2021-05-01

Applying Computational Methods To Study The Interactions Between Sars-Cov-2 And Hace2

Yixin Xie
University of Texas at El Paso

APPLYING COMPUTATIONAL METHODS TO STUDY THE

INTERACTIONS BETWEEN SARS-COV-2 AND hACE2

YIXIN XIE

Master's Program in Computational Science

APPROVED:

_____
Lin Li, Ph.D., Chair

_____
Jorge A. Muñoz, Ph.D.

_____
Jianjun Sun, Ph.D.

_____
Stephen L. Crites, Ph.D.
Dean of the Graduate School

to my

FAMILY & FRIENDS

with love

APPLYING COMPUTATIONAL METHODS TO STUDY THE INTERACTIONS BETWEEN SARS-COV-2 AND hACE2

by

YIXIN XIE, M.A.

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Computational Science Program

THE UNIVERSITY OF TEXAS AT EL PASO

May 2021

ACKNOWLEDGEMENTS

In this research, firstly I would like to thank my advisor Dr. Lin Li, Assistant Professor from Department of Physical Sciences, for his unconditional support, timely help and patient guidance. Dr. Li has inspired me a lot in both my research and my attitude to my future, which encourages me to become a more determined and passionate person.

And I would like to express gratitude to my committee members Dr. Jianjun Sun from Department of Biological Sciences Biosciences and Dr. Jorge Muñoz from Department of Physical Sciences for their kind help and advice.

Meanwhile, I am also grateful for my family and friends for their wonderful love. Thank them for making me become who I am.

ABSTRACT

The ongoing outbreak of COVID-19 has been a serious threat to human health worldwide. The virus SARS-CoV-2 initiates its infection to the human body via the interaction of its spike (S) protein with the human Angiotensin-Converting Enzyme 2 (ACE2) of the host cells. Here we implemented multi-scale computational approaches to study the binding mechanisms of human ACE2 and S proteins of both SARS-CoV and SARS-CoV-2. Electrostatic features, including electrostatic potential, electric field lines and electrostatic forces of SARS-CoV and SARS-CoV-2 were calculated and compared in detail. The results demonstrate that SARS-CoV and SARS-CoV-2 S proteins are both attractive to ACE2 by electrostatic forces even at different distances. However, the residues contributing to the electrostatic features are quite different due to the mutations between SARS-CoV S protein and SARS-CoV-2 S protein. Compared to SARS-CoV, the SARS-CoV-2 binds with ACE2 using a more robust strategy: The electric field line related residues are distributed quite differently, which results in a more robust binding strategy of SARS-CoV-2. Also, SARS-CoV-2 has a higher electric field line density than that of SARS-CoV, which indicates stronger interaction between SARS-CoV-2 and ACE2, compared to that of SARS-CoV. Key residues involved in salt bridges and hydrogen bonds are identified in this study, which may help the future drug design against COVID-19.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1 Motivation

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

1.2 Objectives

Our research goal is to use multi-scale computational methods to study the mechanisms of both SARS-CoV and SARS-CoV-2 S proteins binding with human ACE2.

1.3 Thesis statement and research questions

Our work aims to answer the following questions:

1. Do S proteins of SARS-CoV and SARS-CoV-2 have similarities or differences in structural and sequential features? What are them?

2. How do the MD simulations perform to simulate the binding processes between S protein and hACE2 protein?

3. Do SARS-CoV and SARS-CoV-2 have different electrostatic potential values? What are the differences? What is the best way to visualize the values?

4. What are the electrostatic forces between S proteins and hACE2? Do they change the values or directions in the process of binding?

5. How many strong salt bridges between S proteins and hACE2? What about hydrogen bonds?

6. Can we identify several key residues in order to help the development of the future drug design for COVID-19?

7. What can we learn from SARS-CoV-2 to be prepared for future coronavirus-caused diseases?

1.4 Thesis map

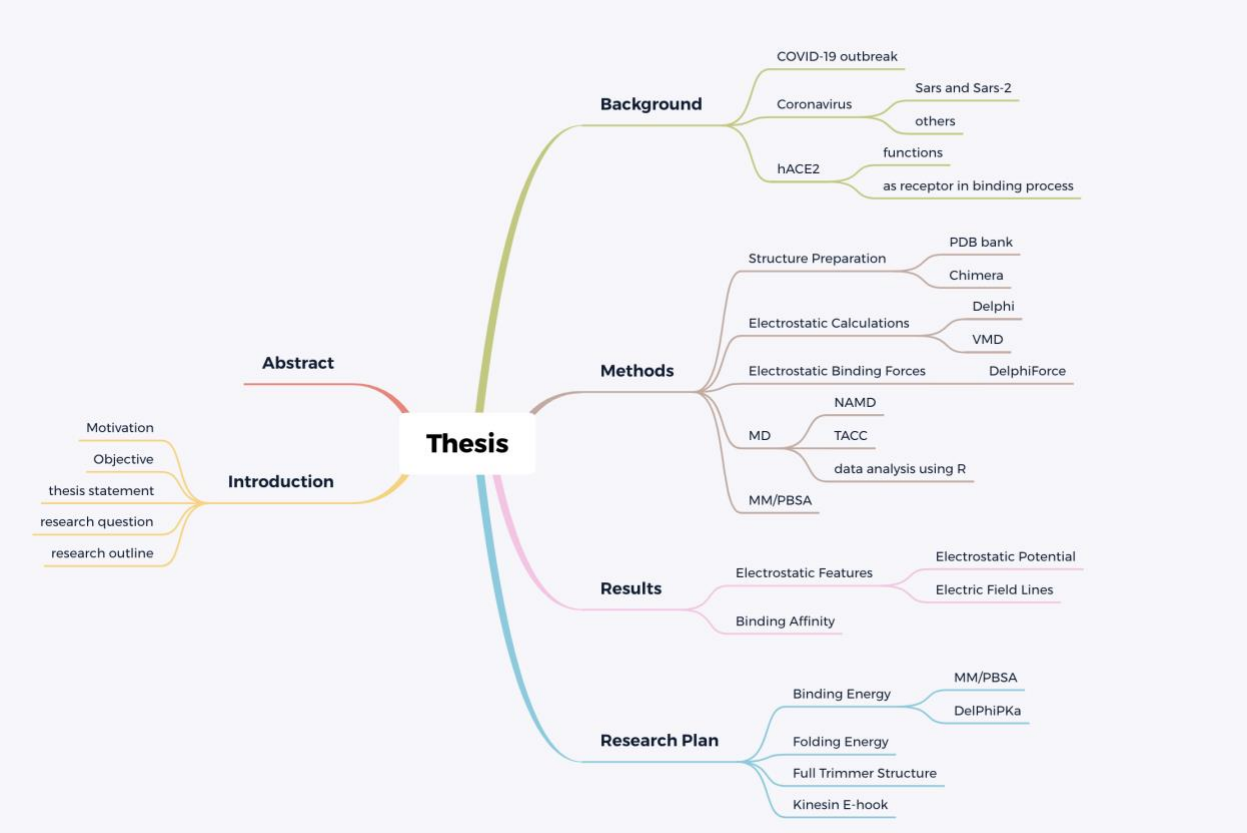

Figure 1: thesis map

2. BACKGROUND

2.1 COVID-19 Outbreak

Recently, the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) is raging throughout the world. This is the seventh member of the Coronaviridae family found which is able to affect human health. Among these seven coronaviruses, four of them (HCoV-229E, HCoV-OC43, HCoV-NL63, HKU1)[7] can only cause mild symptoms, while the other three can cause death-leading diseases. Previously to SARS-CoV-2 that started in 2019, two known respiratory coronaviruses can cause serious respiratory syndromes, that are, the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) [8](broke out in late 2003), and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) [9] (broke out in 2012). The ability of animal-to-human and human-to-human transmission of the 2003 SARS and the following 2012 MERS, resulted in server pandemics that infected over 8000 and 2400 reported infected cases including 774 and 858 death cases, respectively. Compared to SARS-CoV and MERS-CoV, SARS-CoV-2 is causing an even more severe pandemic due to its spreading speed and the population affected.

2.2 Coronaviruses

Four main structural proteins are found in coronaviruses, including spike (S), envelope (E), nucleocapsid (N) and membrane (M) proteins. Giving special attention to the S protein, of which the main function is to bind to the receptor Angiotensin-Converting Enzyme 2 (ACE2), and entering the host cell after binding [10]. Therefore,

the S protein plays a crucial role in the first step of infections for disease-causing coronaviruses. Besides, as reported, the S protein-ACE2 interaction is an easy target for drugs or vaccines. Many efforts have been contributed to investigate S proteins and their receptors, such as ACE2 [11-14]. Even though S proteins of SARS-CoV and SARS-CoV-2 share very similar structures, the binding affinities of S protein and ACE2 of SARS-CoV-2 are much higher than SARS-CoV [15]. This might be the key reason for SARS-CoV-2's faster-spreading speed, comparing to SARS-CoV. In this case, revealing the differences in SARS-CoV and SARS-CoV-2 should provide a deep understanding of how coronaviruses affect human health. Due to the essential role of S proteins, this work reveals some mechanisms of SARS-CoV and SARS-CoV-2's S proteins binding to the ACE2 from biophysics perspectives using multi-scale computational approaches.

2.3 Human ACE2 Protein

Human hACE2 (hACE2) is an enzyme located widely in the human body, including the lungs, kidneys, adipose tissue, central nervous system and cardiovascular system[13, 16, 17] and it has multiple essential functions such as the regulation of amino acid transport in the kidney controlling the blood pressure, and viral receptors including both SARS-CoV-2 and SARS-CoV[17]. Since it is of extreme importance to human health, there are numerous research groups have been or are currently working on S proteins and hACE2 using various approaches.

3. METHODS

3.1 Structure Preparation

The complex structures of SARS-CoV and ACE2 were downloaded from the Protein Data Bank (PDB ID 6ACG [18]). Several SARS-CoV-2 S protein RBD/ACE2 complex structures were determined [12, 13]. In this work, we used 6VW1 [19] as our complex structure to study the electrostatic binding interactions between SARS-CoV-2 S protein RBD and ACE2. However, the 6VW1 only contains the binding domains of SARS-CoV-2 S protein and ACE2. To study the electrostatic features for the overall structure of SARS-CoV-2 S protein, SWISS model [20] was used to model the whole structure of the SARS-CoV-2 S protein trimmer binding with ACE2 based on the template of 6ACG. The sequence of SARS-CoV-2 was obtained from Genebank [21], which was from the early patients in December 2019. The modeled RBD structure has less than 1Å RMSD compared to the experimental determined RBD structures, which demonstrates that the modeled RBD structure is very reliable (as shown in figure 2). When studying the electrostatic interactions, we mainly focused on the receptor binding domain (RBD) of S protein and the binding domain of ACE2. To understand the mechanisms of S protein binding to ACE2, S protein RBD was separated from the ACE2 binding domain by a distance from 5Å to 40Å with a step of 1Å, when analyzing the electrostatic binding forces.

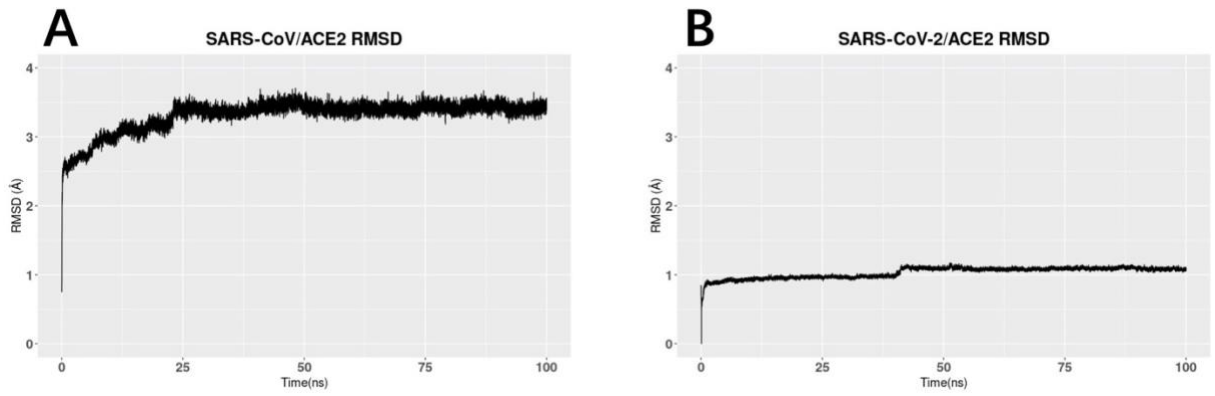

Figure 2. RMSD comparison of SARS-CoV/ACE2 and SARS-CoV-2/ACE2 complex structure

3.2 Electrostatic Calculations using DelPhi

In order to study the electrostatic features, DelPhi is utilized to calculate the electrostatic potential for the S protein RBD and ACE2 binding domain. In the framework of continuum electrostatics, DelPhi calculates the electrostatic potential ϕ (in systems comprised of biological macromolecules and water in the presence of mobile ions) by solving the Poisson-Boltzmann equation (PBE):

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

where ϕ(r) is the electrostatic potential, ϵ(r) is the dielectric distribution, ρ(r) is the charge density based on the atomic structures, κ is the Debye-Huckel parameter, k_B is the Boltzmann constant, and T is the temperature. Due to the irregular shape of macromolecules, DelPhi uses a finite difference (FD) method to solve the PBE.

The electrostatic potential of RBDs of SARS-CoV and SARS-CoV-2 with ACE2 were calculated by Delphi. The calculated electrostatic potential on the surface was

visualized with Chimera. In order to visualize electric field lines between SARS-CoV and ACE2 and between SARS-CoV-2 and ACE2, Visual Molecular Dynamics (VMD) [22] was implemented based on the electrostatic potential map from DelPhi calculations. The color scale range was set from −1.0 to 1.0 kT/Å. In order to clearly show the difference between the surface of SARS-CoV-2 and of SARS, the difference has been calculated and visualized by subtracting the electrostatic potential values of SARS-CoV-2 by that of SARS-CoV. Since the surface structures are not completely the same, the surface of SARS-CoV-2 was used as the model to visualize the difference of charge distribution.

In the process of DelPhi calculations, the PQR file of each capsomer was generated by PDB2PQR [23]. During DelPhi calculations, the resolution was set as 1 grids/Å. The dielectric constants were set as 2.0 for protein and 80.0 for the water environment, respectively. The protein filling percentage of Delphi calculation box (perfil) was set to be 70.0. The probe radius for generating the molecular surface was 1.4 Å. Salt concentration was set as 0.15 M. The boundary condition for the Poisson Boltzmann equation was set as a dipolar boundary condition. The calculated electrostatic potential on the surface was visualized with Chimera. VMD was used to illustrate electric field lines between S protein and ACE2. Finally, the color scale range was set to be from -1.0 to 1.0 kT/Å.

3.3 Electrostatic Binding Forces

To compare the strengths and directions of electrostatic forces between RBDs of SARS-CoV-2 and SARS-CoV with ACE2, DelphiForce [28, 29] was implemented to perform the force calculations. As mentioned above, the structures at each distance of S protein and ACE2 protein were used to calculate binding forces. The electrostatic binding forces calculated by DelphiForce were visualized with VMD and represented by arrows. Forces are shown with different S protein RBD-ACE2 distances from 5Å to 40Å with a step of 2Å. The S protein RBD and ACE2 are separated in the direction of their mass centers connection line. For better visualization of force directions in VMD [35], arrows were normalized to be of the same size at variable distances, which shows only the direction of each force without considering its strength by sizes. The magnitudes of the electrostatic binding forces were illustrated and the trends of forces change of the total binding forces are shown.

3.4 Molecular Dynamic (MD) Simulations for SARS-CoV and SARS-CoV-2 RBDs

To simulate the dynamic interactions between S proteins and ACE2 protein, MD simulations were carried out on GPUs using Lonestar5 clusters at the Texas Advanced Computing Center (TACC https://www.tacc.utexas.edu/). A 2000-step minimization was performed for each simulation, followed by a 100 million steps, during which 20,000 frames were saved from two 100ns simulations of both SARS-CoV and SARS-CoV-2 separately (1.0 fs per step, 1 frame at each 5000 steps, 100 million steps in total).

The RMSDs of the SARS-CoV and SARS-CoV-2 trajectories are about 3.4Å and 1.1 Å, respectively. During the MD simulations, the temperature was set to be 300 K, and the pressure was set to be standard using the Langevin dynamics. For PME, which is set for full-system periodic electrostatics, with the grid size (86, 88, 132) as (x, y, z) value respectively. In those two simulations, atoms that are not located in binding domains were constrained within a margin of 10.0 Å of their natural movement maximum length values. In order to get a more accurate result of the simulation, data of the last 50 ns of simulations were selected and used for later data analysis, since the structure of the first 50 ns is not as stable as the last 50 ns of simulations. The simulation processes are visualized in movies 1 and 2, generated by VMD.

To analyze the interaction between S proteins and ACE2, the salt bridges that formed within the distance of 4 Å were extracted from the last 10,000 frames of simulations, and for hydrogen bonds the cutoff was 4 Å. The several top-strongest salt bridges in each binding domain were determined by calculating their formation frequency (the occupancy in figure 3) during MD simulation.

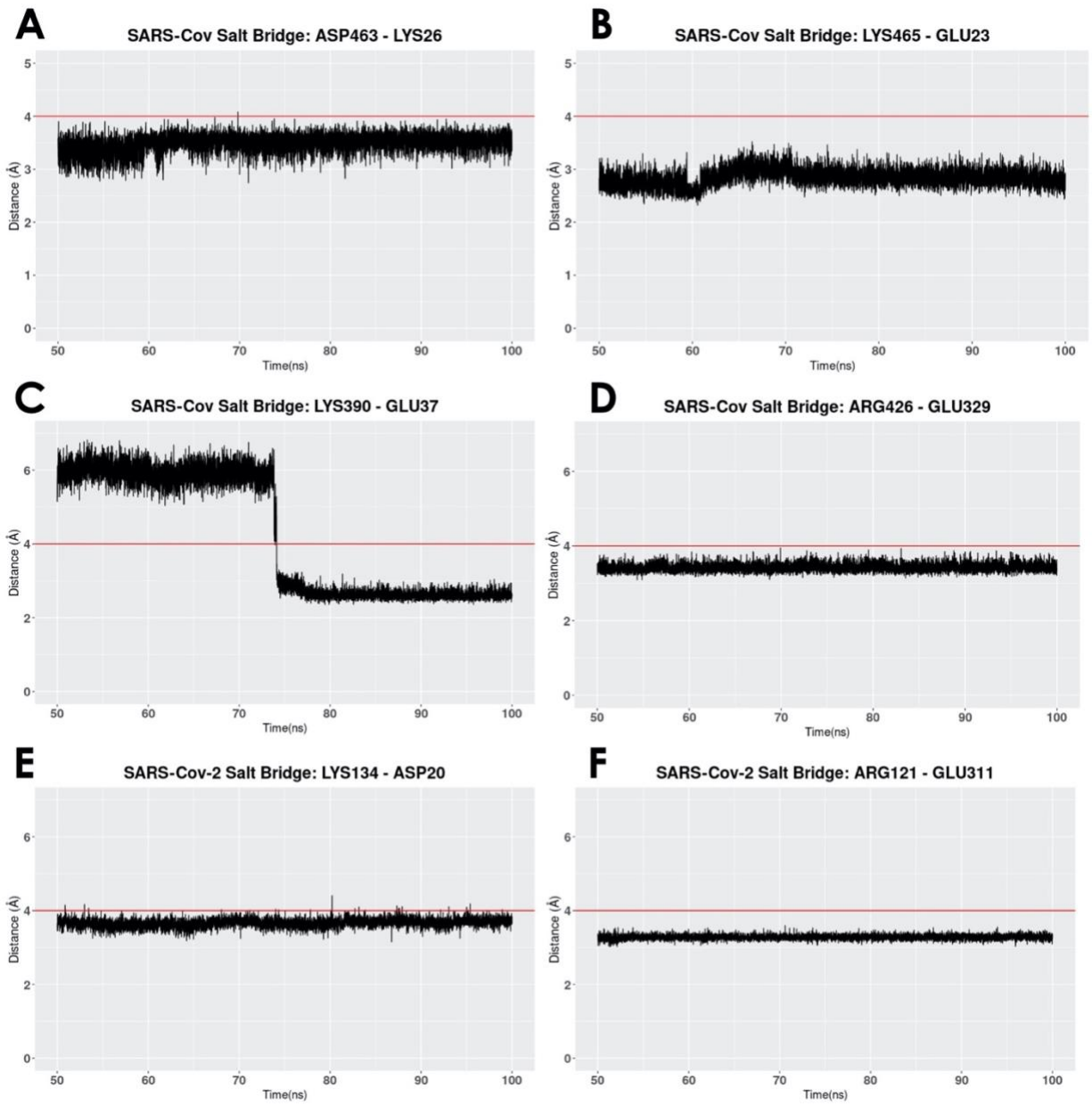

Figure 3. Key residues forming salt bridges at interfaces of S protein RBDs and ACE2 protein binding domain. Red line is the cutoff value of 4 Å which is chosen when calculating the salt bridge. (A) Distance of the ASP463 – LYS26 salt bridge found on the interface of SARS-CoV and ACE2; (B) Distance of the GLU23 – LYS465 salt bridge found on the interface of SARS-CoV and ACE2, (C) Distance of the GLU37 – LYS390 salt bridges found on the interface of SARS-CoV and ACE2; (D) Distance of the GLU329 –

ARG426 salt bridges found on the interface of SARS-CoV and ACE2; (E) Distance of the ASP20 - LYS134 salt bridge found on the interface of SARS-CoV-2 and ACE2; (F) Distance of the GLU311 – ARG121 salt bridges found on the interface of SARS-CoV-2 and ACE2.

4. RESULTS

4.1 The Mutations On SARS-Cov-2

To analyze the overall sequence and structural differences between SARS-CoV and SARS-CoV-2, the sequences of SARS-CoV and SARS-CoV-2 studied in this work are aligned using clustal omega [24]. The result is shown in the figure 4. The positions of those mutations are mapped to SARS-CoV-2 structure as labeled in in four colors on a single chain of S protein: Red represents residues which are mutated to be more negative; Blue represents residues which are mutated to be more positive; yellow represents residues which are mutated from polar to hydrophobic; cyan represents residues which are mutated from hydrophobic to polar. It is found that most of the mutations distribute on the surface of the S protein. We observed that the mutations in the RBD region locate close to the interface by facing to the ACE2. This observation indicates that the mechanisms of S protein binding to ACE2 between SARS-CoV and SARS-CoV-2 may be quite different. Therefore, we performed comprehensive analysis of the binding interfaces to investigate their different binding mechanisms. Furthermore, it is obvious that some mutations are located on the hinge, which links the RBD and other regions of the S protein, as shown in the green circle of figure 5. It suggests that the flexibility of the RBD may also be different between those two viruses, which might open an avenue for our future research on coronaviruses.

Figure 4. Genome sequence alignment of SARS-CoV and SARS-CoV-2 S protein RBDs

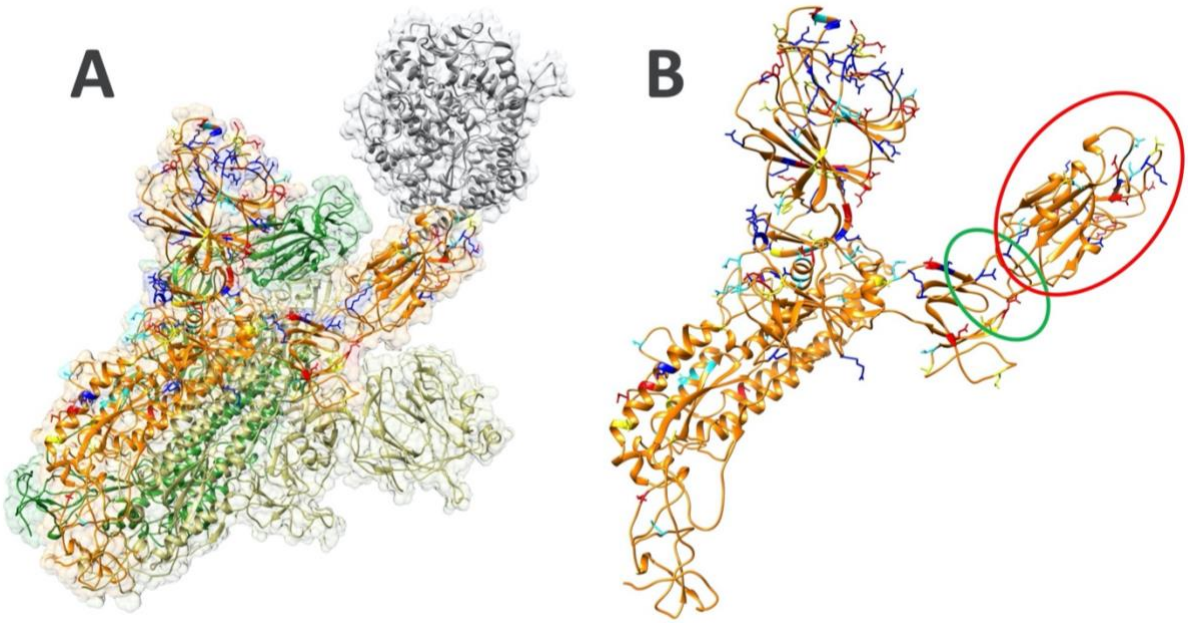

Figure 5. Structure of SARS-CoV-2 S proteins and ACE2 binding domain. (A) The structure of S protein trimer binding with ACE2 binding domain. ACE2 is shown in grey color. The three S protein monomers are represented in yellow, orange and green colors, respectively. The mutations from SARS-CoV to SARS-CoV-2 in this study are labeled in four colors on a single chain of S protein: Red represents residues which are mutated to be more negative; Blue represents residues which are mutated to be more positive; yellow represents residues which are mutated from polar to hydrophobic; cyan represents residues which are mutated from hydrophobic to polar. (B) Structure of a single S protein monomer. The RBD shown in red circle is flipping out to reach ACE2. The green circle region highlights the hinge between RBD and the rest of S protein.

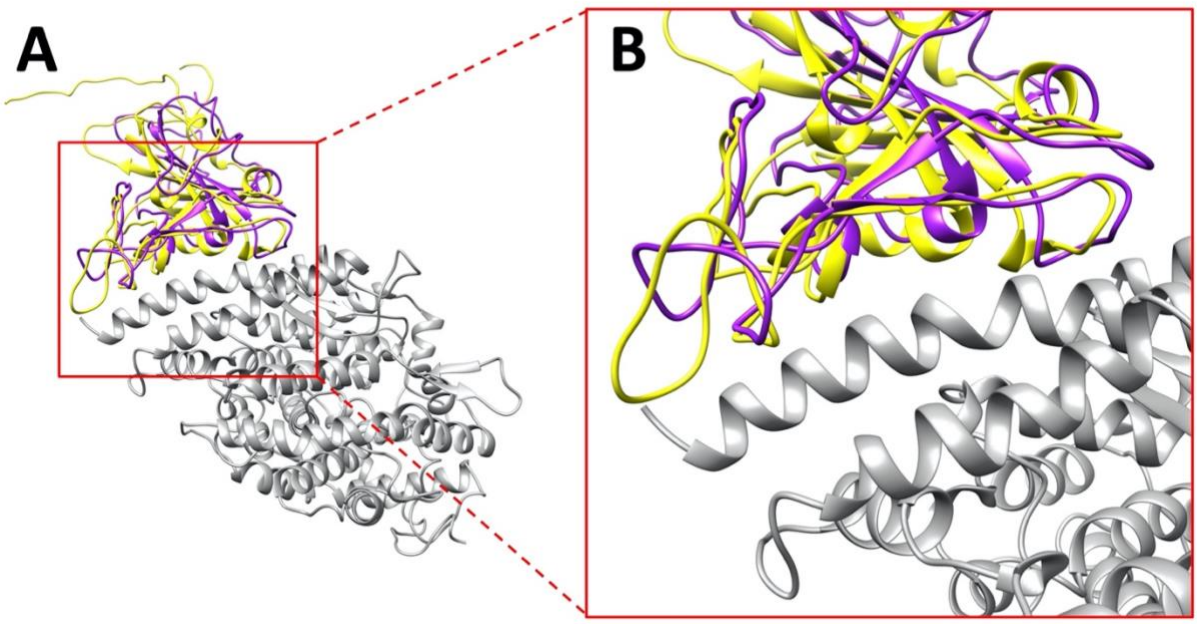

Figure 6. Complex structures of S protein RBDs and ACE2 protein. (A) SARS-CoV S protein RBD (purple) and SARS-CoV-2 S protein RBD (yellow) bind with human ACE2 binding domain (gray); (B) A closeup view of figure 2A, the interface area of SARS-CoV S protein RBD (purple) and SARS-CoV-2 S protein RBD (yellow) with human ACE2 binding domain (gray).

4.2 Electrostatic Surfaces And Field Lines

We compared the structure of the S protein RBDs of SARS-CoV with the same part of SARS-CoV-2. As shown in figure 2A, SARS-CoV S protein RBD (purple) and SARS-CoV-2 S protein RBD (yellow) are aligned based on their structures, while human ACE2 binding domain (gray) are bound with S protein RBDs. The overall RBDs structures of SARS-CoV and SARS-CoV-2 S proteins are very similar, with the RMSD [25] of 0.965 Å, but some differences can still be noticed in several loops of the RBDs

(figure 2B), which is due to two factors: 1. The high flexibility of the loops; 2. The amino acid differences between SARS-CoV and SARS-CoV-2. The variation of binding mechanisms between the two viruses could be caused by the differential residues rather than the whole structures.

4.3 Electrostatic Potential

To study the electrostatic features, DelPhi is utilized to calculate the electrostatic surfaces of the S protein RBDs and ACE2. The charge distribution on SARS-CoV S protein RBD is rendered by Chimera, with a color scale from -1.0 to 1.0 kT/Å. The charge distribution on SARS-CoV-2 S protein RBD is rendered by Chimera, with a color scale from -1.0 to 1.0 kT/Å. Negatively and positively charged areas are colored in red and blue, respectively. The electrostatic surfaces shown that the binding interface of ACE2 is dominantly negative, while the binding interfaces of S protein RBDs are all dominantly positive.

The difference of the electrostatic potential (which are generated from DelPhi) between SARS-CoV and SARS-CoV-2 S protein RBDs was calculated and mapped on the surface of SARS-CoV-2, as shown in figure 7C. From the presentation in figure 7C, an area of positive charge is observed, which also shows that SARS-CoV-2 S protein RBD is more attractive than the SARS-CoV to ACE2, since ACE2 has an overall negative charged surface, as shown in figure 7D. Therefore, we expect the SARS-CoV-2 S protein RBD may form more non-covalent bonds with ACE2, such as hydrogen bonds and salt bridges. In the later sections we demonstrate that besides salt bridge residue pairs, the

SARS-CoV-2 utilizes a cluster of residues to interact with ACE2, which is more robust than individual salt bridges.

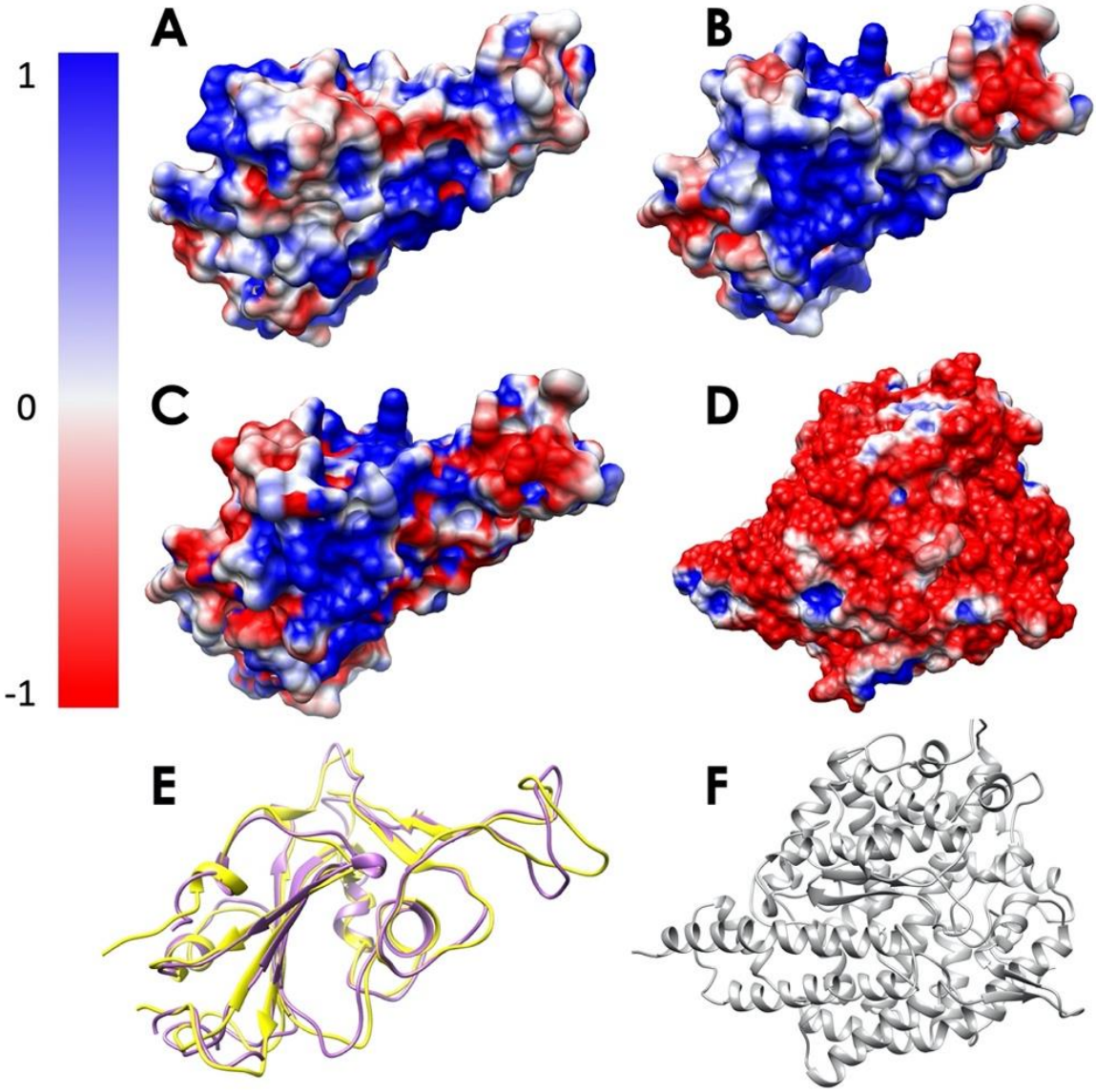

Figure 7. Electrostatic surfaces of SARS-CoV S protein RBD, SARS-CoV-2 S protein RBD and ACE2 RBD. (A) Electrostatic surface of SARS-CoV S protein RBD; (B) Electrostatic surface of SARS-CoV-2 S protein RBD; (C) Electrostatic difference between SARS- CoV

and SARS-CoV-2 S protein RBD, by subtracting electrostatic values of SARS-CoV-2 by SARS-CoV, and mapped the values on the surface of SARS-CoV-2; (D) Electrostatic surface of human ACE2 RBD; (E) Structure comparison of SARS-CoV S protein RBD and SARS-CoV-2 S protein RBD, colored with purple and yellow, respectively; (F) The structure of human ACE2 binding domain, colored with gray.

4.4 Electric Filed Lines

Electric field lines that surround the binding interfaces are calculated using Delphi. To better visualize the field lines between interfaces and show its interaction area with a clear representation, S protein RBDs were separated from ACE2 by 20Å (figure 8). In figure 8, densities of field lines represent the strengths of interactions. Higher density indicates stronger interaction in the region.

Shown in figure 8A and 8B, we see the similarity in field lines of SARS-CoV and SARS-CoV-2 complex structures. In those two panels, the field lines that connect S proteins and ACE2 are clearly shown with high densities all around the surfaces. This fact shows that both SARS-CoV and SARS-CoV-2 S protein RBDs have strong attractive binding forces to ACE2 protein, and the further discussions on binding forces are in the later section of electrostatic forces (figure 9).

However, there are still several remarkable differences if we take a closer look at the interface areas, as shown in figure 8C-H. The first difference is the distribution dissimilarity of electric field line related residues. The residues forming salt bridges are distributed differently in SARS-CoV (figure 8C and 8E) compared to SARS-CoV-2

(figure 8D and 8F). The salt bridge residues of SARS-CoV are clearly shown in the front view of the complex (figure 8C). In contrast, the salt bridge residues of SARS-CoV-2 are mainly observed in the back view of the complex (figure 8F). This indicates that the salt bridge residues are distributed on the opposite sides of the S protein RBDs for SARS-CoV and SARS-CoV-2. Besides the front and back views, we also rendered the bottom views of SARS-CoV(figure 8G) and SARS-CoV(figure 8H) with colorful patches, where green patches represent salt bridge residues, purple patches represent hydrogen bonds, and yellow patches represent special regions that form high-density field lines but do not belong to salt bridges nor hydrogen bonds. By comparing those patches, SARS-CoV-2(figure 4H) has a bigger and more joint hydrogen bond distribution (purple patches) than SARS-CoV(figure 8G); salt bridges(green patches) in SARS-CoV-2(figure 8H) are more concentrated in the distribution, while salt bridges(green patches) in SARS-CoV-2(figure 8G) are distributed more separately; and SARS-CoV-2(figure 8H) has 5 major special regions (yellow patches), while SARS-CoV(figure 8G) has only 2 major special regions (yellow patches). The second difference is about density. Figures 8G and 8H have the same representation setting of field lines with the gradient values of 2.39 kT/(eÅ), it is obvious that SARS-CoV-2(figure 8H) has several higher-density field line regions than SARS-CoV(figure 8G). Since the higher density indicates the stronger interactions, SARS-CoV-2 definitely has stronger attractive interaction than SARS-CoV.

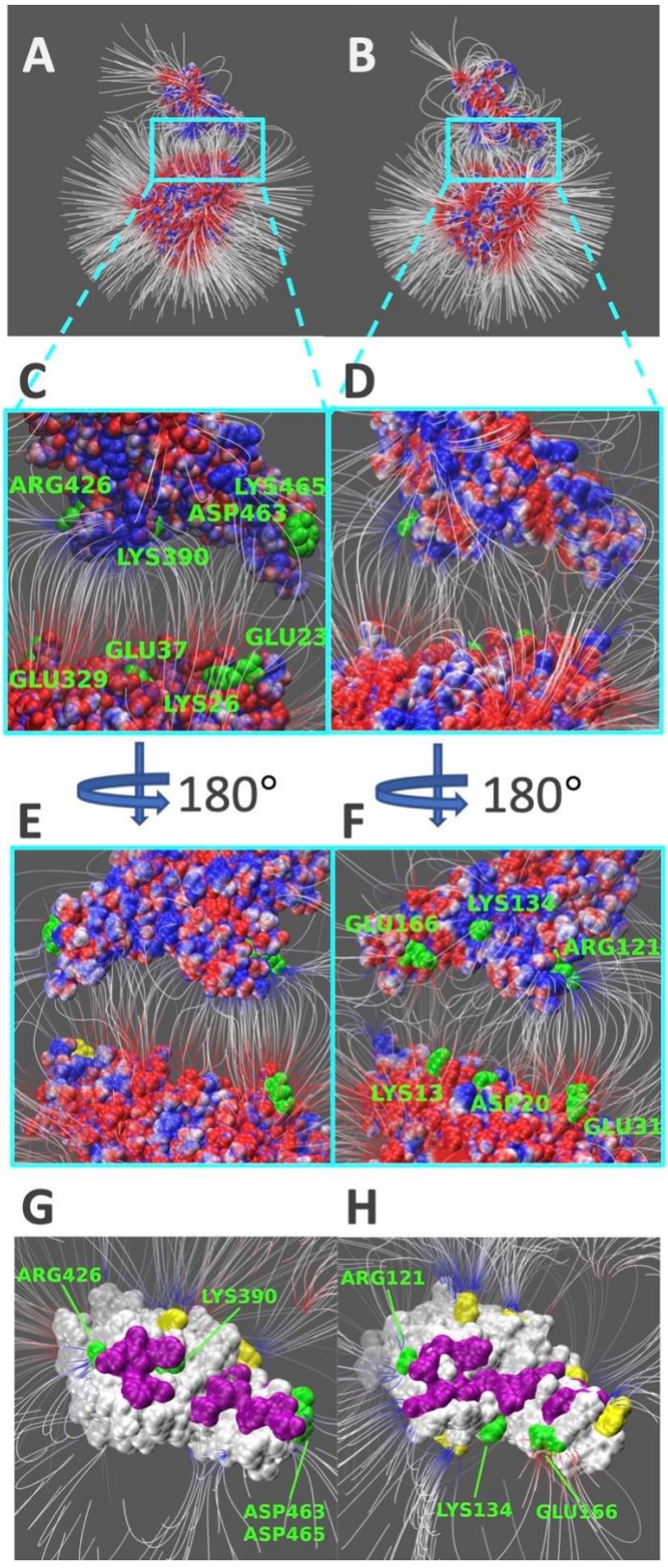
A
B
C
D
ARG426
LYS465
ASP463
LYS390
GLU37
GLU23
GLU329
LYS26
180°
180°
E
F
LYS134
GLU166
ARG121
LYS13
ASP20
GLU31
G
H
ARG426
LYS390
ASP463
ASP465
ARG121
LYS134
GLU166

Figure 8. Electric filed lines at the interfaces of S protein RBDs and ACE2. (A) An overview of electric filed lines between SARS-CoV S protein RBD and ACE2 RBD; (B) An overview of electric field lines between SARS-CoV-2 S protein RBD and ACE2 RBD; (C) A closeup view of figure 8A, with marked key residues that form salt bridges (ARG426-GLU329, LYS390-GLU37, ASP463-LYS26, LYS465-GLU23); (D) A closeup view of figure 8B; (E) The back view of 4C; (F) The back view of 4D, with marked key residues that form salt bridges (GLU166-LYS13, LYS134-ASP20, ARG121-GLU31). The electrostatic surfaces and field lines are rendered by Visual Molecular Dynamics (VMD) [35] with a color scale from -1.0 to 1.0 kT/Å. To present field lines in the clearest way, we adjusted gradient values to 2.39 kT/(eÅ), in (A-B, E-F), and 2.08 in (C-D, G-H). Negatively and positively charged areas are colored in red and blue, respectively; (G) The bottom view of SARS-CoV S protein RBD, salt bridge involved residues are marked green, hydrogen bond involved residues are marked purple, and yellow regions are special residues that have high density of field lines but they are not involved in salt bridges nor hydrogen bonds; (H) The bottom view of SARS-CoV-2 S protein RBD, salt bridge involved residues are marked green, hydrogen bond involved residues are marked purple, and yellow regions are special residues that have high density of field lines but they are not involved in salt bridges nor hydrogen bonds.

4.5 Electrostatic Forces

Electrostatic forces of SARS-CoV and SARS-CoV-2 S protein RBDs at different distances with human ACE2 binding domain are calculated by DelPhiForce to model the binding forces when S proteins bind to ACE2 (figure 9). Arrows in figure 9A and 9B are shown to visualize the net forces between proteins by shifting the S proteins away from ACE2 by a distance ranging from 5 Å to 40 Å with a step of 2 Å. The directions of arrows represent the force directions at different distances. To better visualize the directions of the net forces, the magnitudes of net forces are normalized to be of the same size at different distances, which means that the size of arrows do not represent the force strength. Comparing figure 9A and 9B, as we expected, the overall binding forces are all shown to be attractive for both viruses. As for the force directions, only slight differences were found at different distances. From figures 9C and 9D that represent the force on every residue in the RBDs at a distance of 5 Å, with the arrow sizes representing the force magnitudes, a conclusion can be drawn that SARS-CoV-2 has quite a different force distribution on individual residues with SARS-CoV. A closeup view of the difference is noticed by comparing figure 9E and 9F. The salt bridge involved residues are labeled in the figure 9E and 5F, which confirms that the salt bridge residues do provide significant attractive forces in the interaction process.

The directions of the net forces are shown in figure 9A and 9B, while the magnitudes of the net forces are shown in figure 10. The magnitudes of the net forces on the directions of mass center lines, x, y, z directions are shown in figure 11. For both SARS-CoV-2 and SARS-CoV, the net forces are enhanced when the distance is

decreased from 40 Å to 5 Å, which is expected to see because the main force is a type of electrostatic force. Based on the Coulomb's law, when the charges on the RBD interfaces get closer to the charged residues on the interface of ACE2, the force increases significantly. Besides, by comparing figure 10A and 10B, the net force of SARS-CoV S protein RBD is actually stronger than that of SARS-CoV-2, which might be due to the charge distribution differences between those two binding domains. Even though the attractive force is weaker, the SARS-CoV-2 may still be easier to bind with ACE2. Because there are sequence differences at the hinge which connects the RBD and other parts of the S protein. Such sequence differences may make the RBD more flexible and easier to open and bind with ACE2. We will study the flexibilities of the RBDs from SARS-CoV and SARS-CoV-2 in future work.

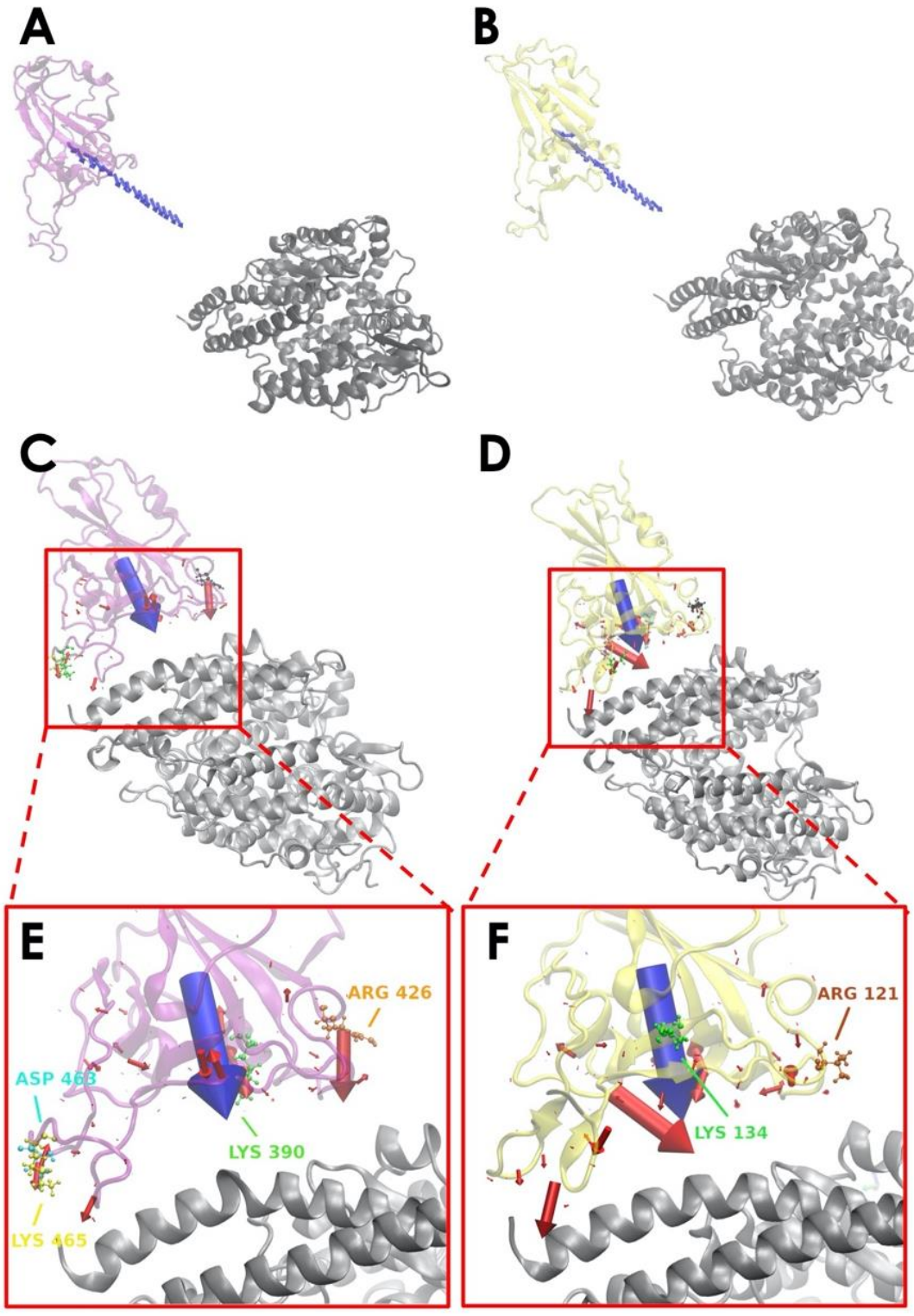
A
B
C
D
E
F
ARG 426
ASP 463
LYS 390
LYS 465
ARG 121
LYS 134

Figure 9. Electrostatic forces of SARS-CoV S protein RBD and SARS-CoV-2 S protein RBD at variable distances with human ACE2 binding domain. (A) Electrostatic forces of SARS-CoV S protein RBD with human ACE2 RBD at a different distance, from 5 Å to 40 Å with a step of 2 Å, where blue arrows show the net force directions. (B) Electrostatic forces of SARS-CoV-2 S protein RBD with human ACE2 RBD at different distance, from 5 Å to 40 Å with a step of 2 Å, where blue arrows show the net force directions. (C) Electrostatic forces of SARS-CoV S protein RBD with human ACE2 RBD at a distance 5 Å, where the blue arrow shows the total net force between S protein and ACE2, and red arrows represent individual forces between single residues of S proteins in interface area and ACE2. (D) Electrostatic forces of SARS-CoV-2 S protein RBD with human ACE2 RBD at a distance of 5 Å, where the blue arrow shows the total net force between S protein and ACE2, and red arrows represent individual forces between single residues of S proteins in interface area and ACE2. (E) A closeup view of (C) in the interface. (F) A closeup view of (D) in the interface.

Salt-bridge-involved residues on SARS-CoV RBD are marked with different colors in figure 9C and labeled with their residue types and sequence numbers. As shown in figure 9E, for SARS-CoV RBD, four salt bridge residues, ARG426 (orange), ASP468 (cyan), LYS390 (green) and LYS465 (yellow), are labeled. Among them, ARG426 provides a strong attractive force to ACE2 while LYS465 results in a repulsive force to ACE2, due to the fact that the LYS465 faces a positively charged region at the ACE2 interface. However, a negatively charged residue, ASP463 (yellow), is located in a neighborhood which results in attractive force that overcome the repulsive force from

LYS465. Also, as shown in figure3, the h-bond formed by LYS465 has a 52.32% appearance occupancy and ASP463 has 48.50%. These results indicate that even though LYS465 has repulsive force to ACE2, still the nearby region has attractive force to ACE2. Note that this calculation is based on the structure of S protein separated from ACE2 by 5 Å. When S protein binds to ACE2, the sidechain of LYS465 on S protein changes the configuration to form a salt bridge with GLU23 on ACE2, which is demonstrated in the later section of salt bridges.

Salt bridge residues of SARS-CoV-2 RBD are marked with different colors in figure 9D and labeled with their residue types and sequence numbers in figure 9F. For SARS-CoV-2 RBD, two strong salt bridge residues ARG121(brown) and LYS134(green) are observed. As the red arrows shown in figure 9F all have the dirction pointing to ACE2, we can conclude that those two residues are all attractive to ACE2, among which LYS134 has a stronger attractive force strength based on the comparison of arrow sizes.

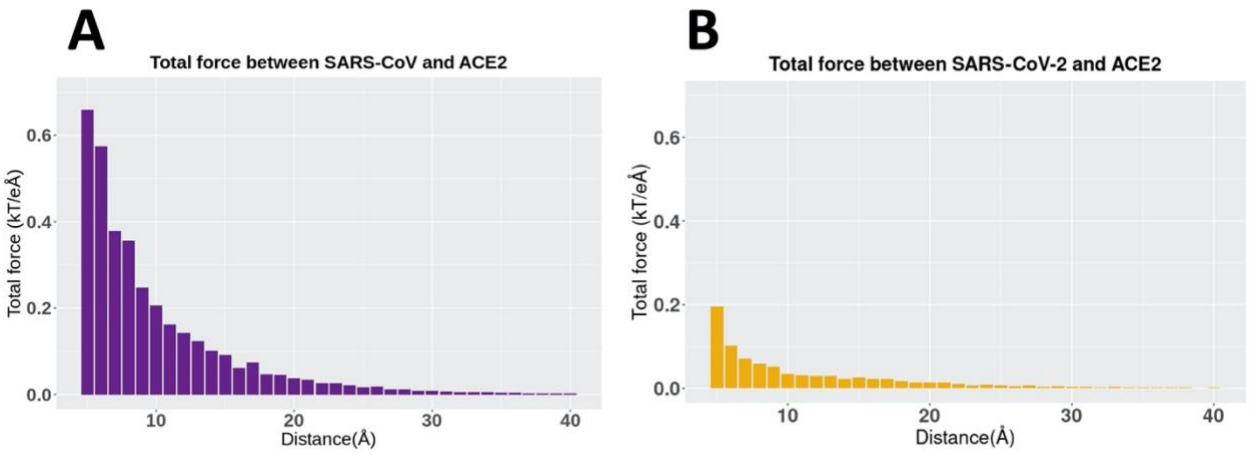

Figure 10. The trends of total electrostatic forces between S protein RBDs and human ACE2 RBD at different distances from 5 Å to 40 Å. (A) Total electrostatic force between

SARS-CoV S protein RBD and human ACE2 binding domain. (B) Total electrostatic force between SARS-CoV-2 S protein RBD with human ACE2 RBD.

In terms of the total electrostatic forces between S protein RBDs and human ACE2 RBD, it should be noticed that SARS-CoV-2 has relatively lower values than SARS-CoV, especially when the distance has a smaller value. Note here that in this comparison, we only take the force strength into consideration rather than the directions of forces, and directions can also play an essential role in the comparison.

4.6 Salt Bridges

Salt bridges at the interfaces of S protein RBDs and ACE2 are analyzed based on the MD simulation results and shown in figure 3. Four pairs of salt bridges have been observed between SARS-CoV RBD and ACE2 RBD, comparing to two pairs between SARS-CoV-2 RBD and ACE2 RBD. Among the four pairs of SARS-CoV salt bridges, as shown in figure 3(A-D), three of them (ASP463 – LYS26, GLU23 – LYS465, GLU329 – ARG426) are strong salt bridge pairs during 50 – 100 ns time duration, as the distance is all below 4Å , which is the selected cut-off value for salt bridge calculations; while the fourth salt bridge (GLU37 – LYS390) performs interestingly: at the beginning, GLU37 and LYS390 keep a distance of about 6 Å, from 73 ns, it suddenly becomes the strongest pair with the shortest distance (about 2.75Å) among those four observations. This change is due to the side chain flexibilities.

Speaking of the two observations of SARS-CoV-2 salt bridges, as shown in figure 3(E-F), they are all strong pairs (ASP20 – LYS134 and GLU311 – ARG121) during the

whole 50ns. While there is a special pair (GLU166 – LYS13) which has been observed that is included in figure 11 of supplementary material. This special pair has a strong salt bridge feature during the first 30ns of the whole 100ns simulation, while those two residues apart from each other to a distance over 7.5Å after 30ns. Since we only took the last 50ns for our data analysis, this special pair is not considered as a salt bridge pair in this work. However, we can still draw a conclusion that some residues involved in the binding process between SARS-CoV-2 and ACE2 are flexible.

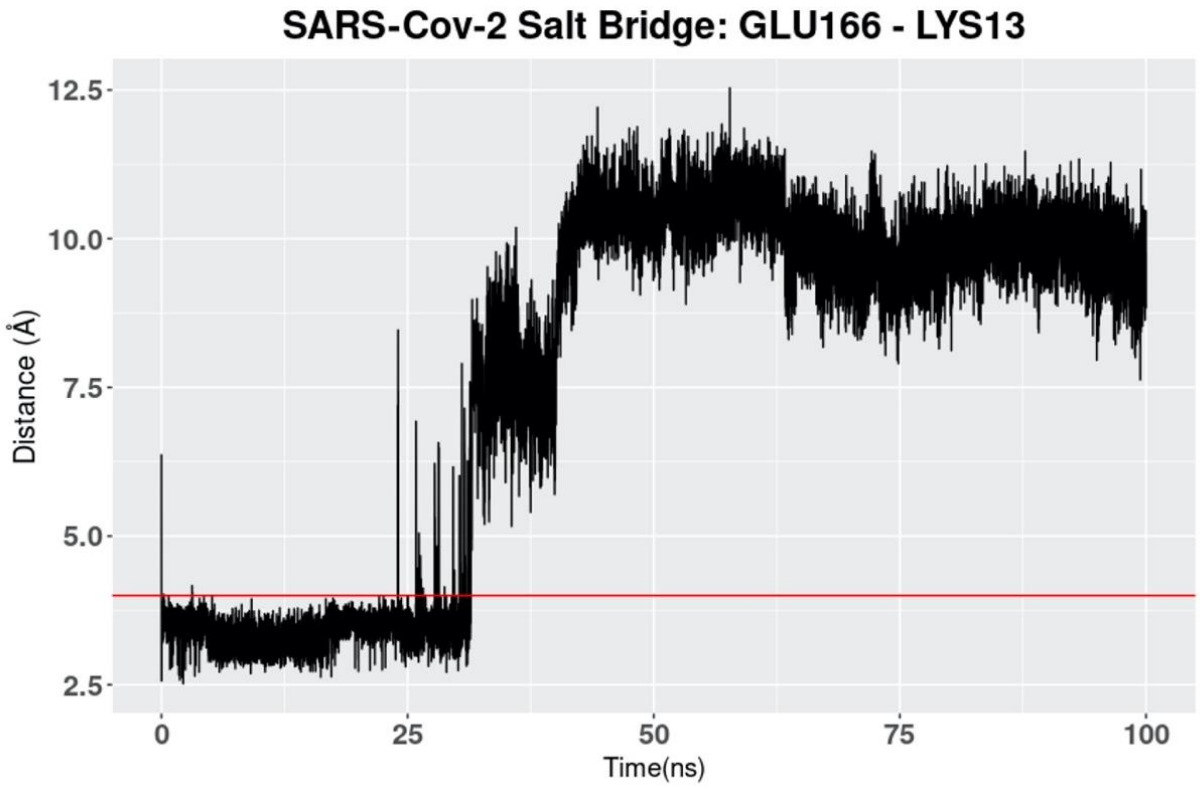

Figure 11. A special salt bridge (GLU166 – LYS13) in SARS-CoV-2/hACE2 complex

4.7 Hydrogen Bonds

Hydrogen bonds at the interfaces of S protein RBDs and ACE2 are also calculated based on the MD simulations as shown in Figure 12. By comparing Figures 12A and 12B, the average numbers of hydrogen bonds at the same time between SARS-CoV and SARS-CoV-2 S protein RBDs and ACE2 are very similar, with the mean values of 25.90 and 21.85, respectively (marked as the red lines). While by comparing Figures 12C and 12D, the details of hydrogen bonds with the occupancies above 30% are quite different. The first difference to notice is the highest occupancy of each complex structure, where we find that SARS-CoV-2 has the highest occupancy of 98.98%, compared to 90.91% for SARS-CoV. Besides, if you pick the 90% as a cutoff value, SARS-CoV-2 has 3 pairs, compared to only 1 pair in SARS-CoV analysis, which means SARS-CoV-2 has more extremely high occupancy hydrogen bonds than SARS-CoV. This fact is also an evidence to show the more robust binding strategy of SARS-CoV-2. And it might be another reason why the COVID-19 is spreading easier and faster than SARS in 2003.

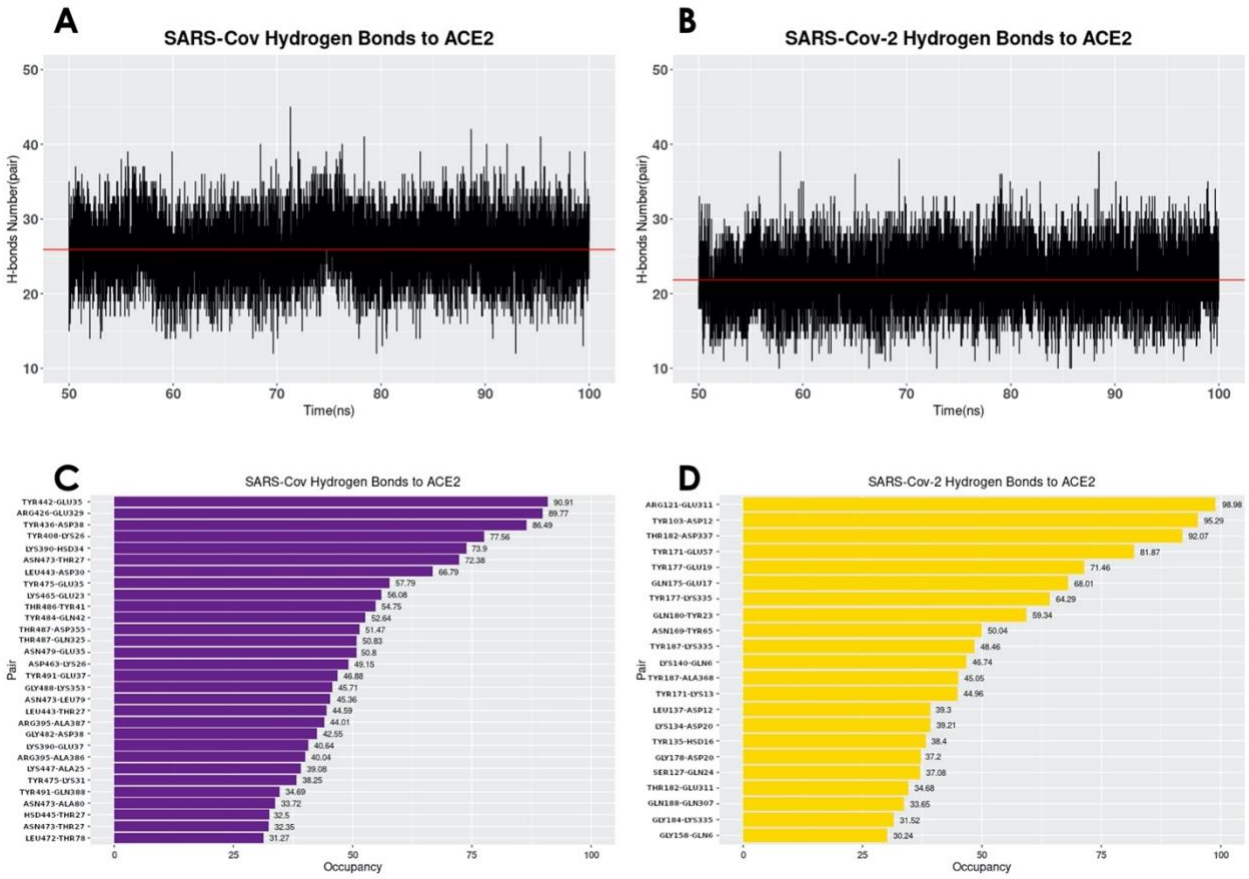

Figure 12. Hydrogen bonds at interfaces of S protein RBDs and ACE2 RBD with their occupancies. (A) Number of hydrogen bonds between SARS-CoV S protein RBD and ACE2 binding domain during the MD simulation. The average number of hydrogen bonds is shown as a red line, with the value of 25.90 pairs. (B) Number of hydrogen bonds between SARS-CoV-2 S protein RBD and ACE2 binding domain in the MD simulation. The average number of hydrogen bonds is 21.85, shown as the red line; (C) Occupancies of 30 pairs of hydrogen bonds (with a cutoff value of 30%) forming at the interface of SARS-CoV S protein RBD and ACE2 protein binding domain. (D) Occupancies of 22 pairs of hydrogen bonds forming at the interface of SARS-CoV-2 S

protein RBD and ACE2 protein binding domain. For each hydrogen bond pair, the residue on the left is from S protein RBD while that on the right is from ACE2.

4.8 Key Residues involved in Salt Bridges and Hydrogen Bonds

The residues involved in salt bridges and hydrogen bonds are identified as the key residues which may significantly contribute to the binding affinity. Figure 13 illustrates the key residues involved in salt bridges observed. As shown in figure 13A and 13B, key residues are mostly around the edges of the binding interfaces rather than the center of the interfaces, and most of the key residues are positive in S protein RBD and negative in ACE2, except for the pair ASP463 – LYS26 in SARS-CoV S protein RBD and the special pair GLU166 – LYS13 in SARS-CoV-2 S protein RBD. Such salt bridges also play significant roles in binding forces and electric field line distributions.

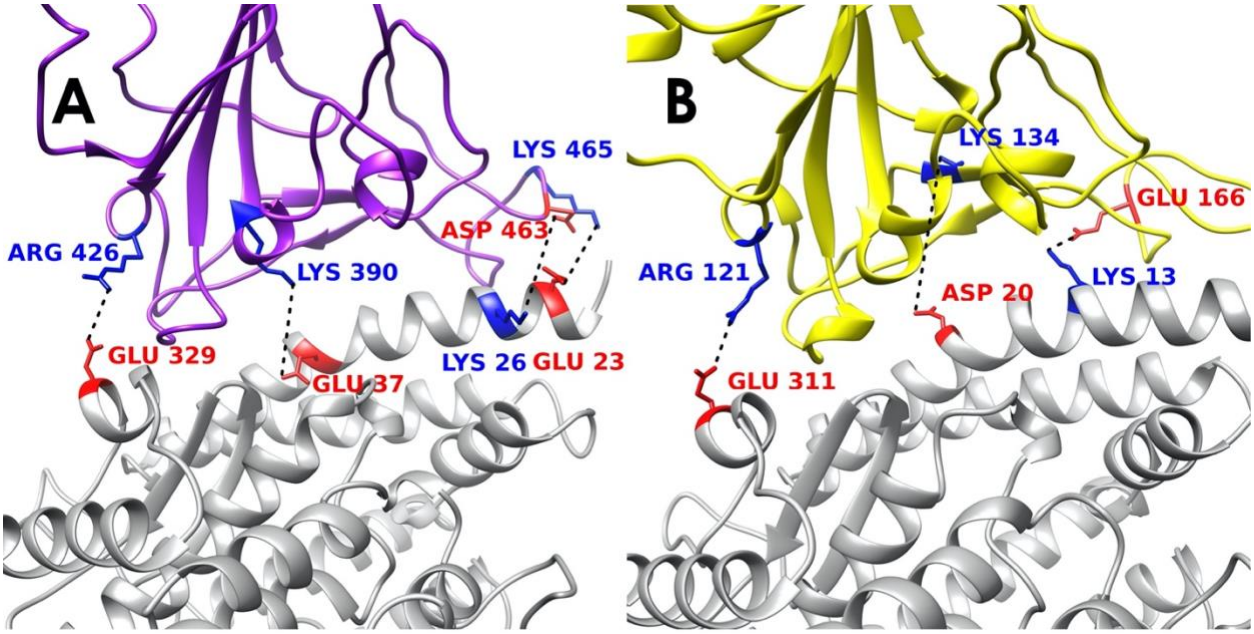

Figure 13. Structural demonstration of key residues that form salt bridges in the interface on both virus S protein RBDs and ACE2 RBD. (A) SARS-CoV S protein RBD (purple) and human ACE2 RBD (gray). (B) SARS-CoV-2 S protein RBD (yellow) and human ACE2 RBD (gray). Blue stands for positively charged key residues while red represents the negatively charged key residues.

5. RESEARCH PLAN

5.1 Binding energy and folding energy calculations using DelPhiPKa

5.1.1Binding energy calculations methods

For the binding energy calculation, we involved two methods, which are DelPhiPKa and MM/PBSA[39]. To calculate binding energy using DelPhiPKa, the following equation was used:

$$\Delta N(pH_{binding}) = 2.3RT \int_{pH_i}^{pH_f} (Q_t(pH) - Q_n(pH) - Q_r(pH))d(pH) \qquad (4)$$

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

5.1.2 Folding energy calculation methods

I plan to use DelPhiPKa [37, 38] to calculate pKa values of S proteins and hACE2 RBDs, with the pH ranging from 0 to 14 with the step size of calculation at 0.5. The net charges of proteins at the unfolded state were calculated using this equation:

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

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

$$\Delta N(pH_{folding}) = 2.3RT \int_{pH_i}^{pH_f} (Q_f(pH) - Q_u(pH)d(pH)) \qquad (3)$$

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

5.2 Electrostatic Features Of Full Trimmer Structures

Since I already did the electrostatic features of the RBDs of S proteins, it is curious for me to see how the full trimmer structure will be like. So in the proposed work, I will complete the calculations of electrostatic potential of the full trimmer structures as well as the field lines between hACE2 and the full structures.

5.3 Preliminary Results

To study the electrostatic features, DelPhi was utilized to calculate the electrostatic potential on surfaces of the S protein trimmer (full structure) and hACE2 RBD. The electrostatic potential distribution on SARS-CoV S protein trimmer structure is showed in figure 14BEH, which were rendered by Chimera with a color scale from -1.0 to 1.0 kT/e. The charge distribution on SARS-CoV-2 S protein trimmer structure is shown in figure 14CFI, which were rendered by Chimera with a color scale from -1.0 to 1.0 kT/e as well, for the comparison. Negatively and positively charged areas are colored in red and blue, respectively.

By comparing the electrostatic potential on surfaces of two trimmer structures, it is obvious that the charge distribution of SARS-CoV and SARS-CoV-2 S proteins are different. From the top view (figure 14A-C) and the bottom view (figure 14G-I), we

noticed that SARS-CoV has slightly more positively charged area (blue), compared to SARS-CoV-2. It indicates that the SARS-CoV may attract the hACE2 more easily, since the hACE2 binding interface is overall negatively charged (movie 3). Such finding supports the previous studies of our research group[26, 27]. The electrostatic distribution differences observed from front views (figure 14D-F) of the S proteins demonstrate that the electrostatic features may have impacts on the stabilities of the trimmers. Here it was not investigated several details about the binding stabilities among monomers in an S protein, due to the scope of this work that mainly focusses on the binding between S protein and hACE2. The electrostatic distributions on S protein RBDs show that the SARS-CoV RBD is more positive, which is consistent with the top view (figure 14BC). The bottom of the SARS-CoV (Figure 14EH) has more positive potential than SARS-CoV-2 (figure 14FI), which may help the S protein of SARS-CoV binding to the membrane of the virus more easily, since the membrane is negatively charged.

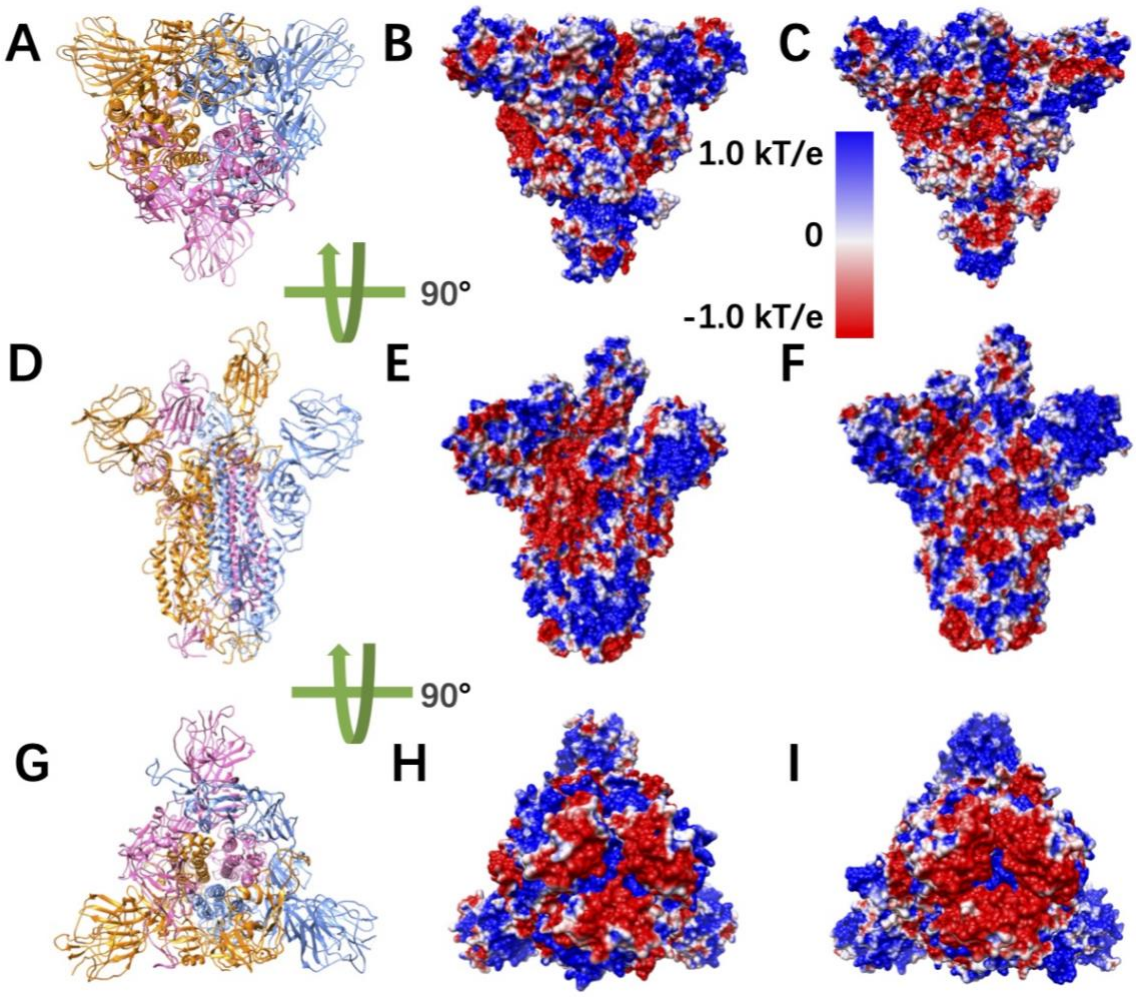

Figure 14. Electrostatic potential on surfaces of SARS-CoV and SARS-CoV-2 S proteins. (A) Top view of S protein; (B,C) Top views of electrostatic potential on surfaces of SARS-CoV and SARS-CoV-2 S protein, respectively; (D) Front view of S protein; (E,F) Front views of electrostatic potential on surfaces of SARS-CoV and SARS-CoV-2 S protein, respectively; (G) Bottom view of S protein; (H,I) Bottom views of electrostatic potential on surfaces of SARS-CoV and SARS-CoV-2 S protein, respectively. Negatively

and positively charged areas are colored in red and blue respectively, with the color scale from -1.0 to 1.0 kT/e.

REFERENCES

1. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S.C. and Di Napoli, R., *Features, evaluation, and treatment of coronavirus (COVID-19)*. Statpearls [internet], 2021.
2. Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B. and Huang, C.-L., *A pneumonia outbreak associated with a new coronavirus of probable bat origin.* nature, 2020. **579**(7798): p. 270-273.
3. Van Der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J. and Berkhout, B., *Identification of a new human coronavirus.* Nature medicine, 2004. **10**(4): p. 368-373.
4. Hadi, A., Werge, M., Kristiansen, K.T., Pedersen, U.G., Karstensen, J.G., Novovic, S. and Gluud, L.L., *Coronavirus disease-19 (COVID-19) associated with severe acute pancreatitis: case report on three family members.* Pancreatology, 2020. **20**(4): p. 665-667.
5. Liu, D.X., Liang, J.Q. and Fung, T.S., *Human coronavirus-229E,-OC43,-NL63, and-HKU1.* Reference Module in Life Sciences, 2020.
6. Loeffelholz, M.J. and Tang, Y.-W., *Laboratory diagnosis of emerging human coronavirus infections–the state of the art.* Emerging microbes & infections, 2020. **9**(1): p. 747-756.
7. Fehr, A.R. and Perlman, S., *Coronaviruses: an overview of their replication and pathogenesis*, in *Coronaviruses*. 2015, Springer. p. 1-23.

8. van der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J. and Berkhout, B., *Identification of a new human coronavirus.* Nat Med, 2004. **10**(4): p. 368-73.

9. Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D. and Fouchier, R.A., *Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia.* N Engl J Med, 2012. **367**(19): p. 1814-20.

10. Li, X., Geng, M., Peng, Y., Meng, L. and Lu, S., *Molecular immune pathogenesis and diagnosis of COVID-19.* Journal of Pharmaceutical Analysis, 2020.

11. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H. and Nitsche, A., *SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor.* Cell, 2020.

12. Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y. and Zhou, Q., *Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2.* Science, 2020. **367**(6485): p. 1444-1448.

13. Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y. and Yuen, K.-Y., *Structural and functional basis of SARS-CoV-2 entry by using human ACE2.* Cell, 2020.

14. Zhang, H., Penninger, J.M., Li, Y., Zhong, N. and Slutsky, A.S., *Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target.* Intensive care medicine, 2020. **46**(4): p. 586-590.

15. Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., Zhou, Y. and Du, L., *Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine.* Cellular & molecular immunology, 2020. **17**(6): p. 613-620.

16. Brielle, E.S., Schneidman-Duhovny, D. and Linial, M., *The SARS-CoV-2 exerts a distinctive strategy for interacting with the ACE2 human receptor.* Viruses, 2020. **12**(5): p. 497.

17. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q. and Zhang, L., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor.* Nature, 2020. **581**(7807): p. 215-220.

18. Song, W., Gui, M., Wang, X. and Xiang, Y., *Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2.* PLoS pathogens, 2018. **14**(8): p. e1007236.

19. Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A. and Li, F., *Structural basis of receptor recognition by SARS-CoV-2.* Nature, 2020. **581**(7807): p. 221-224.

20. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M. and Bordoli, L., *SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information.* Nucleic acids research, 2014. **42**(W1): p. W252-W258.

21. Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H. and Pei, Y.-Y., *A new coronavirus associated with human respiratory disease in China.* Nature, 2020. **579**(7798): p. 265-269.

22. Humphrey, W., Dalke, A. and Schulten, K., *VMD: visual molecular dynamics.* Journal of molecular graphics, 1996. **14**(1): p. 33-38.

23. Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A., *PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations.* Nucleic acids research, 2004. **32**(suppl_2): p. W665-W667.

24. Omega, C., *accurate alignment of very large numbers of sequences; F Sievers, DG Higgins.* Methods in Molecular Biology: p. 105-116.

25. Maiorov, V.N. and Crippen, G.M., *Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins.* 1994.

CURRICULUM VITA

Yixin Xie (yixin.xie@outlook.com) graduated in 2014 from Kennesaw State University, Marietta, GA and got her Bachelor of Science in Computer Science with a minor in Technical Communication. During the bachelor study, she received International Student Honor Award (top 10%), Outstanding Undergraduate Student Award (top 5%) and was recorded in Computer Engineering School Dean's List.

After her graduation in 2014, she moved to Australia for Graduate study. Motivated by her favorite movie 歩いても 歩いても(keep walking), she decided to study film production and in 2016 she got a Master Degree in Media Arts and Production at University of Technology, Sydney.

During 2016 – 2018, she worked as Digital Media Producer in Sydney, Australia and then became an Assistant Documentary Producer at China Central Television (CCTV-1) in Beijing, China. She enjoyed media industry until she got a chance to become Computer Vision Engineer at Shanghai DeepBlue Artificial Intelligence Co. During the experience in AI industry, she found herself interested in Computational Studies, so based on this fact, she started her research career in Computational Science at UTEP, and has been under the supervision of Dr. Lin Li since September 2019.

Her current research interests are Computational biophysics, Molecular biology, Protein interactions, etc. Her future career goal is to become a well-known researcher in computational science and she will try her best make contributions to cancer treatment or prevention.