Molecular simulation of SARS-CoV-2 spike protein binding to pangolin ACE2 or human ACE2 natural variants reveals altered susceptibility to infection

Jingfang Wang†, Xintian Xu‡, Xinbo Zhou‡, Ping Chen‡, Huiying Liang‡, Xuan Li‡, Wu Zhong‡ and Pei Hao†,‡,§

Abstract

We constructed complex models of SARS-CoV-2 spike protein binding to pangolin or human ACE2, the receptor for virus transmission, and estimated the binding free energy changes using molecular dynamics simulation. SARS-CoV-2 can bind to both pangolin and human ACE2, but has a significantly lower binding affinity for pangolin ACE2 due to the increased binding free energy (9.5 kcal mol\(^{-1}\)). Human ACE2 is among the most polymorphic genes, for which we identified 317 missense single-nucleotide variations (SNVs) from the dbSNP database. Three SNVs, E329G (rs143936283), M82I (rs267606406) and K26R (rs46466116), had a significant reduction in binding free energy, which indicated higher binding affinity than wild-type ACE2 and greater susceptibility to SARS-CoV-2 infection for people with them. Three other SNVs, D355N (rs961360700), E37K (rs146676783) and I21T (rs1244687367), had a significant increase in binding free energy, which indicated lower binding affinity and reduced susceptibility to SARS-CoV-2 infection.

As of 13 May 2020, coronavirus disease 2019 (COVID-19) has caused close to 4170000 infections and more than 287000 deaths worldwide [1]. We first reported that the novel coronavirus SARS-CoV-2 (2019-nCoV), the causative agent for the epidemic, is transmitted between humans via the spike protein–angiotensin-converting enzyme II (ACE2) interaction pathway, which is the mechanism through which 2003 SARS-CoV infects human cells [2]. Human ACE2 was previously found to be the functional receptor for SARS-CoV transmission between humans [3]. A recent published study showed that SARS-CoV-2 infected HeLa cells in an ACE2 expression-dependent manner [4]. The C-terminal RBD domain of the spike protein binds to human ACE2 through interactions with its N-terminal peptidase M2 domain [5]. In ACE2, 18 amino acid residues located in 4 regions around the N-terminal alpha-helical lobes were critical for interaction with the RBD domain of SARS-CoV spike protein [6]. Coronaviruses closely related to SARS-CoV-2 were identified in pangolins (*Manis javanica*), which were considered to be a candidate for an intermediate host [7]. Our analysis showed that pangolin-associated betacoronaviruses assembled from metagenomics sequences had ~90% sequence identity to SARS-CoV-2, lower than the 96.2% sequence identity between SARS-CoV-2 and the bat coronavirus RaTG13 detected in *Rhinolophus affinis* [4]. To address the question of whether pangolins might be the intermediate host for SARS-CoV-2, we first examined whether SARS-CoV-2 spike protein can bind to pangolin ACE2, the orthologue of the human receptor for SARS-CoV-2. If it can, how strong is the interaction compared to that between SARS-CoV-2 spike protein and human ACE2?

We first constructed the complex models of SARS-CoV-2 spike protein binding to pangolin and human ACE2 based
on sequence alignment (Fig. S1, available in the online version of this article), as described previously [8] (Fig. 1a, middle and left panels). The computational model of the human ACE2–spike complex was in excellent agreement with the recent crystal structures (PDB entry 6vw1) [9] and cryo-EM structures (PDB entry 6m17) [10]. The backbone root-mean-square deviation (RMSD) of the computational model for ACE2 and spike was only about 0.83 and 1.52 Å from the cryo-EM structure and 1.45 and 1.40 Å from the crystal structure (Fig. S2). To estimate the binding free energy for the complexes, we used molecular dynamics simulation (MDS) to refine the structure of the complexes.

In brief, the initial structure of the complexes was treated using the LEaP module in the Amber 18 suite and parameterized using Amber ff14SB force field parameters [11]. After being neutralized and solvated in explicit TIP3P water surroundings, the complexes were subjected to energy minimization. MDS was performed for 1 ns with periodic boundary conditions and NPT ensemble (310 K and 1 bar) with the Amber 18 package [12]. The RMSD values for the backbone atoms of both ACE2 and spike proteins were calculated, which indicated that the simulation systems were equilibrated (Table S1). The binding free energy of the complexes was computed using molecular mechanics.
through the generalized Born surface area (MM-GB/SA) method [13]. The binding free energy for pangolin and human ACE2–spike binding complexes was averaged from 100 simulated configurations.

The binding free energy for human ACE2 and the SARS-CoV spike was $-60.64 \pm 9.10$ kcal mol$^{-1}$, nearly 2.1 kcal mol$^{-1}$ lower than that of human ACE2 and the SARS-CoV-2 spike ($-58.55 \pm 8.75$ kcal mol$^{-1}$). As $\Delta G = RT \ln K_d$, the binding free energy change of 2.1 kcal mol$^{-1}$ corresponds to a nearly 30-fold dissociation constant, $K_d$. The surface plasmon resonance results indicated that the binding affinity of ACE2 to the SARS-CoV-2 spike was 20-fold higher than that for ACE2 binding to the SARS-CoV spike. Thus, the computational $\Delta G$ is in good agreement with the experimental binding affinity [14].

As a result, compared to the human ACE2–spike binding complex, the pangolin ACE2–spike complex had a binding free energy increase of 9.5 kcal mol$^{-1}$. This indicates that compared to human ACE2, pangolin ACE2 had a significantly lower binding affinity to the SARS-CoV-2 spike protein. In addition, the $\Delta \Delta G$ we calculated for the pangolin ACE2–spike complex is consistent with the results calculated independently by a different study [15]. Looking more closely into the contribution of different residues to the binding free energy increase in the pangolin ACE2–spike complex, the mutations E24 and E30 increased binding free energy by 2.14 and 1.12 kcal mol$^{-1}$ (Fig. 1a, left panel). Although mutation S40 had very little contribution to binding free energy increase, it caused neighbouring Y41 and Q42 to both have increased binding free energy. These accurate results indicated that SARS-CoV-2 can bind to both pangolin and human ACE2, but has a significantly lower binding capacity for pangolin ACE2. These results would indicate that pangolins are less susceptible to SARS-CoV-2 than the human host. If pangolins were the intermediate host, this might suggest that SARS-CoV-2 may have adapted to the human host during animal-to-human transmission.

While the COVID-19 epidemic was spreading globally, there was an indication of likely differences in viral transmissibility, vulnerability and clinical severity for SARS-CoV-2 in different countries or regions. For example, very few locally transmitted COVID-19 cases have surfaced across African region to date, whereas the number of confirmed cases rose quickly in other regions [1]. Genetic variation is one of the important factors affecting the spread and severity of epidemics. In addition to genetic variation, people's social behaviour, virus strain and mutation rate also have significant impacts on viral transmission rates. While those involved in SARS-CoV-2 transmission, clinical symptoms and prognosis have been poorly understood, the genetic variants of ACE2, the human receptor for SARS-CoV-2, stood out as probably the most influential genetic factor for vulnerability to SARS-CoV-2 and clinical progress of COVID-19. The natural genetic variants of ACE2 may alter human susceptibility or response to SARS-CoV-2 infection, which is an important question to address in this the health emergency.
to the binding free energy from $-0.06$ to $3.26$ kcal mol$^{-1}$. Similarly, D355N mutation increased the electrostatic contribution to the binding free energy from $-7.02$ (wild-type) to $-2.50$ kcal mol$^{-1}$, an increase of $-4.52$ kcal mol$^{-1}$. Based on GnomAD exomes data, SNV rs146676783 (E37K) had an AF of 0.0001 in both European and African populations. It would reduce susceptibility to SARS-CoV-2 infection for people who have it.

Additionally, we searched in the 2019 Novel Coronavirus Resource at the National Genomics Data Center (NGDC) (https://bigd.big.ac.cn/ncov/variation/annotation) for mutant SARS-CoV-2 strains. We found that to date SARS-CoV-2 mutants have mostly occurred in the orf1ab gene, with very few in the spike gene. Although the chance of them directly affecting interaction with the ACE2 variant is small, it would be interesting to model how the spike mutant residues affect spike interaction with human ACE2 and ACE2 variants when more spike mutation data have been accumulated.

In summary, we constructed the complex models of SARS-CoV-2 spike protein binding to pangolin or human ACE2, and estimated the binding free energy changes using MDS. SARS-CoV-2 can bind to both pangolin and human ACE2, but has a significantly lower binding affinity for pangolin ACE2, resulting from the increased binding free energy of 9.5 kcal mol$^{-1}$ for pangolin ACE2 compared to human ACE2. ACE2 is among the most polymorphous genes in human populations, for which we identified 317 missense SNVs from dbSNP database. Several missense ACE2 SNVs were found to significantly alter binding capacity to the SARS-CoV-2 spike compared to wild-type ACE2. Three SNVs, E329G (rs143936283), M82I (rs267606406) and K26R (rs4646116), had a significant reduction in binding free energy, which indicates higher binding affinity than wild-type ACE2 and greater susceptibility to SARS-CoV-2 infection. Three other SNVs, D355N (rs961360700), E37K (rs146676783) and I21T (rs1244687367), had a significant increase in binding free energy, indicating lower binding affinity and reduced susceptibility to infection. These findings via molecular simulation provide a framework for future studies to experimentally validate the susceptibility of ACE2 natural variants to SARS-CoV-2 infection.

Acknowledgement
We would like to thank Dr Weiping Zhou for helpful discussions concerning analysis of human ACE2 genetic variants.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References
1. WHO. Coronavirus disease (COVID-19) Situation Report – 114. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200513-covid-19-sitrep-114.pdf 2020.
2. Xu X, Chen P, Wang J, Feng J, Zhou H et al. 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 2020.
3. Kuhn JH, Li W, Choe H, Farzan M. Angiotensin-Converting enzyme 2: a functional receptor for SARS coronavirus. Cell Mol Life Sci 2004,61:2738–2743.
4. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020;579:270–.
5. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 2005;309:1864–1868.
6. Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends Microbiol 2015;23:468–478.
7. Liu P, Chen W, Chen J-P. Viral metagenomics revealed sentai virus and coronavirus infection of malayan pangolins (manis javanica). Viruses,11:979.
8. Xu X, Chen P, Wang J, Feng J, Zhou H et al. 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 2020;63:457–460.
9. Shang J, Ye G, Shi K, Wan Y, Luo C et al. Structural basis of receptor recognition by SARS-CoV-2. Nature 2020;581:221–.
10. Yan R, Zhang Y, Li Y, Xia L, Guo Y et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 2020;367:1444–1448.
11. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 2015;11:3696–3713.
12. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R et al. The amber biomolecular simulation programs. J Comput Chem 2005;26:1668–1688.
13. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H et al. MMPBSA.py: an efficient program for end-state free energy calculations. J Chem Theory Comput 2012;8:3314–3321.
14. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L et al. Cryo-Em structure of the 2019-nCoV spike in the prefusion conformation. Science 2020;367:1260–1263.
15. Ou J, Zhou Z, Zhang J, Lan W, Zhao S et al. Rbd mutations from circulating SARS-CoV-2 strains enhance the structure stability and infectivity of the spike protein. bioRxiv 2020:2020.03.15.991845.
16. Sherry ST. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–311.
17. Lu N, Yang Y, Wang Y, Liu Y, Fu G et al. ACE2 gene polymorphism and essential hypertension: an updated meta-analysis involving 11,051 subjects. Mol Biol Rep 2012;39:6581–6589.