

## Fast assessment of human receptor-binding capability of 2019 novel coronavirus (2019-nCoV)

Qiang Huang<sup>\*1</sup>, and Andreas Herrmann<sup>\*2</sup>

<sup>1</sup> State Key Laboratory of Genetic Engineering, MOE Engineering Research Center of Gene Technology, School of Life Sciences, Fudan University, Shanghai, 200438, China

<sup>2</sup> Institute for Biology and IRI Lifesciences, Humboldt-Universität zu Berlin, 10115 Berlin, Germany

\*To whom correspondence should be addressed. E-mail: huangqiang@fudan.edu.cn (Q.H.); andreas.herrmann@rz.hu-berlin.de (A.H.)

### Abstract

The outbreaks of 2002/2003 SARS, 2012/2015 MERS and 2019/2020 Wuhan respiratory syndrome clearly indicate that genome evolution of an animal coronavirus (CoV) may enable it to acquire human transmission ability, and thereby cause serious threats to global public health. It is widely accepted that CoV human transmission is driven by the interactions of its spike protein (S-protein) with human receptor on host cell surface; so, quantitative evaluation of these interactions may be used to assess the human transmission capability of CoVs. However, quantitative methods directly using viral genome data are still lacking. Here, we perform large-scale protein-protein docking to quantify the interactions of 2019-nCoV S-protein receptor-binding domain (S-RBD) with human receptor ACE2, based on experimental SARS-CoV S-RBD-ACE2 complex structure. By sampling a large number of thermodynamically probable binding conformations with Monte Carlo algorithm, this approach successfully identified the experimental complex structure as the lowest-energy receptor-binding conformations, and hence established an experiment-based strength reference for evaluating the receptor-binding affinity of 2019-nCoV via comparison with SARS-CoV. Our results show that this binding affinity is about 73% of that of SARS-CoV, supporting that 2019-nCoV may cause human transmission similar to that of SARS-CoV. Thus, this study presents a method for rapidly assessing the human transmission capability of a newly emerged CoV and its mutant strains, and demonstrates that post-genome analysis of protein-protein interactions may provide early scientific guidance for viral prevention and control.

## Introduction

Coronaviruses (CoV) are positive-stranded enveloped RNA viruses belonging to the Nidovirales order, and are classified into four genera:  $\alpha$ ,  $\beta$ ,  $\delta$  and  $\gamma$ . Two of the  $\beta$ -CoVs, the severe acute respiratory syndrome CoV (SARS-CoV) and the Middle East respiratory syndrome CoV (MERS-CoV), have caused serious epidemic-like outbreaks. In 2002, SARS emerged in southern China and was spreading in 2003 around the world resulting in ~8,000 cases with a fatality rate of about 10%(WHO, 2004). Ten years later, MERS emerged in Saudi Arabia and spread to different countries with a fatality rate of about 35%(WHO, 2016). Both viruses are of zoonotic nature: SARS-CoV originates from bats/civet cats(Ge et al., 2013), and dromedary camels are the major reservoir host of MERS(Reusken et al., 2016). Nevertheless, both viruses are known to transmit from human to human(de Wit et al., 2016). A few other CoVs of zoonotic origin are circulating endemic in the human population, which cause mild infections of the respiratory tract, such as HCoV-OC43 and HCoV-HKU1(Gaunt et al., 2010).

In December, 2019, a novel CoV (2019-nCoV) has emerged from Wuhan in China, and spread rapidly from this area to the globe(Holshue et al., 2020; Li et al., 2020). Although its original host remains unknown, all available data point again to a zoonotic, a wild animal source. And, it is evident that a human-to-human transmission is causing the rapid spreading of this virus. Viral strains from infected persons of the area have been sequenced arguing for a common ancestor as there was only little genetic variation found(Rambaut, 2020). Full-length genome analysis of various 2019-nCoV patient samples revealed a 79.5% and 96% sequence identity to SARS-CoV and a bat SARSr-CoV (severe acute respiratory syndrome-related CoV), respectively; however, sequences of the seven conserved viral replicase domains in ORF1ab show 94.6% identity between 2019-nCoV and SARS-CoV(Zhou et al., 2020)

It is widely accepted that the CoV human transmission is driven by the interactions of its trimeric transmembrane spike glycoprotein (S-protein) with the human receptors on the host cell surface. The S-protein consists of two subunits (i.e., S1 and S2) and is essential for the early steps of host cell infection: S-protein of a virion binds to the host cell receptors via its S1 receptor-binding domain (S-RBD) and then fuses the viral membrane with cellular membranes. The host cell receptors differ among human CoVs: SARS-CoV binds to exopeptidase angiotensin-converting enzyme 2(ACE2)(Li et al., 2005; Li et al., 2003), and MERS-CoV to dipeptidyl peptidase 4 (DPP4)(Raj et al., 2013). Other CoVs such as HCoV-HKU1 use O-acetylated sialic acid as a receptor (Huang et al., 2015).

Sequence alignment of S1-proteins has revealed a sequence identity of about 75% between that of 2019-nCoV and those of several other CoVs including SARS-CoV(Zhou et al., 2020). Comparing with that of SARS-CoV, the S-RBD sequence of 2019-nCoV S-protein is very conserved, implying that ACE2 is the specific receptor for this virus to bind to the human host cell(Lu et al., 2020; Xu et al., 2020). Indeed, recent experimental evidences also supported that ACE2 is the human receptor of 2019-nCoV(Zhao et al., 2020; Zhou et al., 2020). Consequently, it becomes very interesting to characterize the human receptor-binding capability of 2019-nCoV by analyzing the S-protein interactions with ACE2, in order to assess the potential of this newly emerged CoV for the human-to-human transmission.

Here, we employ protein-protein docking approach to quantify the interactions of 2019-nCoV S-RBD with ACE2, based on the experimentally determined complex structure of SARS-CoV S-RBD complexed with ACE2(Li et al., 2005) and on the predicted structure of 2019-nCoV S-protein RBD according to gene sequence. Computational protein-protein docking with accurate, physics-based energy functions is able to reveal the native-like, low-energy protein-protein complex from the unbound structures of two individual, interacting protein components(Chaudhury et al., 2011), such as antibody-antigen interacting structures(Weitzner et al., 2017), and thereby to calculate the binding free energy of the investigated protein-protein interactions. To this end, large-scale local protein-protein docking was first carried out to sample a large number of thermodynamically probable binding conformations for the interaction of S-RBD with ACE2. Next, the binding free energy of RBD to ACE2 was calculated according to the thermodynamic average of the sampled, low-energy binding conformations. Our results show that the human-receptor binding affinity of 2019-nCoV is about 73% of that of SARS-CoV, suggesting that 2019-nCoV is able to bind to ACE2, and thereby to cause human infections similar to that of SARS-CoV.

## Results and Discussion

### Binding interactions of SARS-CoV S-RBD with human ACE2

To quantify the binding interactions of CoV S-protein with ACE2, we employed the well-established method of local protein-protein docking implemented in Rosetta software suite(Das and Baker, 2008). Using this approach, we sampled a large number of low-energy conformations of S-protein RBD in complex with ACE2 by performing multiple, independent docking runs (i.e., simulations). Physically,

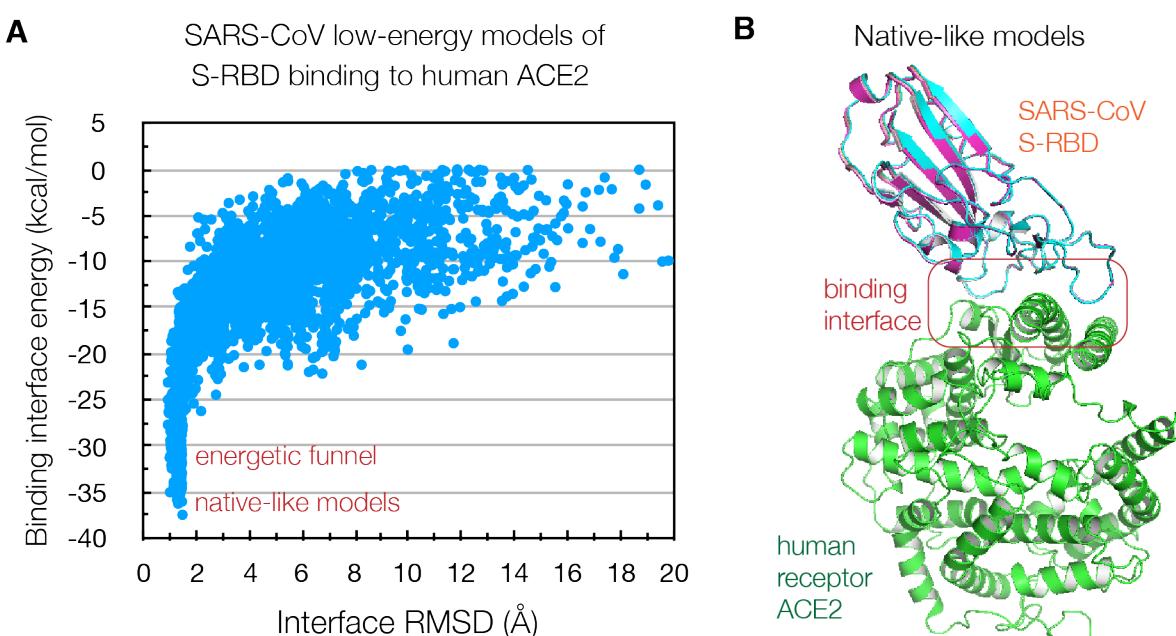
the sampled conformations are considered to be accessible in the real-world process of S-protein binding to ACE2, according to their thermodynamic probabilities. The latter are determined by the conformational energies with the physically realistic, all-atom energy function for macromolecular modeling and design implemented in Rosetta(Alford et al., 2017).

To validate the docking approach, we first carried out local protein-protein docking to investigate the binding interactions of SARS-CoV S-RBD with ACE2, in order to test whether this approach could identify the lowest-energy conformation of S-protein RBD complexed with ACE2 identical or close to the experimental structure(Li et al., 2005). The latter is considered to be the most probable, native binding conformation in reality. As described in Methods, similar to the studies of (Chaudhury et al., 2011), the co-crystal structure of SARS-CoV S-RBD in complex with ACE2 (PDB 2AJF) was used as the starting complex for the local protein-protein docking. Then, more than 3,000 independent docking runs were conducted to sample the low-energy binding conformations by optimizing rigid-body orientation and side-chain conformations, according to the docking protocols in Methods. The binding free-energy scores of the obtained complex models are shown in the scatter plot in Fig. 1A, against their RMSD (root mean square deviation) values of interface residues with respect to those in the starting crystal structure (PDB 2AJF).

As seen, a deep energetic funnel exhibits in the plot, and therefore clearly indicates that the docking simulations are very successful, because the presence of such a funnel is the most robust measure of successful docking simulations(Chaudhury et al., 2011). Due to their locations at the funnel bottom, the top-scoring models have better binding free-energy scores than other low-energy models, and therefore are considered to be the most probable, native-like binding conformations in thermodynamics. Significantly, the number of the native-like models among the top five scores ( $N_5$ ) is 5; also, their interface RMSDs are within 1.5 Å. Therefore, the lowest-energy conformations identified by the docking approach are almost identical to the experimental structure (PDB 2AJF). This provides a very solid basis for using this approach to investigate the S-RBD-ACE2 interactions of 2019-nCoV, in which no experimental complex structure is available because of time limitation, but only the first-determined genome data that can be used to predict the S-protein structure.

The successful identification of the experimental complex structure as the lowest-energy binding conformations also validates the accuracy of the Rosetta all-atom energy function for quantifying the interactions of S-protein RBD with ACE2. Therefore, we could use the average binding energy scores

of these conformations as an experiment-based energy reference to characterize the binding strength of 2019-nCoV S-RBD to ACE2 and to deduce its binding strength relative to that of SARS-CoV. Obviously, such a comparison assessment avoids using the calculated, absolute free-energy scores to correlate with the binding strength, which might systematically differ from the actual measurement values because of the intrinsic, theoretical approximations of the used Rosetta all-atom energy function (Alford et al., 2017). We want to point out that the docking simulations for SARS-CoV and 2019-nCoV use the same protocols and a consistent all-atom energy function. So, even the calculated energy scores contain certain errors caused by theoretical approximations, determination of the relative binding strength by comparing the calculated scores of two viruses should be very accurate.

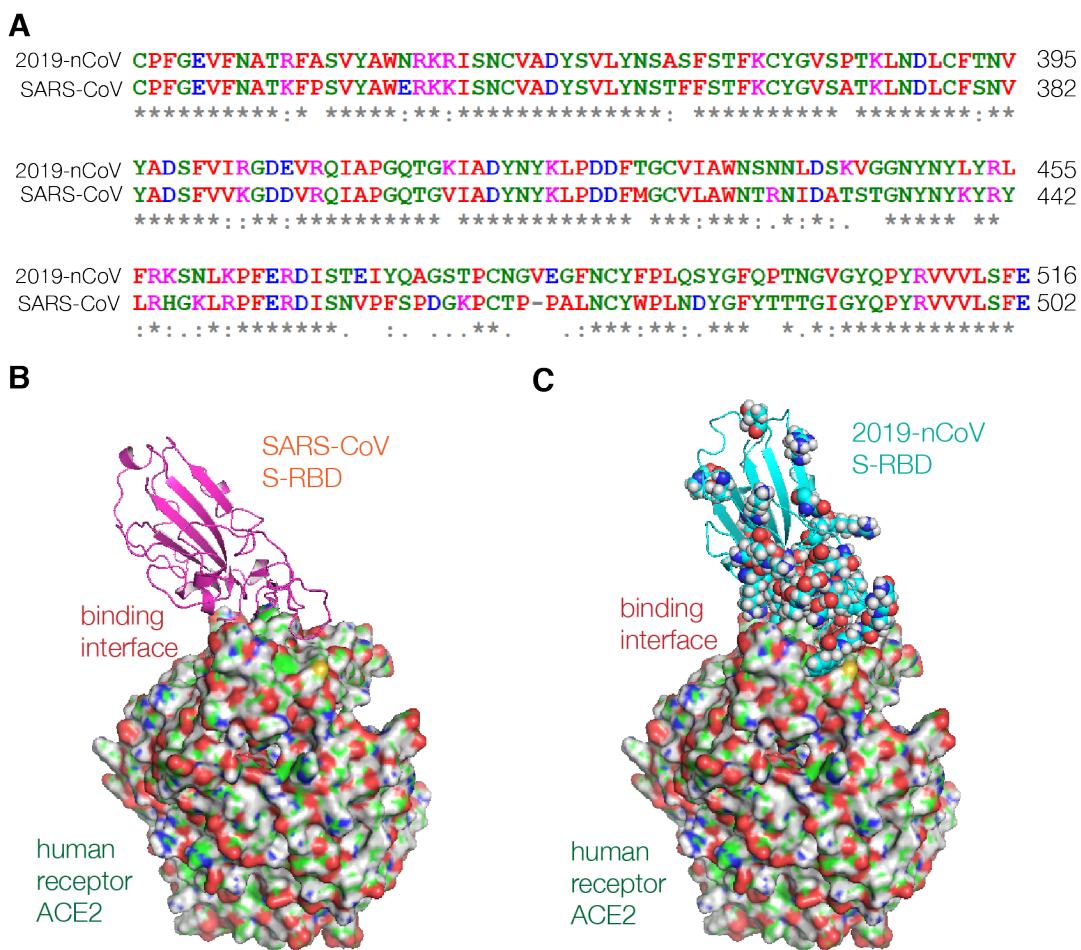


**Figure 1 | Binding interactions between SARS-CoV S-RBD and the human receptor ACE2.** (A) Scatter plot of Rosetta binding free-energy scores versus RMSDs of interface residues of about 3,000 thermodynamically probable, low-energy binding models, where the RMSD reference is the experimental complex structure (PDB 2AJF). (B) The five top-scoring binding models. They are almost identical to the experimental structure with RMSDs < 1.5 Å.

### High-resolution prediction of 2019-nCoV S-RBD structure

Because the 2019-nCoV genome has just been sequenced for a very short period, an experimental 3D structure for performing protein-protein docking of ACE2 to 2019-nCoV S-RBD is not yet available. Alternatively, computational methods might be used to predict the 3D structure (Schwede et al., 2009). Fortunately, the 2019-nCoV genome has shown that the amino-acid sequence of the S-protein is about

75% identical to that of the SARS-CoV S-protein (Fig. 2A). Therefore, with the SARS-CoV S-RBD structure as a template, it is possible to use high-resolution homology modeling method to accurately predict the 3D model(Qian et al., 2007).



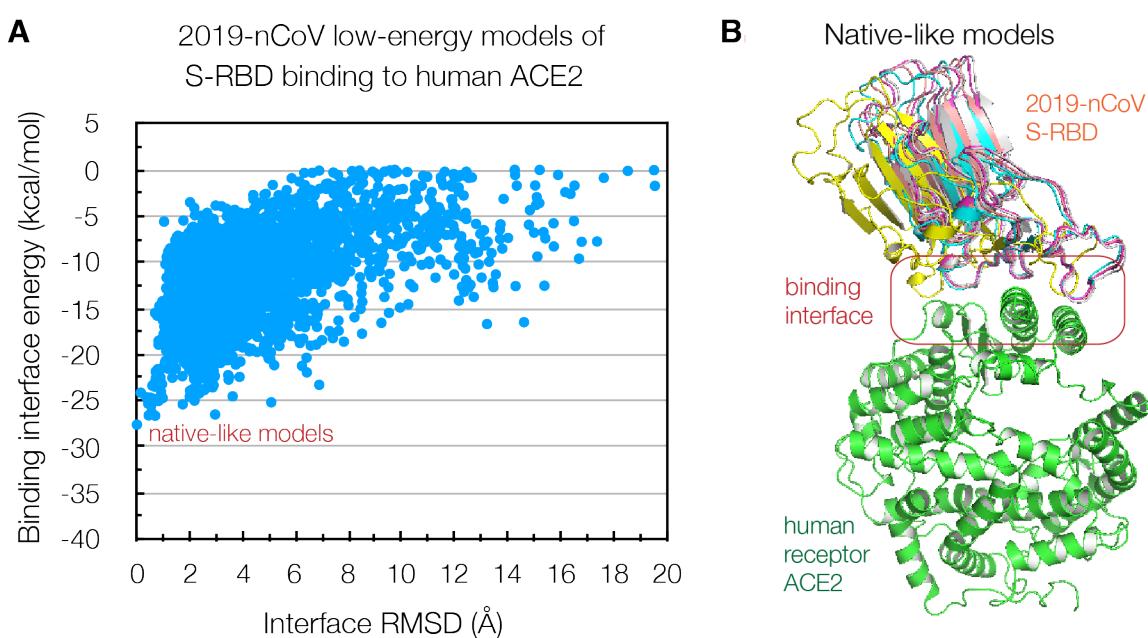
**Figure 2 | Homology modeling of 2019-nCoV S-RBD structure.** (A) Amino-acid sequence alignment of 2019-CoV and SARS-CoV S-RBDs with an identity of 72.8%. (B) SARS-CoV S-RBD-ACE2 complex (PDB 2AJF). ACE2 is shown as surface model. (C) Starting complex of 2019-nCoV S-RBD-ACE2 for docking. For comparison, 50 residues that are different from those of SARS-CoV S-RBD (see also Fig. 2A) are shown as sphere models.

To the end, we first generated an initial model for the 2019-nCoV RBD via SWISS-MDOEL web site, with the experimental S-RBD structure of SARS-CoV as the template (PDB 2AJF). Then, we performed all-atom refinements to optimize this initial model with Rosetta protocols (see Methods). The top-1 scoring model was used as the 3D structure of the 2019-nCoV S-RBD for the protein-protein docking with the experimental structure of ACE2 from PDB 2AJF. By superimposing this model onto

the SARS-CoV S-RBD structure (Fig. 2B), we obtained the starting configuration of S-RBD complexed with ACE2 for the 2019-nCoV docking (Fig. 2C). From this complex, we found that most of the residues that are different from those of SARS-CoV S-RBD (see also Fig. 2A) are close to or located in the binding interface. As we will show, such a residue distribution could lead to a different binding affinity from that of SARS-CoV S-RBD binding to ACE2.

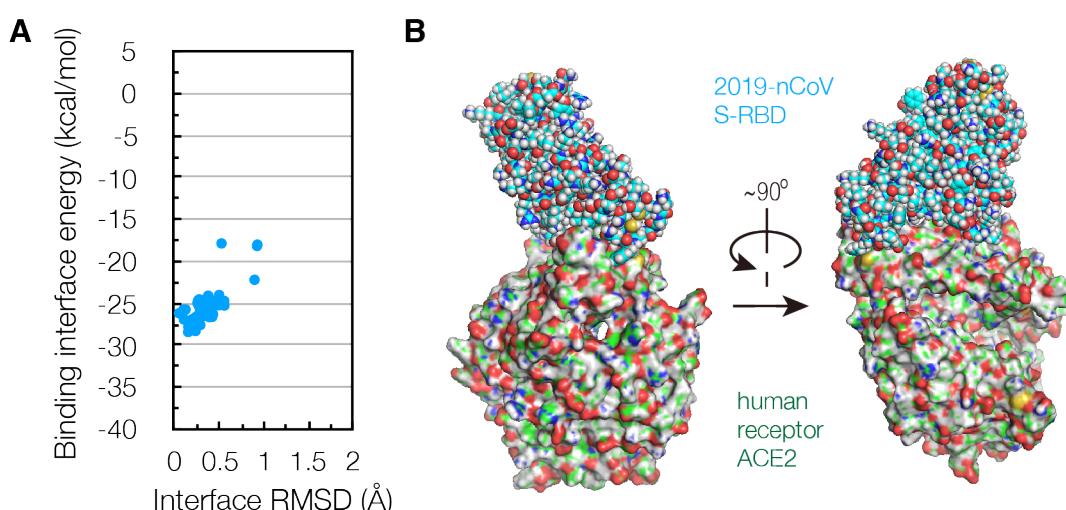
### Binding interactions of 2019-nCoV S-RBD with human ACE2

With the complex model in Fig. 2C, more than 3,000 independent docking runs were performed to identify thermodynamically probable, low-energy binding conformations. The results are illustrated in Fig. 3. As shown in the scatter plot in Fig. 3A, the docking simulations also achieved an energetic funnel, in which four models among the top-5 scores have interface RMSDs within 0.7 Å. This demonstrates that the docking has identified the native-like interacting conformations for the 2019-nCoV S-RBD binding to ACE2. We consider these lowest-energy conformations at the energetic funnel bottom as those being engaged in the receptor-binding function.



**Figure 3 | Binding interactions between 2019-nCoV S-RBD and the human receptor ACE2.** (A) Scatter plot of Rosetta binding free-energy scores versus RMSDs of interface residues of the thermodynamically probable, low-energy binding models, where the RMSD reference is the lowest-energy obtained in the first-phase docking simulations. (B) The five top-scoring binding models.

When comparing the SARS-CoV results in Fig. 1A, we found that there is no binding conformation with a binding energy score which is lower than that of the lowest-energy conformation of SARS-CoV derived from experimentally determined structure (-37.4 kcal·mol<sup>-1</sup> in Fig. 1A). To be sure, we also carried out high-resolution all-atom refinements to further optimize the lowest-energy binding conformation by 200 independent runs (see Methods). As indicated the scatter plot in Fig. 4A, the lowest-energy value of the 200 refined models remains about -28.4 kcal·mol<sup>-1</sup> (see model in Fig. 4B). Considering that the protein-protein docking has already sampled a large number of thermodynamically probable, low-energy binding conformations, these results indicate that the existence of an ACE2-binding conformation for 2019-nCoV with an energy value < -37.4 kcal·mol<sup>-1</sup> is very limited. So, the binding strength of 2019-nCoV S-RBD to ACE2 is likely lower than that of SARS-CoV S-RBD.



**Figure 4 | All-atom refinements of the lowest-energy binding model.** (A) Scatter plot of binding free-energy scores versus RMSDs of interface residues of the refined models, where the RMSD reference is the lowest-energy conformation in Fig. 3A. (B) The lowest-energy model of S-RBD binding to ACE2 with an energy score of -28.4 kcal·mol<sup>-1</sup>.

### Comparison of human receptor-binding capability of 2019-nCoV with SARS-CoV

The successful identifications of the native-like binding conformations of S-RBDs with ACE2 enable a comparison of the receptor-binding capability of 2019-nCoV with that of SARS-CoV. Especially, the obtained energy scores based on the experimentally determined structure of SARS-CoV defined a binding free-energy scale, on which the relative receptor-binding strength of a newly emerged CoV

could be evaluated.

According to statistical thermodynamics, we considered that the sampled binding conformations in the Monte Carlo docking simulations are thermodynamically probable in the binding process of S-protein to ACE2 and, based on definition of binding free energy (i.e., binding affinity) is indeed a thermodynamic average over all possible binding modes of two interacting components. Thus, the binding free energy is approximated by the thermodynamic average of the binding energies of all the sampled low-energy conformations as:

$$\langle \Delta G \rangle \approx \frac{\sum_i^N \Delta G_i \exp\left(-\frac{\Delta G_i}{RT}\right)}{\sum_i^N \exp\left(-\frac{\Delta G_i}{RT}\right)} \quad (1)$$

where  $N$  is the total number of the low-energy conformations sampled in the Monte Carlo docking simulations,  $R$  is the ideal gas constant, and  $T$  is the absolute temperature (i.e., 298.15 K for the room temperature). Note that, when  $N$  approaches the number of all thermodynamically probable conformations,  $\langle \Delta G \rangle$  becomes the exact value.

With equation 1 and all the sampled binding conformations in Figs. 1A and 3A, the binding free energy of SARS-CoV S-RBD to ACE2 is:

$$\langle \Delta G^{SARS} \rangle = -36.6 \text{ kcal}\cdot\text{mol}^{-1} \quad (2)$$

and that of 2019-nCoV S-RBD to ACE2 is:

$$\langle \Delta G^{2019} \rangle = -27.2 \text{ kcal}\cdot\text{mol}^{-1} \quad (3)$$

By comparing the values, we found that 2019-nCoV S-RBD reaches about 73% of the binding strength of SARS-CoV S-RBD to ACE2. Since the receptor-binding of S-protein is the major driving force for the viral transmission to the human host cell, such a strength order implies that 2019-nCoV might be able to infect humans similar to as SARS-CoV.

Just after the 2019-nCoV genome sequencing, Xu et al. also used structural modeling approach to predict the binding affinity of S-RBD to ACE2(Xu et al., 2020). In their study, two single complex conformations (i.e., crystal and predicted structures) were employed to calculate the binding free energies of two CoVs, and showed that the affinity of 2019-nCoV S-RBD to ACE2 was in the order of ~65% of that for SARS-CoV. So, their predicted strength is about 10% lower than our estimates. Certainly, although our values were calculated using a large-number of thermodynamically probable, low-energy binding conformations, further experimental measurements will be needed to testify the accuracy of both methods.

## Conclusions

We have performed large-scale protein-protein docking to evaluate the binding affinities of SARS-CoV and 2019-nCoV S-protein RBDs to the human receptor ACE2, in order to quantitatively assess the human receptor-binding capability of 2019-nCoV. Using the well-investigated SARS-CoV as the reference, our results showed that 2019-nCoV reaches about 73% of the ACE2 receptor-binding strength of SARS-CoV. This supports that the 2019-nCoV S-protein encoded by its genome has processed human receptor-binding ability close to that of SARS-CoV, and thus this newly emerged virus is likely able to bind to ACE2 and then to drive human-to-human transmission. However, the docking sampling for a large number of thermodynamically probable conformations did not find any binding conformation with an energy score lower than that of the lowest-energy conformation of SARS-CoV derived from experimentally determined structure (Figs. 1A and 3A), offering molecular evidence that the receptor-binding ability of 2019-nCoV is less than that of SARS-CoV.

Of course, ultimate confirmation of computational predictions requires experimental verification, which will also be very important for improving our method in the future. However, experimental approaches for deeply understanding the human transmission of viruses in terms of receptor-binding interactions and specificity is rather time consuming. Sometimes, it might be even impossible to employ experimental methods to thoroughly examine every evolved, mutant strain of a newly emerged virus, such as 2019-nCoV in our case. As an alternative, using a high-performance computer, our approach may assess the receptor-binding capability of a newly identified strain within one day. Thus, our method can be useful for the post-genome analysis for a newly identified CoV and its mutant strains, with the first-determined genome data. Note that, although viral genome data provide useful sequence information for comparison-based assessments with previously identified viruses, these data themselves do not offer hints about how the encoded viral proteins carry out their functions which may lead to serious human infections and diseases. Hence, post-genome analysis of viral protein functions is valuable not only for understanding the viruses themselves, but also for offering useful scientific supports for making effective, early decision about viral prevention and control.

In summary, this study presents a fast method for accurately assessing the human transmission capability of newly emerged CoVs, and demonstrates that post-genome analysis of CoV protein-protein interactions may provide early scientific guidance for the viral prevention and control.

## Methods

**Homology modeling of 2019-nCoV S-RBD structure.** The amino-acid sequence of 2019-nCoV S-protein was taken from the first genome sequence of a 2019-nCoV isolate (Wuhan-Hu-1) deposited in GenBank on January 12, 2020 (GenBank No: MN908947.3)(Wu et al., 2020). The initial structure of S-RBD was built according to standard modeling procedure on the SWISS-MODEL server at <https://swissmodel.expasy.org>, by submitting its amino-acid sequence in Fig. 2A and then selecting the SARS-CoV S-RBD structure (PDB 2AJF) as the modeling template. With the obtained model, high-resolution all-atom refinements were then carried out to optimize the conformations of the protein backbone and side-chains with the relax program of the Rosetta software suite(Das and Baker, 2008). And the Rosetta all-atom energy function 2015 (REF15) was employed to calculate the conformational energy(Alford et al., 2017). The all-atom relax protocols are:

```
relax.linuxgccrelease /  
-database $database_path /  
-s $start_model.pdb /  
-relax:thorough /  
-nstruct $model_number
```

**Local protein-protein docking.** The docking program of Rosetta(Das and Baker, 2008) was used to conduct the local protein-protein docking with the starting conformation of a given S-RBD-ACE2 complex. The Rosetta energy function 2015 (REF15) was employed in the docking to evaluate the interaction energy of two docking components (i.e., S-RBD and ACE2). Docking protocols with full receptor flexibility were used. For each S-RBD-ACE2 pair, at least 3,000 independent docking simulations were performed to generate a large number of thermodynamically probable, low-energy binding conformations for evaluating the binding free-energy. The top-scoring, lowest-energy conformations were identified as the native-like conformations of the S-RBD binding to ACE2. The local protein-protein docking protocols are:

```
docking_protocol.linuxgccrelease /  
-database $database_path /  
-s $start_complex.pdb /
```

```
-partners $ACE_chainID_$S-RBD_chainID /  
-dock_pert 3 8 /  
-ex1 /  
-ex2aro /  
-nstruct $model_number
```

**All-atom refinement of the lowest-energy conformation.** The docking program of Rosetta was used to carry out the high-resolution refinement stage of docking for further optimizing the lowest-energy binding conformation, with the Rosetta energy function 2015 (REF15). At least 200 independent docking runs were performed to generate refined binding conformations. The all-atom refinement protocols are:

```
docking_protocol.linuxgccrelease /  
-database $database_path /  
-s $lowest_complex.pdb /  
-docking_local_refine /  
-use_input_sc /  
-ex1 /  
-ex2aro /  
-nstruct $model_number
```

## References

- Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.K., Kappel, K., *et al.* (2017). The Rosetta all-atom energy function for macromolecular modeling and design. *J Chem Theory Comput* 13, 3031-3048.
- Chaudhury, S., Berrondo, M., Weitzner, B.D., Muthu, P., Bergman, H., and Gray, J.J. (2011). Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS One* 6, e22477.
- Das, R., and Baker, D. (2008). Macromolecular modeling with rosetta. *Annu Rev Biochem* 77, 363-382.
- de Wit, E., van Doremale, N., Falzarano, D., and Munster, V.J. (2016). SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 14, 523-534.
- Gaunt, E.R., Hardie, A., Claas, E.C., Simmonds, P., and Templeton, K.E. (2010). Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *J Clin Microbiol* 48, 2940-2947.

- Ge, X.Y., Li, J.L., Yang, X.L., Chmura, A.A., Zhu, G., Epstein, J.H., Mazet, J.K., Hu, B., Zhang, W., Peng, C., *et al.* (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* *503*, 535-538.
- Holshue, M.L., DeBolt, C., Lindquist, S., Lofy, K.H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., *et al.* (2020). First case of 2019 novel coronavirus in the United States. *N Engl J Med*, DOI: 10.1056/NEJMoa2001191.
- Huang, X., Dong, W., Milewska, A., Golda, A., Qi, Y., Zhu, Q.K., Marasco, W.A., Baric, R.S., Sims, A.C., Pyrc, K., *et al.* (2015). Human coronavirus HKU1 spike protein uses O-acetylated sialic acid as an attachment receptor determinant and employs hemagglutinin-esterase protein as a receptor-destroying enzyme. *J Virol* *89*, 7202-7213.
- Li, F., Li, W., Farzan, M., and Harrison, S.C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* *309*, 1864-1868.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., *et al.* (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*, DOI: 10.1056/NEJMoa2001316.
- Li, W., Michael J. Moore, Natalya Vasilieva, Jianhua Sui, Swee Kee Wong, M.A.B., Mohan Somasundaran, John L. Sullivan, Katherine Luzuriaga, Thomas C. Greenough, Hyeryun Choe, and Farzan, M. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* *426*, 450-454.
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., *et al.* (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, DOI: 10.1016/S0140-6736(20)30251-30258.
- Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A.J., Read, R.J., and Baker, D. (2007). High-resolution structure prediction and the crystallographic phase problem. *Nature* *450*, 259-264.
- Raj, V.S., Mou, H., Smits, S.L., Dekkers, D.H.W., Müller, M.A., Dijkman, R., Muth, D., Demmers, J.A.A., Zaki, A., Fouchier, R.A.M., *et al.* (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* *495*, 251-254.
- Rambaut, A. (2020). Phylogenetic analysis of 23 nCoV-2019 genomes, 2020-01-23 (<http://virological.org/t/phylogenetic-analysis-of-23-ncov-2019-genomes-2020-01-23/335>).
- Reusken, C.B., Raj, V.S., Koopmans, M.P., and Haagmans, B.L. (2016). Cross host transmission in the emergence of MERS coronavirus. *Curr Opin Virol* *16*, 55-62.
- Schwede, T., Sali, A., Honig, B., Levitt, M., Berman, H.M., Jones, D., Brenner, S.E., Burley, S.K., Das, R., Dokholyan, N.V., *et al.* (2009). Outcome of a workshop on applications of protein models in biomedical research. *Structure* *17*, 151-159.
- Weitzner, B.D., Jeliazkov, J.R., Lyskov, S., Marze, N., Kuroda, D., Frick, R., Adolf-Bryfogle, J., Biswas, N., Dunbrack, R.L., Jr., and Gray, J.J. (2017). Modeling and docking of antibody structures with Rosetta. *Nat Protoc* *12*, 401-416.
- WHO (2004). Summary of probably SARS cases with onset of illness from 1 November 2002 to 31 July 2003 ([http://www.who.int/csr/sars/country/table2004\\_04\\_21/en/](http://www.who.int/csr/sars/country/table2004_04_21/en/)).
- WHO (2016). Coronavirus infections: disease outbreak news (<http://www.who.int/csr/don/26-april-2016-mers-saudi-arabia/en/>).
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Hu, Y., Song, Z.-G., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., *et al.* (2020). Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome (<https://www.ncbi.nlm.nih.gov/nuccore/MN908947>).
- Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., and Hao, P. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *SCIENCE CHINA Life Sciences* (<https://doi.org/10.1007/s11427-020-1637-5>).
- Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., and Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the

putative receptor of Wuhan 2019-nCov. BioRxiv (<https://doi.org/10.1101/20200126919985>).

Zhou, P., Xing-Lou Yang, Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., *et al.* (2020). Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. BioRxiv (<https://doi.org/10.1101/20200122914952>).

## Acknowledgements

The methodology development in this work was supported by the grant from the National Natural Science Foundation of China (No. 91430112).

## Additional information

Correspondence and requests for materials should be addressed to Q.H. or A.H.

## Competing financial interests

The authors declare no competing financial interests.