Insights on cross-species transmission of SARS-CoV-2 from structural modeling

João PGLM Rodrigues[1*], Susana Barrera-Vilarmau[#2], João MC Teixeira[#3], Elizabeth Seckel[4], Panagiotis Kastritis[5], Michael Levitt[1]

Affiliations:
1. Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305
2. Institute of Advanced Chemistry of Catalonia (IQAC), CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain
3. Program in Molecular Medicine, Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada
4. Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, CA 94305
5. ZIK HALOMEM & Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Biozentrum, Weinbergweg 22, 06120, Halle (Saale), Germany

* correspondence should be addressed to joaor@stanford.edu
[#] indicates equal contribution

Abstract
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for the ongoing global pandemic that has infected more than 6 million people in more than 180 countries worldwide. Like other coronaviruses, SARS-CoV-2 is thought to have been transmitted to humans from wild animals. Given the scale and widespread geographical distribution of the current pandemic, the question emerges whether human-to-animal transmission is possible and if so, which animal species are most at risk. Here, we investigated the structural properties of several ACE2 orthologs bound to the SARS-CoV-2 spike protein. We found that species known not to be susceptible to SARS-CoV-2 infection have non-conservative mutations in several ACE2 amino acid residues that disrupt key polar and charged contacts with the viral spike protein. Our models also predict affinity-enhancing mutations that could be used to design ACE2 variants for therapeutic purposes. Finally, our study provides a blueprint for modeling viral-host protein interactions and highlights several important considerations when designing these computational studies and analyzing their results.

Introduction

SARS-CoV-2, a novel betacoronavirus first identified in China in late 2019, is responsible for the ongoing global pandemic that has infected more than 6 million people worldwide and killed nearly 400.000 [1]. Based on comparative genomics, SARS-CoV-2 is thought to have been transmitted to humans from an animal host, most likely bats or pangolins [2]. Given the widespread human-to-human transmission across the globe, the question emerges whether humans can infect other animal species with SARS-CoV-2, namely domestic and farm animals. Identifying potential intermediate hosts that can act as reservoirs for the virus has both important global health, animal welfare, and ecological implications.

During the course of this pandemic, there have been several news reports of domestic, farm, and zoo animals testing positive for SARS-CoV-2 infection. Belgium [3] and New York [4] reported positive symptomatic cases in cats, The Netherlands reported infection of minks in farms [5], and the Bronx Zoo in New York reported infections in lions and tigers [6]. In all these cases, the vehicle of transmission appears to be an infected human owner or handler. More importantly, in the case of the mink farms in The Netherlands, there is evidence of human-to-animal-to-human transmission. In addition to these reported cases, several groups put forward both pre-prints and peer-reviewed studies on animal susceptibility to SARS-CoV-2 under controlled laboratory conditions [7–9], two of which are of particular interest. The first study showed that cats, civets, and ferrets are susceptible to infection; pigs, chickens, and ducks are not, while the results for dogs were inconclusive [7]. A second study, using human cells expressing recombinant SARS-CoV-2 receptor proteins showed that camels, cattle, cats, horses, sheep, and rabbit can be infected with the virus, but not chicken, ducks, guinea pigs, pigs, mice, and rats [8]. Together, these studies provide a dataset of confirmed susceptible and non-susceptible species that we can use to find molecular discriminants between the two groups. For simplicity, from here on we will refer to susceptible and non-susceptible species as SARS-CoV-2^{pos} and SARS-CoV-2^{neg}, respectively.

Like SARS-CoV-1 before, SARS-CoV-2 infection starts with the binding of the viral spike protein to the extracellular protease domain of angiotensin-converting enzyme 2 (ACE2) [10], a single-pass transmembrane protein expressed on the surface of a variety of tissues, including along the respiratory tract and the intestine. Several biophysical and structural studies identified helices α1 and α2, as well as a short loop between strands β3 and β4 in ACE2 as the interface for the viral spike protein [10–13]. These studies also identified key differences between the sequences of the receptor binding domains (RBD) of SARS-CoV-1 and SARS-CoV-2, which explain the stronger interaction of the latter with human ACE2. As such, we can reasonably assume that sequence variation across ACE2 orthologs might explain why some animal species are susceptible to infection while others are not. In addition, combining structural and binding data with the natural diversity of ACE2 across species is likely to shine a light on the key aspects that drive ACE2 interaction to viral RBDs and ultimately help guide the development of therapeutic molecules against SARS-CoV-2.

Unsurprisingly, several groups already published, or made available as preprints, multiple sequence and structure-based analyses of how sequence variation affects ACE2 binding to SARS-CoV-2 RBD [14–17]. Two recent preprints, specifically, focus on the effects of ACE2 variation on RBD binding. The first used an ACE2 sequence library to select for mutants that bind RBD with high affinity, identifying several mutants that enhance or decrease affinity to the viral protein and providing a blueprint for engineering proteins and peptides with therapeutic purposes [14]. While useful, we note that the authors carried out a single round of selection as opposed to the multiple rounds commonly carried out in similar studies. The second study used computational modeling to predict ΔΔG of mutations in 215 animal species and assess their risk for infection [15]. In addition, the authors also identified a number of locations on ACE2 that contribute to binding the viral RBD, in particular residues 31, 38, 353, as well as a cluster of N-terminal hydrophobic amino acid residues.

In this study, we aimed to leverage structural, binding, and sequence data to investigate how different ACE2 orthologs bind to SARS-CoV-2 RBD. We selected 29 animal species likely to encounter humans in a variety of residential, industrial, and commercial settings. For each of these species, we generated 3D models of ACE2 bound to RBD and refined these models using short molecular dynamic simulations. After refinement, we found that models of SARS-CoV-2^{pos} species generally have a lower (better) score than those of SARS-CoV-2^{neg} species. Further, we carried out a per-residue energy analysis that identified key locations in ACE2 that are consistently

mutated across SARS-CoV-2^{neg} species. Collectively, our results provide a structural framework that explains why certain animal species are not susceptible to SARS-CoV-2 infection, and also suggests potential mutations that can enhance binding to the viral RBD.

Results

Sequence conservation of ACE2 orthologs

We analyzed the sequence conservation of ACE2 across our dataset, with respect to the entire sequence (591 residues) and to the interface residues computed from a structure of ACE2 bound to RBD (PDB ID: 6m17) (22 residues) (Table 1). All orthologs are reasonably conserved, with global similarity values to the human ACE2 sequence (hACE2) ranging from 72% (goldfish) to 99.5% (chimpanzee) (Figure 1, left panel). All species coarsely cluster in three classes consistent with evolutionary distance to humans: primates have the highest similarity values, followed by other mammals, birds and reptiles, and finally fish. Zooming in on the interface residues, we find substantially more variation (Figure 1, right panel). Similarity values for these residues range from 50% (crocodile) to 100% (all 3 primates) but, despite an overall correlation (Pearson R^2 of 0.69), do not always match global similarities. Hedgehogs and sheep, for example, share 86.7% and 86.4% global similarity with hACE2, respectively, but 59% and 95.5% for the interface region. In absolute numbers, these similarities mean that sheep share 21 out of 22 residues with hACE2 at the interface with RBD, while hedgehogs share 13. The horseshoe bat, one of the proposed animal reservoirs for SARS-CoV-2, shares 72.2% interface similarity with hACE2, a comparable value to the 77.3% of the SARS-CoV-2^{neg} mouse sequence. Altogether, these results prompt two observations. First, neither global nor interface sequence similarity is predictive of SARS-CoV-2 susceptibility. Second, that the interface of the viral RBD is substantially plastic and able to bind to sufficiently different ACE2 orthologs.

Refinement of the hACE2:RBD complex

In order to validate the refinement protocol used in our analysis, we created and refined models of human ACE2 (hACE2) bound to SARS-CoV-2 RBD. We used the cryo-EM structure of full-length human ACE2 bound to the RBD, in the presence of the amino acid transporter B^0AT1 (PDB ID: 6m17). Compared to a high-resolution crystal structure of the same complex (PDB ID: 6m0j), the cryo-EM structure lacks several key contacts between our two proteins of interest, which we attribute to poor density for side-chain atoms at the interface region. Our refinement protocol restores the majority of these contacts (Table S1), yielding an average HADDOCK score of -116.2 (arbitrary units) for the 10 best models of the best cluster. See Materials and Methods for further details on the protocol. These negative HADDOCK scores suggest a favorable interface and agree with scores calculated for a reference set of transient protein-protein interactions (N=144, HADDOCK score=-124.9 ± 53.4) [18]. The interfaces in our models are dominated by hydrogen bond interactions involving the ACE2 α1 helix and a small loop between strands β3 and β4. There is one single salt-bridge involving hACE2 D30 and RBD K417 consistently present in all our hACE2 models. These observations all agree with the published crystal structure. Further, the buried surface area of the refined models is also in agreement with published crystal structures (~1800 Å^2). As such, considering the low quality of the interface region in our template structure, we are confident that our modeling and refinement protocol is robust enough to model all ACE2 orthologs.

Refinement of orthologous ACE2:RBD complexes

We modeled and refined complexes for all 29 ACE2 orthologs in our dataset (Table 1) using the same protocol as above. The representative models for each species (10 best models of the best cluster) are available for visualization and download at https://joaorodrigues.github.io/ace2-animal-models/. The HADDOCK scores of all 30 ACE2 complexes (including hACE2) range from -137.5 (dog) to -93.2 (mouse), a significant range that indicates substantial differences between these interfaces (Table 2 and Figure 2). The average HADDOCK score is -116.4, very close to that of the human complex (-116.2). Overall, models of SARS-CoV-2^{pos} species have consistently lower (better) scores than those of SARS-CoV-2^{neg} species. Although it is well-known that docking scores do not quantitatively correlate with experimental binding affinities [19], these scores suggest that SARS-CoV-2^{neg} species lack one or more key ACE2 residues that contribute significantly to the interaction with RBD.

To understand what forces drive the interactions between ACE2 and SARS-CoV-2 RBD, we quantified the contribution of each component of the HADDOCK scoring function to the overall score (Figure 3). The HADDOCK score is a linear combination of van der Waals, electrostatics, and desolvation energy terms. In our models, electrostatics are the most discriminatory component (Pearson R^2 of 0.60), followed by desolvation (0.31), and finally van der Waals (0.08). These correlations suggest that differences between the models of the different species originate primarily in polar and charged residues, in agreement with observations from experimental structures. In addition, the buried surface area of the models also correlates quite strongly with the HADDOCK score (Pearson R^2 of 0.66), which is unsurprising since larger interfaces tend to make more contacts. Most models bury between 1700 and 1850 $Å^2$, in agreement with the crystal and cryo-EM structures, while the top-scoring species (dog and goldfish) bury nearly 2000 $Å^2$ and the lowest-scoring (mouse) bury only 1600 $Å^2$. Finally, there is a weak correlation between the average HADDOCK score of the representative models and the sequence similarity of the ACE2 interface residues (Pearson R^2 of 0.18) (Figure S1).

Per-residue energetics of the ACE2:RBD interface

To gain further insight on how ACE2 sequence variation across the different orthologs affects binding to SARS-CoV-2 RBD, we calculated HADDOCK scores for each interface residue in the refined models (Figure 4). This high-resolution analysis reveals several sites that discriminate between SARS-CoV-2^{pos} and SARS-CoV-2^{neg} species.

The first and most relevant of these sites is amino acid 30, which in hACE2 (D30) interacts with RBD K417 to form the only intermolecular salt-bridge of the interface. In all 12 SARS-CoV-2^{pos} species, this site is occupied by a negatively charged amino acid residue. In contrast, 4 out of 5 SARS-CoV-2^{neg} species have a hydrophobic or polar residue at this position. The goldfish ACE2 sequence is an interesting outlier, with the second-best HADDOCK score despite having a lysine at position 30 that breaks the intermolecular salt-bridge. The loss of such an important site is compensated by the introduction of an alternative salt-bridge between E34 and RBD R403. Finally, the sequences of the top-scoring models also suggest that between aspartate and glutamate, the latter results in a stronger interaction, likely due to a stabilizing effect of the longer side-chain.

The second site is amino acid 31, a lysine in hACE2, and in nearly all of the SARS-CoV-2^{pos} species, that interacts both with ACE2 E35 and RBD Q493. The only exceptions are the civet and dromedary sequences, mutated to threonine and glutamate, respectively. In the case of the civet, our models show that T31 can still hydrogen bond with both E35 and RBD Q493. Dromedaries, on the other hand, share E31 with chickens, guinea pigs, and ducks, all SARS-CoV-2^{neg} species. However, and quite beautifully, dromedaries compensate the possible electrostatic repulsion between E31 and E35 with a lysine at position 76 (Q76 in hACE2) leading to the formation of an additional intramolecular salt-bridge that stabilizes the fold of ACE2 and frees E35 to hydrogen bond with Q493 in 90% of our models. All three SARS-CoV-2^{neg} species have an additional charge-reversal mutation at position 35, although with different outcomes in our models. In both chicken and duck ACE2, E31 is locked in an intramolecular salt-bridge with R35 in all of our models, losing the intermolecular hydrogen bond with RBD Q493. Lastly, guinea pigs compensate K31E with E35K and remain able to hydrogen bond with RBD.

The third discriminatory site between SARS-CoV-2^{pos} and SARS-CoV-2^{neg} species is amino acid 34, a histidine in hACE2 and a polar residue in all SARS-CoV-2^{pos} species. In our hACE2 models, H34 is doubly-protonated and forms an intramolecular salt-bridge/hydrogen bond with E37 and an intermolecular hydrogen bond with the hydroxyl group of RBD Y453. In addition, in most of our models, the aromatic ring of H34 is close enough (<4.5 Å) to the aliphatic side-chain of RBD L455 to form productive hydrophobic interactions. Our energetic analysis shows that substituting H34 by polar (serine, threonine) or hydrophobic (leucine, valine) residues destabilizes the interface, while substitution by a tyrosine substantially contributes to a stronger interaction. SARS-CoV-2^{neg} species except mouse and rat have hydrophobic residues at position 34, losing the ability to hydrogen bond with RBD Y453. In addition, the side-chain of RBD L455 is often out of range of hydrophobic interactions. In contrast, the H34Y substitution in the dog, ferret, and civet sequences loses the intramolecular hydrogen bond with E37 but compensates by hydrophobic interactions with nearby RBD residues and hydrogen bonds with RBD R403 (ferret), S494 (civet) or Y495 (dog). In addition, the loss of aromatic residues at position 34 leads to a steep decrease in desolvation energy of the models(Figure S2).

Besides these three major discriminatory sites, we identified three other sites that are systematically mutated in SARS-CoV-2^{neg} species. The first of these sites is K353 (in hACE2), which is involved in an intramolecular salt-bridge with D38 and an additional backbone hydrogen bond with RBD G502. In rat and mouse ACE2, both SARS-CoV-2^{neg} species, this residue is mutated to a histidine, which weakens the interaction with D38, possibly leading to increased conformational dynamics of the β3-β4 loop and consequently lower binding affinity. Then, position 42, a glutamine in hACE2 and in most other species, hydrogen bonds with RBD Y449 in the majority of our models. In canary, chicken, pigeon, hedgehog, duck, and crocodile ACE2 sequences, this amino acid is mutated to a glutamate. This substitution introduces the possibility of an additional intramolecular salt-bridge with K68, in ACE2 helix α2, which we observe in some of our models, preventing the formation of the intermolecular hydrogen bond. Finally, Y83 in hACE2 is mutated to phenylalanine in canary, chicken, rat, duck, and mouse ACE2, mostly SARS-CoV-2^{neg} species. Although our models do not offer a clear reason as to why this mutation could be damaging to RBD binding, the loss of the terminal hydroxyl group could have two negative consequences. First, there is the clear loss of two possible hydrogen bonds, with ACE2 Q24 and RBD N487. Then, the gain in hydrophobicity could lead the aromatic moiety to bury between both α1 and α2 helices, causing RBD F486 to lose a valuable interaction partner.

Discussion

All our models and the scoring statistics are available for visualization and download at https://joaorodrigues.github.io/ace2-animal-models/.

Can structural modeling predict cross-species transmission of SARS-CoV-2?

Our computational modeling of 30 vertebrate ACE2 orthologs bound to SARS-CoV-2 RBD discriminates between previously reported SARS-CoV-2^{pos} and SARS-CoV-2^{neg} species. Models of SARS-CoV-2^{neg} species – chicken, duck, guinea pig, mouse, and rat – generally have higher (worse) HADDOCK scores than average (Figure 2), suggesting that these species' non-susceptibility to infection could stem from deficient RBD binding to ACE2. Despite this clear trend, there are two notable outliers. Our modeling ranks guinea pig ACE2 (SARS-CoV-2^{neg}) as a better receptor for SARS-CoV-2 RBD than for example, human, cat, horse, or rabbit ACE2 (all SARS-CoV-2^{pos} species), despite experiments showing that there is negligible binding between the two proteins [8]. Then, the goldfish ACE2 sequence ranks second among all models, despite reports that fishes are unlikely to be susceptible to infection due to their physiology and environment [20].

These two results highlight the need for critical thinking when evaluating predictions from computational models. As noted earlier in the introduction, SARS-CoV-2 infection is a complex multi-step process [21]. Thus, while we can assume that impaired ACE2 binding decreases odds of infection, we cannot state that ACE2 binding is predictive of infection. For instance, experiments with recombinant ACE2 show that the pig ortholog binds SARS-CoV-2 RBD and leads to entry of the virus in host cells [8], but tests in live animals returned negative results [7]. In addition, our modeling protocol makes assumptions about the bound state of the two proteins, starting from the cryo-EM template structure. However, cryo-EM structures of the full-length SARS-CoV-2 spike protein [22] highlight multiple unbound conformations for RBD, and coarse-grained simulations of the hACE2:RBD complex show that there is substantial flexibility in some of the interfacial RBD loops [23]. Altogether, these limitations show that computational models alone cannot predict whether certain animal species are at risk of infection. What our models do predict, however, is that there are distinctive molecular features characteristic of SARS-CoV-2^{neg} species. As the adage goes, 'all models are wrong, but some are useful.'

SARS-CoV-2^{neg} species lack important polar and charged ACE2 residues

On further inspection, we find that SARS-CoV-2^{neg} models rank worse due to a substantial decrease in electrostatic energy (Figure 3), indicating loss of polar interface contacts, namely hydrogen bonds and salt-bridges (Figure 4). Indeed, models of mouse, duck, rat, and chicken lack the ability to form an intermolecular salt-bridge with RBD due to the loss of hACE2 D30. These predictions are supported by experimental work, where mutants lacking a negative charge at this position are largely unable to bind RBD [14]. Non-conservative

mutations at other sites on ACE2 also contribute negatively to the interface scores. Residues K31 and H34 (hACE2) engage multiple neighboring residues in both intra- and intermolecular hydrogen bonds, contributing both to ACE2 fold stability and RBD binding, respectively. Our models suggest that the introduction of a negatively charged residue at position 31 is disruptive to binding, in agreement with experiments [14]. In SARS-CoV-2^{pos} species, like dromedary camels, this mutation is more likely to be tolerated due to additional compensatory mutations that stabilize the ACE2 fold and still allow for contacts with RBD. In all SARS-CoV-2^{neg} species except guinea pig however, there are no additional mutations to compensate for this substitution. As for position 34, our predictions contrast with experimental measurements [14], which show that mutation to a hydrophobic residue improves binding between ACE2 and RBD. In our models, the preference seems to be for aromatic residues (histidine, tyrosine) capable of both hydrogen bonds and hydrophobic interactions. We note, however, that our coverage of sequence space is limited to naturally occurring variants. Unlike in the work referenced before [14], where the selection driver is RBD binding, natural selection of ACE2 might impose additional constraints on sequence variability. Finally, our models suggest that reduced flexibility of ACE2 might be a positive contributor to RBD binding affinity. Disrupting an intramolecular salt-bridge between D38 and K353 by substituting K353 with a shorter polar amino acid residue is a consistent feature in mice and rats, both SARS-CoV-2^{neg} species. These results support other computational modeling work [17] that suggest that RBD mutants G496D bind worse to ACE2 because of the disruption of this intramolecular salt-bridge.

Natural variants of ACE2 encode potential affinity-enhancing mutations for SARS-CoV-2 RBD

In addition to identifying mutations that impair binding of SARS-CoV-2 RBD, our models suggest several hACE2 variants that could be used to enhance affinity between the two proteins. The clearest affinity enhancer seems to be D30E, a variant observed in 6 of the 8 best scoring species (Figure 4) and shown in experiments to increase binding to RBD [8,14]. The longer side-chain of a glutamate residue can help strengthen and stabilize the intermolecular salt-bridge with RBD K417. The impact of such conservative mutations in stabilizing protein interactions has been reported previously for other systems [24]. The second predicted enhancer is H34Y, which as we discussed above, contrasts with experimental measurements. In addition to maintaining hydrogen bonds and hydrophobic contacts, our models show that this mutation results in a substantial increase in desolvation energy (Figure S2).

In summary, our protocol combines structural, sequence, and binding data to create a structure-based framework to understand SARS-CoV-2 susceptibility across different animal species. Our models help rationalize the impact of naturally-occurring ACE2 mutations on SARS-CoV-2 RBD binding and explain why certain species are not susceptible to infection with the virus. In addition, we propose possible affinity-enhancing mutants that can help guide engineering efforts for the development of ACE2-based antiviral therapeutics. Despite the aforementioned limitations, our protocol and models can easily be replicated using freely-available tools and web servers and serve as a blueprint for future modeling studies on ACE2 interactions with coronaviruses' RBDs. Finally, to prevent human-to-animal transmission, we recommend following the World Organization for Animal Health guidelines: people infected with COVID-19 should limit contact with their pets, as well as with other animals (including humans).

Materials and Methods

Sequence Alignment of ACE2 Orthologs

Sequences of ACE2 orthologs from 28 species were retrieved from NCBI using the human gene as a reference (Gene ID: 59272, updated on 20-Apr-2020) and the query term "ortholog_gene_59272[group]". Other species, such as *Rhinolophus sinicus*, were manually included using custom queries. The sequences were aligned with MAFFT version 7 [25,26], using the alignment method FFT-NS-i (Standard). Some sequences had undefined amino acids ('X'), which we converted to glycine to allow modeling without any bias for amino acid identity. All species and the respective protein identifiers are listed in Table 1.

Definition of Sequence Similarity

All calculations were based on the alignments from MAFFT, restricted to the region used for modeling (residues 21-600). To calculate sequence similarity, we considered the following groups based on physico-chemical properties: charged-positive (Arg, Lys, His), charged-negative (Asp, Glu), aromatic (Phe, Tyr, Trp), polar (Ser, Thr, Asn, Gln), and apolar (Ala, Val, Ile, Met). Cys, Gly, and Pro residues were considered individual classes.

Modeling of ACE2 Orthologs

The modeling of ACE2 orthologs was carried out using MODELLER 9.24 [27] and custom Python scripts (available upon request). We used the cryo-EM structure of the SARS-CoV-2 RBD bound to human ACE2 (PDB ID: 6M17) [12] as a template for all our subsequent models, including all glycans and the coordinates of RBD. To save computational resources, we modeled only the extracellular domain of ACE2, specifically residues 21-600, which are known to be sufficient to bind to RBD. To avoid unwanted deviation from the initial cryo-EM structure, we restricted the optimization and refinement of the models to the coordinates of atoms of mutated or inserted residues. We used the *fastest* library schedule for model optimization and the *very_fast* schedule for model refinement. For each species, we generated 10 backbone or loop models and selected the one with the lowest normalized DOPE score as a representative. These final models were then processed to remove any sugar molecules in species where the respective asparagine residue had been mutated.

Refinement of ACE2:RBD complexes

The initial complex models were prepared for refinement using the pdb-tools suite [28]. Each chain was separated into a different PDB file (pdb_selchain) and standardized with TER and END statements (pdb_tidy). We used HADDOCK 2.4 [29] to carry out the refinement of the models. The protein molecules were parameterized using the standard force field in HADDOCK, while the sugars were parameterized using updated parameters for carbohydrates [30]. We used a modified version of the topology generation scripts to allow automatic detection of N-linked glycans and expand the range of the interface refinement (10 Å distance cutoff). Each initial homology model was refined through 50 independent short molecular dynamics simulations in explicit solvent (solvshell=True). These refined models were then clustered using the FCC algorithm [31] with default parameters and scored using the HADDOCK score, a linear combination of van der Waals, electrostatics, and desolvation. A lower HADDOCK score is better. The top 10 models of the top scoring cluster, ranked by its average HADDOCK score, were selected as representatives of the complex.

Analysis of interface contacts of refined ACE2:RBD complexes

We used the interfacea analysis library (version 0.1) (http://doi.org/10.5281/zenodo.3516439) to identify intermolecular contacts between hACE2 and RBD, specifically hydrogen bonds, salt bridges, and aromatic ring stacking. Hydrogen bonds were defined between any donor atom (nitrogen, oxygen, or sulfur bound to a hydrogen atom) within 2.5 Å of an acceptor atom (nitrogen, oxygen, or sulfur), if the donor-hydrogen-acceptor angle was between 120 and 180 degrees. Salt bridges were defined between two residues with a pair of cationic/anionic groups within 4 Å of each other. Finally, two aromatic residues were defined as stacking if the centers of mass of the aromatic groups were within 7.5 Å (pi-stacking) or 5 Å (t-stacking) and the angle between the planes of the rings was between 0 and 30 degrees (pi-stacking) or between 60 and 90 degrees (t-stacking). Additionally, for pi-stacking interactions, the projected centers of both rings must fall inside the other ring. For each modelled species, we took the 10 best models of the best cluster, judged by their HADDOCK score, and aggregated all their contacts together. Contacts present in at least 5 models were considered representative.

Per-residue decomposition of HADDOCK scores

We used a custom CNS [32] script to calculate the HADDOCK score of each residue at the interface between ACE2 and RBD. Briefly, the protocol was the following. For each model, since HADDOCK uses a united-atom force field, we first added missing hydrogen atoms and minimized their coordinates, keeping all other atoms fixed. We marked a residue of ACE2 as part of the interface if any of its atoms were within 5 Å of any atom of RBD, and *vice-versa*. We then calculated the electrostatics, van der Waals, and desolvation energies for each of these residues, considering only atoms belonging to the other protein chain. Note that this protocol does not account for intramolecular effects of mutations. Finally, we calculated the HADDOCK score per residue, using the default scoring function weights, and averaged per-residue values for the best 10 models of the best cluster of each species.

Author Contributions

Conceptualization: João P. G. L. M. Rodrigues
Funding Acquisition: João P. G. L. M. Rodrigues, Michael Levitt
Investigation: João P. G. L. M. Rodrigues, Susana Barrera-Vilarmau, João M. C. Teixeira, Panagiotis Kastritis
Writing – Original Draft: João P. G. L. M. Rodrigues, João M. C. Teixeira, Susana Barrera-Vilarmau, Elizabeth Seckel
Writing – Review and Editing: João P. G. L. M. Rodrigues, Susana Barrera-Vilarmau, João M. C. Teixeira, Elizabeth Seckel, Panagiotis Kastritis

Acknowledgements

JPGLMR acknowledges support from the Molecular Sciences Software Institute (ACI-1547580). JPGLMR and ML acknowledge funding from the National Institutes of Health USA (R35GM122543). PLK acknowledges funding from the Federal Ministry for Education and Research (BMBF, ZIK program) (02Z22HN23) and the European Regional Development Funds for Saxony-Anhalt (EFRE: ZS/2016/04/78115). The authors thank T. Dots, K. Lindorff-Larsen, J. Puglisi, and R. Fernandes for feedback and encouragement during the course of the project.

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;20: 533–534. doi:10.1016/S1473-3099(20)30120-1

2. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020;26: 450–452. doi:10.1038/s41591-020-0820-9

3. A cat appears to have caught the coronavirus, but it's complicated. In: Science News [Internet]. 31 Mar 2020 [cited 2 Jun 2020]. Available: https://www.sciencenews.org/article/cats-animals-pets-coronavirus-covid19

4. Confirmation of COVID-19 in Two Pet Cats in New York | CDC Online Newsroom | CDC. 22 Apr 2020 [cited 2 Jun 2020]. Available: https://www.cdc.gov/media/releases/2020/s0422-covid-19-cats-NYC.html

5. Mink infected two humans with coronavirus: Dutch government. Reuters. 25 May 2020. Available: https://www.reuters.com/article/us-health-coronavirus-netherlands-mink-idUSKBN23112K. Accessed 2 Jun 2020.

6. Seven more big cats test positive for coronavirus at Bronx Zoo. In: Animals [Internet]. 22 Apr 2020 [cited 2 Jun 2020]. Available: https://www.nationalgeographic.com/animals/2020/04/tiger-coronavirus-covid19-positive-test-bronx-zoo/

7. Shi J, Wen Z, Zhong G, Yang H, Wang C, Huang B, et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2. Science. 2020;368: 1016–1020. doi:10.1126/science.abb7015

8. Li Y, Wang H, Tang X, Ma D, Du C, Wang Y, et al. Potential host range of multiple SARS-like coronaviruses and an improved ACE2-Fc variant that is potent against both SARS-CoV-2 and SARS-CoV-1. Microbiology; 2020 Apr. doi:10.1101/2020.04.10.032342

9. Chan JF-W, Zhang AJ, Yuan S, Poon VK-M, Chan CC-S, Lee AC-Y, 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. [cited 2 Jun 2020]. doi:10.1093/cid/ciaa325

10. Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell. 2020;181: 894-904.e9. doi:10.1016/j.cell.2020.03.045

11. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;581: 215–220. doi:10.1038/s41586-020-2180-5

12. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. 2020;367: 1444–1448. doi:10.1126/science.abb2762

13. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020;581: 221–224. doi:10.1038/s41586-020-2179-y

14. Procko E. The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2. Biochemistry; 2020 Mar. doi:10.1101/2020.03.16.994236

15. Lam S, Bordin N, Waman V, Scholes H, Ashford P, Sen N, et al. SARS-CoV-2 spike protein predicted to form stable complexes with host receptor protein orthologues from mammals. Bioinformatics; 2020 May. doi:10.1101/2020.05.01.072371

16. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J Virol. 2020;94. doi:10.1128/JVI.00127-20

17. Armijos-Jaramillo V, Yeager J, Muslin C, Perez-Castillo Y. SARS-CoV-2, an evolutionary perspective of interaction with human ACE2 reveals undiscovered amino acids necessary for complex stability. Evol Appl. n/a. doi:10.1111/eva.12980

18. Kastritis PL, Rodrigues JPGLM, Folkers GE, Boelens R, Bonvin AMJJ. Proteins Feel More Than They See: Fine-Tuning of Binding Affinity by Properties of the Non-Interacting Surface. J Mol Biol. 2014;426: 2632–2652. doi:10.1016/j.jmb.2014.04.017

19. Kastritis PL, Bonvin AMJJ. Are Scoring Functions in Protein−Protein Docking Ready To Predict Interactomes? Clues from a Novel Binding Affinity Benchmark. J Proteome Res. 2010;9: 2216–2225. doi:10.1021/pr9009854

20. Bondad-Reantaso MG, Fisheries and Aquaculture Department, Food and Agriculture Organization of the United Nations, Rome, Italy, Mackinnon B, Bin H, Jie H, Tang-Nelson K, et al. Viewpoint: SARS-CoV-2 (The Cause of COVID-19 in Humans) is Not Known to Infect Aquatic Food Animals nor Contaminate Their Products. Asian Fish Sci. 2020 [cited 3 Jun 2020]. doi:10.33997/j.afs.2020.33.1.009

21. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. Nat Rev Immunol. 2020;20: 363–374. doi:10.1038/s41577-020-0311-8

22. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020;181: 281-292.e6. doi:10.1016/j.cell.2020.02.058

23. De Sancho D, Gavira JA, Pérez-Jiménez R. Coarse-grained molecular simulations of the binding of the SARS-CoV 2 spike protein RBD to the ACE2 cell receptor. Biophysics; 2020 May. doi:10.1101/2020.05.07.083212

24. van Wijk SJL, Melquiond ASJ, de Vries SJ, Timmers HThM, Bonvin AMJJ. Dynamic Control of Selectivity in the Ubiquitination Pathway Revealed by an ASP to GLU Substitution in an Intra-Molecular Salt-Bridge Network. PLoS Comput Biol. 2012;8. doi:10.1371/journal.pcbi.1002754

25. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30: 3059–3066. doi:10.1093/nar/gkf436

26. Kuraku S, Zmasek CM, Nishimura O, Katoh K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 2013;41: W22-28. doi:10.1093/nar/gkt389

27. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993;234: 779–815. doi:10.1006/jmbi.1993.1626

28. Rodrigues JPGLM, Teixeira JMC, Trellet M, Bonvin AMJJ. pdb-tools: a swiss army knife for molecular structures. F1000Research. 2018;7. doi:10.12688/f1000research.17456.1

29. van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016;428: 720–725. doi:10.1016/j.jmb.2015.09.014

30. Feng Y. Compatible topologies and parameters for NMR structure determination of carbohydrates by simulated annealing. PLoS ONE. 2017;12. doi:10.1371/journal.pone.0189700

31. Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis P, Karaca E, Melquiond ASJ, et al. Clustering biomolecular complexes by residue contacts similarity. Proteins Struct Funct Bioinforma. 2012;80: 1810–1817. doi:10.1002/prot.24078

32. Brunger AT. Version 1.2 of the Crystallography and NMR system. Nat Protoc. 2007;2: 2728–2733. doi:10.1038/nprot.2007.406

Figures and Tables

Table 1. List of Species

Scientific Name	Common Name	NCBI Protein ID
Homo sapiens	Human	NP_001358344.1
Anas platyrhynchos	Duck	XP_012949915.2
Bos taurus	Cow	XP_005228485.1
Camelus dromedarius	Dromedary	XP_010991717.1
Canis lupus familiaris	Dog	NP_001158732.1
Capra hircus	Goat	NP_001277036.1
Carassius auratus	Goldfish	XP_026131313.1
Cavia porcellus	Guinea pig	XP_023417808.1
Columba livia	Pidgeon	XP_021154486.1
Crocodylus porosus	Crocodile	XP_019384826.1
Equus asinus	Donkey	XP_014713133.1
Equus caballus	Horse	XP_001490241.1
Erinaceus europaeus	Hedgehog	XP_007538670.1
Felis catus	Cat	XP_023104564.1
Gallus gallus	Chicken	XP_416822.2
Macaca mulatta	Macaque	NP_001129168.1
Manis javanica	Pangolin	XP_017505746.1
Mesocricetus auratus	Hamster	XP_005074266.1
Mus musculus	Mouse	NP_081562.2
Mustela putorius furo	Ferret	NP_001297119.1
Oryctolagus cuniculus	Rabbit	XP_002719891.1
Ovis aries	Sheep	XP_011961657.1
Paguma larvata	Civet	AAX63775.1
Pan troglodytes	Chimpanzee	XP_016798468.1
Panthera tigris altaica	Siberian Tiger	XP_007090142.1
Pongo abelii	Orangutan	NP_001124604.1
Rattus norvegicus	Rat	NP_001012006.1
Rhinolophus sinicus	Horseshoe bat	AGZ48803.1
Serinus canaria	Canary	XP_009087922.1
Sus scrofa	Pig	NP_001116542.1

Table 2. Statistics of top 10 refined models for each species

Species	HADDOCK Score (a.u.)	van der Waals (kcal/mol)	Electrostatics (kcal/mol)	Desolvation (a.u.)	Buried Surface Area (Å^2)
Human	-116,2 ± 3,2	-54,4 ± 2,8	-221,8 ± 15,1	-17,5 ± 3,2	1781 ± 37
Duck	-98,8 ± 2,3	-63,3 ± 1,9	-99,5 ± 6,0	-15,5 ± 3,1	1782 ± 55

Cow	-115,3 ± 4,1	-54,9 ± 1,5	-212,1 ± 19,1	-18,0 ± 4,0	1766 ± 41
Dromedary	-113,1 ± 2,2	-56,0 ± 2,0	-181,3 ± 16,3	-20,8 ± 2,7	1733 ± 65
Dog	-137,5 ± 3,7	-64,5 ± 4,0	-230,3 ± 18,4	-26,9 ± 2,5	1903 ± 44
Goat	-108,5 ± 4,0	-53,5 ± 3,6	-189,0 ± 10,5	-17,2 ± 3,1	1745 ± 64
Goldfish	-131,2 ± 5,9	-68,2 ± 3,8	-189,3 ± 19,2	-25,1 ± 3,2	1925 ± 67
Guinea pig	-118,0 ± 3,3	-62,3 ± 2,6	-163,3 ± 12,2	-23,0 ± 2,5	1868 ± 41
Pidgeon	-106,0 ± 2,0	-57,3 ± 2,7	-189,7 ± 11,9	-10,7 ± 3,0	1730 ± 28
Crocodile	-96,0 ± 2,5	-64,9 ± 2,7	-90,4 ± 18,1	-13,1 ± 2,6	1690 ± 49
Donkey	-119,9 ± 1,9	-58,3 ± 1,7	-220,5 ± 13,5	-17,6 ± 3,3	1767 ± 44
Horse	-116,8 ± 4,2	-55,4 ± 3,7	-220,0 ± 14,0	-17,4 ± 2,4	1762 ± 34
Hedgehog	-103,5 ± 3,2	-57,3 ± 5,3	-160,9 ± 38,9	-14,0 ± 3,8	1726 ± 48
Cat	-117,2 ± 1,7	-55,2 ± 2,0	-202,4 ± 9,8	-21,5 ± 3,0	1752 ± 42
Chicken	-111,8 ± 1,4	-64,8 ± 1,6	-132,3 ± 15,0	-20,5 ± 3,7	1802 ± 44
Macaque	-120,8 ± 3,2	-56,5 ± 2,1	-217,7 ± 18,6	-20,7 ± 2,3	1841 ± 37
Pangolin	-127,7 ± 3,3	-59,9 ± 3,6	-233,8 ± 16,0	-21,0 ± 2,0	1854 ± 32
Hamster	-119,1 ± 3,1	-57,6 ± 2,8	-242,8 ± 23,8	-13,0 ± 2,5	1821 ± 51
Mouse	-93,2 ± 2,6	-53,8 ± 2,6	-93,1 ± 14,4	-20,8 ± 2,5	1598 ± 65
Ferret	-127,2 ± 1,8	-60,6 ± 2,6	-195,8 ± 16,8	-27,5 ± 3,4	1824 ± 70
Rabbit	-116,4 ± 5,6	-55,6 ± 2,7	-222,6 ± 18,2	-16,2 ± 2,7	1827 ± 55
Sheep	-126,7 ± 2,5	-61,4 ± 2,2	-231,2 ± 26,8	-19,0 ± 3,6	1876 ± 25
Civet	-123,6 ± 4,2	-56,4 ± 4,5	-193,6 ± 27,9	-28,6 ± 3,4	1786 ± 56
Chimpanzee	-121,2 ± 3,4	-57,5 ± 3,7	-213,5 ± 17,6	-21,0 ± 3,1	1816 ± 33
Siberian Tiger	-126,0 ± 4,8	-60,7 ± 3,9	-196,3 ± 27,8	-26,0 ± 3,3	1804 ± 37
Orangutan	-118,5 ± 4,5	-62,5 ± 2,7	-182,3 ± 21,1	-19,5 ± 3,8	1837 ± 49
Rat	-108,9 ± 3,5	-55,3 ± 4,5	-162,4 ± 16,0	-21,2 ± 2,7	1777 ± 58
Horseshoe bat	-125,8 ± 2,7	-60,7 ± 3,2	-259,4 ± 17,5	-13,2 ± 3,2	1819 ± 52
Canary	-112,5 ± 2,8	-61,5 ± 5,0	-170,7 ± 33,7	-16,8 ± 3,8	1822 ± 73
Pig	-113,3 ± 1,1	-53,8 ± 2,1	-202,7 ± 12,6	-18,9 ± 2,5	1715 ± 32

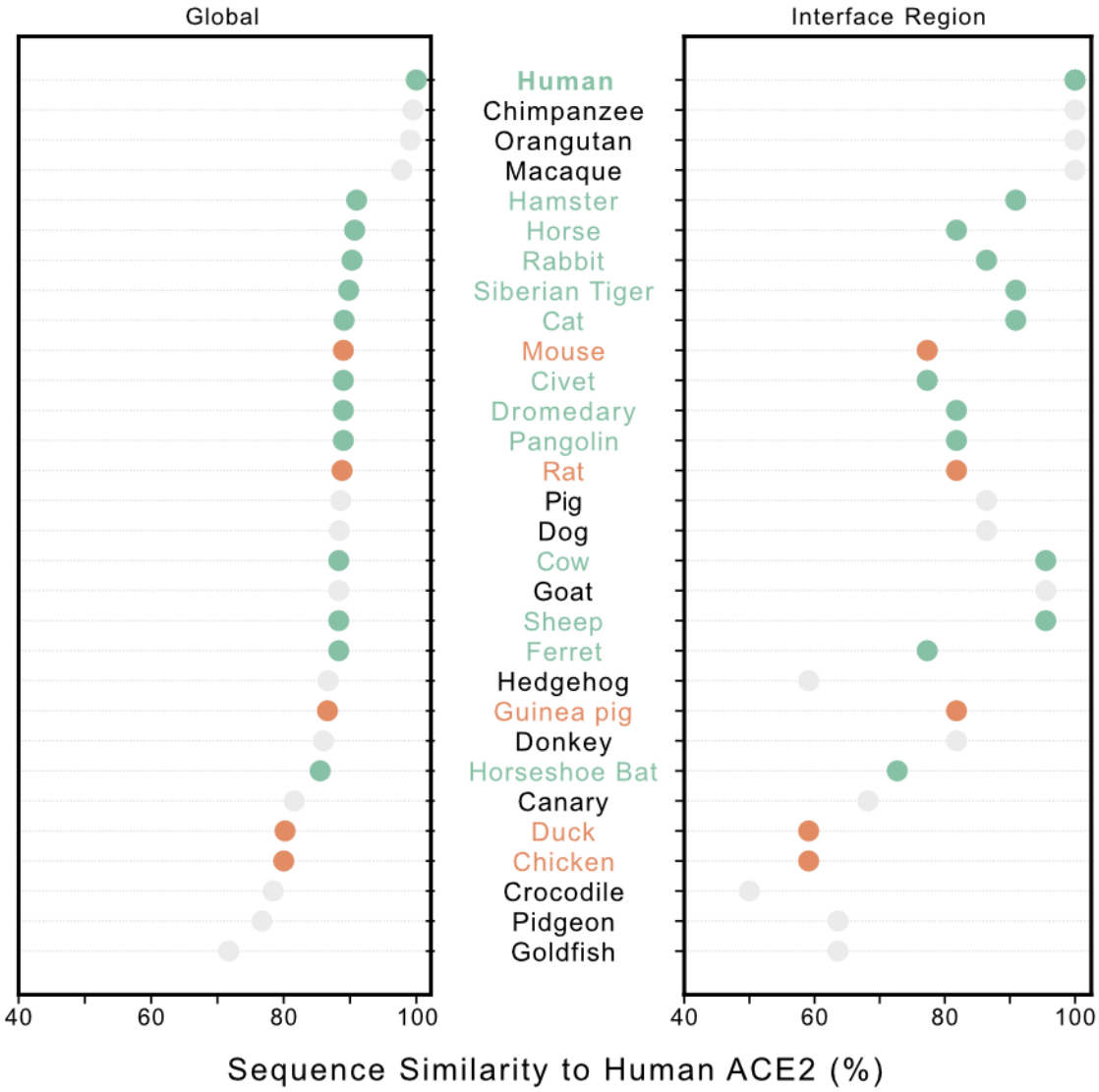

Figure 1. Sequence similarity of ACE2 orthologs to human ACE2. Global sequence similarity values range from 70-95%, while similarity values for residues interacting with SARS-CoV-2 RBD (derived from 6m17) range from 50-100%. Species are ordered by decreasing global sequence similarity to human ACE2. Colors indicate known susceptibility to infection: SARS-CoV-2^{pos} species in green, SARS-CoV-2^{neg} species in red, others in gray.

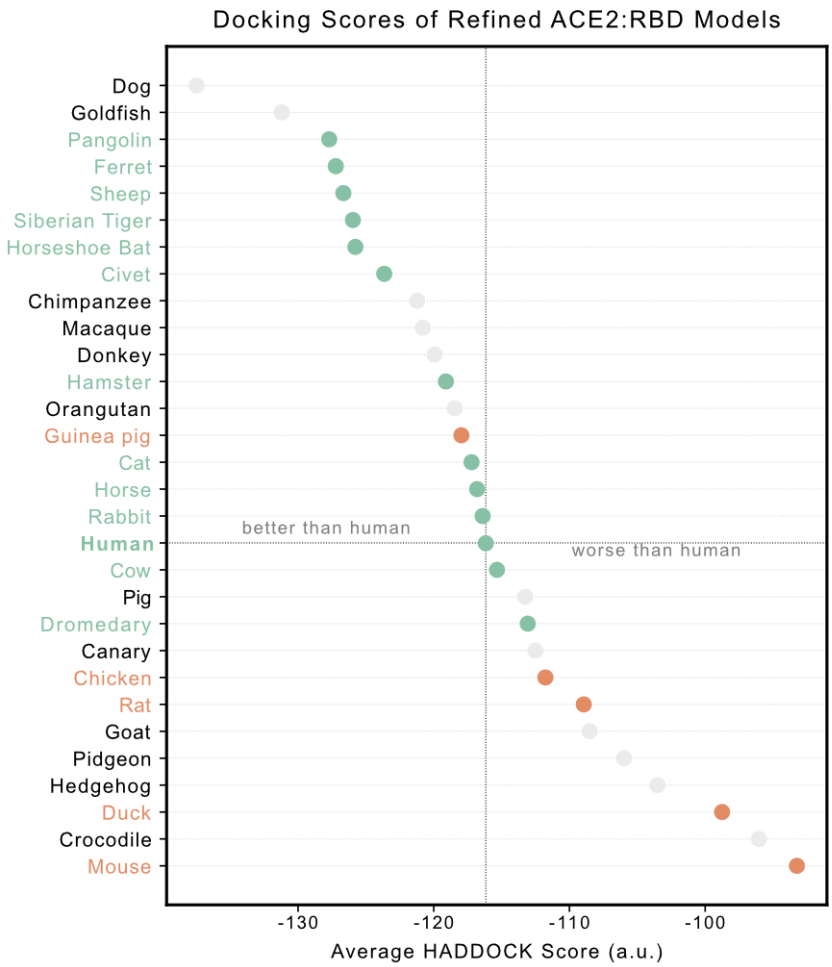

Figure 2. HADDOCK scores of modeled ACE2 orthologs bound to SARS-CoV-2 RBD. The HADDOCK score predicts the strength of the interaction between proteins. Models of SARS-CoV-2^{pos} species (green) generally have better (more negative) scores than SARS-CoV-2^{neg} species (red), suggesting that impaired binding between the two proteins might explain differences in viral susceptibility. The scores shown here are the average of the 10 best models for each ACE2 ortholog.

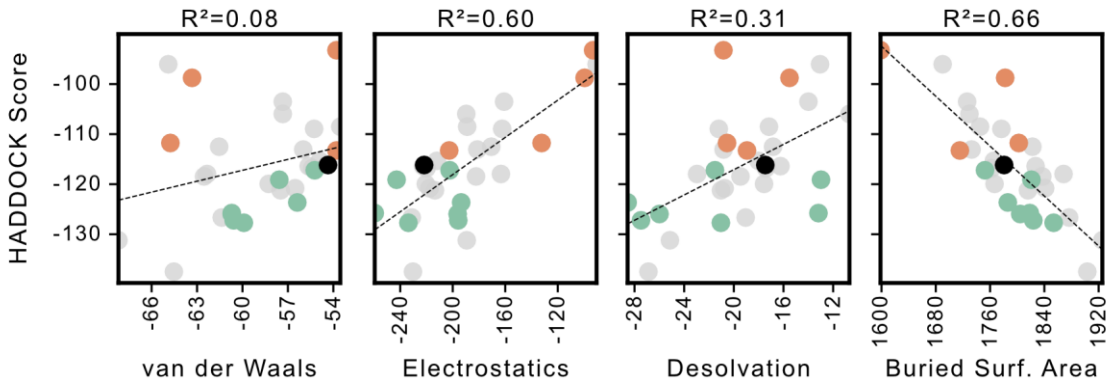

Figure 3. Correlation of HADDOCK score with individual energy terms and structural features. Differences in electrostatics energy contribute the most towards discriminating SARS-CoV-2^{pos} species (green) from SARS-CoV-2^{neg} species (red), supporting observations of hydrogen bonding networks and charged interactions in experimental structures. The buried surface area of the models is also correlated with their HADDOCK score. The units for van der Waals and electrostatics energies, desolvation, and buried surface area are kcal.mol^{-1}, arbitrary units, and Å^2, respectively. The human complex is shown in black for reference.

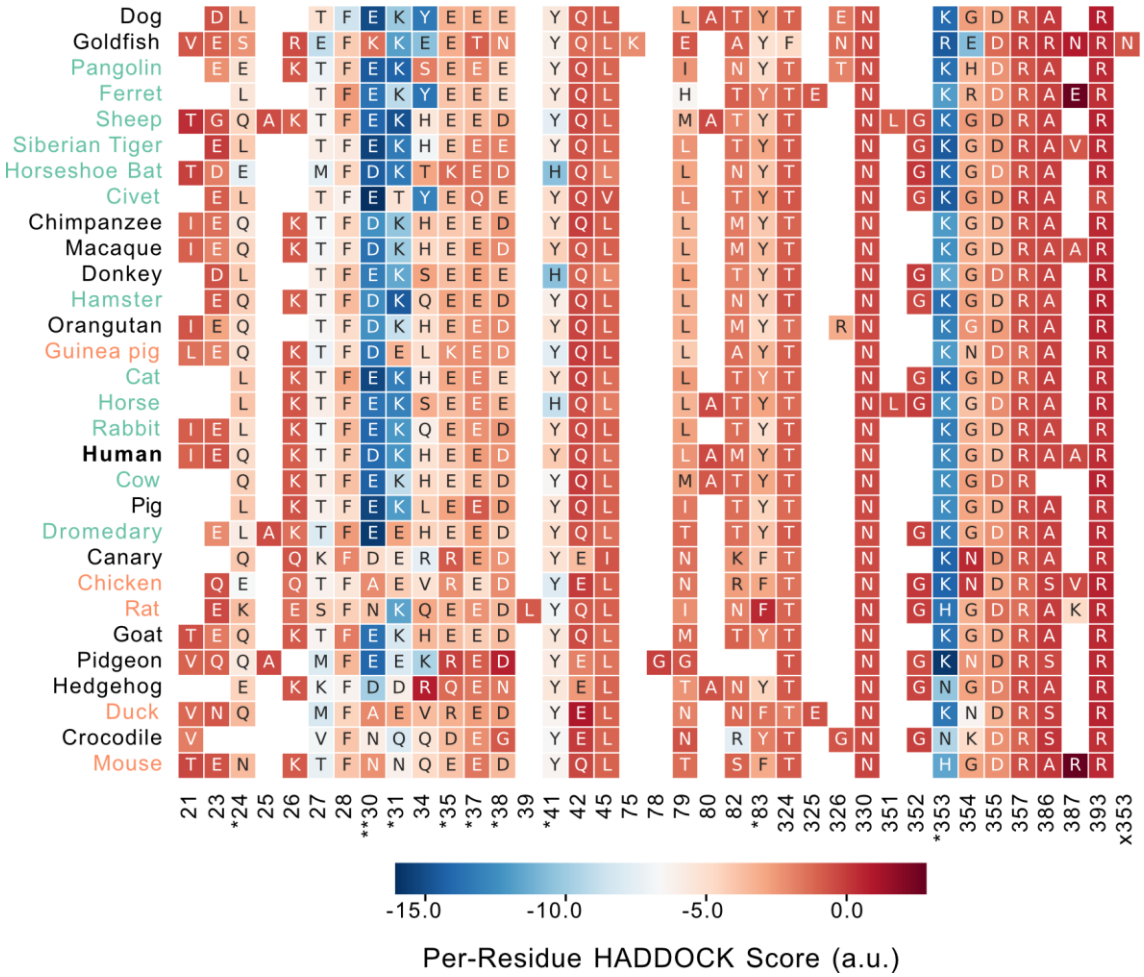

Figure 4. HADDOCK score of individual ACE2 interface residues. Amino acid residues at positions 30, 31, 34, and 353 are predicted to be the largest contributors to the stability of the interface. SARS-CoV-2^{neg} (red labels) species consistently show changes in these positions which could explain their non-susceptibility to the virus. The top-scoring SARS-CoV-2^{pos} (green labels) also suggest that hACE2 D30E and H34Y could potentially act as affinity enhancers. For each species, each block represents an interface residue of ACE2. The identity of the amino acid is shown in one-letter codes. The colors represent the average HADDOCK score of each particular residue over the best 10 models: lower scores (blue) indicate more favorable interactions. Positive scores (dark red) indicate steric clashes or electrostatic repulsion. Blank squares indicate that in that ortholog, that position is not part of the interface of the complex. Residues marked with *, and ** are observed to form hydrogen bonds or salt-bridges in the hACE2:RBD crystal structure, respectively. See Materials and Methods for additional details on definitions.

Table S1. Interface contacts of refined human ACE2:RBD

Type	ACE2 Residue	RBD Residue	Frequency in Top 10 hACE2 Models	In crystal (6m0j)
hbond	Q24	A475	7	
hbond	Q24	N487	6	x
hbond	T27	Y489	9	
ionic	D30	K417	10	x
hbond	K31	F490	4	
hbond	K31	Q493	10	x
hbond	H34	Y453	2	
hbond	E35	Q493	6	x
hbond	E37	Y505	4	x
hbond	D38	Y449	10	x
hbond	Y41	T500	10	x
hbond	Y83	N487	2	x
hbond	K353	Y495	1	
hbond	K353	G496	9	
hbond	K353	G502	10	x

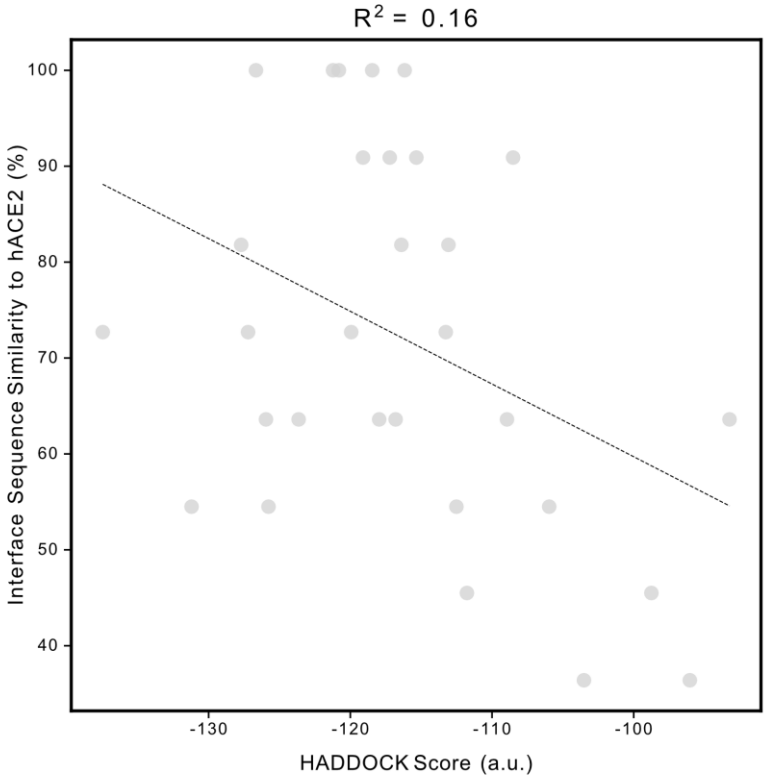

Figure S1. Correlation between interface sequence similarity to hACE2 and HADDOCK score.

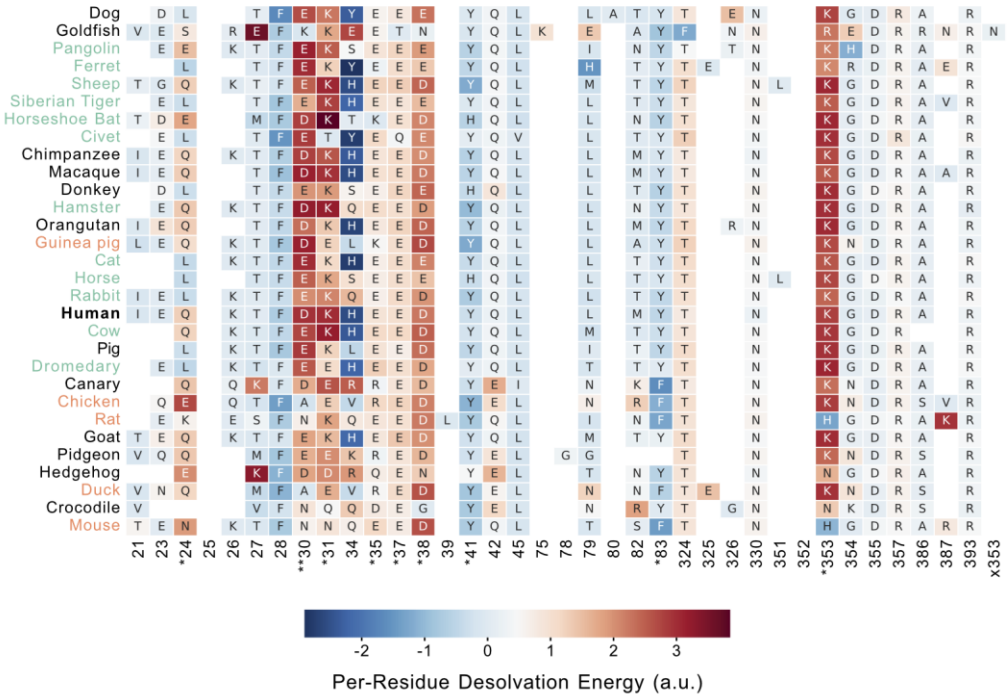

Figure S2. Desolvation energy of individual ACE2 interface residues. Aromatic residues at position 34 contribute the most to the gain in desolvation energy across all species of the complex, indicating that H34Y could be a potential affinity enhancing mutation in hACE2. For each species, each block represents an interface residue of ACE2. The identity of the amino acid is shown in one-letter codes. The colors represent the average desolvation energy of each particular residue over the best 10 models: lower scores (blue) indicate more favorable interactions. Blank squares indicate that in that ortholog, that position is not part of the interface of the complex. Residues marked with *, and ** are observed to form hydrogen bonds or salt-bridges in the hACE2:RBD crystal structure, respectively. See Materials and Methods for additional details on definitions.