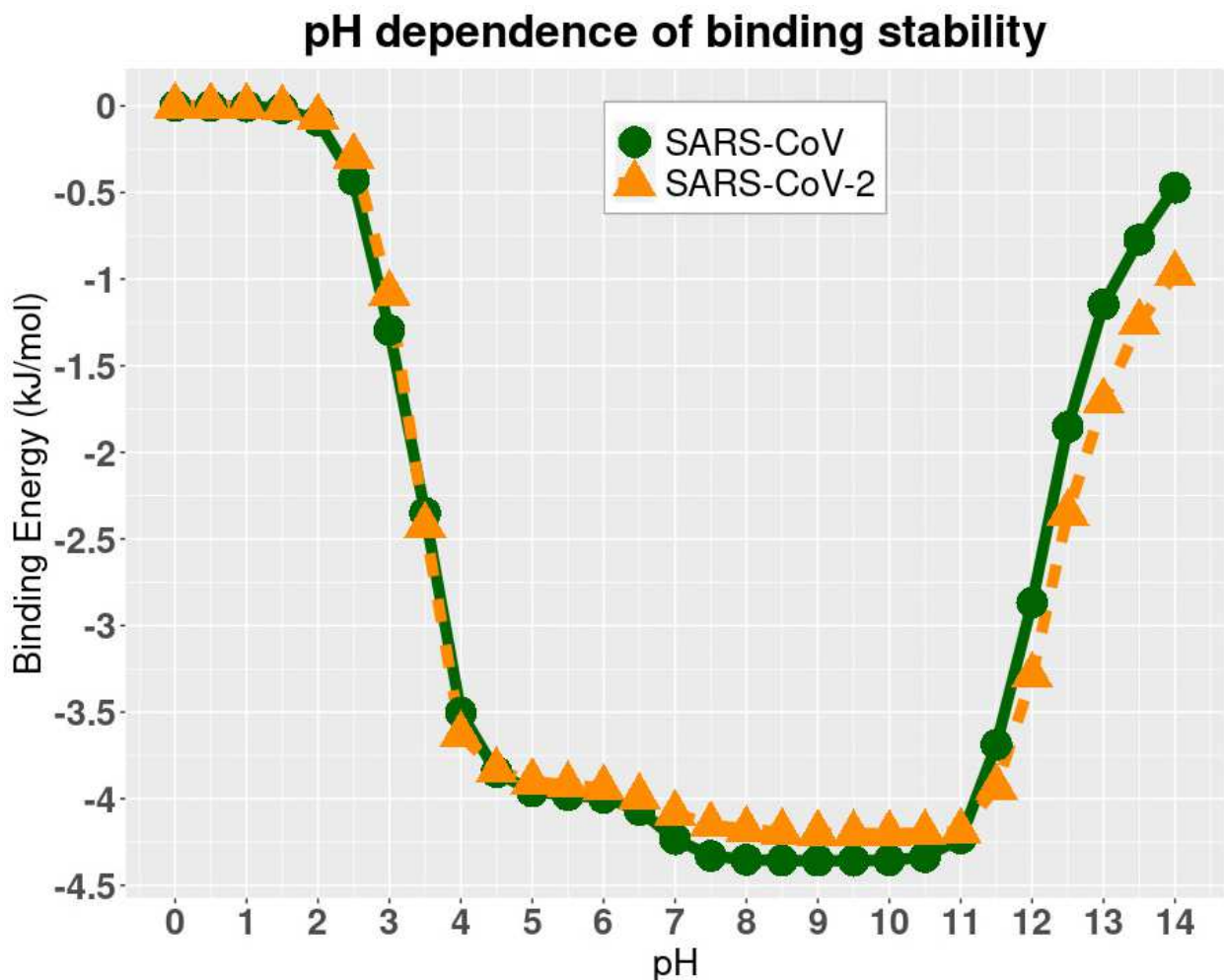
292

293 **Figure 4.** pH-dependence of the relative folding energy of S protein RBDs of SARS-CoV and
294 SARS-CoV-2.

295

296 **3.5 pH-Dependence of Relative Binding Energies**

297 DelPhiPKa was implemented to calculate the binding energies of two complex structures at
298 different pH values. The results are presented in figure 5, where we noticed that the binding free
299 energies of both SARS-CoV and SARS-CoV-2 complexes are stable at the pH values ranging from
300 7.5 to 10.5, which indicates that both SARS-CoV and SARS-CoV-2 have a slight preference of
301 weakly basic environment. Note that the method implementing DelPhiPKa calculates the relative
302 folding and binding energies rather than absolute energies. The folding/binding energy at pH 0 is
303 set as reference, which is 0 kcal/mol. The relative energy profile is used to study the
304 folding/binding energy dependence on pHs. The absolute binding energies was calculated in later
305 section using MM/PBSA method. Combine the folding and binding energy profiles, it is concluded
306 that the best pH environment for both the SARS-CoV and SARS-CoV-2 is from pH 7.5 to 9. Please
307 note that the binding energies in figure 5 are relative values because we set the reference energy
308 to be 0 kJ/mol when pH is equal to 0. We did not calculate the absolute values of binding energies
309 since we focused on the pH dependency of the binding stability.
310



311
312 **Figure 5.** The relative binding energies of complexes at different pH values.

313

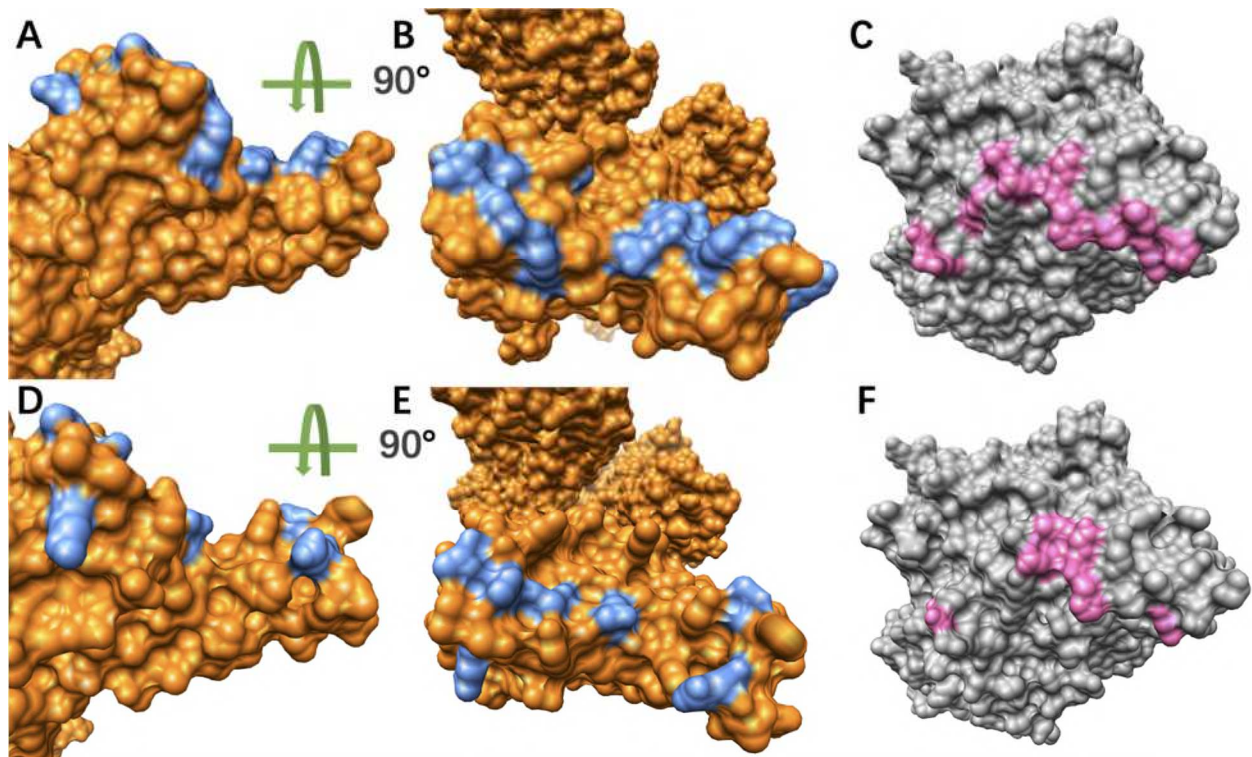
314 **3.6 Hydrogen Bonds Analysis**

315 To analyze the hydrogen bonds distributions on both S proteins RBDs and hACE2 RBD, we
316 colored the residues forming hydrogen bonds which are over 50% frequency during the MD
317 simulations in figure 6. It's obvious that the SARS-CoV S protein has more residues involved in
318 the hydrogen bonds which are over 50%. Accordingly, the hACE also has more residues forming
319 hydrogen bonds (over 50% frequency) with SARS-CoV S protein.

320

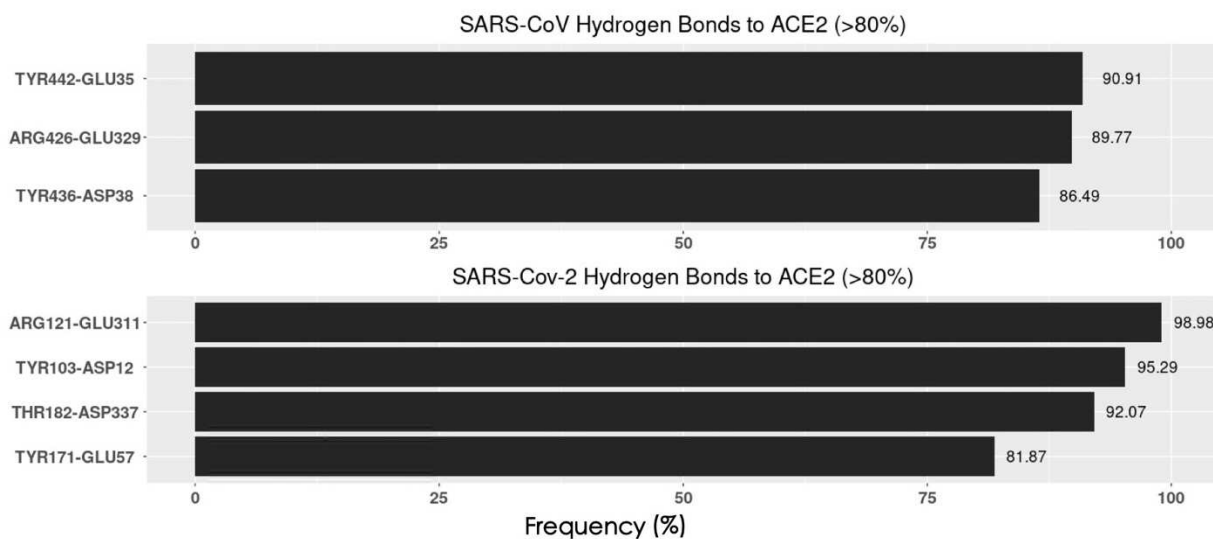
321 In order to consider the most essential hydrogen bonds, which are the hydrogen bonds with
322 relatively high frequencies, we took 80% as a cutoff, which means those hydrogen bonds with
323 80% or higher frequency are considered as the relatively more essential ones. By comparing the
324 figure 7A and 7B, SARS-CoV-2 RBD forms one more essential hydrogen bonds than SARS-
325 CoV RBD when binding to hACE2. The residues involved in forming hydrogen bonds over 50%
326 frequency were colored with their side chains, in which the residues with over 80% frequency
327 hydrogen bonds were labeled and highlighted in grey squares (figure 8CF). From the analyses of
328 figure 6-8, it is revealed that SARS-CoV uses more hydrogen bonds to bind with hACE2.
329 However, more high frequency hydrogen bonds are formed in the SARS-CoV-2/hACE2
330 complex. The key residues forming essential hydrogen bonds from SARS-CoV-2 are: ARG-121,
331 TYR103, THR182 and TYR171. Such residues have significant contributions to the binding of
332 SARS-CoV-2 and hACE2. Therefore, these residues have higher potential to be targets for future
333 drug design.

334



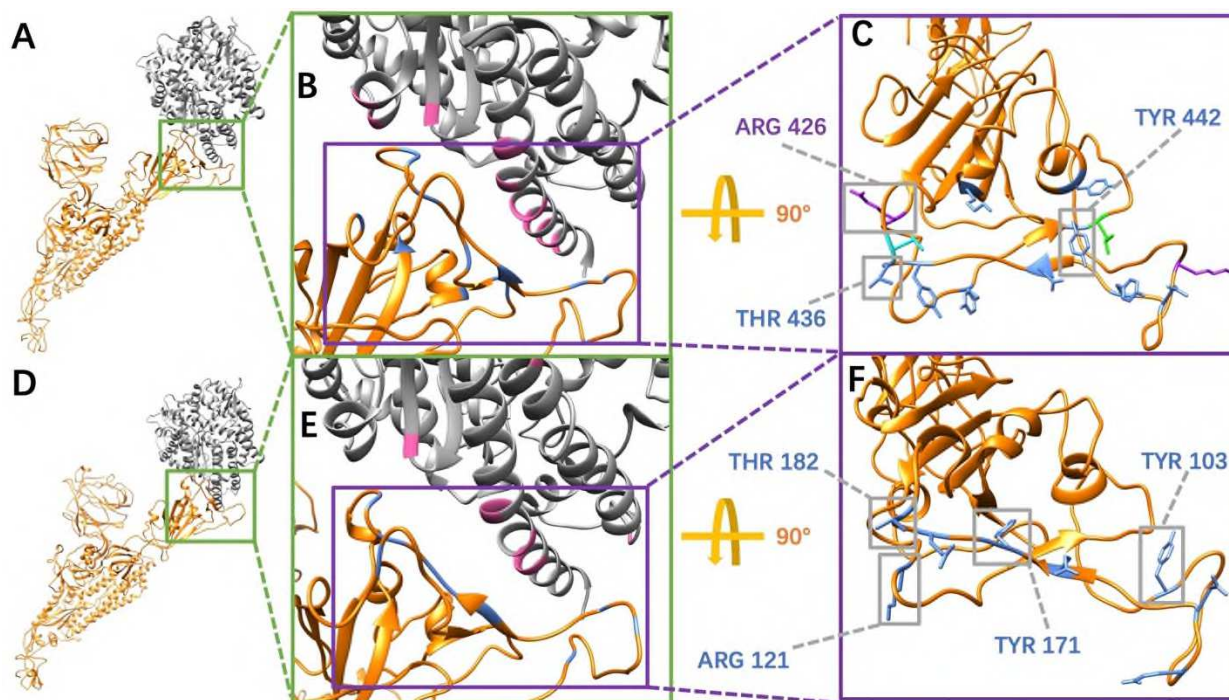
335
 336
 337
 338
 339
 340
 341
 342
 343
 344

Figure 6. Hydrogen bonds distributions at the binding interfaces. (A) Hydrogen bonds distribution (blue) on the interface of SARS-CoV RBD (orange); (B) Turn (A) for 90 degree for the top view, which is the interface that faces hACE2; (C) The hydrogen bonds distribution (pink) at the interface of hACE2 (grey) where SARS-CoV binds; (D) Hydrogen bonds distribution (blue) on the interface of SARS-CoV-2 RBD (orange); (E) Turn (D) for 90 degree for the top view, which is the interface that faces hACE2; (F) The hydrogen bonds distribution (pink) at the interface of hACE2 (grey) where SARS-CoV-2 binds.



345
346
347
348
349
350

Figure 7. Essential Hydrogen bonds at the interfaces between SARS-CoV/SARS-CoV-2 RBDs and hACE2 RBD with the frequency above 80%.



351
352
353
354
355

Figure 8. Key residues involved in essential hydrogen bonds at the interfaces between SARS-CoV/SARS-CoV-2 RBDs and hACE2 RBD with the frequency above 80%. (A) SARS-CoV S protein single chain binds to hACE2; (B) A closeup view of (A) at the binding interface; (C)

356 Labelled key residues that form essential hydrogen bonds (frequency over 80%) at the interface;
357 (D) SARS-CoV-2 S protein single chain binds to hACE2; (E) A closeup view of (D) at the
358 binding interface; (F) Labelled key residues that form essential hydrogen bonds (frequency over
359 80%) at the interface.
360

361 **4 Limitation**

362 The limitation for this work is that we used relative folding energy and binding energy to analyze
363 rather than the absolute values. Since our work is focused on the relative stability under the pH
364 effects, the relative energy calculations do not affect our conclusions.
365

366 **5 Conclusion**

367 In this work, we applied several computational methods, including MD simulations, DelPhi,
368 DelPhiForce and DelPhiPKa to study the electrostatic features of S proteins for SARS-CoV and
369 SARS-CoV-2. From our results, SARS-CoV and SARS-CoV-2 S protein RBDs both have
370 positively charged interfaces, which provides attractive interactions to hACE2 as hACE2 has
371 negatively charged surface.
372

373 Also, we revealed the pH-dependence calculations of relative folding energy for SARS-CoV and
374 SARS-CoV-2 S protein RBDs. The best pH to stabilize SARS-CoV and SARS-CoV-2 S protein
375 RBDs is in the range of 6 to 9. The study on pH dependence of binding energies revealed that the
376 complex structures of hACE2 and S proteins of SARS-CoV/ SARS-CoV-2 are stable from pH
377 7.5 to 10.5. Therefore, SARS-CoV and SARS-CoV-2 survive in a similar pH environment. The
378 pH 7.5 to 9 is the best condition for both SARS-CoV and SARS-CoV-2 to best perform their
379 functions to bind with hACE2.
380

381 Besides, based on 100ns MD simulations, we found that for the essential hydrogen bonds (>80%
382 frequency), SARS-CoV-2 has four pairs while SARS-CoV has three pairs, which indicates the
383 relatively more robust binding strategy of SARS-CoV-2 compared to SARS-CoV. The key
384 residues forming essential hydrogen bonds from SARS-CoV-2 are ARG-121, TYR103, THR182
385 and TYR171, which are potential drug targets for COVID-19 treatments. By using multiple
386 computational approaches, the findings in this work shed light on the current and future
387 treatments of COVID-19 and other coronaviruses-caused diseases.
388

389 **Acknowledgement**

390
391

392 This research is funded by Grant SC1GM132043 from National Institutes of Health (NIH); Grant
393 5U54MD007592 from the National Institutes on Minority Health and Health Disparities
394 (NIMHD), a component of the NIH.

395
396

397 **References**

- 398 1. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S.C. and Di Napoli, R., *Features,*
399 *evaluation, and treatment of coronavirus (COVID-19).* Statpearls [internet], 2021.
- 400 2. Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B. and
401 Huang, C.-L., *A pneumonia outbreak associated with a new coronavirus of probable bat*
402 *origin.* nature, 2020. **579**(7798): p. 270-273.
- 403 3. Van Der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers,
404 K.C., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J. and Berkhout, B.,
405 *Identification of a new human coronavirus.* Nature medicine, 2004. **10**(4): p. 368-373.
- 406 4. Hadi, A., Werge, M., Kristiansen, K.T., Pedersen, U.G., Karstensen, J.G., Novovic, S. and
407 Gluud, L.L., *Coronavirus disease-19 (COVID-19) associated with severe acute*
408 *pancreatitis: case report on three family members.* Pancreatology, 2020. **20**(4): p. 665-
409 667.
- 410 5. Liu, D.X., Liang, J.Q. and Fung, T.S., *Human coronavirus-229E,-OC43,-NL63, and-HKU1.*
411 Reference Module in Life Sciences, 2020.
- 412 6. Loeffelholz, M.J. and Tang, Y.-W., *Laboratory diagnosis of emerging human coronavirus*
413 *infections—the state of the art.* Emerging microbes & infections, 2020. **9**(1): p. 747-756.
- 414 7. Yang, J., Wang, W., Chen, Z., Lu, S., Yang, F., Bi, Z., Bao, L., Mo, F., Li, X. and Huang, Y., *A*
415 *vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity.*
416 Nature, 2020. **586**(7830): p. 572-577.
- 417 8. Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A. and Li, F., *Cell entry*
418 *mechanisms of SARS-CoV-2.* Proceedings of the National Academy of Sciences, 2020.
419 **117**(21): p. 11727-11734.
- 420 9. Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y. and Yuen,
421 K.-Y., *Structural and functional basis of SARS-CoV-2 entry by using human ACE2.* Cell,
422 2020.
- 423 10. Brielle, E.S., Schneidman-Duhovny, D. and Linial, M., *The SARS-CoV-2 exerts a distinctive*
424 *strategy for interacting with the ACE2 human receptor.* Viruses, 2020. **12**(5): p. 497.
- 425 11. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q. and Zhang, L.,
426 *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor.*
427 Nature, 2020. **581**(7807): p. 215-220.

- 428 12. Wang, J., *Fast identification of possible drug treatment of coronavirus disease-19*
429 *(COVID-19) through computational drug repurposing study*. Journal of chemical
430 information and modeling, 2020. **60**(6): p. 3277-3286.
- 431 13. Li, L., Wang, L. and Alexov, E., *On the energy components governing molecular*
432 *recognition in the framework of continuum approaches*. Frontiers in molecular
433 biosciences, 2015. **2**: p. 5.
- 434 14. Li, L., Guo, D., Huang, Y., Liu, S. and Xiao, Y., *ASPDock: protein-protein docking algorithm*
435 *using atomic solvation parameters model*. BMC bioinformatics, 2011. **12**(1): p. 1-9.
- 436 15. Allen, M.P., *Introduction to molecular dynamics simulation*. Computational soft matter:
437 from synthetic polymers to proteins, 2004. **23**(1): p. 1-28.
- 438 16. Liwo, A., He, Y. and Scheraga, H.A., *Coarse-grained force field: general folding theory*.
439 Physical Chemistry Chemical Physics, 2011. **13**(38): p. 16890-16901.
- 440 17. Lopez-Hernandez, A.E., Xie, Y., Guo, W. and Li, L., *The electrostatic features of dengue*
441 *virus capsid assembly*. Journal of Computational Biophysics and Chemistry, 2021.
- 442 18. Guo, W., Xie, Y., Lopez-Hernandez, A.E., Sun, S. and Li, L., *Electrostatic features for*
443 *nucleocapsid proteins of SARS-CoV and SARS-CoV-2*. Mathematical Biosciences and
444 Engineering, 2021. **18**(3): p. 2372-2383.
- 445 19. Peng, Y. and Alexov, E., *Computational investigation of proton transfer, p K a shifts and p*
446 *H - optimum of protein - DNA and protein - RNA complexes*. Proteins: Structure,
447 Function, and Bioinformatics, 2017. **85**(2): p. 282-295.
- 448 20. Wang, L., Witham, S., Zhang, Z., Li, L., Hodsdon, M.E. and Alexov, E., *In silico*
449 *investigation of pH-dependence of prolactin and human growth hormone binding to*
450 *human prolactin receptor*. Communications in computational physics, 2013. **13**(1): p.
451 207.
- 452 21. Xian, Y., Karki, C.B., Silva, S.M., Li, L. and Xiao, C., *The roles of electrostatic interactions in*
453 *capsid assembly mechanisms of giant viruses*. International journal of molecular
454 sciences, 2019. **20**(8): p. 1876.
- 455 22. Salas, G.G.S., Hernandez, A.E.L., He, J., Karki, C., Xie, Y., Sun, S., Xian, Y. and Li, L., *Using*
456 *computational approaches to study dengue virus capsid assembly*. Computational and
457 Mathematical Biophysics, 2019. **7**(1): p. 64-72.
- 458 23. Xian, Y., Xie, Y., Silva, S.M., Karki, C., Qiu, W. and Li, L., *Structure manipulation tool*
459 *structureMan: A structure manipulation tool to study large scale biomolecular*
460 *interactions*. Frontiers in molecular biosciences, 2020. **7**: p. 476.
- 461 24. Karki, C., Xian, Y., Xie, Y., Sun, S., Lopez-Hernandez, A.E., Juarez, B., Wang, J., Sun, J. and
462 Li, L., *A computational model of ESAT-6 complex in membrane*. Journal of Theoretical
463 and Computational Chemistry, 2020. **19**(03): p. 2040002.

- 464 25. Li, C., Li, L., Zhang, J. and Alexov, E., *Highly efficient and exact method for parallelization*
465 *of grid - based algorithms and its implementation in DelPhi*. Journal of computational
466 chemistry, 2012. **33**(24): p. 1960-1966.
- 467 26. Xie, Y., Du, D., Karki, C.B., Guo, W., Lopez-Hernandez, A.E., Sun, S., Juarez, B.Y., Li, H.,
468 Wang, J. and Li, L., *Revealing the mechanism of SARS-CoV-2 spike protein binding with*
469 *ACE2*. Computing in Science & Engineering, 2020. **22**(6): p. 21-29.
- 470 27. Xie, Y., Karki, C.B., Du, D., Li, H., Wang, J., Sobitan, A., Teng, S., Tang, Q. and Li, L., *Spike*
471 *proteins of SARS-CoV and SARS-CoV-2 utilize different mechanisms to bind with human*
472 *ACE2*. Frontiers in molecular biosciences, 2020. **7**.
- 473 28. Hart, O.E. and Halden, R.U., *Computational analysis of SARS-CoV-2/COVID-19*
474 *surveillance by wastewater-based epidemiology locally and globally: Feasibility,*
475 *economy, opportunities and challenges*. Science of the Total Environment, 2020. **730**: p.
476 138875.
- 477 29. Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., Wang, Q., Xu, Y., Li, M. and Li, X.,
478 *Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by*
479 *computational methods*. Acta Pharmaceutica Sinica B, 2020. **10**(5): p. 766-788.
- 480 30. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel,
481 R.D., Kale, L. and Schulten, K., *Scalable molecular dynamics with NAMD*. Journal of
482 computational chemistry, 2005. **26**(16): p. 1781-1802.
- 483 31. Humphrey, W., Dalke, A. and Schulten, K., *VMD: visual molecular dynamics*. Journal of
484 molecular graphics, 1996. **14**(1): p. 33-38.
- 485 32. Song, W., Gui, M., Wang, X. and Xiang, Y., *Cryo-EM structure of the SARS coronavirus*
486 *spike glycoprotein in complex with its host cell receptor ACE2*. PLoS pathogens, 2018.
487 **14**(8): p. e1007236.
- 488 33. Juraszek, J., Rutten, L., Blokland, S., Bouchier, P., Voorzaat, R., Ritschel, T., Bakkers, M.J.,
489 Renault, L.L. and Langedijk, J.P., *Stabilizing the closed SARS-CoV-2 spike trimer*. Nature
490 communications, 2021. **12**(1): p. 1-8.
- 491 34. Webb, B. and Sali, A., *Comparative protein structure modeling using MODELLER*. Current
492 protocols in bioinformatics, 2016. **54**(1): p. 5.6. 1-5.6. 37.
- 493 35. Li, L., Li, C., Sarkar, S., Zhang, J., Witham, S., Zhang, Z., Wang, L., Smith, N., Petukh, M.
494 and Alexov, E., *DelPhi: a comprehensive suite for DelPhi software and associated*
495 *resources*. BMC biophysics, 2012. **5**(1): p. 9.
- 496 36. Li, L., Li, C., Zhang, Z. and Alexov, E., *On the dielectric "constant" of proteins: smooth*
497 *dielectric function for macromolecular modeling and its implementation in DelPhi*.
498 Journal of chemical theory and computation, 2013. **9**(4): p. 2126-2136.
- 499 37. Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A., *PDB2PQR: an automated*
500 *pipeline for the setup of Poisson-Boltzmann electrostatics calculations*. Nucleic acids
501 research, 2004. **32**(suppl_2): p. W665-W667.

- 502 38. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A., *Development and*
503 *testing of a general amber force field*. Journal of computational chemistry, 2004. **25**(9):
504 p. 1157-1174.
- 505 39. Li, L., Jia, Z., Peng, Y., Chakravorty, A., Sun, L. and Alexov, E., *DelPhiForce web server:*
506 *electrostatic forces and energy calculations and visualization*. Bioinformatics, 2017.
507 **33**(22): p. 3661-3663.
- 508 40. Wang, L., Zhang, M. and Alexov, E., *DelPhiPKa web server: predicting pKa of proteins,*
509 *RNAs and DNAs*. Bioinformatics, 2015. **32**(4): p. 614-615.
- 510 41. Wang, L., Li, L. and Alexov, E., *pKa predictions for proteins, RNAs, and DNAs with the*
511 *Gaussian dielectric function using DelPhi pKa*. Proteins: Structure, Function, and
512 Bioinformatics, 2015. **83**(12): p. 2186-2197.
- 513 42. Massova, I. and Kollman, P.A., *Combined molecular mechanical and continuum solvent*
514 *approach (MM-PBSA/GBSA) to predict ligand binding*. Perspectives in drug discovery and
515 design, 2000. **18**(1): p. 113-135.
- 516 43. Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner,
517 J., Jeong, J.C. and Qi, Y., *CHARMM-GUI input generator for NAMD, GROMACS, AMBER,*
518 *OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force*
519 *field*. Journal of chemical theory and computation, 2016. **12**(1): p. 405-413.
- 520

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S1.jpg](#)
- [S2.jpg](#)
- [SI.docx](#)