

# **In silico comparison of spike protein-ACE2 binding affinities across species; significance for the possible origin of the SARS-CoV-2 virus**

Sakshi Piplani<sup>1,2</sup>, Puneet Kumar Singh<sup>2</sup>, David A. Winkler<sup>3-6\*</sup>, Nikolai Petrovsky<sup>1,2\*</sup>

<sup>1</sup> College of Medicine and Public Health, Flinders University, Bedford Park 5046, Australia

<sup>2</sup> Vaxine Pty Ltd, 11 Walkley Avenue, Warradale 5046, Australia

<sup>3</sup> La Trobe University, Kingsbury Drive, Bundoora 3042, Australia

<sup>4</sup> Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, Australia

<sup>5</sup> School of Pharmacy, University of Nottingham, Nottingham NG7 2RD. UK

<sup>6</sup> CSIRO Data61, Pullenvale 4069, Australia

\*Joint senior authors

## **Abstract**

The devastating impact of the COVID-19 pandemic caused by SARS–coronavirus 2 (SARS-CoV-2) has raised important questions on the origins of this virus, the mechanisms of any zoonotic transfer from exotic animals to humans, whether companion animals or those used for commercial purposes can act as reservoirs for infection, and the reasons for the large variations in SARS-CoV-2 susceptibilities across animal species. Traditional lab-based methods will ultimately answer many of these questions but take considerable time. Increasingly powerful *in silico* modeling methods provide the opportunity to rapidly generate information on newly emerged pathogens to aid countermeasure development and also to predict potential future behaviors. We used an *in silico* structural homology modeling approach to characterize the SARS-CoV-2 spike protein which predicted its high affinity binding to the human ACE2 receptor. Next we sought to gain insights into the possible origins and transmission path by which SARS-CoV-2 might have crossed to humans by constructing models of the ACE2 receptors of relevant species, and then calculating the binding energy of SARS-CoV-2 spike protein to each of these. Notably, SARS-CoV-2 spike protein had the highest overall binding energy for human ACE2, greater than all the other tested species including bat, the postulated source of the virus. This indicates that SARS-CoV-2 is a highly adapted human pathogen. Of the species studied, the next highest binding affinity after human was pangolin, which is most likely explained by a process of convergent evolution. Binding of SARS-CoV-2 for dog and cat ACE2 was similar to affinity for bat ACE2, all being lower than for human ACE2, and is consistent with only occasional observations of infections of these domestic animals. Snake ACE2 had low affinity for spike protein, making it highly improbable that snakes acted as an intermediate vector. Overall, the data indicates that SARS-CoV-2 is uniquely adapted to infect humans, raising important questions as to whether it arose in nature by a rare chance event or whether its origins might lie elsewhere.

## Introduction

The devastating impact of COVID-19 infections caused by SARS–coronavirus 2 (SARS-CoV-2) has stimulated unprecedented international activity to discover effective vaccines and drugs for this and other pathogenic coronaviruses.<sup>1-16</sup> It has also raised important questions on the mechanisms of zoonotic transfer of viruses from animals to humans, questions as to whether companion animals or those used for commercial purposes can act as reservoirs for infection, and the reasons for the large variations in SARS-CoV-2 susceptibility across animal species.<sup>17-19</sup> Understanding how viruses move between species may help us prevent or minimize these pathways in the future. Elucidating the molecular basis for the different susceptibilities of species may also shed light on the differences in susceptibilities in different sub-groups of humans.

Very recently, Shi et al. published the results of experiments to determine the susceptibility to SARS-CoV-2 of ferrets, cats, dogs, and other domesticated animals.<sup>20</sup> They showed that SARS-CoV-2 virus replicates poorly in dogs, pigs, chickens, and ducks, but ferrets and cats are permissive to infection. Other studies have reported the susceptibility of other animal species to SARS-CoV-2.<sup>17,20,21</sup> Susceptible species such as macaques, hamsters and ferrets are used as animal models of SARS-CoV-2 infection.<sup>22-24</sup> In the absence of purified, isolated ACE2 from all the relevant animal species that could be used to measure the molecular affinities to spike protein experimentally, computational methods offer considerable promise for determining the rank order of affinities across species, as a method to impute which species may be permissive to SARS-CoV-2.

Here we show how computational chemistry methods from structure-based drug design can be used to determine the relative binding affinities of the SARS-CoV-2 spike protein for its receptor, angiotensin converting enzyme (ACE)-2, a critical initiating event for SARS-CoV-2 infection, across multiple common and exotic animal species.<sup>25-27</sup> The aim of these studies was to better understand the species-specific nature of this interaction and see if this could help elucidate the origin of SARS-CoV-2 and the mechanisms for its zoonotic transmission.

## **Materials and Methods**

### *Homology modelling of structures*

To construct the three-dimensional structure of the SARS-CoV-2 spike protein, the sequence was retrieved from NCBI Genbank Database (accession number YP\_009724390.1). A PSI-BLAST search against the PDB Database for template selection was performed and the X-ray structure of SARS coronavirus spike template (refcode 6ACC) was selected with 76.4% sequence similarity to SARS-CoV-2 spike protein. The protein sequences of the ACE2 proteins for different species is summarized in Table 11 and full sequence alignment in Supplementary Figure 1. The phylogenetic tree for ACE2 proteins from selected animal species is illustrated in Supplementary Figure 2. The 3D-structures were built using Modeller 9.21 (<https://salilab.org/modeller/>).<sup>28</sup> The quality of the generated models was evaluated using the GA341 score and DOPE scores, and the models assessed using SWISS-MODEL structure assessment server (<https://swissmodel.expasy.org/assess>).<sup>29</sup>

### *Protein Structure preparation*

The X-ray crystal structures of human ACE2 (refcode 3SCI) and human SARS spike protein (refcode 5XLR) were retrieved from Protein Data Bank. Protein preparation and removal of non-essential and non-bridging water molecules for docking studies were performed using the UCSF Chimera package (<https://www.cgl.ucsf.edu/chimera/>).<sup>30</sup>

### *Molecular Docking*

These modelled structures were docked against SARS-CoV-2 spike protein structure using the HDock server (<http://hdock.phys.hust.edu.cn/>).<sup>31,32</sup> Molecular docking was performed on the homology modelled SARS-CoV-2 spike protein with human and animal ACE2 proteins. The SARS-CoV spike protein was also docked with human ACE2 protein to obtain the docking pose for binding energy calculations. The docking poses were ranked using an energy-based scoring function. The docked structures were analyzed using UCSF Chimera software.

### *Molecular Dynamics Simulation*

The final models were optimized using the AMBER99SB-ILDN force field in Gromacs2020 (<http://www.gromacs.org/>).<sup>33</sup> Docked complexes (SARS-CoV-2 spike with human ACE2, human

SARS-CoV spike with human ACE2, SARS-CoV-2 spike with bat ACE2 etc) were used as starting geometries for MD simulations. Simulations were carried out using the GPU accelerated version of the program with the AMBER99SB-ILDN force field I periodic boundary conditions on an Oracle Cloud Server. Docked complexes were immersed in a truncated octahedron box of TIP3P water molecules. The solvated box was further neutralized with Na<sup>+</sup> or Cl<sup>-</sup> counter ions using the tleap program. Particle Mesh Ewald (PME) was employed to calculate the long-range electrostatic interactions. The cutoff distance for the long-range van der Waals (VDW) energy term was 12.0 Å. The whole system was minimized without any restraint. The above steps applied 2500 cycles of steepest descent minimization followed by 5000 cycles of conjugate gradient minimization. After system optimization, the MD simulations was initiated by gradually heating each system in the NVT ensemble from 0 to 300 K for 50 ps using a Langevin thermostat with a coupling coefficient of 1.0/ps and with a force constant of 2.0 kcal/mol·Å<sup>2</sup> on the complex. Finally, a production run of 100 ns of MD simulation was performed under a constant temperature of 300 K in the NPT ensemble with periodic boundary conditions for each system. During the MD procedure, the SHAKE algorithm was applied for the constraint of all covalent bonds involving hydrogen atoms. The time step was set to 2 fs. The structural stability of the complex was monitored by the RMSD and RMSF values of the backbone atoms of the entire protein. Finally, the free energies of binding were calculated for all the simulated docked structures. Calculations were also performed for up to 500 ns on human ACE2 to ensure that 100ns is sufficiently long for convergence. Duplicate production runs starting with different random seeds were also run to allow estimates of binding energy uncertainties to be determined for the strongest binding ACE2 structures.

#### *MM-PBSA binding free energy study*

The binding free energies of the protein-protein complexes were evaluated in two ways. The traditional method is to calculate the energies of solvated SARS-CoV-2 spike and ACE2 proteins and that of the bound complex proteins and derive the binding energy by subtraction.

$$\Delta G (\text{binding, aq}) = \Delta G (\text{complex, aq}) - (\Delta G (\text{spike, aq}) + \Delta G (\text{ACE2, aq})) \quad (1)$$

We also calculated binding energies using the molecular mechanics Poisson Boltzmann surface area (MM-PBSA) tool in GROMACS that is derived from the nonbonded interaction energies of the complex.<sup>34,35</sup> The method is also widely used method for binding free energy calculations. The binding free energies of the protein complexes were analyzed during equilibrium phase from the output files of 100 ns MD simulations. The `g_mmpbsa` tool in GROMACS was used after molecular dynamics simulations, the output files obtained were used to post-process binding free energies by the single-trajectory MM-PBSA method. Specifically, for a non-covalent binding interaction in the aqueous phase the binding free energy,  $\Delta G$  (bind,aq), is: –

$$\Delta G (\text{bind,aqu}) = \Delta G (\text{bind,vac}) + \Delta G (\text{bind,solv}) \quad (2)$$

where  $\Delta G$  (bind,vac) is the binding free energy in vacuum, and  $\Delta G$ (bind,solv) is the solvation free energy change upon binding: –

$$\Delta G (\text{bind,solv}) = \Delta G (\text{R:L, solv}) - \Delta G (\text{R,solv}) - \Delta G (\text{L,solv}) \quad (3)$$

where  $\Delta G$  (R:L,solv),  $\Delta G$  (R,solv) and  $\Delta G$  (L,solv) are solvation free energies of complex, receptor and ligand, respectively.

Free energy decomposition analyses were also performed by MM-PBSA decomposition to get a detailed insight into the interactions between the ligand and each residue in the binding site. The binding interaction of each ligand–residue pair includes three terms: the van der Waals contribution, the electrostatic contribution, and the solvation contribution.

### *Interaction energy*

Another estimate of the strength of the interaction between protein-protein complex can be obtained from the non-bonded interaction energy between the complex. GROMACS has the ability to decompose the short-range nonbonded energies between any number of defined groups. To compute the interaction energies as a part of our analysis, we reran the trajectory files obtained during simulation to recompute energies using `-rerun` command. The interaction energy is the

combination of short range Coulombic interaction energy (Coul-SR:Protein-Protein) and the short-range Lennard-Jones energy (LJ-SR:Protein-Protein (see Table 3).

While this paper was being prepared, a paper by Guterres and Im described a substantial improvement in protein-ligand docking results using high-throughput MD simulations.<sup>36</sup> They employed docking using AutoDock Vina, followed by MD simulation using CHARMM. The parameters they advocated were very similar to those used in our study. Proteins were solvated in a box of TIP3P water molecules extending 10 Å beyond the proteins and the particle-mesh Ewald method was used for electrostatic interactions. Nonbonded interactions over 10 and 12 Å were truncated. Their systems were minimized for 5000 steps using the steepest descent method followed by 1 ns equilibration with an NVT setting. For each protein-ligand complex, they ran 3 × 100 ns production runs from the same initial structure using different initial velocity random seeds and an integration step size of 2 fs.

## Results

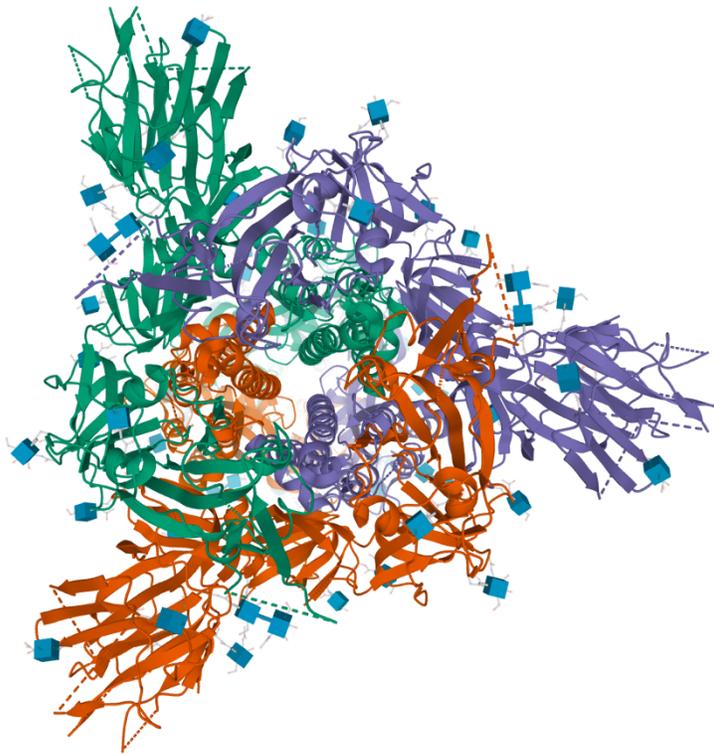
The ancestry of SARS-CoV-2 traces back to the human, civet and bat SARS-CoV strains, which all use the same ACE2 proteins for cellular entry.<sup>37-39</sup> The similarities and variations in sequences for both the SARS-CoV-2 spike protein and the SARS spike protein were determined from sequences retrieved from NCBI GenBank Databank and aligned using CLUSTALW. The spike protein receptor binding domain (RBD) region showed a 72% identity between the two viruses (Figure 1).

SARS_spike	RV <b>V</b> PSGDVVRFPNITNLCPFGVEFNAT <b>K</b> FP <b>S</b> VYAW <b>E</b> R <b>K</b> ISNCVADYSVLYNS <b>T</b> FFSTFK
COVID-19_spike	RV <b>Q</b> P <b>T</b> ESIVRFPNITNLCPFGVEFNAT <b>R</b> F <b>A</b> S <b>V</b> YAW <b>N</b> R <b>K</b> RISNCVADYSVLYNS <b>A</b> SFSTFK
SARS_spike	CYGV <b>S</b> ATKLNDLCF <b>S</b> NVYADSFV <b>V</b> KGDDVRQIAPGQ <b>T</b> G <b>V</b> IADYNYKLPDDF <b>M</b> GCV <b>L</b> AW <b>N</b> T
COVID-19_spike	CYGV <b>S</b> P <b>T</b> KLNDLCF <b>T</b> NVYADSFV <b>I</b> R <b>G</b> D <b>E</b> VRQIAPGQ <b>T</b> G <b>K</b> IADYNYKLPDDF <b>T</b> GCV <b>I</b> AW <b>S</b>
SARS_spike	<b>R</b> N <b>I</b> D <b>A</b> T <b>S</b> TGN <b>Y</b> N <b>K</b> Y <b>R</b> Y <b>L</b> R <b>H</b> G <b>K</b> L <b>R</b> P <b>F</b> ERDIS <b>N</b> V <b>P</b> F <b>S</b> P <b>D</b> G <b>K</b> P <b>C</b> T- <b>P</b> PALNCY <b>W</b> PLNDYGFY
COVID-19_spike	<b>N</b> N <b>L</b> D <b>S</b> K <b>V</b> G <b>G</b> N <b>Y</b> N <b>L</b> Y <b>R</b> L <b>F</b> R <b>K</b> S <b>N</b> L <b>K</b> P <b>F</b> ERDIST <b>E</b> I <b>Y</b> Q <b>A</b> G <b>S</b> T <b>P</b> C <b>N</b> G <b>V</b> E <b>G</b> F <b>N</b> C <b>Y</b> F <b>P</b> L <b>Q</b> S <b>Y</b> G <b>F</b> Q
SARS_spike	<b>T</b> T <b>T</b> G <b>I</b> G <b>Y</b> Q <b>P</b> Y <b>R</b> V <b>V</b> L <b>S</b> F <b>E</b> L <b>L</b> N <b>A</b> P <b>A</b> T <b>V</b> C <b>G</b> P <b>K</b> L <b>S</b> T <b>D</b> L <b>I</b> K <b>N</b> Q <b>C</b> V <b>N</b> F
COVID-19_spike	<b>P</b> T <b>N</b> G <b>V</b> G <b>Y</b> Q <b>P</b> Y <b>R</b> V <b>V</b> L <b>S</b> F <b>E</b> L <b>L</b> H <b>A</b> P <b>A</b> T <b>V</b> C <b>G</b> P <b>K</b> K <b>S</b> T <b>N</b> L <b>V</b> K <b>N</b> K <b>C</b> V <b>N</b> F

**Figure 1.** SARS and SARS-CoV-2 receptor binding domain sequence alignment showing conserved regions in black and non-conserved residues in red and blue.

*Spike Protein-ACE2 interaction*

The three-dimensional structure of SARS-CoV-2 Spike protein (Figure 2) used was PDB refcode 6VXX. ACE2 is the receptor for SARS-CoV-2 and is needed for its cellular entry. The low affinity binding of SARS spike protein for mouse ACE2 has been postulated as the reason mice are largely non-permissive to SARS infection. Similarly, we postulated that variation in ACE2 between species may determine the affinity of binding of SARS-CoV-2 Spike protein and hence determine which species are permissive to SARS-CoV-2 infection. The sequence alignment of the selected ACE2 species is shown in Supplementary Figure 1 and the phylogenetic tree showing relatedness of ACE2 proteins across selected animal species is illustrated in Supplementary Figure 2. ACE2 modelled structures were generated for each relevant species and the ones with the lowest DOPE (Discrete Optimized Protein Energy) score were chosen and refined by energy minimization and used for further analysis.



**Figure 2.** 3D structure of SARS-CoV-2 spike protein (PDB refcode 6VXX)

The Ramachandran scores of the modeled ACE2 structures for selected species are summarized in Table 1 with the actual predicted ACE2 structures and Ramachandran plots for each selected species shown in Supplementary Figure 3.

**Table 1.** Ramachandran scores for ACE2 modelled structures for selected species

Species	MolProbity Score	Ramachandran Score (favoured region)
<i>Rhinolophus sinicus</i> (bat)	2.45	97.8%
<i>Mus musculus</i> (mouse)	2.49	98.2%
<i>Mustela putorius furo</i> (ferret)	2.5	98.3%
<i>Mesocricetus auratus</i> (hamster)	2.59	98.7%
<i>Felis catus</i> (cat)	2.42	98.5%
<i>Paguma larvata</i> (civet)	2.52	97.3%
<i>Macaca fascicularis</i> (monkey)	2.59	98.2%
<i>Manis javanica</i> (pangolin)	2.42	98.3%
<i>Ophiophagus hannah</i> (king cobra)	2.93	96.0%
<i>Canis lupus familiaris</i> (dog)	2.9	97.1%
<i>Equus caballus</i> (horse)	2.29	96.3%
<i>Panthera tigris altaica</i> (tiger)	2.25	96.6%
<i>Bos Taurus</i> (cattle)	2.70	96.2%

The three-dimensional structures of ACE2 receptor of selected species were created using Modeller9.23. The templates used to model the various ACE2 structures were 1R42 (human ACE2), 3CSI (glutathione transferase) and 3D0G (spike protein receptor-binding domain from the 2002-2003 SARS coronavirus human strain) (Table 2). Similarity of the template and query is important while building the model, the query sequence of *Macaca fascicularis* (monkey) accession number A0A2K5X283 was found to be 96.9% similar to the structure of human ACE2 (PDB id: 1R42) whereas *Ophiophagus hannah* (king cobra) had a much lower similarity of 61.42% to its template, 1R42 (Table 2). All structures were subjected to energy minimization and those with the lowest DOPE score energy minimized structures were selected for further study (Supplementary Figure 3).

**Table 2.** Template structures used to model selected ACE2 species and similarity scores of each ACE2 sequence to the selected template used

<b>Species</b>	<b>Accession No.</b>	<b>Template</b>	<b>Similarity</b>
<i>Rhinolophus sinicus</i> (Bat)	AGZ48803.1	3CSI	79.73%
<i>Mus musculus</i> (Mouse)	Q8R0I0	1R42	84.27%
<i>Mustela putorius furo</i> (Ferret)	Q2WG88	1R42	83.44%
<i>Mesocricetus auratus</i> (Hamster)	A0A1U7QTA1	1R42	87.58%
<i>Felis catus</i> (Cat)	Q56H28	1R42	85.93%
<i>Canis luparis</i> (Dog)	J9P7Y2	1R42	84.93%
<i>Paguma larvata</i> (Civet)	Q56NL1	3D0G	86.77%
<i>Macaca fascicularis</i> (Monkey)	A0A2K5X283	1R42	96.91%
<i>Manis javanica</i> (Pangolin)	XP_017505752.1	1R42	85.57%
<i>Ophiophagus hannah</i> (King cobra)	V8NIH2	1R42	61.42%
<i>Equus caballus</i> (Horse)	F6V9L3	6MI7	85.91%
<i>Panthera tigris altaica</i> (Tiger)	XP_007090142.1	6MI7	85.91%
<i>Bos taurus</i> (Cow)	NP_001019673.2	6MI7	80.30%

### *Structure Quality Assessment*

The modelled structures were further assessed for quality control using Ramachandran Plot and molprobity scores in SWISSModel structure assessment. The Ramachandran Plot checks the stereochemical quality of a protein by analyzing residue-by-residue geometry and overall structure geometry, is also a way to visualize energetically allowed regions for backbone dihedral angles  $\psi$  against  $\phi$  of amino acid residues in protein structure. The Ramachandran score of SARS-CoV-2 spike protein was 90% in the binding region and molprobity scores that provide an evaluation of model quality at both the global and local level was 3.17. Further, the ACE2 modelled structures from selected species were also assessed using Swiss model structure assessment. Ramachandran score or the percentage of amino acid residues falling into the energetically favored region for all species ranged from 96-99% (Table 1). These Ramachandran

and Molprobit scores show that all the built structures were of good quality and were suitable for use in further studies. The Ramachandran graphs are presented in Supplementary Figure 3.

#### *Molecular Docking analysis*

The Receptor Binding Domain (RBD) of SARS-CoV-2 spike protein was docked against ACE2 receptor of various species using HDOCK server. The interacting residues of ACE2 and SARS-CoV-2 spike protein are depicted in Table 3. We found certain key amino acids in the receptor binding motif (RBM) that were in accordance with previous studies<sup>40</sup>. Certain amino acids including PHE28, ASN330, ASP355 and ARG357 were conserved in ACE2 of most of the selected species and were observed to take part in the interaction with spike protein. TYR41, LYS353, ALA386 and ARG393 also interacted with spike protein residues and were highly conserved across all species except bat, mouse, ferret and pangolin, respectively. Spike protein interacting residues with *Ophiophagus hannah* (King cobra) were least common amongst all the ACE2 species included in the study, consistent with its low sequence similarity to human ACE2.

**Table 3.** Protein interacting residues in selected ACE2 species when docked with SARS-CoV-2 spike protein. Residues common to spike interacting residues in human ACE2 are labelled red.

Species	Accession Number	Position															
		19	24	27	28	30	31	34	37	38	41	42	79	83	330	353	393
<i>Homo sapiens (Human)</i>	Q9BYF1	S	Q	T	F	D	K	H	E	D	Y	Q	L	Y	N	K	R
<i>Macaca fascicularis (Monkey)</i>	A0A2K5X283	S	Q	T	F	D	K	H	E	D	Y	Q	L	Y	N	K	R
<i>Rhinolophus sinicus (Bat)</i>	U5WHY8	S	E	M	F	D	K	T	E	D	H	Q	L	Y	N	K	R
<i>Mustela putorius furo (Ferret)</i>	Q2WG88	D	L	T	F	E	K	T	E	E	Y	Q	-	Y	N	K	R
<i>Manis javanica (Pangolin)</i>	XP_017505752.1	-	E	T	F	E	K	S	E	E	Y	Q	I	Y	N	K	R
<i>Ophiophagus hannah (Snake)</i>	ETE61880.1	Q	V	K	F	E	Q	A	-	D	Y	N	N	F	N	L	R
<i>Canis luparis (Dog)</i>	J9P7Y2	-	L	T	F	E	K	Y	E	E	Y	Q	L	Y	N	K	R
<i>Mesocricetus auratus (Hamster)</i>	A0A1U7QTA1	S	Q	T	F	D	L	Q	E	D	Y	Q	L	Y	N	K	R
<i>Panthera tigris (Tiger)</i>	XP_007090142.1	S	L	T	F	D	K	H	E	E	Y	Q	L	Y	K	K	R
<i>Bos Taurus (Cow)</i>	NP_001019673.2	S	Q	T	F	E	K	H	E	D	Y	Q	M	Y	N	K	R
<i>Mus musculus (Mouse)</i>	Q8R0I0	S	N	T	F	N	N	Q	E	D	Y	Q	Y	F	N	K	R
<i>Paguma larvata (Civet)</i>	Q56NL1	S	L	T	F	E	K	Y	E	Q	Y	Q	L	Y	N	K	R
<i>Felis catus (Cat)</i>	Q56H28	S	L	T	F	E	K	H	E	E	Y	Q	L	Y	N	K	R
<i>Equus ferus caballus (Horse)</i>	F6V9L3	S	L	T	F	D	K	S	E	E	H	Q	L	Y	N	K	R

### *Molecular dynamics simulations*

The molecular dynamics simulation of complexes of SARS-CoV-2 spike protein and ACE2 receptors of various species were performed for 100ns. All complexes became stable during simulation with RMSD fluctuations converging to a range of 0.5 to 0.8 nm from the original position. The calculated binding energies for the interactions of SARS-CoV-2 with ACE2 from the species studied are presented in Table 4. The MMPBSA binding energies are summarized for comparison. The energies calculated by equation 1 and those from the MMPBSA algorithm correlate strongly ( $r^2=0.92$ ). The binding energies (Table 2) correlate poorly with sequence similarity ( $r^2=0.27$ ). Also shown is the observed and predicted SARS-CoV-2 susceptibilities of the species from analysis of phylogenetic clustering and sequence alignment with currently known ACE2s utilized by SARS-CoV-2, as described by Qiu et al.<sup>41</sup> (Supplementary Figure 2) Also listed is the interaction energies of the spike and ACE2 proteins for each species calculated by Wu et al. using an automatic docking method, ICM-Pro<sup>42</sup>. Table 4 also includes observational *in vivo* data on SARS-CoV-2 infectivity and disease symptoms in the species where this has been reported.

**Table 4.** Binding energies of SARS-CoV-2 spike protein to ACE2 of selected species and potential species susceptibilities from other studies

<b>Species</b>	<b>Binding energy kcal/mol</b>	<b>MMPBSA energy kcal/mol</b>	<b>COVID infectivity</b> n/a = not assessed
<i>Homo sapiens</i> (human)	-52.8	-57.6	Permissive, high infectivity, severe disease in 5-10%,
<i>Manis javanica</i> (pangolin)	-52.0	-56.5	Permissive
<i>Canis luparis</i> (dog)	-50.8	-49.5	Permissive, low infectivity, no overt disease
<i>Macaca fascicularis</i> (monkey)	-50.4	-50.8	Permissive, medium infectivity, lung disease
<i>Mesocricetus auratus</i> (hamster)	-49.7	-50.0	Permissive, high infectivity, lung disease
<i>Mustela putorius furo</i> (ferret)	-48.6	-49.2	Permissive, medium infectivity, mild disease
<i>Felis catus</i> (cat)	-47.6	-48.9	Permissive, high infectivity, lung disease
<i>Panthera tigris</i> (tiger)	-47.3	-42.5	Permissive, overt respiratory symptoms
<i>Rhinolophus sinicus</i> (bat)	-46.9	-49.6	n/a
<i>Paguma larvata</i> (civet)	-45.1	-46.1	n/a
<i>Equus ferus caballus</i> (horse)	-44.1	-49.2	Permissive, low infectivity, no overt disease
<i>Bos taurus</i> (cow)	-43.6	-42.5	n/a
<i>Ophiophagus hannah</i> (snake)	-39.5	-52.5	n/a
<i>Mus musculus</i> (mouse)	-38.8	-39.4	n/a

## Discussion

Knowing which species may be permissive to SARS-CoV-2 is very important, both in respect of identifying intermediate hosts and potential source of the original virus, as well as helping to identify suitable species for use as infection models to allow testing of COVID-19 drugs and vaccines. Directly testing broad ranges of species for virus susceptibility is difficult and time consuming and in the cases of very rare species, may not even be practicable. A less direct method to try and obtain the same information might involve measurement of binding affinity of the SARS-CoV-2 spike protein to the target ACE2 receptor from different species. This can be done, for example, through use of cell lines transfected with ACE2 receptors from each individual species, but this would again be time consuming and not practicable. A third approach, adopted here, is to use fast, efficient *in silico* structural modelling and docking algorithms, using available genomic and structural biology data, to generate relevant ACE2 structural models and use molecular dynamics to calculate the binding energies. Notably, this approach surprisingly revealed that the binding energy between SARS-CoV-2 spike protein and ACE2 was highest for humans out of all species tested, suggesting that SARS-CoV-2 spike protein is uniquely evolved to bind and infect cells expressing human ACE2. This finding is particularly surprising as, typically, a virus would be expected to have highest affinity for the receptor in its original host species, e.g. bat, with a lower initial binding affinity for the receptor of any new host, e.g. humans. However, in this case, the affinity of SARS-CoV-2 is higher for humans than for the putative original host species, bats, or for any potential intermediary host species.

The calculated binding energies of SAR-Cov2 spike protein with ACE2 from selected species are in general agreement with the relatively limited amount of published information. SAR-Cov2 binding affinity to monkey ACE2 was lower than for human ACE2. Young cynomolgus macaques when infected expressed viral RNA in nasal swabs but did not develop overt clinical symptoms whereas aged animals showed higher viral RNA loads and some weight loss and rapid respiration associated with moderate interstitial pneumonia and virus replication in upper and lower respiratory tract.<sup>23,48</sup> Syrian hamsters and ferrets are highly susceptible to SARS-CoV infection and are used as animal models of the disease. Mahdy published a preprint reviewing SARS-CoV-2 infection in animals and suggested that SARS-CoV-2 might recognize ACE2 from a variety of animal species, including palm civet, the intermediate host for SARS-CoV.<sup>45</sup>

Although bats carry many coronaviruses including SARS-CoV, a relative of SARS-CoV-2, direct evidence for existence of SARS-CoV-2 in bats has not been found. As highlighted by our data, the binding strength of SARS-CoV-2 for bat ACE2 is considerably lower than for human ACE2,

suggesting that even if SARS-CoV-2 did originally arise from a bat precursor it must later have adapted its spike protein to optimise its binding to human ACE2. There is no current explanation for how, when or where this might have happened. Instances of direct human infection by coronaviruses or other bat viruses is rare with transmission typically involving an intermediate host. For example, lyssaviruses such as Hendra are periodically transmitted from bats to horses and then to humans who contact the infected horse. Similarly, SARS-CoV was shown to be transmitted from bats to civet cats and from them to humans. To date, a virus identical to SARS-CoV-2 has not been identified in bats or any other non-human species, making its origins unclear. To date, the most closely related coronavirus to SARS-CoV-2, is the bat coronavirus, BatCoV RaTG1, which has 96% whole-genome identity to SARS-CoV-2.<sup>50</sup> The fact that SARS-CoV-2 has also not been found in any likely intermediate host raises questions of the origins of the original SARS-CoV-2 virus that infected human case zero in late 2019. Wuhan, the epicentre of the outbreak, hosts China's only BSL4 facility and is the site of considerable bat coronavirus research. Identification of an intermediate animal host in which SARS-CoV-2 might have adapted to a human ACE2 permissive form would go a long way to alleviating concerns that SARS-CoV-2 is not a natural virus. Lam et al.<sup>44</sup> made confused public claims of finding SARS-CoV-2 in Malayan pangolins, suggesting that pangolins were an intermediate vector for SARS-CoV-2. However, further sequence analysis of these claims by Zhang et al. established that Pangolin-CoV was a very different coronavirus that had modest at best ~90% sequence similarity to SARS-CoV-2. While Pangolin-CoV spike RBD shared some similarities to SARS-CoV-2, its spike protein did not share the furin cleavage site that was a prominent feature of SARS-CoV-2 spike protein.<sup>51</sup> Hence, any similarity of Pangolin-CoV to SARS-CoV-2 was restricted to the residues in the RBD and RBM. Overall, Pangolin-CoV is only a distant relative of SARS-CoV-2.

Based on our data, the similarity of Pangolin-CoV to SARS-CoV-2 in the spike RBM could be a case of convergent evolution of the two viruses, whereby the close similarity of the structure of the pangolin ACE2 spike binding domain (SBD) to the same region of human ACE2, drove the convergent evolution of the spike protein RBD of both viruses, allowing them to bind to pangolin and human ACE2, respectively. Such close similarity of pangolin and human ACE2 SBD could make it easy for any pangolin CoV to cross from pangolins to humans. Nevertheless, with no virus matching the SARS-CoV-2 sequence identified in pangolins this makes it less likely that SARS-CoV-2 has a pangolin origin. However, this does call for more intensive survey of coronaviruses in pangolin populations, should such viruses pose future human pandemic threats. This finding also supports the need for strict enforcement of a worldwide ban on any trafficking of pangolins to reduce risks of pangolin coronaviruses crossing to humans and becoming a trigger for a future

coronavirus pandemic. However, there continues to be a lack of evidence to indicate that SARS-CoV-2 is a pangolin-derived virus that first crossed from pangolins to humans in late 2019.

Early in the COVID-19 outbreak it was suggested that snakes may be an intermediate vector. ACE2 of turtle and snake has very low homology to human ACE2 and SARS-CoV-2 was shown to not bind to reptile ACE2, making snakes unlikely as possible hosts for SARS-CoV-2.<sup>46</sup> This is consistent with our own data predicting moderate to low binding of SARS-CoV-2 to snake ACE2. Rodents varied widely in their permissiveness to SARS-CoV with mice being resistant and hamsters permissive. Similarly there is inefficient virus replication of SARS-CoV-2 in mice making them unsuitable as models to test SARS or COVID-19 vaccines or drugs.<sup>47</sup> This reflects the low predicted binding energy of SARS-CoV-2 spike protein for mouse ACE2, as demonstrated by our model data. Mice only became permissive for SARS infection when made transgenic for human ACE2, with the same likely to be true for SARS-CoV-2. By contrast, hamsters were highly permissive for SARS-CoV infection. Our model predicted that hamster should be permissive to SARS-CoV-2 infection based on the predicted strong binding of SARS-Cov-2 spike protein for hamster ACE2 (Table 4). Consistent with our model data, Syrian hamsters have been shown to exhibit clinical and histopathological responses to SARS-CoV-2 that closely mimic human upper and lower respiratory tract infections, with high virus shedding and ability to transmit to naïve contact animals.<sup>24</sup> Ferrets were another permissive model of SARS CoV infection, and our modelling data indicated that SARS-CoV-2 has a similar binding energy to ferret, as it does to hamster, ACE2 (Table 4). Consistent with our model data, ferrets have been shown to be permissive to infection with SARS-CoV-2, with high virus titre in the upper respiratory tract, virus shedding, infected ferrets showing acute bronchiolitis but without severe disease or death, and active transmission to naïve ferrets through direct contact.<sup>20,22</sup>

Cat and tiger ACE2 were shown by our model to have similar binding affinity for SARS-CoV-2 spike protein and both these species have been shown to be permissive for SARS-CoV-2 infection. Similarly our data suggests that SARS-CoV-2 binds with moderate affinity to dog ACE2 suggesting that dogs may be susceptible to infection by SARS-CoV-2. In the case of companion animals that live in close proximity to humans, Shen et al. state that SARS-CoV-2 can be efficiently transmitted in cats and dogs while Shi et al. found that SARS-CoV-2 replicates poorly in dogs, pigs, chickens, and ducks, but that ferrets and cats were permissive to infection. Temmam et al. tested 9 cats and 12 dogs living in close contact with their owners, two of whom tested positive for SARS-CoV-2 and 11 of 18 others showed clinical signs of COVID-19 but no antibodies against SARS-CoV-2 were detectable in their blood using an immunoprecipitation assay. Goumeniu et al. published an

editorial querying the role of dogs in the Lombardy COVID-19 outbreak and recommended use of computational docking experiments to provide evidence for or against infection of dogs.<sup>43</sup> Hence, our model data suggesting that dog ACE2 might be permissive for SARS-CoV-2 binding and infection is therefore consistent with anecdotal reports of dogs being infected with SARS-CoV-2. Hence other genetic factors could underlie the apparent lack of susceptibility of dogs to COVID-19 clinical infection.

It is known that gain of function (GOF) mutations occur in viruses that can lead to pandemics. GOF means viruses gain a new property e.g. in influenza virus GOF has been associated with the acquisition of a new function, such as mammalian transmissibility, increased virulence for humans, or evasion of existing host immunity.<sup>49</sup> The conditioning of viruses to humans as pandemics progress is well recognized. However, the SARS-CoV-2 structures and sequences that we employed were from viruses collected very early in the pandemic. It is therefore not clear how SARS-CoV-2 could have developed such a high affinity for human ACE2, notably higher than for those of putative zoonotic sources for SARS-CoV-2, unless it has been previously selected on human ACE2 or an ACE2 of another species bearing a closely homologous spike protein binding domain. Interestingly, pangolin ACE2 bears some similarities in its SBD to human ACE2. This marries with the fact that Pangolin-CoV shares a highly similar RBD to SARS-CoV-2, although their remaining sequence has only 90% similarity. This could be consistent with a process of convergent evolution whereby human and pangolin coronaviruses infecting via ACE2, have come to the same solution in respect of evolving an optimal spike RBD for binding of either human or pangolin ACE2, respectively. Our data does indicate that humans might be permissive to pangolin CoVs that use ACE2 for cell entry, a fact that needs to be borne in mind in respect of future potential coronavirus pandemic sources. However, this does not mean that pangolin ACE2 was the receptor on which the SARS-CoV-2 spike protein RBD was initially selected, with the strength of binding to pangolin ACE2 lower than binding to human ACE2. This makes it unlikely that pangolins are the missing intermediate host. If SARS-CoV-2 spike was selected on pangolin ACE2, then given the higher affinity of SARS-CoV-2 for human ACE2 than for bat ACE2, SARS-CoV-2 would have to have circulated in pangolins for a long period of time for this evolution and selection to occur and to date there is no evidence of a SARS-CoV-2 like virus circulating in pangolins.

Another possibility would be a short term evolutionary step where a pangolin was recently co-infected with a bat ancestor to SARS-CoV-2 at the same time as it was infected by a pangolin CoV allowing a recombination event to occur whereby the spike RBD of the pangolin virus was inserted into the bat CoV, thereby conferring the bat CoV with high binding for both pangolin and human

ACE2. Such recombination events are known to occur with other RNA viruses and can explain creation of some pandemic influenza strains<sup>49</sup>. Nevertheless, such events are by necessity rare as they require coinfection of the one host at exactly the same time. Most importantly, if such a recombination event had occurred in pangolins it might have been expected to have similarly triggered an epidemic spread of the new highly permissive SARS-CoV-2 like virus among pangolin populations, such as we now see occurring across the human population. Currently there is no evidence of such a pangolin SARS-CoV-2 like outbreak, making this whole scenario less likely. Indeed, pangolins might be protected from SARS-CoV-2 infection due to the existence of cross-protective spike RBD neutralising antibodies induced by exposure to pangolin CoV, given the RBD similarity of these two viruses. Another possibility which still cannot be excluded is that SARS-CoV-2 was created by a recombination event that occurred inadvertently or consciously in a laboratory handling coronaviruses, with the new virus then accidentally released into the local human population.

Given the seriousness of the ongoing SARS-CoV-2 pandemic, it is imperative that all efforts be made to identify the original source of the SARS-CoV-2 virus. In particular, it will be important to establish whether COVID-19 is due to a completely natural chance occurrence where a presumed bat virus was transmitted to humans via an intermediate animal host or whether COVID-19 has alternative origins. This information will be of paramount importance to help prevent any similar human coronavirus outbreak in the future.

### **Acknowledgements**

We would like to thank Harinda Rajapaksha for assistance to optimise Gromacs for this project. We would like to thank Oracle Corporation for providing their Cloud computing resources through the Oracle for Research program for the modelling studies described herein. In particular, we wish to thank Peter Winn, Dennis Ward, and Alison Derbenwick Miller for facilitating this access. S.P., P.K.S and N.P. are supported by the National Institute of Allergy and Infectious Disease of the National Institutes of Health under Contracts HHSN272201400053C and HHSN272201800044C. This publication's contents are solely the responsibility of the authors and do not necessarily represent the official views of their affiliated institutions, funding bodies or Oracle Corporation.

## References

- 1 Zhang, T. et al. Clinical trials for the treatment of Coronavirus disease 2019 (COVID-19): A rapid response to urgent need. *Sci China Life Sci*, doi:10.1007/s11427-020-1660-2 (2020).
- 2 Zhang, J. et al. Progress and Prospects on Vaccine Development against SARS-CoV-2. *Vaccines (Basel)* 8, doi:10.3390/vaccines8020153 (2020).
- 3 Yavuz, S. & Unal, S. Antiviral Treatment of Covid-19. *Turk J Med Sci*, doi:10.3906/sag-2004-145 (2020).
- 4 Whitworth, J. COVID-19: a fast evolving pandemic. *Trans R Soc Trop Med Hyg* 114, 241-248, doi:10.1093/trstmh/traa025 (2020).
- 5 Thanh Le, T. et al. The COVID-19 vaccine development landscape. *Nat Rev Drug Discov*, doi:10.1038/d41573-020-00073-5 (2020).
- 6 Sohrabi, C. et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int J Surg* 76, 71-76, doi:10.1016/j.ijisu.2020.02.034 (2020).
- 7 Schlagenhauf, P., Grobusch, M. P., Maier, J. D. & Gautret, P. Repurposing antimalarials and other drugs for COVID-19. *Travel Med Infect Dis*, 101658, doi:10.1016/j.tmaid.2020.101658 (2020).
- 8 Sanders, J. M., Monogue, M. L., Jodlowski, T. Z. & Cutrell, J. B. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*, doi:10.1001/jama.2020.6019 (2020).
- 9 Rosales-Mendoza, S., Marquez-Escobar, V. A., Gonzalez-Ortega, O., Nieto-Gomez, R. & Arevalo-Villalobos, J. I. What Does Plant-Based Vaccine Technology Offer to the Fight against COVID-19? *Vaccines (Basel)* 8, doi:10.3390/vaccines8020183 (2020).
- 10 Rosa, S. G. V. & Santos, W. C. Clinical trials on drug repositioning for COVID-19 treatment. *Rev Panam Salud Pub* 44, e40, doi:10.26633/RPSP.2020.40 (2020).
- 11 Olsen, M., Cook, S. E., Huang, V., Pedersen, N. & Murphy, B. G. Perspectives: Potential Therapeutic Options for SARS-CoV-2 Patients Based on Feline Infectious Peritonitis Strategies: Central Nervous System Invasion and Drug Coverage. *Int J Antimicrob Agents*, 105964, doi:10.1016/j.ijantimicag.2020.105964 (2020).
- 12 Mendes, A. Research towards treating COVID-19. *Br J Community Nurs* 25, 204-205, doi:10.12968/bjcn.2020.25.4.204 (2020).
- 13 Lu, S. Timely development of vaccines against SARS-CoV-2. *Emerg Microbes Infect* 9, 542-544, doi:10.1080/22221751.2020.1737580 (2020).
- 14 Jiang, S. Don't rush to deploy COVID-19 vaccines and drugs without sufficient safety guarantees. *Nature* 579, 321, doi:10.1038/d41586-020-00751-9 (2020).

- 15 Ciotti, M. et al. COVID-19 Outbreak: An Overview. *Chemother*, 1-9, doi:10.1159/000507423 (2020).
- 16 Berkley, S. COVID-19 needs a big science approach. *Science*, doi:10.1126/science.abb8654 (2020).
- 17 Almendros, A. & Gascoigne, E. Can companion animals become infected with COVID-19? *Vet Rec* 186, 419-420, doi:10.1136/vr.m1322 (2020).
- 18 Almendros, A. Can companion animals become infected with COVID-19? *Vet Rec* 186, 388-389, doi:10.1136/vr.m1194 (2020).
- 19 Ahmad, T. et al. COVID-19: Zoonotic aspects. *Travel Med Infect Dis*, 101607, doi:10.1016/j.tmaid.2020.101607 (2020).
- 20 Shi, J. et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science*, doi:10.1126/science.abb7015 (2020).
- 21 Temmam, S. et al. Absence of SARS-CoV-2 infection in cats and dogs in close contact with a cluster of COVID-19 patients in a veterinary campus. *bioRxiv* (2020) doi: 10.1101/2020.04.07.029090..
- 22 Kim, Y. I. et al. Infection and Rapid Transmission of SARS-CoV-2 in Ferrets. *Cell Host Microbe*, doi:10.1016/j.chom.2020.03.023 (2020).
- 23 Yu, P. et al. Age-related rhesus macaque models of COVID-19. *Animal Model Exp Med* , 1-5, doi:10.1002/ame2.12108 (2020).
- 24 Chan, J. F. et al. Simulation of the clinical and pathological manifestations of Coronavirus Disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease pathogenesis and transmissibility. *Clin Infect Dis*, doi:10.1093/cid/ciaa325 (2020).
- 25 Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367, 1444-1448, doi:10.1126/science.abb2762 (2020).
- 26 Touyz, R. M., Li, H. & Delles, C. ACE2 the Janus-faced protein - from cardiovascular protection to severe acute respiratory syndrome-coronavirus and COVID-19. *Clin Sci (Lond)* 134, 747-750, doi:10.1042/CS20200363 (2020).
- 27 Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature*, doi:10.1038/s41586-020-2179-y (2020).
- 28 Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234, 779-815, doi:10.1006/jmbi.1993.1626 (1993).
- 29 Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinform* 27, 343-350, doi:10.1093/bioinformatics/btq662 (2011).

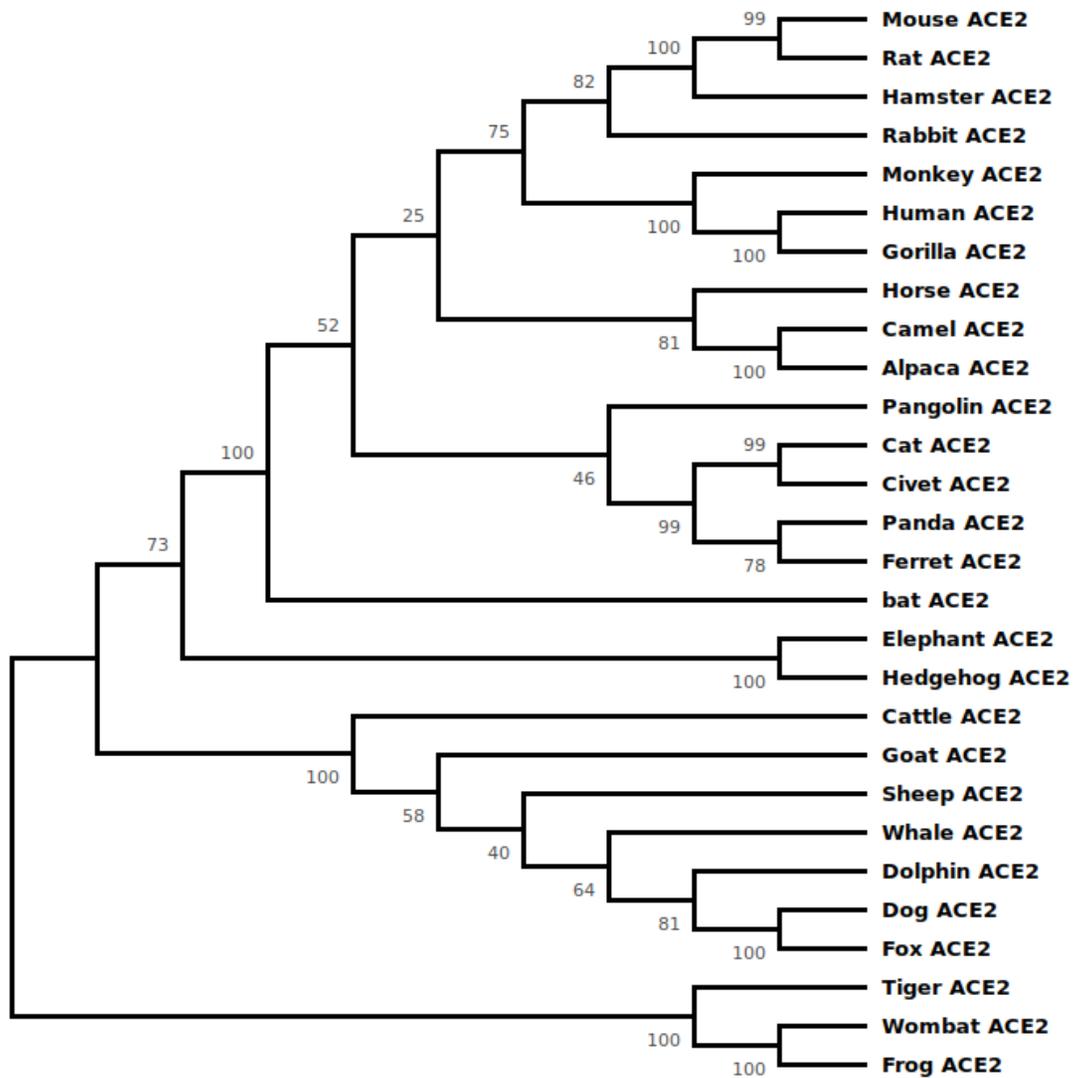
- 30 Pettersen, E. F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612, doi:10.1002/jcc.20084 (2004).
- 31 Yan, Y., Zhang, D., Zhou, P., Li, B. & Huang, S. Y. HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res* 45, W365-W373, doi:10.1093/nar/gkx407 (2017).
- 32 Yan, Y., Tao, H., He, J. & Huang, S. Y. The HDock server for integrated protein-protein docking. *Nat Protoc*, doi:10.1038/s41596-020-0312-x (2020).
- 33 Abraham, M. J. et al. GROMACS; High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2, 19-25 (2015).
- 34 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98, 10037-10041, doi:10.1073/pnas.181342398 (2001).
- 35 Kumari, R., Kumar, R., Open Source Drug Discovery, C. & Lynn, A. g\_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* 54, 1951-1962, doi:10.1021/ci500020m (2014).
- 36 Guterres, H. & Im, W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J Chem Inf Model*, doi:10.1021/acs.jcim.0c00057 (2020).
- 37 Ge, X. Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 503, 535-538, doi:10.1038/nature12711 (2013).
- 38 Li, W. et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450-454, doi:10.1038/nature02145 (2003).
- 39 Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J Virol* 94, doi:10.1128/JVI.00127-20 (2020).
- 40 Yuan, M. et al. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science*, doi:10.1126/science.abb7269 (2020).
- 41 Qiu, Y. et al. Predicting the angiotensin converting enzyme 2 (ACE2) utilizing capability as the receptor of SARS-CoV-2. *Microbes Infect*, doi:10.1016/j.micinf.2020.03.003 (2020).
- 42 Wu, C. et al. In Silico Analysis of Intermediate Hosts and Susceptible Animals of SARS-CoV-2. *ChemRxiv* (2020) doi: 10.26434/chemrxiv.12057996.v1.
- 43 Goumenou, M., Spandidos, D. A. & Tsatsakis, A. [Editorial] Possibility of transmission through dogs being a contributing factor to the extreme Covid19 outbreak in North Italy. *Mol Med Rep* 21, 2293-2295, doi:10.3892/mmr.2020.11037 (2020).
- 44 Lam, T. T. et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature*, doi:10.1038/s41586-020-2169-0 (2020).

- 45 Mahdy, M. An Overview of SARS-CoV-2 and Animal Infection. Preprints (2020) doi: 10.20944/preprints202004.0192.v1.
- 46 Luan, J., Jin, X., Lu, Y. & Zhang, L. SARS-CoV-2 spike protein favors ACE2 from Bovidae and Cricetidae. *J Med Virol*, doi:10.1002/jmv.25817 (2020).
- 47 Bao, L. et al. The pathogenicity of 2019 novel coronavirus in hACE2 transgenic mice. *bioRxiv* (2020) doi: 2020/2002.2007.939389.
- 48 Rockx, B. et al. Comparative pathogenesis of COVID-19, MERS And SARS in a non-human primate model. *bioRxiv* (2020) doi: 2020/2003.2017.995639.
- 49 Casadevall, A. & Imperiale, M. J. Risks and benefits of gain-of-function experiments with pathogens of pandemic potential, such as influenza virus: a call for a science-based discussion. *mBio* 5, e01730-01714, doi:10.1128/mBio.01730-14 (2014).
- 50 Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. Published online February 3, 2020. <https://doi.org/10.1038/s41586-020-2012-7>.
- 51 Zhang, T.; Wu, Q.; Zhang, Z. 2020, *Current Biology* 30, 1346–1351, April 6, 2020

# SUPPLEMENTARY INFORMATION

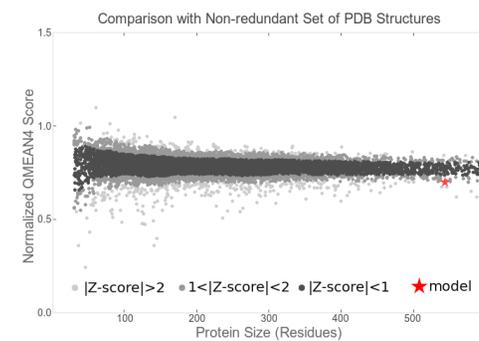
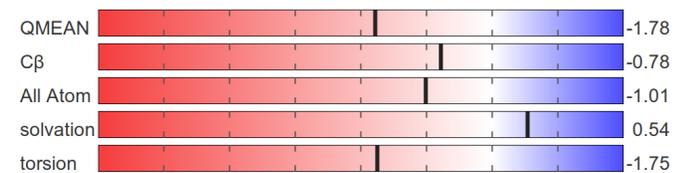
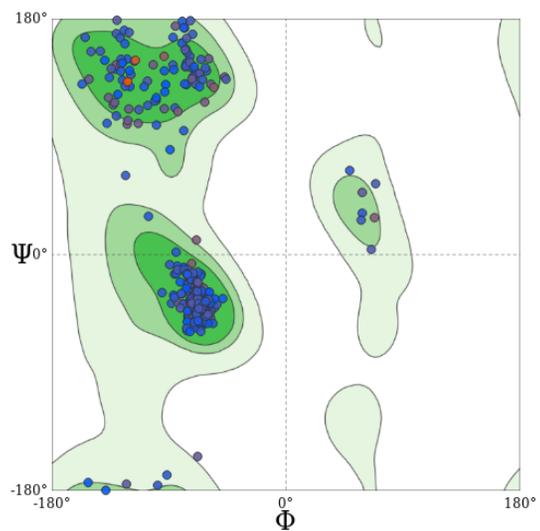
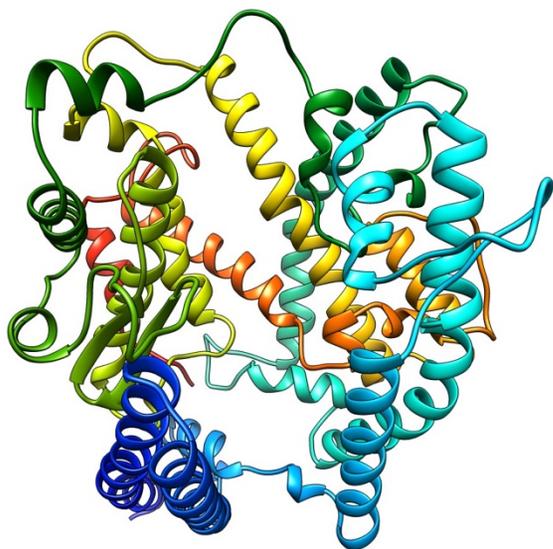
Ophiophagus	MLMKQAPVRKPSRSRSTHFAFFDLKGNMLTWLCLTWSLVVLAQA-DETKVATKFLQFD	59
Mesocricetus	-----MSSSWL LLS--LVAVTTAQSIIEEQAKTFLDKFN	33
Mus	-----MSSSWL LLS--LVAVTTAQSLTEENAKTFLNMFN	33
Rhinolophus	-----MSGSWL LLS--LVAVTTAQSTTEDEAKMFLDKFN	33
Bos	-----MTGSFWL LLS--LVAVTTAAQSTTEEQAKTFLKFN	33
Homo	-----MSSSWL LLS--LVAVTTAAQSTIEEQAKTFLDKFN	33
Macaca	-----MSGSWL LLS--LVAVTTAAQSTIEEQAKTFLDKFN	33
Mustela	-----MLGSSWL LLS--LAALTTAAQSTTEDLAKTFLKFN	33
Canis	-----MSGSWL LLS--LAALTTAAQS-TEDLVKTFLEKFN	32
Paguma	-----MSGFWL LLS--FAALTTAAQSTTEELAKTFLKFN	33
Felis	-----MSGFWL LLS--FAALTTAAQSTTEELAKTFLKFN	33
Panthera	-----LS--FAALTTAAQSTTEELAKTFLKFN	25
Manis	-----MSGSWL LLS--LVAVTTAAQSTDEEAKTFLKFN	33
Equus	-----MSGSWL LLS--LVAVTTAAQSTTEDLAKTFLKFN	33
	* : : : : * * : : : * * : : : *	
Ophiophagus	ARATDLYYNASIASWNYNTNLTEENAKIMHEKDNIFSKFYGEACRNAMFNVNHTDETI	119
Mesocricetus	QEAEADLSYQSALASWNYNTNI TEENAQKMEAAAKWSAFYEEQSKLAKNYSLQEVQNLTI	93
Mus	QEAEADLSYQSSLASWNYNTNI TEENAQKMEAAAKWSAFYEEQSKTAQFSLQEIQTPII	93
Rhinolophus	TKAEADLSHQSSSLASWNYNTNI DENVQKMEAGAKWSAFYEEQSKLAKNYSLEIQNVTV	93
Bos	HEAEADLSYQSSLASWNYNTNI DENVQKMEAGAKWSAFYEEQSKMAKTYSLLEIQNLT	93
Homo	HEAEADLFYQSSSLASWNYNTNI TEENVQNMNAGDKWSAFLEKQSTLAQMYPLQEIQNLTV	93
Macaca	HEAEADLFYQSSSLASWNYNTNI TEENVQNMNAGDKWSAFLEKQSTLAQMYPLQEIQNLTV	93
Mustela	YEAEELSYQSSLASWNYNTNI DENIQKMNAGAKWSAFYEEESQHKATYPLLEIQDPDI	93
Canis	YEAEELSYQSSLASWNYNTNI DENVQKMNAGAKWSAFYEEQSKLAKTYPLLEIQDSTV	92
Paguma	YEAEELSYQSSVASWNYNTNI DENAKNMNAGAKWSAFYEEQSKLAQTYPLAEIQDAKI	93
Felis	HEAEELSYQSSLASWNYNTNI DENVQKMEAGAKWSAFYEEQSKLAKTYPLAEIHNTTV	93
Panthera	HEAEELSYQSSLASWNYNTNI DENVQKMEAGAKWSAFYEEQSKLAQTYPLAEIHNTTV	85
Manis	SEAEELSYQSSLASWNYNTNI DENVQKMNAGAKWSAFYEEQSKLAKTYPLAEIQNDTI	93
Equus	SEAEELSHQSSSLASWNYNTNI DENVQKMEAGARWSAFYEEQCKLAKTYPLLEIQNLT	93
	* : * : : : : * * * * * : * * * * * : * : * : * : : : *	
Ophiophagus	KLQIRLLQSGSTD---STKQDLDVLHKMSTLYS-----LDDIMA	156
Mesocricetus	KRQLQALQSGSSALSADKNKQLNLTNTMSTIYSTGKVCNPKNPQECLEPGLDDIMA	153
Mus	KRQLQALQSGSSALSADKNKQLNLTNTMSTIYSTGKVCNPKNPQECLEPGLDDIMA	153
Rhinolophus	KLQQLQALQSGSPVLSADKSKRLNLTNTMSTIYSTGKVCNPKNPQECLEPGLDNIMG	153
Bos	KRQLKALQSGSALSADKSKRLNLTNTMSTIYSTGKVLDPN-TQECALALEPGLDDIME	152
Homo	KLQQLQALQSGSSVLSADKSKRLNLTNTMSTIYSTGKVCNPNPQECLEPGLNEIMA	153
Macaca	KLQQLQALQSGSSVLSADKSKRLNLTNTMSTIYSTGKVCNPNPQECLEPGLNEIME	153
Mustela	KRQLRALQSGSSVLSADKSKRLNLTNTMSTIYSTGKACNPNPQECLEPGLDDIME	153
Canis	KRQLRALQSGSSVLSADKSKRLNLTNTMSTIYSTGKACNPNPQECLEPGLDDIME	152
Paguma	KRQLQALQSGSSVLSADKSKRLNLTNTMSTIYSTGKACNPNPQECLEPGLDNIME	153
Felis	KRQLQALQSGSSVLSADKSKRLNLTNTMSTIYSTGKACNPNPQECLEPGLDDIME	153
Panthera	KRQLQALQSGSSVLSADKSKRLNLTNTMSTIYSTGKACNPNPQECLEPGLDDIME	145
Manis	KRQLQALQSGSSALSADKSKRLNLTNTMSTIYSTGKVCNPNPQECLEPGLDNIME	153
Equus	KRQLQALQSGSSVLSADKSKRLNLTNTMSTIYSTGKVCNPNPQECLEPGLDAIME	153
	* * * * * : * : * * * * * : * : * * * * * : * * * * *	
Ophiophagus	NNWNYPERLWAWEGWRANVGGKMRPLYETVYELKNKYARLGRYADYGDYWRANVEVDLP	216
Mesocricetus	TSTDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYAEAGD	213
Mus	TSTDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYAEAGD	213
Rhinolophus	TSKDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYETEESP	213
Bos	NSRDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYEVGAG	212
Homo	NSLDYNERLWAWESWRSEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYEVGND	213
Macaca	KSLDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYEVGND	213
Mustela	NSKDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYEEWAD	212
Canis	NSKDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYEEWEN	213
Paguma	NSKDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYEEWGT	213
Felis	NSKDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYEEWTD	213
Panthera	NSKDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYEEWTD	205
Manis	SSKDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYAEAGN	213
Equus	NSKDYNERLWAWEGWRAEVGKQLRPLYEYVVLKNEMARANNYEDYDWRGDEYAEAGPS	213
	.. : * * * * * : * * * * * : * * * * * : * * * * *	
Ophiophagus	KFYQREQLITDVESTFKQ-----QLHAYVRHLLKRYGPELINPEGAIPAHLLGDMWG	270
Mesocricetus	GNYNNGNQLIEDVERTFKEIKPLYEHLHAYVRKLMNTY-PSYISPTGCLPAHLLGDMWG	272
Mus	GNYNNGNQLIEDVERTFAEIKPLYEHLHAYVRKLMNTY-PSYISPTGCLPAHLLGDMWG	272
Rhinolophus	GPYSDRQLMKDVERIFTEIKPLYEHLHAYVRKLMNTY-PFHISPTGCLPAHLLGDMWG	272
Bos	DYDYSRDLMKDVERIFTEIKPLYEHLHAYVRKLMHTY-PSYISPTGCLPAHLLGDMWG	272
Homo	GDYSDRQLIEDVEHTFEEIKPLYEHLHAYVRKLMNAY-PSYISPTGCLPAHLLGDMWG	271
Macaca	GDYSDRQLIEDVERTFEEIKPLYEHLHAYVRKLMNAY-PSYISPTGCLPAHLLGDMWG	272
Mustela	GYSYRNLQIEDVEHTFTQIKPLYEHLHAYVRKLMADY-PSRISPTGCLPAHLLGDMWG	272
Canis	GNYSDRNLQIEDVEHTFTQIMPLYQHLYHAYVRKLMNTY-PSYISPTGCLPAHLLGDMWG	271
Paguma	GNYSDRNLQIEDVEHTFEIKPLYEHLHAYVRKLMNTY-PSRISRTGCLPAHLLGDMWG	272
Felis	GNYSDRNLQIEDVEHTFTQIKPLYEHLHAYVRKLMNTY-PSRISPTGCLPAHLLGDMWG	272
Panthera	GNYSDRNLQIEDVEHTFTQIKPLYEHLHAYVRKLMNTY-PSRISPTGCLPAHLLGDMWG	264
Manis	GNYSDRDLIEDVEHTFTEIKPLYEHLHAYVRKLMNDY-PSHISPTGCLPAHLLGDMWG	272
Equus	GDYSDRQLIEDVERTFAEIKPLYEHLHAYVRKLMNTY-PSHINPTGCLPAHLLGDMWG	272
	* : : * * * * * : * : * * * * * : * : * * * * * : * : * * * * *	
Ophiophagus	RFWNTLYPLMVPYPNKTSIDVTSAMEKKKWTNSIFKAAEHFFISIGLFNMTVGFWKNMS	330
Mesocricetus	RFWNTLYPLTVPFQKPNIDVTDAMVQGWNAERIFKEAEKFFVSVGLPYMTQGFWNSM	332
Mus	RFWNTLYPLTVPFQKPNIDVTDAMVQGWDAERIFKEAEKFFVSVGLPNMTQGFWNSM	332
Rhinolophus	RFWNTLYPLTVPFQKPNIDVTEMLKQGWDAERIFKEAEKFFVSVGLPNMTQGFWNSM	332
Bos	RFWNTLYPLTVPFQKPNIDVTEKMNQSWDAERIFKEAEKFFVSVGLPNMTQGFWNSM	332
Homo	RFWNTLYPLTVPFQKPNIDVTDAMVDQAWDAERIFKEAEKFFVSVGLPNMTQGFWNSM	331
Macaca	RFWNTLYPLTVPFQKPNIDVTDAMVQAWNAERIFKEAEKFFVSVGLPNMTQGFWNSM	332
Mustela	RFWNTLYPLMVPFRQKPNIDVTDAMVQSWDARRIFEEAEFTFVSVGLPNMTQGFWNSM	332
Canis	RFWNTLYPLTVPFQKPNIDVTDAMVQSWDARRIFKEAEKFFVSVGLPNMTQGFWNSM	332
Paguma	RFWNTLYPLTVPFQKPNIDVTDAMVQSWDARRIFKEAEKFFVSVGLPNMTQGFWNSM	331
Felis	RFWNTLYPLTVPFQKPNIDVTDAMVQSWDARRIFKEAEKFFVSVGLPNMTQGFWNSM	332
Panthera	RFWNTLYPLTVPFQKPNIDVTDAMVQSWDARRIFKEAEKFFVSVGLPNMTQGFWNSM	324
Manis	RFWNTLYPLTVPFQKPNIDVTDAMVQSWDARRIFKEAEKFFVSVGLPKMTQGFWNSM	332
Equus	RFWNTLYPLTVPFQKPNIDVTDAMVQSWDAKRIFEEAEKFFVSVGLPNMTQGFWNSM	332
	***** * * * : * * * * * : * : * * * * * : * * * * * : * * * * *	
Ophiophagus	LEEPKGRKVVCHPTAWDMGKEDYRIKMKCTKINMEDFLTAHHEMGIHYDMAYANQPFL	390
Mesocricetus	LTPDGDGRKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYATQPFL	392
Mus	LTPADGRKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYARQPFL	392
Rhinolophus	LTEPGDGRKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYASQPFL	392
Bos	LTEPGDGRKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYAAQPFL	391
Homo	LTPDGNVQKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYAAQPFL	392
Macaca	LTPDGNVQKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYAAQPFL	392
Mustela	LTEPGDNKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYAEQPFL	392
Canis	LTEPDSRNVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYAAQPFL	391
Paguma	LTEPGDGRKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYAAQPFL	392
Felis	LTEPGDSRNVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYAVQPFL	392
Panthera	LTEPGDNKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYAVQPFL	384
Manis	LTEPGDGRKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYANQPFL	392
Equus	LTEPGDGRKVVCHPTAWDLGKGFRIKMKCTKVTMDNFLTAHHEMGIHYDMAYAVQPFL	392
	* : * : * : * * * * * : * : * * * * * : * : * * * * * : * * * * *	



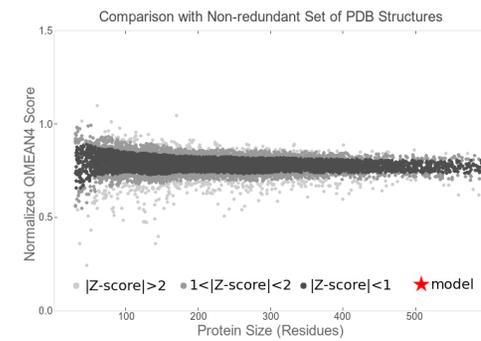
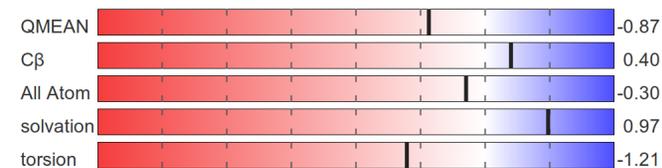
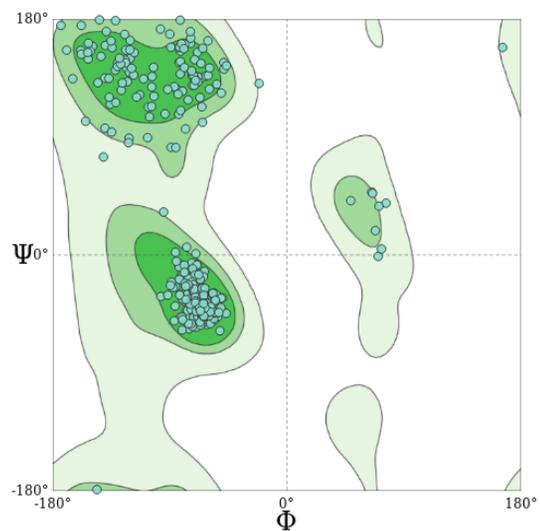
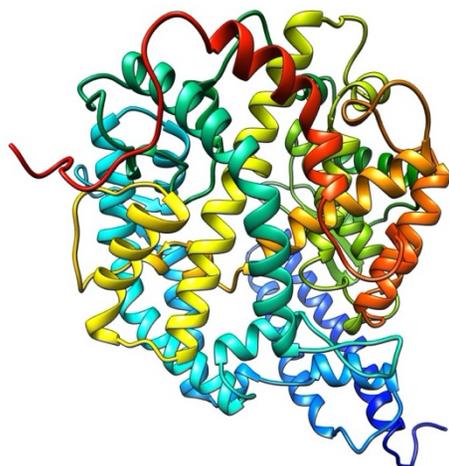


Supplementary Figure 2. Phylogenetic tree showing relatedness of sequences of ACE2 proteins from selected species.

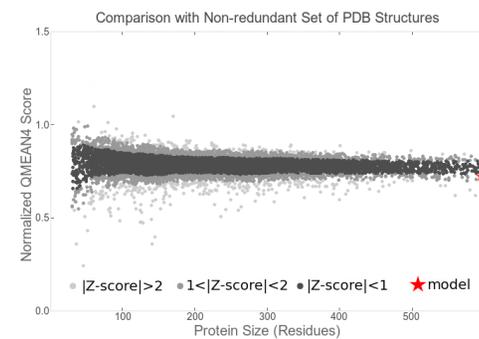
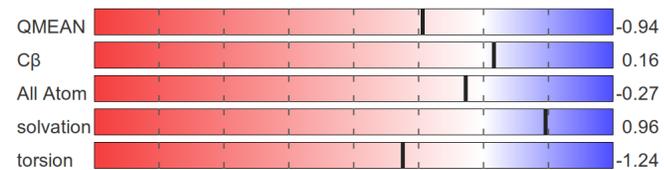
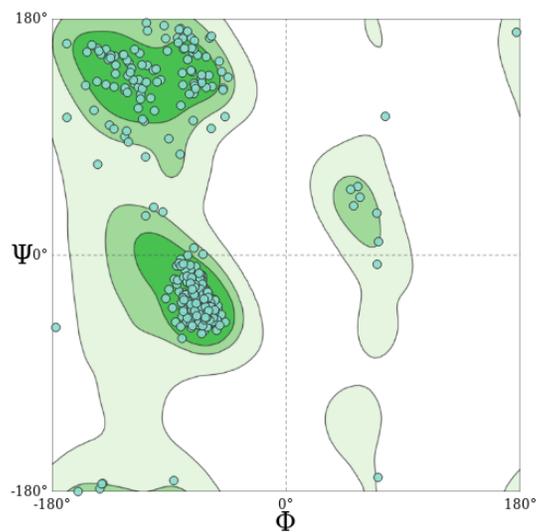
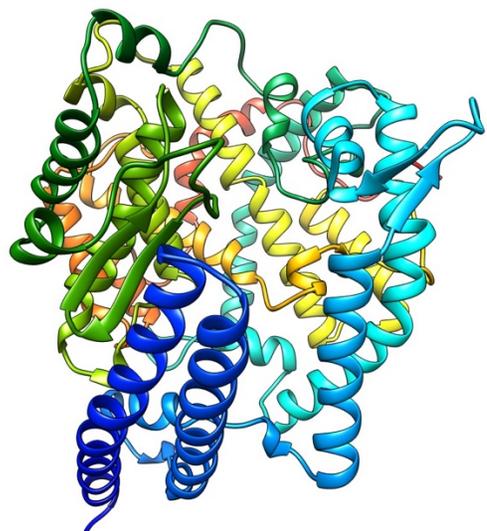
## BAT ACE2



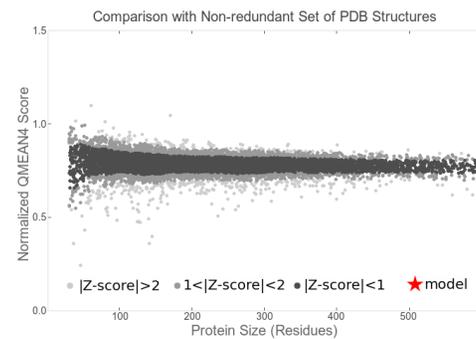
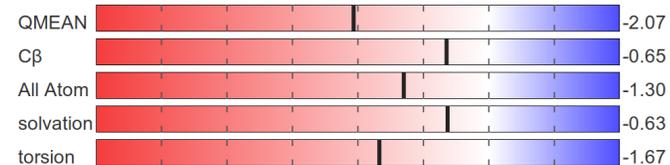
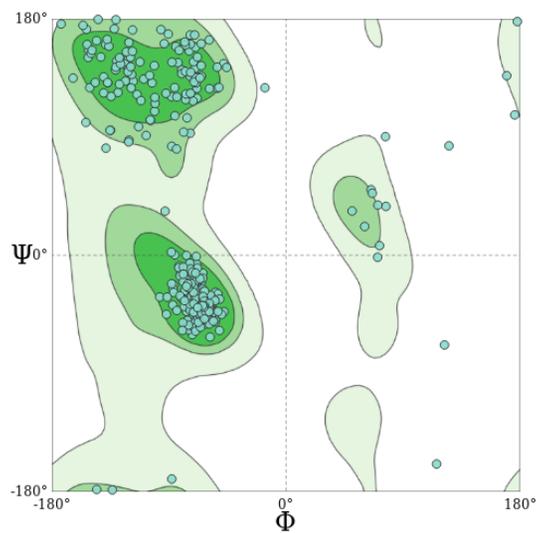
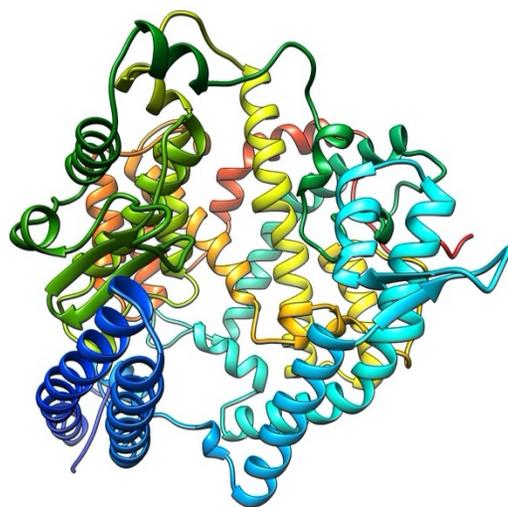
## CAT ACE2



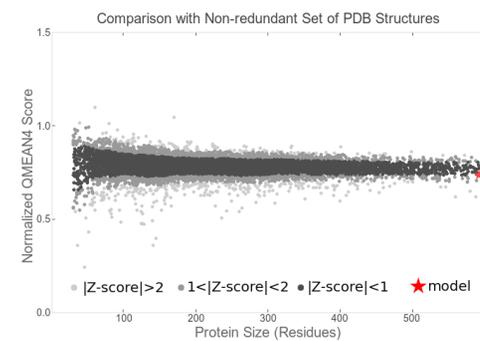
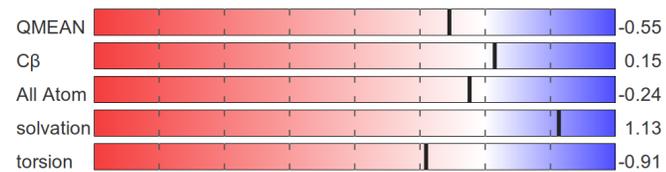
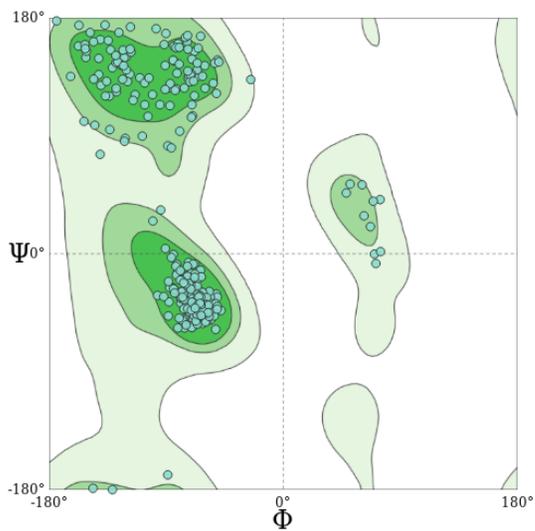
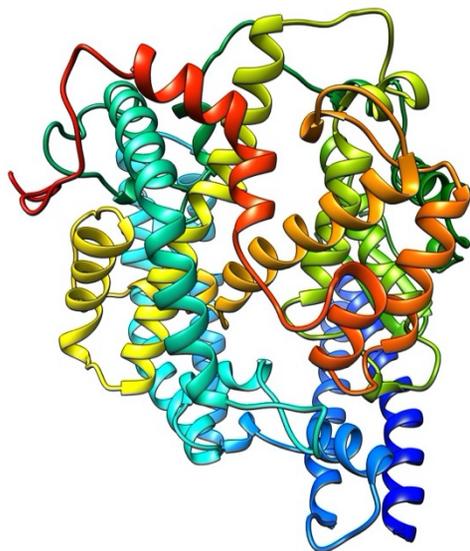
## CIVET ACE2



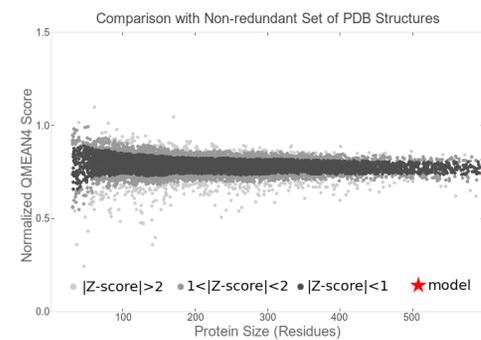
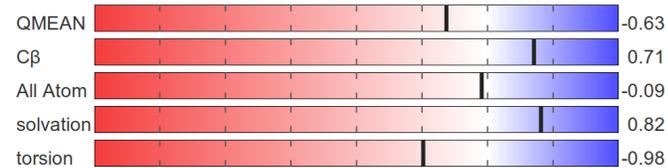
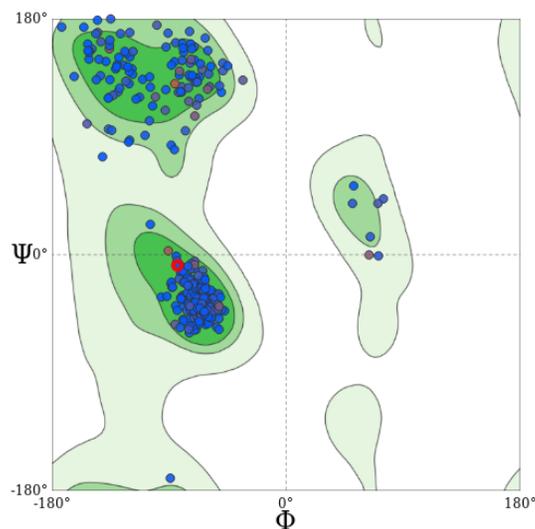
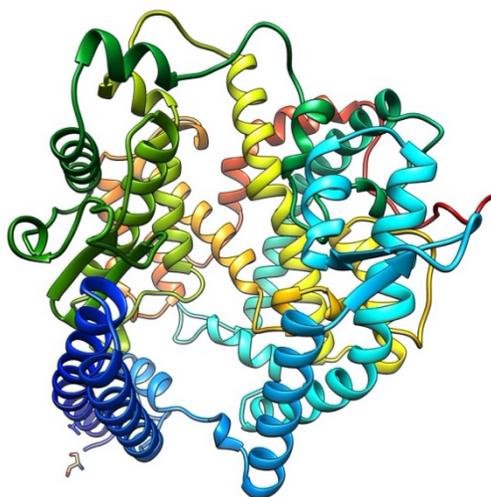
## DOG ACE2



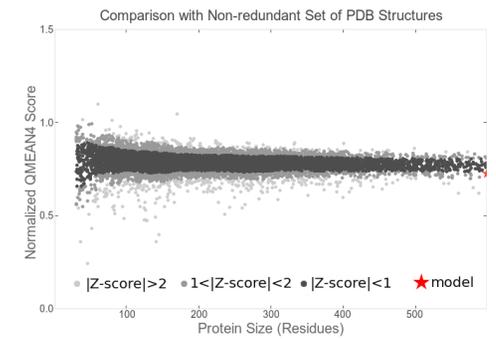
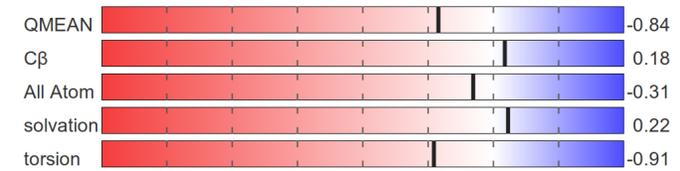
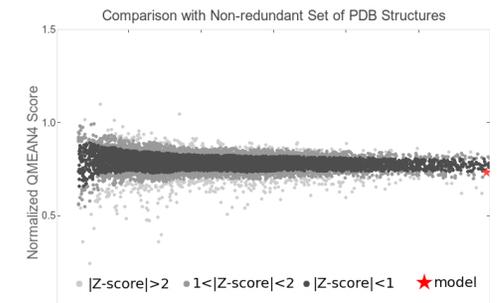
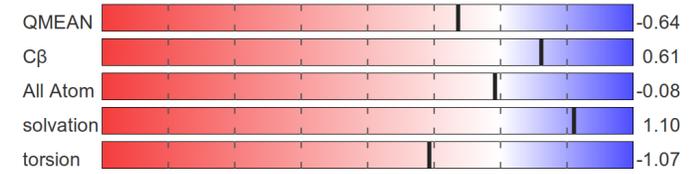
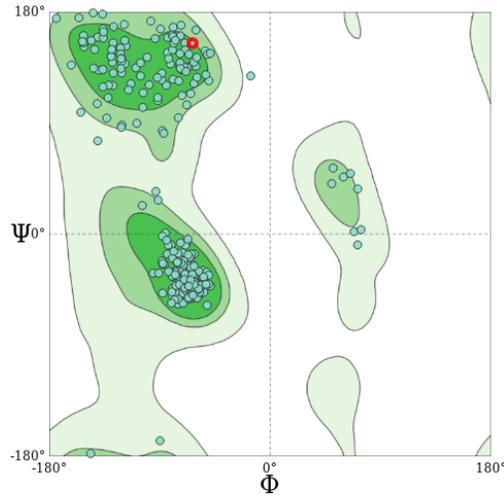
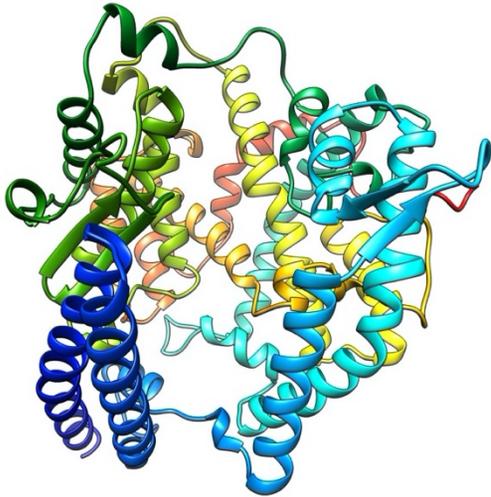
## FERRET ACE2



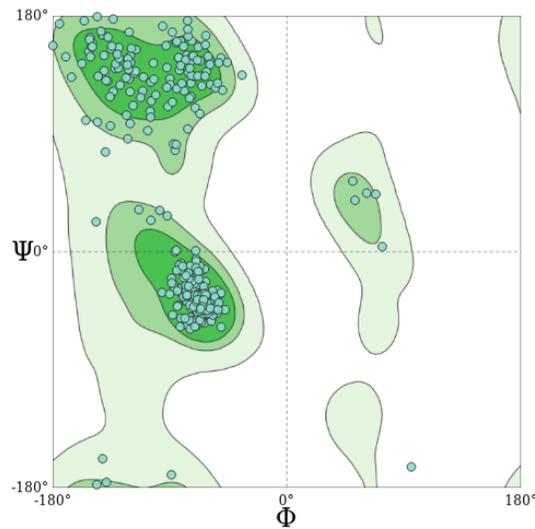
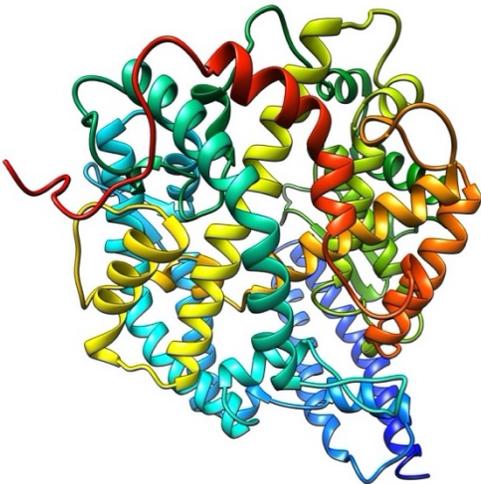
## HAMSTER ACE2



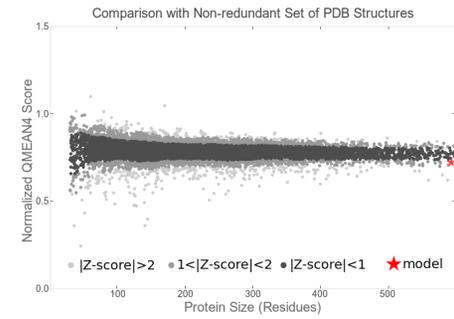
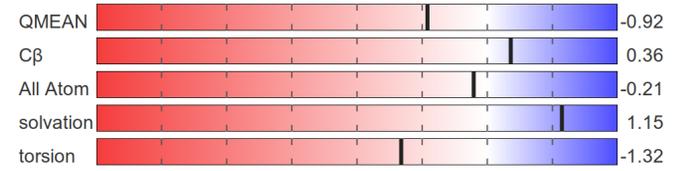
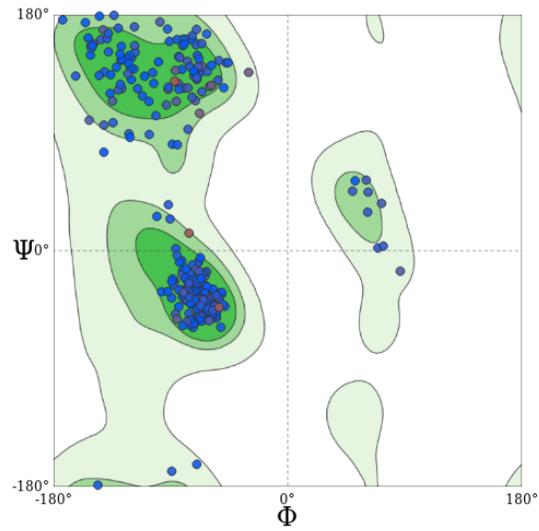
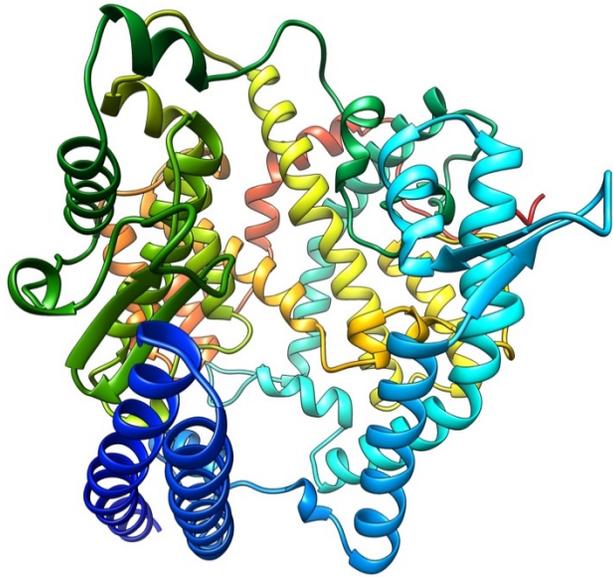
## MONKEY ACE2



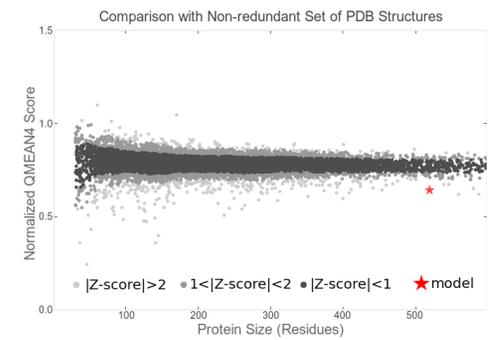
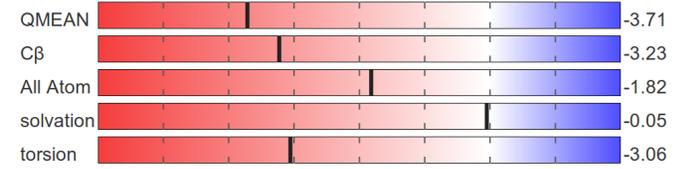
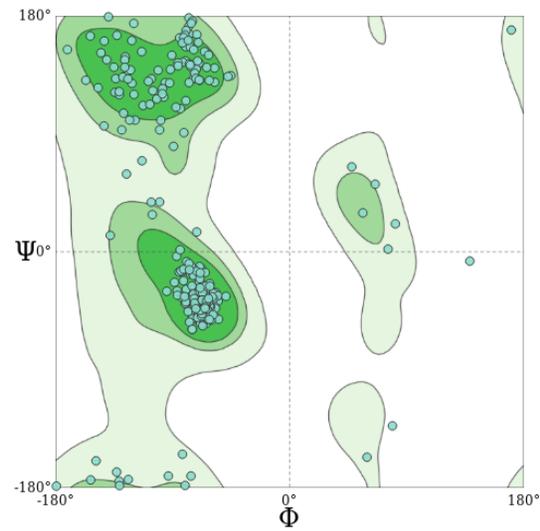
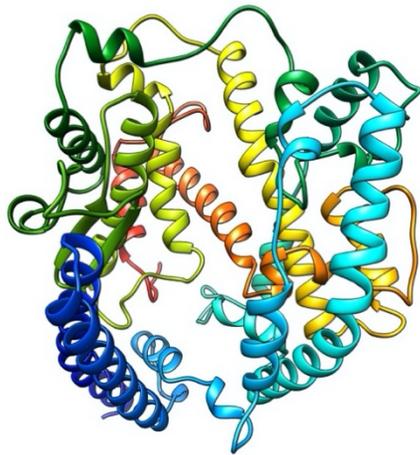
## MOUSE ACE2



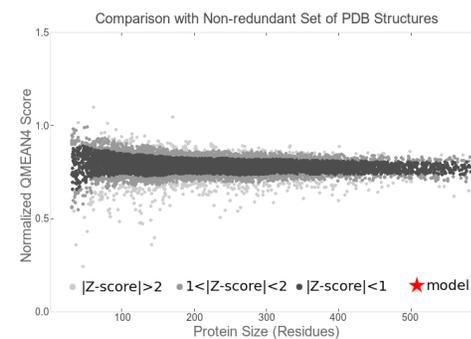
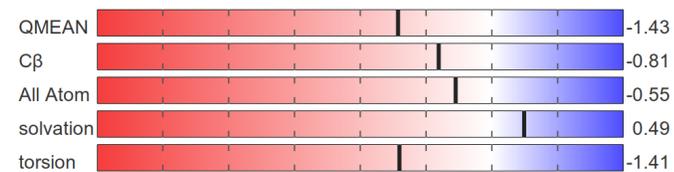
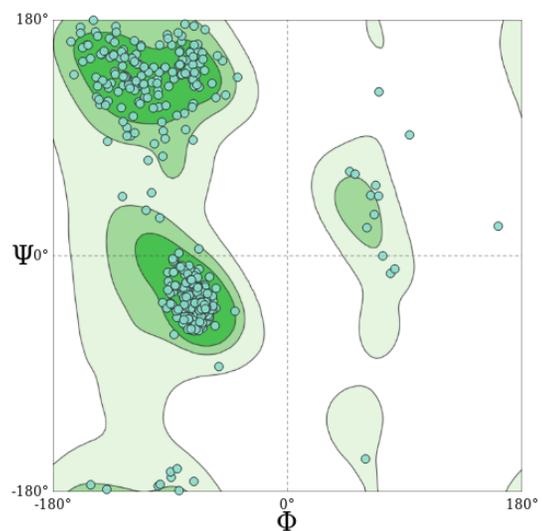
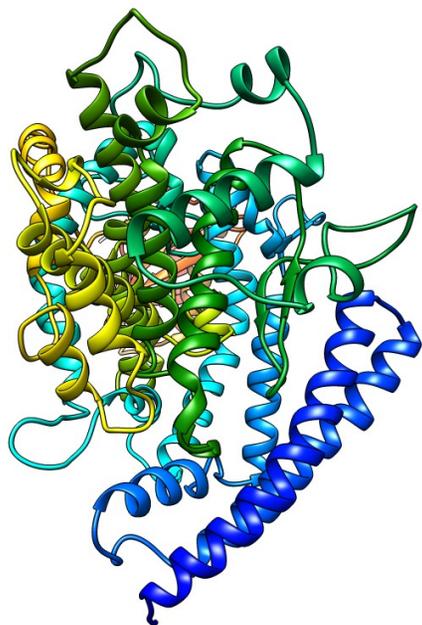
### PANGOLIN ACE2



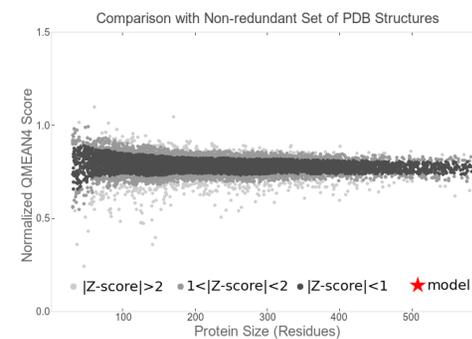
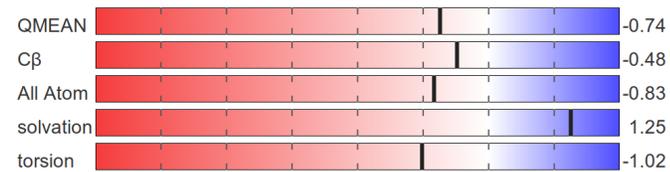
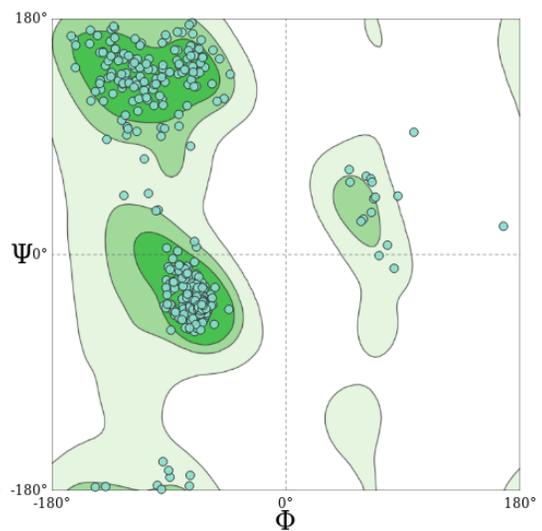
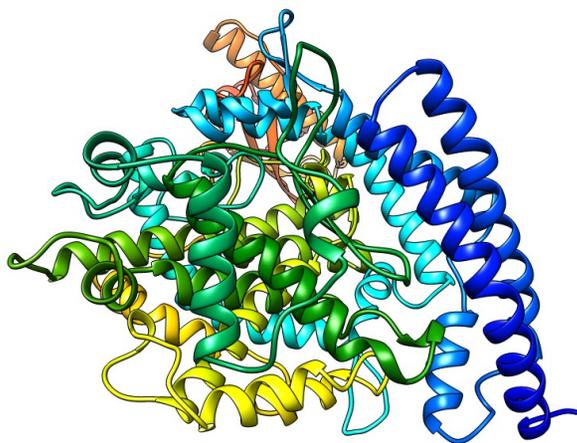
### SNAKE ACE2



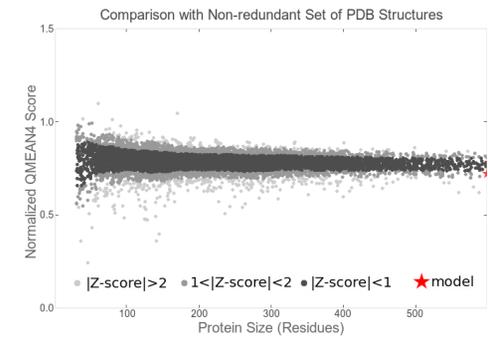
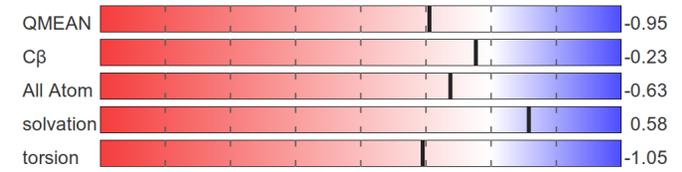
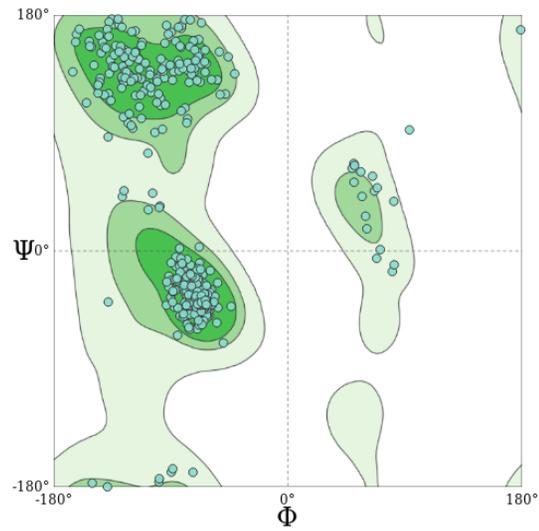
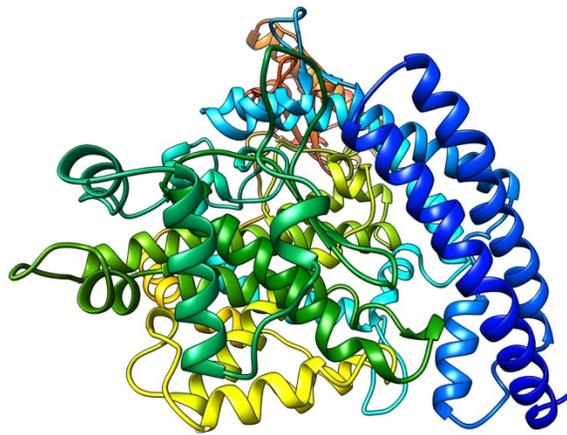
## HORSE ACE2



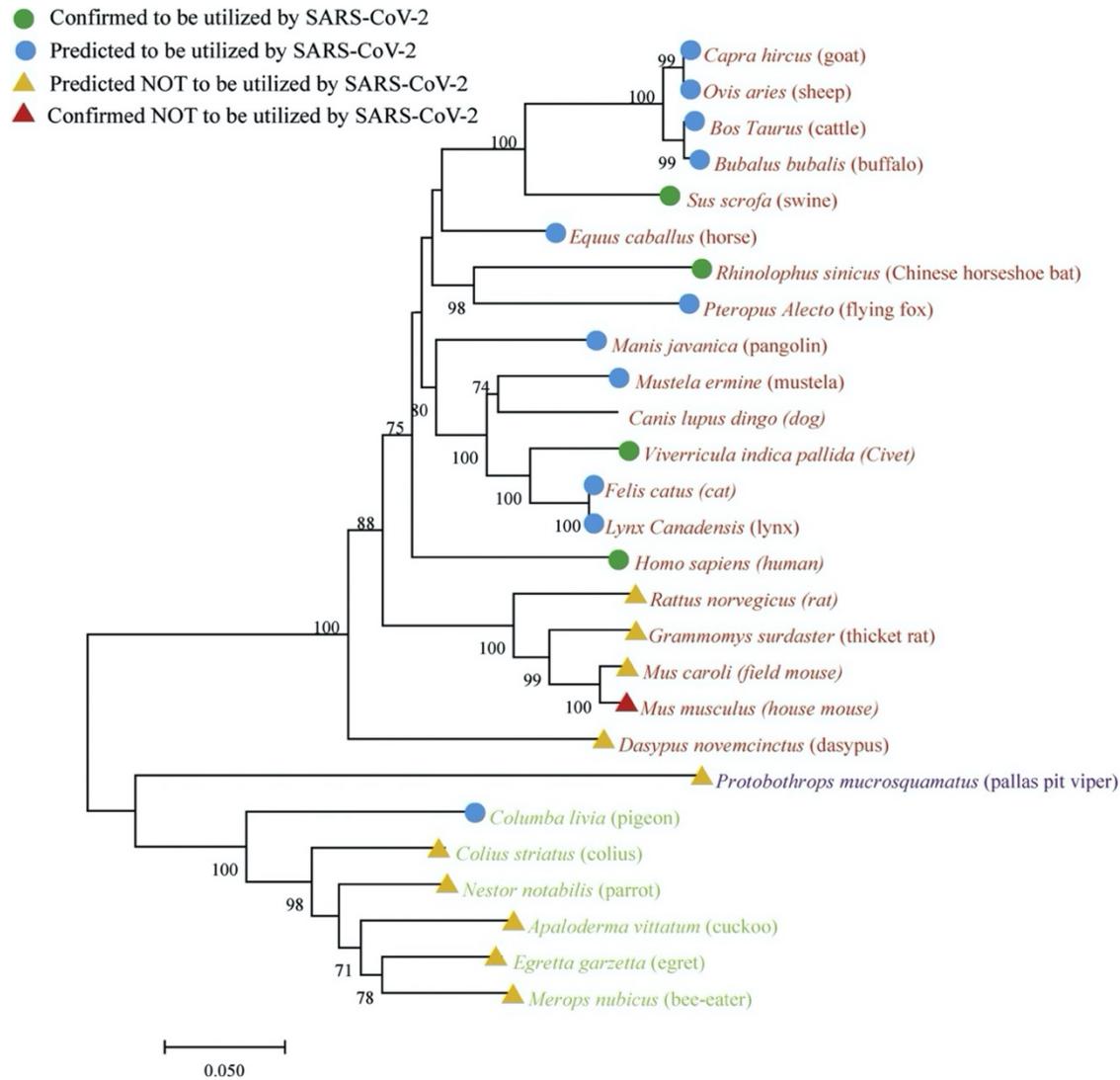
## CATTLE ACE2



## TIGER ACE2



Supplementary Figure 3. Modelled ACE2 structures for selected species, with Ramachandran plots and quality metrics



Supplementary Figure 4. Predicted and confirmed utilization of ACE2 receptors by the SARS-Cov-2 spike protein based on sequence homology from Qui et al.<sup>41</sup>