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.
Molecular dynamic simulation analysis of SARS-CoV-2 spike mutations and evaluation of ACE2 from pets and wild animals for infection risk

Ping Chen\textsuperscript{a,b,1}, Jingfang Wang\textsuperscript{b,1}, Xintian Xu\textsuperscript{b,e}, Yuping Li\textsuperscript{c}, Yan Zhu\textsuperscript{d}, Xuan Li\textsuperscript{d,e,*}, Ming Li\textsuperscript{c,**}, Pei Hao\textsuperscript{a,b,e,***}

\textsuperscript{a} Guangzhou Women and Children’s Medical Center, Guangzhou Medical University, Guangzhou 510623, China
\textsuperscript{b} Institute Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai 200031, China
\textsuperscript{c} West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu 610041, China
\textsuperscript{d} Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Science, Chinese Academy of Sciences, Shanghai 200032, China
\textsuperscript{e} University of Chinese Academy of Sciences, Beijing 100039, China

\begin{abstract}
Coronavirus Disease 2019 (COVID-19) is an ongoing global health emergency that has caused tremendous stress and loss of life worldwide. The viral spike glycoprotein is a critical molecule mediating transmission of SARS-CoV-2 by interacting with human ACE2. However, through the course of the pandemics, there has not been a thorough analysis of the spike protein mutations, and on how these mutants influence the transmission of SARS-CoV-2. Besides, cases of SARS-CoV-2 infection among pets and wild animals have been reported, so the susceptibility of these animals requires great attention to investigate, as they may also link to the renewed question of a possible intermediate host for SARS-CoV-2 before it was transmitted to humans. With over 226,000 SARS-CoV-2 sequences obtained, we found 1573 missense mutations in the spike gene, and 226 of them were within the receptor-binding domain (RBD) region that directly interacts with human ACE2. Modeling the interactions between SARS-CoV-2 spike mutants and ACE2 molecules showed that most of the 74 missense mutations in the RBD region of the interaction interface had little impact on spike binding to ACE2, whereas several within the spike RBD increased the binding affinity toward human ACE2 thus making the virus likely more contagious. On the other hand, modeling the interactions between animal ACE2 molecules and SARS-CoV-2 spike revealed that many pets and wild animals’ ACE2 had a variable binding ability. Particularly, ACE2 of bamboo rat had stronger binding to SARS-CoV-2 spike protein, whereas that of mole, vole, \textit{Mus pahari}, palm civet, and pangolin had a weaker binding compared to human ACE2. Our results provide structural insights into the impact on interactions of the SARS-CoV-2 spike mutants to human ACE2, and shed light on SARS-CoV-2 transmission in pets and wild animals, and possible clues to the intermediate host(s) for SARS-CoV-2.
\end{abstract}

1. Background

The coronavirus disease 2019 (COVID-19) is an ongoing global health emergency (Wang et al., 2020). According to the World Health Organization (WHO) report, as of 5 October 2021, coronavirus disease 2019 (COVID-19) has caused over 234 million confirmed infections and more than 4.8 million cumulative number of deaths worldwide (WHO: COVID-19, 2021). This outbreak of respiratory disease is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), an enveloped virus with a positive-sense and single-stranded RNA genome.
This coronavirus belongs to the beta-coronavirus (beta-CoV) genera, which also includes human coronavirus OC43 (HCoV-OC43), severe acute respiratory syndrome coronavirus (SARS-CoV), and the Middle East Respiratory Syndrome coronavirus (MERS-CoV). HCoV-OC43 engages sugars for the host cell attachment (Li et al., 2005), and MERS-CoV binds to the human dipeptidyl peptidase 4 (hDPP4) receptor (Raj et al., 2013). On the other hand, SARS-CoV-2 shares the same mechanism for SARS-CoV to infect human cells, i.e., utilizing the angiotensin-converting enzyme 2 (ACE2) as a cellular receptor for virus entry (Xu et al., 2020). Viral infection studies found that SARS-CoV-2 exploited the ACE2 receptor for virus entry into the target cells, suggesting that SARS-CoV-2 targets a similar spectrum of cells as known coronaviruses led to a suggestion that the snake was a possible intermediate host for SARS-CoV-2 (Liu et al., 2019; Zhang et al., 2020).

The spike glycoprotein of SARS-CoV-2 (also known as the S protein) is envelope-embedded and surface-located, functioning to mediate membrane fusion and virus entry. After being attached to the host cells, the spike protein will be cleaved into two subunits (named S1 and S2) by the host proteases (e.g., TMPRSS2) (Hoffmann et al., 2020), with responsibility for receptor recognition and membrane fusion, respectively. The S1 subunit contains two independent domains, the N-terminal domain (NTD) and receptor-binding domain (RBD). Sequence alignment indicated that the SARS-CoV-2 spike protein showed a nearly 70% sequence identity with the SARS-CoV spike (Xu et al., 2020) and 98% sequence identity with the bat CoV RaTG13 spike (Zhou et al., 2020). Flow cytometry assay, crystal, and cryo-EM studies showed that SARS-CoV-2 interacted with the human ACE2 via the RBD domain of the S1 subunit (Wang et al., 2020; Ian et al., 2020; Yan et al., 2020). According to surface plasmon resonance (SPR) results, the SARS-CoV-2 spike protein had a nearly 15 nM affinity to the human ACE2, which was approximately 20-fold higher than that of the SARS-CoV spike protein binding to the human ACE2 (Wrapp et al., 2020).

First, as the pivotal viral molecule for human infection, any changes with the SARS-CoV-2 spike protein sequence, especially in its RBD region, could alter the interaction between SARS-CoV-2 and the human ACE2. Thus it is critical for us to investigate the mutational changes with the SARS-CoV-2 spike protein, and understand how the mutations potentially affect the viral infection in humans. Throughout COVID-19 global pandemics, more than 96,000,000 human infections were recorded. However, only a small fraction of the SARS-CoV-2 viruses from human patients have been sequenced. While mutations have been found in all SARS-CoV-2 coding genes, there has not been a thorough analysis on the spike protein mutations, and on how these mutations potentially affect the interaction between spike protein and human ACE2 molecule, hence as a consequence changing SARS-CoV-2 infect risk.

Second, while many cases of animal infections, including pets and wild animals, by SARS-CoV-2, have been reported (Gaudreault et al., 2020; Sit et al., 2020; McAloose et al., 2020; Shi et al., 2020), the susceptibility of these animals requires great attention to investigate, as they may also link to the renewed question of what is the possible intermediate host(s) for SARS-CoV-2 before it transmitted in human populations. Most of the coronaviruses causing human illness have animal origins. Early studies indicated that SARS-CoV and MERS-CoV were transmitted from palm civet and dromedary camel to human, respectively (Guan et al., 2003; Madani et al., 2014; Azhar et al., 2014). HCoV-OC43 and HKU1 are considered to have originated from rodents (Forni et al., 2017; Cui et al., 2019). However, whether there is and what is the possible intermediate host(s) for SARS-CoV-2 remains a big unknown. While candidate hosts for SARS-CoV-2 have been proposed so far. The metagenomic and evolutionary analysis showed that pangolin coronavirus was over 90% identical at a whole-genome level to SARS-CoV-2 and BatCoV RaTG13, indicating pangolin as a likely intermediate host for SARS-CoV-2 (Liu et al., 2019; Zhang et al., 2020). Sequence and coding analysis of SARS-CoV-2 in comparison to other known coronaviruses led to a suggestion that the snake was a possible intermediate host (Ji et al., 2020). A deep learning study on the virus-host prediction found that Neovison vison coronavirus had a close infectivity pattern to SARS-CoV-2 (Guo et al., 2020).

To answer these important questions, in the current study we used more than 220,000 SARS-CoV-2 virus sequences collected from patients worldwide and identified over 2700 mutations in the spike protein gene. Using molecular dynamic simulation, we have analyzed the non-synonymous mutations and assessed their effect on interactions with the human ACE2. Then to study the interactions of ACE2 from pets and wild animals to SARS-CoV-2 spike protein, we constructed structural models of ACE2 from twenty-five different animals and used molecular dynamic simulation to optimize their binding complex models. To assess the binding affinity of these complexes, we also employed the molecular mechanics Poisson–Boltzmann surface area (MM-PB/SA) approach to calculate the binding free energies of the SARS-CoV-2 spike to the animals’ ACE2 molecules. Our results provide structural insights into the interactions of the animal ACE2 to SARS-CoV-2 spike mutants and shed light on the transmission of SARS-CoV-2 in pets and wild animals, and possible clues to the intermediate host(s) for SARS-CoV-2.

2. Results and discussion

2.1. Identification of mutations of SARS-CoV-2 spike protein through COVID-19 pandemics

Among the reported cases of human infections by SARS-CoV-2, the causative agent responsible for the COVID-19 pandemics, only a fraction of the SARS-CoV-2 viruses from human patients have been sequenced. We obtained the SARS-CoV-2 virus sequence data from a number of resources, including the NCBI Virus portal (Hatcher et al., 2017), GISAID (Elbe and Buckland-Merrett, 2017), and China National Center for Bioinformation (CNCB) (Song et al., 2020). There were a total number of 226,484 SARS-CoV-2 sequences, of which 213,705 are complete viral genome sequences from human sources (Additional file 1: Table S1). The spike gene sequences from complete SARS-CoV-2 sequences were extracted and mapped to the SARS-CoV-2 spike gene reference, for which 2782 mutations were found within the spike gene (Additional file 2: Table S2). These mutations were further analyzed, and 1573 of them were found to be missense mutations (Additional file 3: Table S3). Among the missense mutations, 226 of them were in critical RBD region of spike protein (viral genome position 22517–23185) (Fig. 1 and Additional file 3: Table S3). The RBD domain is located at the C-terminal of the S1 subunit that directly interacts with host receptors hence facilitating virus infection (Li, 2012).

Fig. 1. Identified missense mutations of the SARS-CoV-2 spike gene. The distribution of 1573 missense single nucleotide variations (SNVs) in different domains of the SARS-CoV-2 spike gene is illustrated. The domain structure of spike protein is denoted on the top, which shows the S1 and S2 subunits of the spike protein.
2.2. Modeling of missense mutations in SARS-CoV-2 spike RBD region for interaction with human ACE2

To detect the structural influence of the aforementioned 226 RBD missense mutations, we modeled these missense mutations based on the crystal structure of SARS-CoV-2 spike RBD (Lan et al., 2020). Among the RBD missense mutations, 74 are found to be located in the region of interaction interface between RBD and ACE2 (Additional file 3: Table S3). Modeling and MM-PB/SA free energy calculations indicated that most of the missense mutations in the interface had little impact on the spike binding to ACE2. Interestingly, two mutations, Y449F and Y453F, caused the loss of hydrogen bond interactions between RBD and ACE2 (Fig. 2A). However, the binding free energies for Y449F (−59.55 kcal/mol) and Y453F (−59.17 kcal/mol) did not show a significant reduction, compared to that of the wild-type spike protein (−60.64 kcal/mol), which helps explain why they maintained infection in people. These results indicate a compensation mechanism may exist for the viral spike protein interacting with human ACE2 molecule to maintain the stability of the complex. Interestingly, three mutations, E471Q (binding free energy: −62.34 kcal/mol), S477T (−62.62 kcal/mol), and N501Y (−63.48 kcal/mol), were found to have a decrease in binding free energies, which made them bind more tightly to human ACE2. Notably, N501Y is one of the critical mutations recently found in the dominant SARS-CoV2 variants circulating in the UK (B.1.1.7) and in South Africa (501Y.V2) that became more infectious among people. However, the E484K mutation found in the South Africa variant 501Y.V2 did not significantly alter binding free energy (−60.06 kcal/mol) by itself. So over the course of the pandemics, several missense mutations within the spike RBD region increased the binding affinity toward human ACE2 thus making the virus likely more contagious, while most of the mutations had little impact on the spike ACE2 interaction.

2.3. Identification of critical interface groups in human ACE2 binding to SARS-CoV-2 spike by MM-PB/SA approach

Both experimental and theoretical evidence indicated that SARS-CoV-2 shared similar receptor recognition and binding mechanism with SARS-CoV due to the similarity in the spike protein sequence (Xu et al., 2020; Wrapp et al., 2020). To investigate the interaction of ACE2 from pets and wild animals with SARS-CoV-2 spike protein, we first analyzed the crystal structure of the spike-ACE2 complex (PDB entry 6m0j) (Lan et al., 2020) for critical residues for ligand recognition. In the binding interface between the SARS-CoV-2 spike and human ACE2, three groups (left, middle, and right) of critical residues for ligand recognition with the hydrogen bonding interactions were present (Fig. 2A). The hydrogen-bonding interactions between the ACE2 residues Asp38 and Tyr41, and the spike residues Tyr449 and Gln498 defined the recognition group on the left side of the binding interface. These hydrogen bonds are also known to exist in the binding interface of the SARS-CoV spike and human ACE2. The middle recognition group was composed of two hydrogen-bonding interactions, i.e. Tyr453-His34 and Gln493-Glu35. These two hydrogen bonds that also existed in the SARS-CoV spike-ACE2 complex, were essential for the transmission of SARS-CoV from animal to human (Li et al., 2005; Song et al., 2005). Mutations in these residues, for example, N479L and N479R in the SARS-CoV spike protein, significantly reduced the binding affinities to the human ACE2 (Li, 2008; Li et al., 2005). The recognition group on the right side contained two hydrogen bond interactions between the ACE2 residues Ser19 and Tyr83, and the spike residues Ala475 and Tyr489, which were important for the spike to recognize and bind to the human ACE2.

To estimate their binding affinity and the contribution for each of the critical residues, we computed the binding free energies of the SARS-CoV-2 spike to ACE2 by MM-PB/SA approach (see Materials and methods). The binding free energies for SARS-CoV spike binding to human ACE2 were also computed as a comparison. As a result, the binding free energy of SARS-CoV-2 spike to the human ACE2 was −60.64 kcal/mol on average, a little lower than that of the SARS-CoV spike binding to the human ACE2 (−58.55 kcal/mol ± 8.75), indicating that a stronger affinity of SARS-CoV-2 than SARS-CoV binding to the human ACE2. This observation is in good agreement with the results from the surface plasmon resonance (SPR) study, which estimated a value of 14.7 nM for the affinity of SARS-CoV-2 spike to the human ACE2 (Wrapp et al., 2020). Energy decomposition analysis indicated that the ACE2 residues Tyr41 and Tyr83 contributed most to the spike binding affinity for both SARS-CoV-2 and SARS-CoV (Fig. 2B). Basing on the crystal structure analysis, these two tyrosine residues belong to the critical recognition groups in the binding interface of the spike and human ACE2. Besides, Met82, Gly354, and Asp355 also had substantial contributions to the binding free energy between spike protein and human ACE2. These residues were considered to be critical and their alteration in ACE2 of other animals will greatly impact the interaction of SARS-CoV-2 spike and animals’ ACE2 molecules, thus affecting the dynamic of animal infection by SARS-CoV-2.

2.4. Modeling the interactions between animal ACE2 and SARS-CoV-2 spike protein

We obtained the sequences of the ACE2 gene for twenty-five pets or wild animals from GenBank at NCBI (Additional file 4: Table S4), some of which were suggested as the possible intermediate host for transmission to humans. After constructing the structural models for the ACE2 of these animals, we found many critical residues were replaced that resulted in changes in the binding affinity of the spike-ACE2
SARS-CoV-2 spike binding to civet and pangolin ACE2, we calculated the corresponding binding free energies. As a result, the binding free protein Lys353. In the middle recognition region, only one hydrogen-bonding interaction with the ACE2 residue Gln42 (Fig. 3 B). Besides, the viral residue Tyr449, which turned to have a hydrogen-bonding interaction with the ACE2 residue Glu38 did not form hydrogen bonds with the viral residue Gln493. Also, the His-to-Ser transformation at position 34 interaction was detected between the ACE2 residue Ser34 and the viral residue Tyr83-Asn487; middle group, Ser34-Gln493; left group, Lys352-Gln498, and Lys352-Asn501 (Fig. 3 A). This observation indicated that the SARS-CoV-2 spike protein is weakly bound to both civet and pangolin ACE2 receptors. The weak binding of the spike protein to the civet ACE2 was also supported by the virus-infected experiments in HeLa cells (Zhou et al., 2020).

Compared with the human ACE2 molecule, the civet ACE2 had a substitution from phenylalanine to serine at position 40 in the first helix at the N-terminal lobe region. This mutation broke the pi-pi stacking interactions between the N-terminal helices, resulting in the side-chain rearrangement and loose helix-packing structures (Fig. 3 A). These structural alternations disrupted almost all the hydrogen bonds in the left recognition group, only leaving the ones formed by the civet ACE2 residue Lys352 and the viral residues Gln498 and Asn501 (Fig. 3 A). This observation was also supported by the data of residue contribution to the binding free energy. The energy contribution of the civet ACE2 residue Tyr41 to the binding free energy was increased to −2.96 kcal/mol, which was −4.51 kcal/mol in the human ACE2 receptor upon the SARS-CoV-2 spike binding (Additional file 5: Table S5). Additionally, the salt bridge between Lys31 and Glu35 disappeared in the civet ACE2 receptor due to the Asp-to-Glu transformation at position 30, leading to an unfavorable residue contribution of Lys31 to the binding free energy (4.32 kcal/mol for the civet ACE2 receptor and −0.98 kcal/mol in the human ACE2 receptor) (Additional file 5: Table S5).

Similar to the civet ACE2 receptor, there was also an Asp-to-Glu transformation in the pangolin ACE2 (at position 38 in the left recognition region). This transformation had a significant impact on the hydrogen bond network in the left recognition region. In the pangolin ACE2, the ACE2 residue Glu38 did not form hydrogen bonds with the viral residue Tyr449, which turned to have a hydrogen-bonding interaction with the ACE2 residue Gln42 (Fig. 3 B). Besides, the viral residue Gln498, which interacted with the ACE2 residue Tyr41 in the human ACE2, also formed a hydrogen bond with the pangolin ACE2 residue Lys353. In the middle recognition region, only one hydrogen-bonding interaction was detected between the ACE2 residue Ser34 and the viral residue Gln493. Also, the His-to-Ser transformation at position 34 in the pangolin ACE2 reduced its energy contribution to the binding free energy (−2.34 kcal/mol in the pangolin ACE2 and −3.36 kcal/mol in the human ACE2) (Additional file 5: Table S5).

### 2.5. Civet and pangolin ACE2 molecules bind to SARS-CoV-2 spike protein

Based on the receptor-binding complex models constructed for the SARS-CoV-2 spike binding to civet and pangolin ACE2, we calculated the corresponding binding free energies. As a result, the binding free energies of SARS-CoV-2 spike protein to civet and pangolin ACE2 receptors were −56.11 ± 2.42 kcal/mol and −54.78 ± 2.22 kcal/mol, showing a lower binding affinity than human ACE2 (Table 1). This observation indicated that the SARS-CoV-2 spike protein is weakly bound to both civet and pangolin ACE2 receptors. The weak binding of the spike protein to the civet ACE2 was also supported by the virus-infected experiments in HeLa cells (Zhou et al., 2020).

Compared with the human ACE2 molecule, the civet ACE2 had a substitution from phenylalanine to serine at position 40 in the first helix at the N-terminal lobe region. This mutation broke the pi-pi stacking interactions between the N-terminal helices, resulting in the side-chain rearrangement and loose helix-packing structures (Fig. 3 A). These structural alternations disrupted almost all the hydrogen bonds in the left recognition group, only leaving the ones formed by the civet ACE2 residue Lys352 and the viral residues Gln498 and Asn501 (Fig. 3 A). This observation was also supported by the data of residue contribution to the binding free energy. The energy contribution of the civet ACE2 residue Tyr41 to the binding free energy was increased to −2.96 kcal/mol, which was −4.51 kcal/mol in the human ACE2 receptor upon the SARS-CoV-2 spike binding (Additional file 5: Table S5). Additionally, the salt bridge between Lys31 and Glu35 disappeared in the civet ACE2 receptor due to the Asp-to-Glu transformation at position 30, leading to an unfavorable residue contribution of Lys31 to the binding free energy (4.32 kcal/mol for the civet ACE2 receptor and −0.98 kcal/mol in the human ACE2 receptor) (Additional file 5: Table S5).

Similar to the civet ACE2 receptor, there was also an Asp-to-Glu transformation in the pangolin ACE2 (at position 38 in the left recognition region). This transformation had a significant impact on the hydrogen bond network in the left recognition region. In the pangolin ACE2, the ACE2 residue Glu38 did not form hydrogen bonds with the viral residue Tyr449, which turned to have a hydrogen-bonding interaction with the ACE2 residue Gln42 (Fig. 3 B). Besides, the viral residue Gln498, which interacted with the ACE2 residue Tyr41 in the human ACE2, also formed a hydrogen bond with the pangolin ACE2 residue Lys353. In the middle recognition region, only one hydrogen-bonding interaction was detected between the ACE2 residue Ser34 and the viral residue Gln493. Also, the His-to-Ser transformation at position 34 in the pangolin ACE2 reduced its energy contribution to the binding free energy (−2.34 kcal/mol in the pangolin ACE2 and −3.36 kcal/mol in the human ACE2) (Additional file 5: Table S5).

### Table 1
Comparison of binding free energies (kcal/mol) for the ACE2 molecules from pets and wild animals.

| Organism          | Important ACE2 residues at the binding interface with the spike | ΔG (kcal/mol) |
|-------------------|-----------------------------------------------------------------|--------------|
| Human             | 19STIEEQAKTFLDKFNQHEALDFLSQSLASWNYYR52                        | −60.64 ± 3.10|
| Bamboo rat        | 19STIEEQAKTFLDKFNQHEALDFLSQSLASWNYYR52                        | −65.04 ± 3.65|
| Mole              | 19LIEEQAKTFLDKFNQHEALDFLSQSLASWNYYR52                         | −59.67 ± 2.95|
| Vole              | 19STIEEQAKTFLDNKFNQHEALDFLSQSLASWNYYR52                       | −57.18 ± 2.57|
| Mus pahari        | 19STIEEQAKTFLDNKFNQHEALDFLSQSLASWNYYR52                       | −56.88 ± 3.13|
| Palen civet       | 19STIEEQAKTFLDNKFNQHEALDFLSQSLASWNYYR52                       | −56.11 ± 2.42|
| Rat               | 19STIEEQAKTFLDNKFNQHEALDFLSQSLASWNYYR52                       | −56.01 ± 2.55|
| Mus musculus      | 19STIEEQAKTFLDNKFNQHEALDFLSQSLASWNYYR52                       | −55.59 ± 2.79|
| Pangolin          | 19STIEEQAKTFLDNKFNQHEALDFLSQSLASWNYYR52                       | −54.78 ± 2.22|
| Mus caroli        | 19STIEEQAKTFLDNKFNQHEALDFLSQSLASWNYYR52                       | −50.90 ± 2.62|
| Snake             | 44QDETKVATRFLEQDARATDLYNASIASSWQDYNT67                        | > 800         |
| Neovison vison    |                                                                  |              |

* positions of residues in the ACE2 sequence are denoted with numbers before the first and after the last amino acid residues. Sequence accession numbers are included in Additional file 6: Table S6.

Fig. 3. The binding interface of the SARS-CoV-2 spike protein to civet and pangolin ACE2 receptors. (A) SARS-CoV-2 spike RBD (light blue) binding to the civet ACE2 receptor (yellow). The hydrogen-bonding interactions in the recognition groups are labeled as gray dash lines: right group, Ser19-Ala475; middle group, Glu35-Gln493 and His34-Tyr453; left group, Lys352-Gln498, and Lys352-Asn501 (Fig. 3 A).
2.6. Binding affinities of wild mouse ACE2 to SARS-CoV-2 spike protein

Rodents have been regarded as intermediate hosts or carriers of many human pathogens (Luis et al., 2013), and murine hepatitis virus (MHV) has been used as a model system to study coronaviruses (De Albuquerque et al., 2006). Biological studies on alpha- and beta-coronaviruses in rodents indicated the murine origin of some beta-coronaviruses (Lau et al., 2015; Ge et al., 2017). We constructed the ACE2 of the wild mice from Yunnan province, i.e., mole, vole, rat, mice, and bamboo rat, and employed free energy calculations to estimate their binding affinities to the SARS-CoV-2 spike. Most mouse ACE2 receptors (including rats and mice) showed higher binding free energies than the human ACE2 (Table 1), indicating weaker binding to the SARS-CoV-2 spike. However, the mole ACE2 had a similar binding free energy to the human ACE2 due to their high sequence identity at the binding interface to the SARS-CoV-2 spike.

Notably, the bamboo rat ACE2 showed a lower binding free energy (−65.04 kcal/mol on average) than the human ACE2 (−60.64 kcal/mol on average), indicating a stronger inclination for the SARS-CoV-2 spike binding to the bamboo rat ACE2. Looking at their binding interface, the major differences between bamboo rat and human ACE2 molecules were caused by the His-to-Gln transformation at position 34 and the Asp-to-Glu transformation at position 38 in the N-terminal helix (Table 1). These changes in the bamboo rat ACE2 led to a side-chain rearrangement at the binding interface to the SARS-CoV-2 spike. As a result, the hydrogen bonding between the human ACE2 residue Asp38 and the viral residue Tyr449 disappeared in the bamboo rat ACE2 receptor, and instead, the bamboo rat ACE2 residue Gln34 had a hydrogen-bonding interaction with the viral residue Ser452 (Fig. 4). Furthermore, the bamboo rat ACE2 residue Glu35 formed two hydrogen bonds with the viral residues Gln493 and Phe488, and the aromatic sidechain of the bamboo rat ACE2 residue Tyr41 employed a pi-pi stacking interaction with the viral residue Tyr449 (Fig. 4). These additional interactions enhanced the SARS-CoV-2 spike binding to the bamboo rat ACE2.

2.7. Neovison vison or snake ACE2 does not bind to SARS-CoV-2 spike protein

Due to the strong positive binding free energies (> 400 kcal/mol), the SARS-CoV-2 spike protein did not bind to Neovison vison and snake ACE2 (Table 1). The unfavorable binding free energies for Neovison vison and snake ACE2 molecules were caused by their structural gaps in the N-terminal helices, which were essential for receptor recognition. According to the sequence alignment (Additional file 7: Fig. S1), a large segment in the N-terminal region (residues 1–318 in the human ACE2 receptor) was missing in the Neovison vison ACE2. The residues 19–52 in the human ACE2 contained the important residues interacting with the receptor-binding motif of the SARS-CoV-2 spike protein, which was stabilized by the second helix in the N-terminal lobe region (residues 53–84).

The unfavorable binding free energy for the snake ACE2 was caused by the large dissimilarity to the human ACE2 in protein sequence at the N-terminal lobe region (Table 1). In the human ACE2, two virus-binding hot spots have been identified at the binding interface of the SARS-CoV receptor-binding-domain (RBD). One centers on the ACE2 residue Lys31 (hot spot 31), forming a salt bridge with the ACE2 residue Glu35 (Fig. 2B). Another is on the ACE2 residue Lys353 (hot spot 353), forming a salt bridge with the ACE2 residue Asp38. However, the Lys31 and Glu35 in human ACE2 for the virus-binding hot spot 31 were mutated to glutamine (Gln56) and arginine (Arg60) in the snake ACE2, which broke the significant salt bridge. Besides, the side-chain conformation in the snake ACE2 receptor had an adverse effect on the SARS-CoV-2 spike binding. For example, the snake ACE2 residues Asp64 and Glu352 showed strong van der Waals conflicts with Tyr488 and Tyr505 in the SARS-CoV-2 spike.

3. Conclusions

We modeled SARS-CoV-2 spike-ACE2 complexes and calculated binding free energies to evaluate spike protein mutations occurring during COVID-19 pandemics, and assess their binding affinities. We found 1573 missense mutations in the spike gene, of which 226 were within the RBD region that interacts with the human ACE2. We found 74 of them in the region of interaction interface between RBD and ACE2. Modeling the interactions between SARS-CoV-2 spike mutants and ACE2 molecules indicated that several missense mutations within the spike RBD region increased the binding affinity toward human ACE2 thus making the virus likely more contagious, while most of the mutations had little impact on the spike ACE2 interaction. On the other hand, modeling the interactions between animal (pet and wild animal) ACE2 molecules and SARS-CoV-2 spike revealed that many pets and wild animals’ ACE2 had a variable binding ability. Particularly, ACE2 of Bamboo rat had stronger binding to SARS-CoV-2 spike protein, whereas that of mole, vole, Mus pahari, Palm civet, and Pangolin had a weaker binding compared to human ACE2. It is worth stressing that experiments are needed to validate the results from the in silico analysis. Our study provides structural insights into the impact on interactions of the SARS-CoV-2 spike mutants to human ACE2 and sheds light on SARS-CoV-2 transmission in pets and wild animals, and possible clues to the intermediate host(s) for SARS-CoV-2.

4. Methods

4.1. Structural models of SARS-CoV-2 spike mutants from human patients in complex with the ACE2

The structural models of SARS-CoV-2 spike mutants were constructed based on the crystal structure of the wild type SARS-CoV-2 spike (PDB entry 1m0j) by MODELLER. The complexes of SARS-CoV-2 mutants and ACE2 were then modeled by fitting the SARS-CoV-2 spike mutant into the crystal structure of the SARS-CoV-2 spike-ACE2 complex.
human ACE2 receptor (PDB entry 1r42) as a structural template (Towler SWISS-MODEL (Waterhouse et al., 2018) with the crystal structure of the minimized by a short molecular dynamics simulation (about 1 nanosecond (ns)) with the protein backbone fixed, and then without any limitation for another ns.

4.2. Structural models of ACE2 from pets and wild animals in complex with the SARS-CoV-2 spike

The ACE2 protein sequences for human (Homo sapiens), and animals: civet (Paguma larvata), pangolin (Manis javanica), Neovison vison, snake (Ophiophagus hannah), bamboo rat (Rhizomys pruinosus), mole (Fukomys damarensis), vole (Microtus ochrogaster), mice (Mus pahari, Mus caroli, and Mus musculus), rat (Rattus norvegicus), etc, were obtained from the National Center for Biotechnology Information (NCBI) GenBank. Building the structural models of ACE2 molecules interacting with spike protein was previously described (Wang et al., 2020). In brief, the structural models for these ACE2 molecules were constructed by SWISS-MODEL (Waterhouse et al., 2018) with the crystal structure of the human ACE2 receptor (PDB entry 1r42) as a structural template (Towler SWISS-MODEL (Waterhouse et al., 2018)). The SARS-CoV-2 spike binding models were then built by fitting the ACE2 structural models into the crystal structure of the SARS-CoV-2 spike-ACE2 complex (PDB entry 6m0j) (Lan et al., 2020). These complexes were finally minimized by a short molecular dynamics simulation (about 1 nanosecond (ns)) with the protein backbone fixed, and then without any limitation for another ns.

4.3. Molecular dynamics simulation protocols

The initial structures of ACE2-SARS-CoV-2 complexes were prepared by the LEaP module in the Amber 18 suite and parameterized by Amber ff14SB force field parameters (Maier et al., 2015), in which all the non-polar hydrogen atoms were removed. The pKa values for each residue were calculated by Delphi (Georgescu et al., 2002) as a Poisson Boltzmann solver with a dielectric constant of 4.0. The hydrogen atoms were then re-added to the protein structures using the t-leap module implemented in Amber 18 suite based on the computational pKa values. After neutralized and solvated in the explicit TIP3P water surroundings, all the complexes were subjected to energy minimization by steepest descent for 5000 steps, followed by conjugate gradient for the next 5000 steps and equilibrated with all the protein backbone atoms fixed by a short molecular dynamics (MD) simulations (about 1 ns) to reduce the van der Waals conflicts in the simulation systems. 200-ns MD simulations were finally performed with periodic boundary conditions and NPT ensemble (310 K and 1 bar) by Amber 18 package (Case et al., 2005). Langevin thermostat and isotropic pressure scaling were used to control the temperature and the pressure in 310 K and 1 bar with coupling times of 0.1 ps and 1.0 ps. SHAKE algorithm was applied to constrain all the chemical bonds in the simulation systems and the atom velocities for the start-up run were obtained based on the Maxwell distribution at 310 K. The particle mesh Ewald (PME) algorithm was used to treat long-range electrostatic interactions and the van der Waals interactions were calculated with a cut-off of 12 Å. The time step of all the simulations was set to 2 femtoseconds (fs), and the structural coordinates were saved every 1 picosecond (ps).

4.4. Estimation of binding free energy for SARS-CoV-2 spike and ACE2 molecules

The binding affinities of the SARS-CoV-2 spike to the ACE2 receptors were estimated by the MM-PB/SA method (Miller et al., 2012). In the MM-PB/SA method, the binding free energy of the SARS-CoV-2 spike (deltaG) was calculated as the energy difference between the free energies of ACE2, SARS-CoV-2 spike, and their complex (Eq. (1)).

\[ \Delta G = G_{\text{complex}} - (G_{\text{ACE2}} + G_{\text{spike}}) \]  
(1)

As shown in Eq. (2), these free energies were estimated by summing the internal energy in the gas phase \( G_{\text{gas}} \), solvation free energy \( G_{\text{sol}} \), and a vibrational entropy term \( TS_{\text{conf}} \). The internal energy in the gas phase \( G_{\text{gas}} \) was a standard force field energy computed from Eq. (3) by the strain energies from covalent bonds \( E_{\text{bond}} \) and \( E_{\text{ang}} \), torsion angles \( E_{\text{vdw}} \), van der Waals \( E_{\text{ele}} \), and electrostatic energies \( E_{\text{ele}} \). The solvation free energy \( G_{\text{sol}} \) was calculated by electrostatic \( G_{\text{ele}} \) and non-polar \( G_{\text{non-polar}} \) energies (Eq. (4)). The former was obtained from the Poisson–Boltzmann approach, and the latter was estimated by molecular solvent accessible surface areas (SASAs). Because all the simulation systems involved in the current study have similar entropy (due to the similar atom numbers in the simulation system), we neglected the entropy term to compare the binding affinities of SARS-CoV-2 spike in different ACE2 receptors. A total of 500 snapshots were used for MM-GB/SA calculations, which were abstracted from the final 5-ns MD trajectories with an interval of 10 ps.

Funding

This work is supported in part by grants from the National Key Research and Development Program of China (2018YFA0900070), the National Science and Technology Major Projects (2018ZX09711000), the National Natural Science Foundation of China (31701137, 31900470), and funds from Derivative Bank of Chinese Biological Resources from Chinese Academy of Sciences (KFJ-BRP-009-002).

CRediT authorship contribution statement

Pei Hao, Ming Li, Xuan Li: Concepted the project. Ping Chen, Jingfang Wang, Xintian Xu: Analyzed the data and gathered the results. Yuping Li, Yan Zhu: Helped collect data and discuss the results. Ping Chen, Xuan Li, Jingfang Wang, Pei Hao: Wrote the manuscript. All authors have read and approved the final manuscript.

Ethics declarations

Ethics approval and consent to participate Not applicable.

Consent for publication

Not applicable.

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.

Acknowledgments

We would like to thank Dr. Zhang Zhang for the helpful discussion concerning the analysis of SARS-CoV-2 spike mutants. We acknowledge GISAID (https://www.gisaid.org/) for facilitating open data sharing for SARS-CoV-2 sequences. Thanks go to the editor and the anonymous reviewers for their comments and suggestions.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.compbiolchem.2021.107613.
References

Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F., 2020. A novel coronavirus outbreak of global health concern. N. Engl. J. Med. 382, 1953–1954.

WHO COVID-19 Weekly Epidemiological Update - 5 October 2021. (https://www.who.int/publications/m/item/weekly-epidemiological-update-on-2019-ncov-5-october-2021).

Zhao, P., Yang, X.L., Wang, Y.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, X., Chen, Q.J., Deng, F., Liu, H.L., Yan, B., Zhan, F.K., Wang, Y.Y., Xiao, G.F., Shi, Z.L., 2020: A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273.

Li, F., Li, W., Farzan, M., Harrison, S.C., 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309, 1864–1868.

Raj, V.S., Mou, H., Smits, S.L., Dekkers, D.H., Müller, M.A., Dijkman, R., Muth, D., Demmers, J.A., Zaki, A., Fouchier, R.A., Thié, V., Drosten, C., Rottier, P.J., Osterhaus, A.D., Bosch, B.J., Haagmans, B.L., 2013. Dipeptidyl peptide 4 is a functional receptor for the emerging human coronavirus-EMC. Nature 495, 251–254.

Xu, X., Chen, P., Wang, J., Feng, J., Zhu, H., Li, X., Wang, Z., Hao, P., 2020: Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci. China Life Sci. 63, 457–460.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Li, F., Li, W., Farzan, M., Harrison, S.C., 2020. Correlation of receptor-binding and cell entry activities of SARS-CoV-2 spike glycoprotein with disease severity. Science 368, 1016–1020.

P. Chen et al. Computational Biology and Chemistry 96 (2022) 107613

No. of missing pages: 1

Page 7