Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Which species does the virus like most: Binding modes study between SARS-CoV-2 S protein and ACE2 receptor

Sheng Guo a, b, 1, Jing Yang b, 1, Yu Lei a, b, Bin Liu b, Wei Zhang b, Li Zhang a, **, Zhili Zuo a, b, *

a School of Chemical Engineering, Sichuan University of Science & Engineering, 180 Xueyuan Street, Huixing Road, Zigong, Sichuan, 643000, China
b State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, 650201, China

Article info
Article history:
Received 16 November 2020
Received in revised form 17 February 2021
Accepted 1 March 2021
Available online 4 March 2021

Keywords:
SARS-CoV-2
COVID-19
Homology modeling
Molecular dynamics simulation
Binding free energy

Abstract
Background: The emergence of a novel coronavirus (SARS-CoV-2) has been spreading worldwide in 2020. Coronaviruses could mainly cause respiratory tract infections in humans and multiple system infections in many animals. The coronavirus enters the host cell through the binding of surface spike glycoprotein (S Protein) with host angiotensin-converting enzyme-II (ACE2) protein.

Methods: ACE2 sequences of various species were aligned with human ACE2, accordingly, homology models for different species were constructed. Then, S-protein-ACE2 complexes were constructed using the generated homology models. The molecular dynamics simulations and Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) were carried out to study the dynamical behavior of the generated S-ACE2 virtual complexes. Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF) and Radius of Gyration (Rg) were calculated to evaluate protein stability and compactness.

Results: The binding free energies of S protein with ACE2 from Procyon lotor and Camelus dromedarius are about equal to that of humans. By comparing the free binding energies it were possible to identify potential viral hosts that could transmit the virus to human (risk of cross-species transmission). The predication showed that, besides human beings, SARS-CoV-2 may possibly infect Procyon lotor and Camelus dromedarius as well.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

The outbreak of novel coronavirus pneumonia is a great threat to global public health. There are more than 100 million confirmed cases worldwide, with cumulative reported deaths of more than 2 million people by the end of January 2021. The virus is still spreading fast and the data is changing every day. The International Committee on Taxonomy of Viruses named the new coronavirus as Severe Acute Respiratory Syndrome coronavirus-2 (SARS-CoV-2). The World Health Organization named the disease caused by the virus as Coronavirus Disease 2019 (COVID-19) [1]. Prior to SARS-CoV-2, there have been 6 types of coronaviruses discovered that can infect humans. Including two highly lethal coronaviruses, named Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and 4 types of coronaviruses that can cause mild upper respiratory diseases, named HCoV-OC43, HCoV-229E, HCoV-NL63 and HCoV-HKU1 [2,3]. COVID-19 is extremely infectious, whose main clinical manifestations are fever, fatigue, and cough. In addition, the level of inflammatory cytokines in the body fluids of infected persons was significantly increased [4,5]. Organ damage caused by SARS-CoV-2 infection is still the main cause of severe illness [6]. So far, no specific medicine is available to treat this disease. It is difficult to control the spread of the virus. On December 11, 2020, the U.S. Food and Drug Administration (FDA) issued the first emergency use authorization (EUA) for a vaccine to prevent the coronavirus disease 2019 caused by SARS-CoV-2 in individuals 16 years of age and older, named Pfizer-BioNTech COVID-19 Vaccine. On December 18, 2020, FDA approved another dose of vaccine named Moderna COVID-19 Vaccine (https://www.fda.gov). Promising progress has been made in developing vaccines and antiviral drugs [7], but we still need to prevent infections via understanding its possible way.
of spreading via cross-species transmission. SARS-CoV-2 is a single positive-stranded RNA virus, belonging to the genus Coronavirus \( \beta \) [6]. It contains four structural proteins namely surface spike glycoprotein (S Protein), envelope protein, membrane protein, nucleocapsid protein and four non-structural proteins namely 3-chymotrypsin-like protease (3CLpro) which is the main protease (M pro) [8]. Papain-like protease (PL pro), Helicase and RNA-dependent RNA polymerase (RdRp) [9,10]. The complete sequencing of the SARS-CoV-2 genome has been completed, including untranslated regions (UTR) at both ends and a complete open reading frame (ORF) gene [9]. The genome shows that it has high sequence homology with SARS-CoV [11], but its RBD (receptor-binding domain) is obviously different.

SARS-CoV-2 and SARS-CoV share the same angiotensin-converting enzyme –II (ACE2) as cellular receptor via the spike glycoprotein [11–16]. The S protein exists in a metastable pre-fusion conformation before fusion, and undergoes a structural reorganization [17,18]. During the fusion process, the S protein is directly binds ACE2, and S2 is responsible for membrane fusion [21,22]. Then, the virus enters the host cell through the endosomal pathway to release the genome. The 3CLpro and PL pro proteases encoded by ORF1a (open reading frame genes 1a) cleave the polyproteins pp1a (polyprotein 1a) and pp1ab (polyprotein 1 ab) encoded by ORF1b (open reading frame genes 1 b) to produce replication transcription complexes, which are transcribed and replicated in the host cell to express new virus particles and complete cell infection. Finally, the cells are discharged in the form of exocytosis and participate in a new round of virus proliferation [18].

Due to the wide distribution of coronaviruses, genetic diversity, frequent recombination, and the expansion of human activities, the risk of cross-species transmission is increasing. Literature shows that SARS-CoV-2 infection in Rhesus macaques induced humoral and cellular immune responses and provided protective efficacy against SARS–CoV-2 [23,24]. In the event of an outbreak of SARS-CoV-2, will animals exposed to the virus be infected, will animals infected with the virus be lethal, and will these animals become new intermediate hosts to transmit the virus to humans? All of these questions need to be explored.

In this study, considering the important role of an ACE2 in the process of virus transmission, we simulated the binding modes of SARS-CoV-2 S protein with ACE2 from various species to compare the binding free energies to speculate the susceptible species. ACE2 sequences of various species (https://www.ncbi.nlm.nih.gov/) were aligned with human ACE2. The sequence identity from species is more than 60% and sequence similarity is more than 80%. ACE2 proteins from different species were complexed with SARS-CoV-2 S protein and their dynamical behavior was studied using molecular dynamics simulations (MDS) by GROMACS 5.0.2 [25]. RMSD, RMSF and Rg were calculated to evaluate protein stability and compactness. Then, binding free energy between the two proteins was calculated. By comparing the binding free energy of each species, it was possible to identify potential viral hosts that could be infected, consequently, could transmit the virus to human (risk of cross-species transmission). Energy decomposition and binding modes were used for further analysis.

2. Materials and methods

2.1. Protein structures built by homology modeling

The protein sequences of ACE2 of other species were downloaded from the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/), including *Rhinolophus ferrumequinum* (Bat), *Bos taurus* (Cattle), *Camelus dromedarius* (Dromedary), *Canis lupus familiaris* (Dog), *Felis catus* (Cat), *Macaca mulatta* (Rhesus macaque), *Mus musculus* (Mouse), *Paguma larvata* (Beaver), *Pongo abelii* (Chimpanzee), *Procyon lotor* (Raccoon), *Rattus norvegicus* (Rats), *Ovis aries* (Sheep), *Sus scrofa* (Pig), *Tadorna cana* (Duck), *Mains javanica* (Pangolin), *Mustela putorius* (Ferret), *Panthera tigris altaica* (Tiger) and *Gallus* (Chicken). Cryo-EM (cryo-electron microscopy) structure of *Felis catus* ACE2 and SARS-CoV-2 RBD was extracted from the Protein Data Bank (PDB, http://www.rcsb.org/, PDB ID:7C8D [26]). The sequence identity and sequence similarity between the target sequence and the template sequence were obtained using the Macromolecules module in DS 4.0 (Discovery Studio 4.0, https://www.3ds.com/). The homology modeling technique was used to generate the three-dimensional structure of ACE2 for species with unsolved 3D structures. The complex of human SARS-CoV-2 S protein and ACE2 protein is extracted from the PDB (ID: 6LZG [16]). The human ACE2 protein was selected as the template and 10 protein models of each species were generated using the Build Homology Models protocol in DS 4.0. The quality of the generated model was assessed using the Verify Protein (Profiles-3D) protocol. For structure regions that have a low Verify Score and high PDF energies, it is necessary to re-evaluate the input alignment and regenerate the model with refined alignment, or refine the structure using the Refine Loops protocol.

2.2. Molecular dynamics simulation study

The molecular dynamics simulations were carried out to simulate all the complex structures. Initially, the topology files were obtained using the pdb2gmx module in GROMACS 5.0.2. The water model and OPLS-AA force field were defined [27]. The topological structure of the molecule, the location constraint files, and the processed structure files were generated. A periodic box was defined and each system was solvated with a simple point charge (spc 216) water model [27]. Then, the solvated systems were neutralized by adding suitable counterions. The structural optimization is needed before the molecular dynamics simulation for the initially established system often has partial irrationality in structure. Before starting the dynamics simulation, the systems were minimized until the maximum force \( <238.8 \text{ kcal mol}^{-1} \text{ nm}^{-1} \) to ensure that the structure of the system is normal, the distance between atoms is not too close, and the geometric configuration is reasonable. In the first equilibrium stage, the system is gradually heated from 0 K to 300 K under NVT conditions with the solute restrained and then equilibrated under NPT conditions for 1000 ps at 300 K. The close monitoring of the energy, temperature, pressure and density of the system is required in this stage to ensure proper convergence. After the temperature and pressure were equilibrated, the conditional constraints were removed, and each system was subjected to 200 ns free simulation. The time step was set as 2 fs and coordinates were saved every 10 ps.

2.3. Protein-protein binding free energy predication by MM/PBSA

In this study, the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) method [28] was used to estimate the binding free energy between S protein and ACE2 from different species. The binding free energy calculated by the g_mmpbsa [29] tool is composed of three parts, vacuum potential energy (MM), polar solvation energy (PB) and non-polar solvation energy (SA).The MM part consists of molecular mechanics energy, energy in vacuum, van der Waals energy and electrostatic interaction energy. The binding free energy is decomposed according to each residue, \( \Delta G_{\text{MM}} \), \( \Delta G_{\text{polar}} \) and \( \Delta G_{\text{nonpolar}} \) were separately calculated for each residue and then summed up to obtain the contribution of
3. Results and discussion

3.1. ACE2 receptor structures

An alignment of different ACE2 sequences with the template structure (PDB ID: 6LZG) is shown in Figure S1. The sequence identity and sequence similarity were obtained and a total of 10 models were built for each sequence that ranked based on PDF Total Energy and DOPE scores (Discrete Optimized Protein Energy, Table 1). Theoretically, if the sequence homology is over 50%, the structural accuracy of the constructed protein is relatively high. If it is less than 30%, the feasibility is poor and it is difficult to obtain good results [30]. In this study, all species share a high sequence identity and sequence similarity. Thus, the built protein has high accuracy. In addition, the smaller the PDF Total Energy, the better the structure is modeled. The DOPE score can evaluate the credibility of different conformations. The lower the score, the more reliable the model. The most reasonable model was selected from ten models of each species for further study. All the constructed ACE2 proteins were aligned with human ACE2 and are shown in Fig. 1.

To assess the compatibility of an atomic model (3D) with its own amino acid sequence (1D), a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar) was assigned [31]. The compatibility assessment was achieved by calculating a verify score, the verify score takes a value between the highest expected score and the lowest expected score. The closer to the highest expected score the reasonable the established model. More than 80% of the amino acids with scored >0.2 in the 3D-1D profile, also show that the established model is reasonable (https://servicesn.mbi.ucla.edu/Verify3D) (Table 2). It can be seen from Fig. 2 that the modeled protein was very consistent with the template protein.

PROCHECK program (https://servicesn.mbi.ucla.edu/PROCHECK/) was used to check the stereochemical quality of protein structures by Ramachandran Plot [32,33]. The rotation degree of the bond between the α-carbon atom and the carbonyl carbon atom in the peptide chain in the three-dimensional structure of the protein is $\psi$ (psi). The rotation degree of the bond between the α-carbon atom and the nitrogen atom in the peptide chain is $\phi$ (phi). Ramachandran Plot is used to illustrate the relationship between psi and phi, indicating the energetically allowable and forbidden conformations of amino acids in proteins. There are four regions on ramachandran plot includes most favorable regions, additional allowed regions, generously allowed regions and disallowed regions. A good quality model would be expected to have over 90% in the most favorable regions [32](Table 3, Figure S2).

3.2. Molecular dynamics simulation

3.2.1. Construction of simulated systems

A total of 19 complexes were used for molecular dynamics simulations, including the crystal structure (PDB ID: 6LZG, 7CBD) and the constructed complexes of S protein with ACE2 of 17 animals. Before molecular dynamics simulations, the size of the box was defined and side length was about 13 nm. The box was filled with water and sodium ions were added to balance the charge of the system. The box size, number of added water molecules and number of added ions are shown in Table 4. The state of the simulation system is shown in Fig. 3(A). Now, we have added each residue to the binding free energy.

Table 1

| Species                  | Sequence identity | Sequence similarity | PDF Total Energy | DOPE Score |
|--------------------------|-------------------|---------------------|-----------------|------------|
| Rhinolophus ferrumequinum| 76.7%             | 87.9%               | 2529.0          | 80019      |
| Bos taurus               | 80.1%             | 90.2%               | 2532.2          | 78952      |
| Camelus dromedarius      | 80.6%             | 89.7%               | 25843.5         | 78478      |
| Canis lupus familiaris   | 81.5%             | 89.7%               | 558.3           | 78199      |
| Felis catus              | 83.6%             | 90.9%               | —               | —          |
| Macaca mulatta           | 91.8%             | 95.8%               | 2666.2          | 78630      |
| Mus musculus             | 81.8%             | 89.1%               | 2568.0          | 78015      |
| Paguma larvata           | 82.7%             | 90.1%               | 2529.7          | 78158      |
| Pongo abelii             | 95.9%             | 96.6%               | 2580.4          | 79872      |
| Batsus norvegicus        | 81.5%             | 90.1%               | 2614.5          | 77929      |
| Ovis aries               | 80.3%             | 90.4%               | 2580.5          | 79069      |
| Sus scrofa               | 80.3%             | 89.1%               | 2489.1          | 79367      |
| Taurens cana             | 68.4%             | 82.4%               | 2596.6          | 77946      |
| gallus                   | 67.3%             | 81.1%               | 2587.8          | 77766      |
| Mains javanica           | 82.9%             | 89.9%               | 556.5           | 78132      |
| Mustela putorius         | 81.1%             | 89.3%               | 2650.5          | 77911      |
| Panthera tigris altaica  | 85.0%             | 91.9%               | 2644.4          | 77739      |
| Procyon lotor            | 81.8%             | 90.4%               | 2717.0          | 77248      |

*Felis catus does not have PDF Total Energy or DOPE values because the protein was downed from the Protein Data Bank (http://www.rcsb.org/, PDB ID: 7CBD), not a homology model.
solvent molecules and ions to obtain an electrically neutral system. Before starting the dynamics simulation, we ran the minimization stage until the maximum force $< 238.8 \text{ kcal mol}^{-1} \text{ nm}^{-1}$. During the first equilibrium stage, a simulation was carried out for 100 ps at a temperature of 300 K in the NVT (constant number of particles, volume and temperature) ensemble with solute heavy atoms restrained by a force constant of 100 kcal mol$^{-1}$ Å$^{-1}$. The system was heated to 300 K during a 1000 ps simulation in the NPT ensemble, with the same force constant applied to the solute atoms. In the simulation phase, the relaxed systems were simulated without restrained under NPT ensemble conditions for 200 ns. Within this simulation time, when the total energy and RMSD of the protein backbone C$_\alpha$ atoms reached a plateau, the systems were considered equilibrated and suitable for statistical analysis. After molecular dynamics simulations, the average complex structure of the 50 ns equilibrium stage was extracted and aligned with the initial complex structure (Fig. 3(B)), presenting changes in the overall protein. As the alignment RMSD shown, the binding domains of ACE2 protein and S protein had conformational fluctuations, which make the distance between them closer and bind stronger.

### Table 2

| Species                  | Verify Score | Expected High Score | Expected Low Score | 3D-1D score ≥0.2 (%) |
|--------------------------|--------------|---------------------|--------------------|----------------------|
| Rhinolophus ferrumequinum| 259.03       | 272.697             | 122.714            | 94.56                |
| Bos taurus               | 252.41       | 272.236             | 122.506            | 97.72                |
| Camelus dromedarius      | 253.22       | 272.697             | 122.714            | 93.60                |
| Canis lupus familiaris   | 246.99       | 272.697             | 122.714            | 94.69                |
| Felis catus              | 250.10       | 272.697             | 122.714            | 96.21                |
| Macaca mulatta           | 257.86       | 272.697             | 122.714            | 95.70                |
| Mus musculus             | 259.60       | 272.697             | 122.714            | 95.32                |
| Paguma larvata           | 261.30       | 272.697             | 122.714            | 96.84                |
| Pongo abelii             | 250.69       | 272.697             | 122.714            | 97.47                |
| Rattus norvegicus        | 252.90       | 272.697             | 122.714            | 93.93                |
| Ovis aries               | 256.91       | 272.236             | 122.506            | 94.94                |
| Sus scrofa               | 252.63       | 272.697             | 122.714            | 95.70                |
| Tadorana cana            | 251.49       | 273.159             | 122.921            | 93.68                |
| Gallus                   | 254.68       | 273.159             | 122.921            | 94.07                |
| Mains javanica           | 255.70       | 272.697             | 122.714            | 95.58                |
| Mustela putorius         | 255.04       | 272.697             | 122.714            | 94.94                |
| Panthera tigris altaica  | 259.03       | 272.697             | 122.714            | 98.36                |
| Procyon lotor            | 259.55       | 272.697             | 122.714            | 95.45                |

**Fig. 2.** The residual of averaged 3D-1D score.

3.2.2. Radius of gyration about the protein backbone

As shown in Fig. 4, the radius of gyration about the protein backbones maintains relative stability. All the Rg values stabilize at 3.0–3.3 nm, which shows that the protein system was properly folded.

3.2.3. RMSD analysis

In order to evaluate whether the system is stable during the simulation, the RMSD value during the entire simulation process was calculated. The result is shown in Fig. 5. It can be seen from the figure that all systems gradually reach equilibrium, however the time for each system to reach equilibrium was different. This also
demonstrates that although different species share a high degree of sequence similarity, the small differences will lead to varying degrees of impact on the stability of the system.

### 3.2.4. RMSF

The RMSF for each system was calculated to investigate the fluctuation of amino acids. As shown in Fig. 6, part A is a coiled-coil

| Table 3 | Ramachandran Plot statistics for each region. |
|---------|---------------------------------------------|
| Species            | Most favoured regions (%) | Additional allowed regions (%) | Generously allowed regions (%) | Disallowed regions (%) |
| 6LZG     | 90.8                         | 9                              | 0                             | 0.3                   |
| *Rhinolophus ferrumequinum* | 92.1          | 7.4                            | 0.4                           | 0.1                   |
| *Bos taurus*      | 91.9                         | 7.5                            | 0.3                           | 0.3                   |
| *Camelus dromedarius* | 91.6         | 8.3                            | 0                             | 0.1                   |
| *Canis lupus familiaris* | 92.0          | 7.7                            | 0.3                           | 0.3                   |
| *Felis catus*      | 89.0                         | 10.8                           | 0.2                           | 0.3                   |
| *Macaca mulatta*  | 91.7                         | 8.1                            | 0                             | 0.1                   |
| *Mus musculus*    | 91.9                         | 8.0                            | 0                             | 0.1                   |
| *Paguma larvata*  | 91.7                         | 8.0                            | 0                             | 0.3                   |
| *Pongo abelii*    | 92.1                         | 7.8                            | 0.1                           | 0.1                   |
| *Rattus norvegicus* | 91.6          | 8.1                            | 0                             | 0.3                   |
| *Ovis aries*      | 92.3                         | 7.2                            | 0.1                           | 0.3                   |
| *Sus scrofa*      | 92.0                         | 7.9                            | 0                             | 0.1                   |
| *Tadorna cana*    | 91.9                         | 7.6                            | 0.3                           | 0.1                   |
| *Gallus*          | 92.1                         | 7.4                            | 0.1                           | 0.4                   |
| *Macaca mulatta*  | 91.9                         | 7.5                            | 0.1                           | 0.1                   |
| *Mus musculus*    | 91.7                         | 8.0                            | 0                             | 0.3                   |
| *Paguma larvata*  | 91.7                         | 8.0                            | 0                             | 0.3                   |
| *Pongo abelii*    | 92.1                         | 7.8                            | 0                             | 0.3                   |
| *Rattus norvegicus* | 91.6          | 8.1                            | 0                             | 0.3                   |
| *Ovis aries*      | 92.3                         | 7.2                            | 0.1                           | 0.3                   |
| *Sus scrofa*      | 92.0                         | 7.9                            | 0                             | 0.1                   |
| *Tadorna cana*    | 91.9                         | 7.6                            | 0.3                           | 0.1                   |
| *Gallus*          | 92.1                         | 7.4                            | 0.1                           | 0.4                   |

| Table 4 | The box size, number of added water molecules and number of added ions during construct dynamic simulation system. |
|---------|----------------------------------------------------------------------------------------------------------------|
| Species            | Box size (nm) | Added water molecules | Added ions | Alignment RMSD (nm) |
| 6LZG     | 13.15951     | 70220                 | 24         | 2.534               |
| *Rhinolophus ferrumequinum* | 13.25506   | 71137                 | 28         | 1.190               |
| *Bos taurus*      | 13.14124     | 70221                 | 20         | 2.736               |
| *Camelus dromedarius* | 13.26625   | 72122                 | 27         | 1.382               |
| *Canis lupus familiaris* | 13.27732   | 72817                 | 18         | 1.471               |
| *Felis catus*      | 13.18565     | 69790                 | 15         | 2.382               |
| *Macaca mulatta*  | 13.17847     | 69004                 | 22         | 2.017               |
| *Mus musculus*    | 13.32780     | 74508                 | 16         | 1.618               |
| *Paguma larvata*  | 13.16273     | 70228                 | 16         | 2.200               |
| *Pongo abelii*    | 13.14151     | 69873                 | 23         | 1.237               |
| *Rattus norvegicus* | 13.24835   | 71025                 | 17         | 1.899               |
| *Ovis aries*      | 13.25335     | 71126                 | 19         | 2.095               |
| *Sus scrofa*      | 13.20863     | 69741                 | 24         | 1.975               |
| *Tadorna cana*    | 13.30531     | 74148                 | 22         | 2.200               |
| *Gallus*          | 13.27006     | 72287                 | 20         | 1.478               |
| *Macaca mulatica* | 13.25344    | 71138                 | 15         | 1.975               |
| *Mus musculus*    | 13.24950     | 71077                 | 24         | 1.523               |
| *Panthera tigris altaica* | 13.26101  | 71602                 | 15         | 1.306               |
| *Procyon lotor*   | 13.24162     | 70872                 | 23         | 2.040               |

Fig. 3. A. The simulation system of 6LZG filled with water molecules and ions. The green is the protein complex, the orange are water molecules, the blue are sodium ions; B. The alignment of average structure after MDS with the initial complex structure of 6LZG. The structure in green is the initial complex structure, and the blue is the average complex structure. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
α helix where the two helices are combined with each other by the hydrophobic interaction of the hydrophobic side chain, and the free energy is very low, which means it is very stable. Part B consists of α helix, β-sheet and loop region. There is a strong fluctuation from amino acid 336 to 344 for each system, as its solvent exposed and far from the S protein binding area. Part C is a main binding area on S protein which consist of α helix, β-sheet and loop region. The amino acids on the S protein binding region have varying degrees of fluctuation. In general, even different species of ACE2 have some differences in structure, but the ACE2 binding domain of different species is relatively stable. S protein can adjust its structure to bind ACE2 better over induced fit. As shown in Fig. 6, an obvious structural fluctuation for S protein binding.

3.3. Binding energy calculation by MM/PBSA

For all 19 protein complexes, frames were extracted every 1 ns from the last 50 ns trajectory with stable RMSD value to calculate
According to the binding free energy (Fig. 7), these species were grouped into four categories for further research. Procyon lotor \((-309.7 \pm 3.21 \text{ kcal/mol})\), Human \((-309.7 \pm 5.63 \text{ kcal/mol})\) and Camelus dromedarius \((-309.5 \pm 3.74 \text{ kcal/mol})\) were the first category, which had strong binding with S protein. Macaca mulatta \((-272.3 \pm 2.92 \text{ kcal/mol})\), Rhinolophus ferrumequinum \((-269.7 \pm 4.08 \text{ kcal/mol})\), Pongo abelii \((-267.6 \pm 3.50 \text{ kcal/mol})\), Tadorna cana \((-266.0 \pm 5.50 \text{ kcal/mol})\), Mustela putorius \((-262 \pm 3.36 \text{ kcal/mol})\) and Sus scrofa \((-252.6 \pm 3.56 \text{ kcal/mol})\) were the second category. Mains javanica \((-229.9 \pm 2.87 \text{ kcal/mol})\), Canis lupus familiaris \((-224.7 \pm 3.40 \text{ kcal/mol})\), Rattus norvegicus \((-223.2 \pm 3.56 \text{ kcal/mol})\), Ovis aries \((-221.9 \pm 3.99 \text{ kcal/mol})\), Paguma larvata \((-221.2 \pm 4.73 \text{ kcal/mol})\), Panthera tigris altaica \((-220.9 \pm 2.30 \text{ kcal/mol})\) were the third category. The rest fall into another category, which had weak binding between ACE2 and S protein.

### 3.4. Energy decomposition

Less residues from S protein contribute to the binding than ACE2. The energy decomposition of the amino acids in the binding region showed that the contribution per amino acid residue on the S protein (Fig. 8) to the binding energy is much higher than the contribution of the amino acid residues on ACE2 (Fig. 9). The contribution of the amino acid residues on S protein is up to \(-66.72 \text{ kcal/mol}\). Most of the contribution is provided by electrostatic interaction. While the contribution of the amino acid residues on ACE2...
residues on ACE2 is only –14.13 kcal/mol. As Fig. 8 shows, energy decomposition of the amino acid residues on the binding surface, we can conclude that the electrostatic interaction site of Arg685, Arg690, Lys699, Lys706, Lys726, Arg736, Arg739, Lys740, Lys744, Arg748 and Arg791 had contributed the most to the binding free energy. From Fig. 9, Lys 352 on Bos taurus is bad for binding.

3.5. Binding modes analysis

Five average structures based on the last 50 ns equilibrated trajectories of predicted susceptible species were extracted and used to analyze the hydrogen bonds, hydrophobic and salt bridges interactions via DS 4.0 (Fig. 10). Attention on the first category consists of Procyon lotor and Camelus dromedarius. When Procyon lotor binds to S protein, there are 10 hydrogen bonds formed, as well as one salt bridge interaction and 2 hydrophobic interactions. Camelus dromedarius formed 12 hydrogen bonds, as well as 1 salt bridge interaction and 4 hydrophobic interactions. For human, there are 8 hydrogen bonds formed, as well as 2 salt bridge interactions and 4 hydrophobic interactions. Arg685, Arg690, Lys699, Lys706, Lys726, Arg736, Arg739, Lys740, Lys744, Arg748 and Arg791 on S protein are the common amino acids which contribute largely to the binding. Besides that, His 34, Asp 38 and Tyr41 on Camelus dromedarius and human are the same. All these 14 amino acid sites

![Fig. 7. Total binding free energy of each species.](image)

![Fig. 8. The contribution of the amino acid residues from the S protein.](image)

![Fig. 9. The contribution of the amino acid residues on the ACE2 receptor.](image)
are playing a key role in the binding of ACE2 to S protein. When S protein binds ACE2 of *Felis catus*, there are 4 hydrogen bonds, 2 salt bridges and 5 hydrophobic interactions formed. When S protein binds ACE2 of *Bos taurus*, there are 11 hydrogen bonds and 5 hydrophobic interactions formed without salt bridges.

4. Conclusions

CoVs are zoonotic pathogens and infect humans via interspecies transmission. In this study, we built the ACE2 protein structures of 17 different species based on the template structure of human ACE2. Then molecular dynamics simulation and binding free energy calculation were performed based on these structures. The generated results from current study including homology models, RMSD analysis, RMSF analysis, binding free energy analysis, energy decomposition and binding modes analysis aided the prediction of potential intermediate hosts likely to communicate the viral infection to humans. SARS-CoV and MERS-CoV can cross the species barrier and result in human infection. Previous studies have shown that the MERS-CoV viruses first jumped from their natural hosts (bats) [34] to an intermediate adaptive animal (dromedary camels) [35] before infecting humans. Delineating this cross-species transmission route could be highly instructive for disease control. In this study, the calculated binding free energy of virus S protein and ACE2 from different species from high to low is *Felis catus*, *Bos taurus*, *Gallus*, *Mus musculus*, *Panthera tigris altaica*, *Paguma larvata*, *Ovis aries*, *Rattus norvegicus*, *Canis lupus familiaris*, *Mains javanica*, *Sus scrofa*, *Mustela putorius*, *Tadorna cana*, *Pongo abelii*, *Rhinolophus ferrumequinum*, *Macaca mulatta*, *Camelus dromedarius*, and *Procyon lotor*. Strong binding between the S protein and ACE2 from host cells could be a clue that virus may infect the host. From the predicted binding affinity, *Camelus dromedarius* and *Procyon lotor* could be infected by SARS-CoV-2. As reported before [36,37], *Camelus dromedarius* could be infected by the virus, which confirm the predication by this work. Some experiments revealed that *Macaca mulatta* can be infected and produce a cellular immune response, and provide protective efficacy against SARS-CoV-2 rechallenge [23,24]. Besides that, some experiments revealed that SARS-CoV-2 replicates poorly in *Canis lupus familiaris* (dogs), *Gallus* (chickens), *Sus scrofa* (pigs) and *Tadorna cana* (ducks), but efficiently in *Mustela putorius* (ferrets) and *Felis catus* (cats) [38] and Syrian hamster (not the same as *Mus musculus*) [39], which partially validates our predication that the binding energy of *Canis lupus familiaris*, *Gallus*, *Tadorna cana* ACE2 with S protein are quite weak and the binding energy of *Mustela putorius*, *Macaca mulatta* ACE2 with S protein are strong. Beyond that, this work is trying to explore possibility of virus infection through the protein-protein binding energy predication. However, the efficient receptor
binding is only the first step during virus infection. Some more experiments and monitoring should be carried out to explore the species which could be infected by SARS-CoV-2.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project was financially supported by Science and Technology Planning Project of Sichuan Province (Grant No: 2019YFS0095), and the State Key Laboratory of Phytochemistry and Plant Resources in West China (Grant No: P2017-KF07, P2018-KF14). Prof Weiliang Zhu and Cheng Peng from Shanghai Institute of Materia Medica helped a lot on the design and construction of the project. Supercomputing Center of Sichuan University of Science & Engineering Provides Computing Resources.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmgm.2021.107893.

References

[1] A.E. Gorbalenya, et al., The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, Nature Microbiology 5 (2020) 536–544, https://doi.org/10.1038/s41564-020-0659-x.
[2] J. Cui, F. Li, Origin and evolution of pathogenic coronaviruses, Nat. Rev. Microbiol. 17 (2018) 181–192, https://doi.org/10.1038/s41579-018-0118-9.
[3] S. Adachi, T. Koma, N. Doi, M. Nomaguchi, A. Adachi, Commentary: origin and evolution of pathogenic coronaviruses, Front. Immunol. 11 (2020) 811, https://doi.org/10.3389/fimmu.2020.00811.
[4] K. Kuppalli, A. Rasmussen, A glimpse into the eye of the COVID-19 cytokine storm, ElBioMedicine 55 (2020) 102789, https://doi.org/10.1016/j.elbion.2020.102789.
[5] X. Xu, et al., Pathological findings of COVID-19 associated with acute respiratory distress syndrome, The Lancet Respiratory Medicine 8 (2020) 420–422, https://doi.org/10.1016/S2213-2600(20)30076-X.
[6] N. Zhu, et al., A novel coronavirus from patients with pneumonia in China, 2019, N. Engl. J. Med. 382 (2020) 727–733, https://doi.org/10.1056/NEJMoa2001017.
[7] C. Ma, et al., Ehselen, Dieldridrin, Carmofur, PX-12, Tidegulusb, and Shikonin Are Non-Specific Promiscuous SARS-CoV-2 Main Protease Inhibitors, Bioinformatics, 2020, pp. 1256–1277 [Preprint].
[8] R.A.-O.X. Narkhede, A.A.-O. Pise, R.A.-O. Cheke, S.A.-O.X. Shinde, Recognition of natural products as potential inhibitors of COVID-19 main protease (mpro): in-silico evidences, Natural Products and Bioprospecting 10 (5) (2020) 297–306, https://doi.org/10.1007/s11569-020-00253-1.
[9] J. Chan, et al., Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, Emerg. Microb. Infect. 9 (2020) 221–236, https://doi.org/10.1080/22221751.2020.1719902.
[10] Ren, et al., Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study, Chinese Med. J 9 (2020) 1015–1024, https://doi.org/10.1097/cmj.0000000000000722.
[11] R. Lu, et al., Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet 395 (2020) 565–574, https://doi.org/10.1016/S0140-6736(20)30251-8.
[12] A.C. Walls, et al., Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2) (2020) 281–292, https://doi.org/10.1016/j.cell.2020.02.058.
[13] M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses, Nature Microbiology 5 (2020) 562–569, https://doi.org/10.1038/s41564-020-0688-y.
[14] W. Li, et al., Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus, Nature 426 (2003) 450–454, https://doi.org/10.1038/ nature02145.
[15] M. Hoffmann, et al., SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor, Cell 181 (2) (2020) 271–280, https://doi.org/10.1016/j.cell.2020.02.052.
[16] O. Wang, et al., Structural and functional basis of SARS-CoV-2 entry by using human ACE2, Cell 181 (4) (2020) 894–904, https://doi.org/10.1016/j.cell.2020.03.045.
[17] T. Gallagher, M. Buchmeier, Coronavirus spike proteins in viral entry and pathogenesis, Virolology 279 (2001) 371–374, https://doi.org/10.1006/viro.2000.0757.
[18] D. Wrapp, et al., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, BioRxiv 367 (6483) (2020) 1260–1263, https://doi.org/10.1101/2020.02.12.944682.
[19] G. Simmons, P. Zmora, S. Gierer, A. Heurich, S. Pohlmann, Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research, Antiviral Reseach 100 (3) (2013) 605–614, https://doi.org/10.1016/j.antiviral.2013.09.028.
[20] S. Belouzard, V.C. Chu, G.R. Heurich, Structure and function of SARS-CoV-2 spike protein via sequential proteolytic cleavage at two distinct sites, Proc. Natl. Acad. Sci. Unit. States Am. 106 (14) (2009) 5871, https://doi.org/10.1073/pnas.0809524106.
[21] F. Li, W. Li, M. Farzan, S.C. Harrison, Structure of SARS-CoV spike receptor-binding domain complexed with receptor, Science 309 (5742) (2005) 1864–1868, https://doi.org/10.1126/science.1116480.
[22] R.A.-O. Yan, et al., Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, Science 367 (6485) (2020) 1444–1448, https://doi.org/10.1126/scienceabb2762.
[23] A.A.-O. Chandrashekar, S.-L. SARS-CoV-2 infection protects against rechallenge in rhesus macaques, Science 369 (6505) (2020) 812–817, https://doi.org/10.1126/science.abd4776.
[24] Y. Wan, J. Shang, R. Graham, R.S. Baric, F. Li, Receptor recognition by the novel coronavirus from wuhan: an analysis based on decade-long structural studies of SARS coronavirus, J. Virol. 94 (7) (2020) 20181277 [Preprint], https://doi.org/10.1126/science.abc4776.
[25] J. Chan, et al., Simulation of the clinical and pathological manifestations of coronavirus disease 2019 (COVID-19) in a golden Syrian hamster model: implications for disease pathogenesis and transmissibility, Clin. Infect. Dis. 71 (9) (2020) 2428–2446, https://doi.org/10.1093/cid/ciaa325.

Journal of Molecular Graphics and Modelling 105 (2021) 107893