Which animals are at risk? Predicting species susceptibility to Covid-19

Alexander MR*[1,2], Schoeder CT*[3,4], Brown JA[5,6], Smart CD[7], Moth C[3,8], Wikswo JP[5,6,7,9], Capra JA[3,8], Meiler J[#3,4,9,10], Chen W[#7], Madhur MS[#1,2,7,11]

* authors contributed equally

[#] corresponding authors, contributing equally

[1]Department of Medicine, Division of Cardiovascular Medicine, Vanderbilt University Medical Center (VUMC), Nashville, TN, USA; [2]Department of Medicine, Division of Clinical Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA; [3]Center for Structural Biology, Vanderbilt University, Nashville, TN, USA; [4]Department of Chemistry, Vanderbilt University, Nashville, TN, USA; [5]Department of Physics and Astronomy, Vanderbilt University, Nashville, TN, USA; [6]Vanderbilt Institute for Integrative Biosystems Research and Education, Vanderbilt University, Nashville, TN, USA; [7]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, USA; [8]Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA; [9]Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA; [10]Institute for Drug Discovery, Leipzig University Medical School, Leipzig, Germany; [11]Vanderbilt Institute for Infection, Immunology, and Inflammation

Word Count: 4,515

Address for Correspondence:

Jens Meiler
Department of Chemistry
Vanderbilt University
7330 Stevenson Center
Station B 351822
Nashville, TN 37235
615-936-6594
jens@meilerlab.org

Wenbiao Chen
Department of Molecular Physiology and Biophysics
Vanderbilt University
735B Light Hall
2215 Garland Ave
Nashville, TN 37232
615-936-7390
wenbiao.chen@vanderbilt.edu

Meena S. Madhur
Department of Medicine
Vanderbilt University
2215 Garland Avenue
P415D MRB IV
Nashville, TN 37232
615-875-3273
meena.madhur@vanderbilt.edu

Abstract

In only a few months, the novel coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a global pandemic, leaving physicians, scientists, and public health officials racing to understand, treat, and contain this zoonotic disease. SARS-CoV-2 has made the leap from animals to humans, but little is known about variations in species susceptibility that could identify potential reservoir species, animal models, and the risk to pets, wildlife, and livestock. While there is evidence that certain species, such as cats, are susceptible, the vast majority of animal species, including those in close contact with humans, have unknown susceptibility. Hence, methods to predict their infection risk are urgently needed. SARS-CoV-2 spike protein binding to angiotensin converting enzyme 2 (ACE2) is critical for viral cell entry and infection. Here we identified key ACE2 residues that distinguish susceptible from resistant species using in-depth sequence and structural analyses of ACE2 and its binding to SARS-CoV-2. Our findings have important implications for identification of ACE2 and SARS-CoV-2 residues for therapeutic targeting and identification of animal species with increased susceptibility for infection on which to focus research and protection measures for environmental and public health.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the virus responsible for the global pandemic of coronavirus disease-2019 (Covid-19) that is impacting millions of lives and the global economy. Covid-19 is a zoonotic infection capable of crossing the species barrier. SARS-CoV-2 is thought to have originated in bats and subsequently transmitted to humans, perhaps through a secondary host.[1,2] Emerging experimental and observational evidence demonstrates differences in species susceptibility to infection. For example, humans, house cats, tigers, and lions are all susceptible to infection by SARS-CoV-2.[3-6] Golden Syrian hamsters and rhesus monkeys are also capable of being experimentally infected by SARS-CoV-2 and developing Covid-19 pathologies.[7,8] In contrast, observational and experimental studies with direct intranasal inoculation have demonstrated that chickens, ducks, and mice are not susceptible to SARS-CoV-2 infection.[5,9-11] Interestingly however, susceptibility is not dichotomous. Although ferrets are also susceptible to infection, intranasal inoculation failed to result in spread of infection to the lower respiratory tract, significantly limiting symptom development.[5] In addition, although dogs failed to exhibit infection of the respiratory tract and appear asymptomatic, a minority of experimentally or environmentally exposed dogs exhibited evidence of infection by SARS-CoV-2 PCR or SARS-CoV-2 seroconversion with production of SARS-CoV-2-specific antibodies.[5,12] While pigs have not demonstrated evidence of infection after intranasal inoculation, overexpression of swine ACE2 in cultured cells supports some degree of viral entry.[5,9,13] Hence, ferrets, dogs, and pigs are classified as having intermediate susceptibility to infection. Despite these findings, the number of animal species tested for susceptibility to infection in experimental or observational studies is very limited. Thus, methods of determining risk of species with unknown susceptibility are urgently needed to reduce risk of propagating transmission, protect food supplies, identify potential intermediate hosts, and discover animal models for research. Identifying the key residues mediating susceptibility to infection can also guide rational drug design.

SARS-CoV-2 is a member of the coronavirus family of single-stranded RNA viruses.[9] The spike protein on the surface of the SARS-CoV-2 virus mediates interaction with its receptor, angiotensin converting enzyme 2 (ACE2), to promote membrane fusion and virus entry into the cell. The receptor binding domain (RBD) of the spike protein contains a receptor binding motif (RBM) that binds to the peptidase domain of ACE2.[14] Following spike protein cleavage, fusion of the viral and host cell membranes occurs to enable viral entry into the cell.[15] Interaction of the SARS-CoV-2 spike protein RBD and ACE2 is thus critical for viral cell entry and infection.[9] The importance of this interaction in infection is further supported by evidence that exogenous soluble ACE2 limits infection in human organoids,[10] and that overexpression of human ACE2 is necessary to enable viral cell entry in HeLa cells *in vitro* and SARS-CoV-2 infection in mouse models *in vivo*.[9,16]

ACE2 is present in almost all vertebrates, however sequence differences exist that may hold clues to differences in SARS-CoV-2 susceptibility, as has been observed for SARS-CoV.[17,18] Understanding such differences could provide insight into key structural interactions between ACE2 and SARS-CoV-2 RBD important for infection, and permit development of a susceptibility score for estimating the infection risk of various species. In this manuscript we integrate

experimentally validated differences in susceptibility to SARS-CoV-2 infection with ACE2 sequence comparisons and in-depth structural analyses to determine how differences in ACE2 across species influence interaction with SARS-CoV-2 RBD. We identified multiple key residues mediating structural interactions between ACE2 and SARS-CoV-2 RBD and use these residues to generate a susceptibility score to predict animals with elevated risk of infection. We also demonstrate that SARS-CoV-2 is nearly optimal for binding ACE2 of humans compared to other animals, which may underlie the highly contagious nature of this virus amongst humans. Our findings have important implications for identification of ACE2 and SARS-CoV-2 residues for therapeutic targeting and identification of animal species with increased susceptibility for infection on which to focus research and protection efforts.

RESULTS

Susceptibility does not segregate according to phylogeny and ACE2 sequence similarity

Given experimental evidence for susceptibility of humans, house cats, tigers, lions, rhesus macaques, and Golden Syrian hamsters to SARS-CoV-2 infection, and experimental evidence for non-susceptibility of mice, ducks, and chickens,[3-5,7,9-11,19] we performed protein sequence alignment of ACE2 from these organisms using MAFFT (**Extended Data Figure 1**).[20] We also included species with intermediate susceptibility, including dogs, pigs, and ferrets,[5,9,12] as well as species with unknown susceptibility, including camels, horses, Malayan pangolin, and sheep. The degree of similarity of ACE2 protein sequences largely fell along expected phylogenetic relationships among species (**Extended Data Figure 2**). Susceptibility to SARS-CoV-2 infection, however, did not match either phylogenetic relationships or ACE2 sequence similarities across species. For example, mouse (*Mus musculus*) is not susceptible to infection. However, mouse ACE2 sequence is more similar to a susceptible species, Golden Syrian hamster (*Mesocricetus auratus*), than non-susceptible species such as duck (*Aythya fuligula*) or chicken (*Gallus gallus*).[9,16] In addition, mice are phylogenetically more similar to susceptible species such as humans (*Homo sapiens*) and rhesus macaques (*Macaca mulatta*) than non-susceptible species such as ducks and chicken.[9,16] These findings suggest that neither phylogenetic relationships nor overall ACE2 protein sequence similarity across species is able to predict susceptibility to SARS-CoV-2 infection.

Sequence alignment identifies ACE2 residues distinguishing susceptible from non-susceptible species

An alternative approach is to use the experimentally validated differences in infection susceptibility across species to focus on ACE2 amino acids that most differ between susceptible and non-susceptible species. We thus calculated a weighted score of how well the aligned amino acids stratify susceptible versus non-susceptible species, incorporating amino acid similarity. This score, termed GroupSim, permits quantitative determination of which amino acids in the alignment best stratify susceptible from non-susceptible species.[21] This analysis demonstrated that multiple amino acid positions in the ACE2 alignment, including Leu79, His34, Tyr83, and Gln24, are highly similar in susceptible species and quite different in non-susceptible

species (**Extended Data Table 1 and Supplemental Table 1**). When mapping these scores onto the structure of the SARS-CoV-2 RBD and ACE2 complex, multiple residues with high GroupSim scores were present at or near the binding interface including His34, Asp30, Thr92, Gln24, Lys31, and Leu79 (**Figure 1**). We then extended this analysis by focusing on key residues previously demonstrated from prior structural analysis to be important for ACE2 and SARS-CoV-2 RBD interactions (**Table 1**).[7,22-24] Interestingly, this revealed that key amino acids for the ACE2 and SARS-CoV-2 spike protein interaction were enriched among the top scoring GroupSim positions (7 of 35; $p<0.0001$; Fisher's exact test). Such key residues based on structural analysis being over-represented in amino acid positions that best discriminated susceptible from non-susceptible species suggests that structural interactions between ACE2 and SARS-CoV-2 spike protein importantly determine differences in species susceptibility to infection. In addition, these data suggest that certain ACE2 amino acid residues may be particularly important for determining susceptibility, including Leu79, His34, Tyr83, Gln24, Lys31, Asp30, and Glu329.

SARS-CoV-2 has lower predicted binding affinity for ACE2 from non-susceptible avian species

We used homology modeling to identify structural determinants of binding the ACE2 protein from species with known differences in susceptibility to SARS-CoV-2 infection. The models were based on previously reported crystal structures of the human ACE2 in complex with SARS-CoV-2 (PDB: 6LZG and 6M0J).[14] We modeled ACE2 in the presence of the SARS-CoV-2 RBD to allow backbone adjustment to the binder and refined by redocking of the RBD domain to optimize sidechains. Models were selected by overall calculated protein stability of the SARS-CoV-2 RBD complex, predicted binding energy between ACE2 and SARS-CoV-2 RBD, and similarity (as Cα-root mean square deviation [Cα-RMSD], **Extended Data Figure 3 and Extended Data Figure 4**). Based on these models, multiple approaches where undertaken to investigate the structural interactions between SARS-CoV-2-RBD and ACE2.

We evaluated the overall calculated protein stability and predicted binding energy for SARS-CoV-2-RBD and ACE2 complexes for each species. We considered the 100 best models for each species and evaluated evidence for difference in binding energy or stability between susceptible and non-susceptible species. The average mean predicted binding energy and calculated protein stability differs across species (**Figure 2**). Consistent with the lack of susceptibility of chickens (*Gallus gallus)*, chicken ACE2 in complex with SARS-CoV-2-RBD was the lowest scoring, or most energetically unfavorable model. The complex with duck ACE2 (*Aythya fuligula*) shows similarly unfavorable scores, indicating that ACE2 sequence differences leading to a lower structural binding ability in these two avian species may explain their lack of susceptibility to SARS-CoV-2 infection. However, the complex of SARS-CoV-2-RBD and ACE2 of the non-susceptible mouse (*Mus musculus*) exhibits lower binding energy and higher protein stability than several species that are susceptible, including the lion (*Panthera leo*), tiger (*Panthera tigris*), and cat (*Felis catus*). Thus, differences in SARS-CoV-2 and ACE2 complex stability have some discriminative power but are not the sole factor in differences in susceptibility across species.

Homology modeling identifies a link between ACE2 D30 and Y83 and SARS-CoV-2 susceptibility

As a complementary approach to determine whether particular residues may discriminate susceptible from non-susceptible species, we performed energetic modeling of residue-residue interactions in the interface of SARS-CoV-2 and ACE2 using Rosetta. Although the overall interaction pattern across residues is similar between susceptible, non-susceptible, and intermediate susceptibility species, there are significant differences in the magnitude of residue-residue interactions (**Figure 3**). For example, residue 30 (which is an aspartate in all susceptible species) forms a strong ionic interaction with lysine 417 of SARS-CoV-2 RBD and interacts modestly with other residues, including Phe456 and Tyr473. In contrast, in non-susceptible species such as chicken and duck where residue 30 contains an alanine this interaction is no longer present and is not substituted by any other structural rearrangements that might accommodate this change. Mouse (*Mus musculus)* ACE2 contains an asparagine in position 30 instead of an aspartate, which results in lower predicted binding energy due to the lack of an ionic interaction. A close-up view of residue 30 shows the different structural environment available in the non-susceptible species chicken, duck, and mouse as compared to susceptible species, including human (**Figure 4**). This analysis also identifies residue 83 of ACE2 as having differential energetic interactions across species. Residue 83 is a tyrosine in susceptible species and a phenylalanine in non-susceptible species (**Table 1**). Compared to susceptible species, this position exhibits significantly decreased binding energy with residues Asn487 and Tyr489 in SARS-CoV-2 RBD in non-susceptible species (**Figure 3**). Although ACE2 residue 83 also interacts with SARS-CoV-2 RBD phenylalanine 486, this interaction is unlikely to be significantly affected by differences between tyrosine and phenylalanine. However, the hydroxyl group of tyrosine at position 83 forms a hydrogen bond with the backbone oxygen of asparagine 487 that is negatively impacted by substitution to phenylalanine in non-susceptible species (**Figure 5A**). In addition to this residue-residue structural analysis, both ACE2 positions 30 and 83 were identified through the GroupSim analysis described above to be top residues discriminating susceptible from non-susceptible species based on sequence alignment (**Extended Data Table 1**). These results suggest that these amino acid positions of ACE2 may be important mediators of the structural interaction of ACE2 and SARS-CoV-2 RBD and determinants of differences to susceptibility to infection across species.

Multistate design reveals ACE2 G354 as determinant of susceptibility

It is an evolutionary advantage for SARS-CoV-2 to maintain its ability to infect multiple species. Thus, we hypothesized that the sequence of SARS-CoV-2 RBD is not optimized for a single species but is capable of binding ACE2 of multiple species. Multistate design is a computational approach to test this hypothesis. It allows us to determine the sequence of SARS-CoV-2 RBD that is optimal for binding ACE2 of multiple species. We used Restraint Convergence (RECON) multistate design to test this hypothesis. This method determines how many mutations one protein requires to acquire affinity for multiple targets at once.[25,26]

We adapted this strategy to evaluate the ability of the SARS-CoV-2-RBD to bind non-human ACE2 variants starting from the constraint of the known binding to human ACE2. We

hypothesized that engineering a SARS-CoV-2 RBD with binding affinity for ACE2 from non-susceptible species would require more changes to binding interface residues than for susceptible species. To test this hypothesis, we redesigned the SARS-CoV-2 RBD interface sequence using RECON in the presence of the known binder, human ACE2, and ACE2 from other species in turn (**Figure 6A**).

As an initial positive control, the SARS-CoV-2 RBD was redesigned against human ACE2 only. By mutating multiple SARS-CoV-2 RBD residues to improve binding affinity, we tested at each designable position the frequency of native sequence recovery, which measures the fraction of models in which the native SARS-CoV-2 RBD amino acid is retained. This resulted in very few proposed amino acid changes of SARS-CoV-2 RBD to optimally bind human ACE2, indicating that the SARS-CoV-2 RBD sequence overall represents a solution close to optimal (**Figure 6B**). The exception is valine 503, for which more polar amino acids were deemed optimal. This valine, however, is near a glycosylation site at asparagine 322 in ACE2 at the SARS-CoV-2 and ACE2 interface (**Extended Data Figure 5**). Since glycans are not incorporated into the RECON multistate design technique, this valine 503 may have a higher affinity binding partner when considering the presence of ACE2 glycosylation sites..

Designing SARS-CoV-2 RBD in the presence of ACE2 from additional species revealed that ACE2 from a number of species have lower sequence recovery (including non-susceptible species such as duck and chicken, but also hamster, macaque, cat, lion and dog). When evaluating residue-specific interactions based on the native sequence recovery from RECON multistate design, tyrosine 505 shows no sequence recovery in avian species as compared to the human ACE2 control. This tyrosine interacts very prominently with lysine 353 in ACE2, however this residue is highly conserved across all species examined (**Table 1**). Tyrosine 505 also interacts less strongly with glycine 354, which is occupied by an asparagine in the avian species (chicken and duck) (**Table 1 and Figure 5B**). This secondary interaction might explain the differences in native sequence recovery. However, another experimentally verified non-susceptible species, the mouse (*Mus musculus*), has a high degree of sequence recovery, similar to human ACE2. This suggests that other factors beyond residue-residue interactions of ACE2 and SARS-CoV-2 RBD at the interface may determine susceptibility to infection, at least in the mouse, and that differences in RECON multistate design explain only partially differences in species susceptibility to SARS-CoV-2 infection.

ACE2 glycosylation at N90 and N322 as determinants of susceptibility

As a final additional approach to structurally evaluate differences in species susceptibility, we investigated the predicted glycosylation profiles of various species in comparison to human ACE2. Protein glycosylation is increasingly recognized as a critical contributor to receptor-ligand interactions;[27] however, given the challenges in identifying glycans in protein crystal structures, glycosylation has received considerably less attention than SARS-CoV-2 RBD and ACE2 protein-protein interactions. Naturally occurring glycans as posttranslational modifications are not fully visible in crystal structures. Normally only the first *N*-actylglucosamine is visible or no sugar moiety can be observed, or glycosylation sites are mutated prior to crystallization. In the crystal structures of the human ACE2 used here, a sugar moiety bound to an asparagine at a

surface exposed NXT/S sequon was seen three times in proximity to the binding interface on the ACE2. To understand whether the ACE2 of other species have similar glycosylation patterns, glycosylation was predicted using NetNGlyc 1.0, a neural network for predicting N-glycosylation sites, and compared to the glycosylation patterns of human ACE2.[28] Residues 53, 90, 103, and 322 were identified as glycosylation sites in human ACE2, with 53, 90, and 322 demonstrating glycosylation in the crystal structure (PDB: 6M0J and 6LZG)[14] (**Table 2**). Other susceptible species were quite similar to this pattern, except for position 103, which is only predicted to be glycosylated in humans and rhesus macaques. Among known susceptible species, only Golden Syrian hamster ACE2 lacks predicted glycosylation in position 322. At position 90, all susceptible species were predicted to be glycosylated and all non-susceptible and intermediate susceptibility species were non-glycosylated. Interestingly, ACE2 from the non-susceptible mouse, despite not showing significant differences in predicted binding energy or RECON multistate analysis compared to susceptible species, is predicted to lack glycosylation at residues 90 and 322, distinguishing it from ACE2 of nearly all susceptible species. This suggests a potential mechanism by which mice may be non-susceptible despite having similar binding energy and SARS-CoV-2 native sequence recovery to susceptible species.

A SARS-CoV-2 susceptibility score predicts species at risk

Taken together, results of these studies reveal a set of key ACE2 residues important for interaction with SARS-CoV-2 RBD and for which differences help discriminate susceptible from non-susceptible species. These differences include ACE2 amino acid positions 30 and 83, which exhibit differential residue-residue binding energy, position 354, which exhibits low native sequence recovery in interaction with SARS-CoV-2, and positions 90 and 322, which exhibit differences in glycosylation. Using these key residues in aggregate, we developed a SARS-CoV-2 susceptibility score based on similarity to the human ACE2 sequence using the BLOSUM62 similarity matrix (**Table 3**).[29] This analysis revealed that experimentally validated non-susceptible species have in fact the lowest susceptibility scores, while species with previously demonstrated intermediate susceptibility have intermediate susceptibility scores. Using the lowest score of the susceptible species, 23, as the lower cutoff for susceptibility and the highest score of non-susceptible species, 11, as the upper cutoff for non-susceptibility, we extended these results to species with unknown susceptibility. This revealed high scores in the susceptible range for the Chinese horseshoe bat (*Rhinolophus sinicus*), horse (*Equus caballus*), and camels (*Camelus dromedarius* and *Camelus bactrianus*) and intermediate susceptibility scores for the Malayan pangolin (*Manis javanica*), cow (*Bos taurus*), goat (*Capra hircus*), and sheep (*Ovis aries*).

To permit wider use of this susceptibility score for evaluation of additional species with unknown susceptibility, including those species that in the future may be of particular concern, we developed an implementation of the susceptibility score algorithm in R for public use. This implementation takes as input human ACE2 aligned with ACE2 of another species of interest and provides a susceptibility score using differences in ACE2 positions 30, 83, 90, 322, and 354. R code for implementation of this algorithm as a graphical user interface is available in Supplemental Methods.

DISCUSSION

Here we tested the hypothesis that differences in ACE2 proteins across various species alter structural interactions with SARS-CoV-2 RBD, leading to differences in species susceptibility to SARS-CoV-2 infection. Our results, combining prior knowledge of experimentally validated differences in species susceptibility with multiple methods of determining effects on ACE2 structure and interaction with SARS-CoV-2 RBD, reveal five key residues that in aggregate help discriminate susceptibility across species. These include ACE2 positions 30, 83, and 354, which exhibit alterations in binding energy, and positions 90 and 322, which exhibit alterations in glycosylation that likely contribute to differences in interactions at the interface. Taken together, our results provide insight into the molecular determinants of species susceptibility to SARS-CoV-2 infection and have important implications for identification of key residues for therapeutic targeting and determining susceptibility of additional species to infection.

Our study has several unique features that permit rigorous evaluation of differences in species susceptibility to infection. Prior studies have similarly performed ACE2 sequence alignments across species and modeled structural effects of the amino acid changes on the SARS-CoV-2 and ACE2 interface.[7,30-35] However, our study integrates experimentally validated susceptibility to SARS-CoV-2 with in-depth structural analyses to determine critical ACE2 residues for infection. In addition, we performed multiple structural analyses, including residue-residue interactions, RECON multistate design, and glycosylation analysis, to rigorously determine the structural basis for species differences in ACE2 interaction with SARS-CoV-2 RBD. Prior studies of ACE2 sequence alignment with limited structural modeling have suggested that pigs are susceptible to infection,[36] and that hamsters and house cats are in an intermediate risk group.[37] Recent experimental work with direct inoculation, however, has demonstrated that pigs are non-susceptible,[5] and that house cats and Golden Syrian hamsters are susceptible.[5,7] We identified key residues on which to build a susceptibility score that closely matches experimentally verified *in vivo* susceptibility, including predicting an intermediate susceptibility of the pig and higher susceptibility of house cats and Golden Syrian hamsters.

A key principle revealed by our findings is the importance of using multiple methods for determining the structural basis for differences in ACE2 interaction with SARS-CoV-2 RBD. For example, although calculated binding energy, protein stability, and RECON multistate design of SARS-CoV-2 RBD in complex with duck and chicken ACE2 distinguished non-susceptible chicken and duck ACE2 from susceptible species, mouse ACE2 did not fit the pattern of other non-susceptible species. However, analysis of ACE2 protein glycosylation revealed two residues, 90 and 322, for which differences in mouse ACE2 distinguished it from susceptible species. In addition, combining ACE2 sequence alignment, GroupSim calculations, and residue-residue interaction modeling identified residues 30 and 83, which are distinctly different in all non-susceptible compared to susceptible species. Differences in these residues in non-susceptible species result in decreased binding energy with SARS-CoV-2 RBD. Although no single residue appears capable of explaining the difference in susceptibility to SARS-CoV-2 infection across species, in combination amino acid positions 30, 83, 90, 322, and 354 can help distinguish susceptible from non-susceptible species, as reflected by the calculated

susceptibility score, which was lower in non-susceptible species and intermediate in those species with intermediate susceptibility.

Our findings have important implications for determining infectability of animals with heretofore unknown susceptibility to SARS-CoV-2 infection. Determining such susceptibility is critical to prevent disruption to food supplies, identify optimal animal models for research, aid in the search for intermediate hosts, and enhance identification of potential animal reservoirs that can propagate transmission.[38] We applied our infection susceptibility score to several important species with unknown susceptibility to date. These data suggest that cows (*Bos taurus*), Malayan pangolin (*Manis javanica*), and goats (*Capra hircus*) have intermediate susceptibility to infection, while Chinese horseshoe bats (*Rhinolophus sinicus*), horses (*Equus caballus*), and camels (*Camelus dromedarius* and *Camelus bactrianus*) have higher susceptibility. Although the ultimate test is direct exposure of live animals to evaluate infectability and transmissability,[5,7] this is complicated by the need for BSL3 containment and is quite costly and challenging with larger animals. Observational studies and case reports could also help provide evidence of susceptibility. Indeed, our results suggest that horses and camels should be tested and/or closely monitored for evidence of Covid-19 infection. The close interaction of these animals with humans, and the importance of these animals as domestic companions and laborers worldwide make determination of their susceptibility an urgent need. The use of the susceptibility score developed here can also be applied to additional species of interest to help direct resources for focused research and protection efforts in the future.

ACE2 residues identified in this paper that provide a structural basis to differences in species susceptibility to infection reveal important insights into the SARS-CoV-2 RBD and ACE2 structural interaction and potential for therapeutic targeting. By incorporating differences in species susceptibility into the structural analysis, our findings enhance the potential to identify particularly important residues mediating the ACE2 and SARS-CoV-2 RBD interaction. Indeed, although GroupSim scores were not used in the structural analysis, three of the five key identified residues (30, 83, and 90) from the structural modeling are in the top scoring ACE2 positions by GroupSim score. This suggests that the amino acids at these positions in ACE2 differ significantly between susceptible and non-susceptible species, consistent with an important contribution of these residues to differences in susceptibility. Amino acid positions 30 and 83 of ACE2 in particular exhibited large differences in residue-residue interaction binding energies between susceptible and non-susceptible species. Asp30 on ACE2 interacts with residues Lys417, Phe456, and Tyr473 of SARS-CoV-2 RBD, and ACE2 Tyr83 interacts with Asn487 and Tyr489 of SARS-CoV-2 RBD. These amino acids mark sites of SARS-CoV-2 interaction with ACE2 that may be important for development of antibody-based therapies or small molecule inhibitors.

Applying a multistate design algorithm to probe the SARS-CoV-2-RBD interactions for their ability to cross-bind to ACE2 of multiple species yielded several novel observations. First, this technique identified ACE2 position 354 as an important site for differentiating binding and non-binding ACE2 of different species to SARS-CoV-2 RBD. Second, this approach demonstrated that the SARS-CoV-2 RBD sequence is nearly optimal for binding to human ACE2 compared to other species. This is a remarkable finding, and likely underlies the high transmissibility of this

387 virus amongst humans. This finding is also consistent with recent results that compared SARS-
388 CoV and SARS-CoV-2 and determined that a number of differences in the SARS-CoV-2 RBD
389 have made it a much more potent binder to human ACE2 through the introduction of numerous
390 hydrogen bonding and hydrophobic networks.[39]

391 Although ACE2 and SARS-CoV-2 RBD interactions are critical to SARS-CoV-2 infection,[9,10,16]
392 differences in other factors across species may also contribute to differences in susceptibility.
393 This includes differences in ACE2 expression levels[40] and differences in the protein sequence of
394 TMPRSS2, a protein that contributes to viral and host cell membrane fusion through cleavage of
395 spike protein.[15,41] With further experimental and observational data on infectability of currently
396 unknown species, the susceptibility score we have developed can also help determine species
397 for which differences in ACE2 protein may not inadequately predict differences in susceptibility.
398 For these species future studies could compare differences in expression levels of ACE2 and/or
399 differences in TMPRSS2 structure. These structural comparisons of TMPRSS2, however, will
400 require elucidation of the protein crystal structure, which is not yet available.

401

402 **CONCLUSION**

403 We combined in-depth structural analyses with knowledge of varying species susceptibility to
404 SARS-CoV-2 infection to determine key structural determinants of infection susceptibility. First,
405 we identified multiple key residues mediating structural interactions between ACE2 and SARS-
406 CoV-2 RBD. Differences in these residues were used to generate a susceptibility score that can
407 help predict animals with elevated risk of infection for which we do not yet have experimental
408 evidence of susceptibility, including horses and camels. Finally, we have demonstrated that
409 SARS-CoV-2 is nearly optimal for binding ACE2 of humans compared to other animals, which
410 may underlie the highly contagious transmissibility of this virus amongst humans. Taken
411 together, results of these studies identify key structural regions of the ACE2 and SARS-CoV-2
412 interaction for therapeutic targeting and for identifying animal species on which to focus
413 additional research and protection efforts for environmental and public health.

414

References

1 Adhikari, S. P. *et al.* Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. *Infectious diseases of poverty* **9**, 29, doi:10.1186/s40249-020-00646-x (2020).

2 Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Current biology : CB* **30**, 1346-1351.e1342, doi:10.1016/j.cub.2020.03.022 (2020).

3 APHISpress@usda.gov. USDA Statement on the Confirmation of COVID-19 in a Tiger in New York. *United States Department of Agriculture Animal and Plant Health Inspection Service.* https://www.aphis.usda.gov/aphis/newsroom/news/sa_by_date/sa-2020/ny-zoo-covid-19. (April 6, 2020).

4 Mallapaty, S. Coronavirus can infect cats - dogs, not so much. *Nature*, doi:10.1038/d41586-020-00984-8 (2020).

5 Shi, J. *et al.* Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science*, doi:10.1126/science.abb7015 (2020).

6 Halfmann, P. J. *et al.* Transmission of SARS-CoV-2 in Domestic Cats. *New England Journal of Medicine*, doi:10.1056/NEJMc2013400 (2020).

7 Chan, J. F.-W. *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. *Clinical Infectious Diseases*, doi:10.1093/cid/ciaa325 (2020).

8 Chandrashekar, A. *et al.* SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science*, eabc4776, doi:10.1126/science.abc4776 (2020).

9 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).

10 Monteil, V. *et al.* Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell*, doi:10.1016/j.cell.2020.04.004 (2020).

11 Bao, L. *et al.* The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. *Nature*, doi:10.1038/s41586-020-2312-y (2020).

12 Sit, T. H. C. *et al.* Infection of dogs with SARS-CoV-2. *Nature*, doi:10.1038/s41586-020-2334-5 (2020).

13 Liu, Y. *et al.* Functional and Genetic Analysis of Viral Receptor ACE2 Orthologs Reveals Broad Potential Host Range of SARS-CoV-2. *bioRxiv*, 2020.2004.2022.046565, doi:10.1101/2020.04.22.046565 (2020).

14 Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, doi:10.1038/s41586-020-2180-5 (2020).

15 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e278, doi:10.1016/j.cell.2020.02.052 (2020).

16 Bao, L. *et al.* The Pathogenicity of SARS-CoV-2 in hACE2 Transgenic Mice. *bioRxiv*, 2020.2002.2007.939389, doi:10.1101/2020.02.07.939389 (2020).

17 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).

18 Li, F. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Res* **100**, 246-254, doi:10.1016/j.antiviral.2013.08.014 (2013).

19 Yu, P. *et al.* Age-related rhesus macaque models of COVID-19. *Animal models and experimental medicine* **3**, 93-97, doi:10.1002/ame2.12108 (2020).

20 Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research* **47**, W636-W641, doi:10.1093/nar/gkz268 (2019).

21 Capra, J. A. & Singh, M. Characterization and prediction of residues determining protein functional specificity. *Bioinformatics* **24**, 1473-1480, doi:10.1093/bioinformatics/btn214 (2008).

22 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).

23 Wang, Q. *et al.* Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell*, doi:10.1016/j.cell.2020.03.045 (2020).

24 Procko, E. The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2. *bioRxiv*, 2020.2003.2016.994236, doi:10.1101/2020.03.16.994236 (2020).

25 Sevy, A. M., Jacobs, T. M., Crowe, J. E. & Meiler, J. Design of Protein Multi-specificity Using an Independent Sequence Search Reduces the Barrier to Low Energy Sequences. *Plos Computational Biology* **11**, doi:ARTN e100430010.1371/journal.pcbi.1004300 (2015).

26 Sevy, A. M. *et al.* Multistate design of influenza antibodies improves affinity and breadth against seasonal viruses. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 1597-1602, doi:10.1073/pnas.1806004116 (2019).

27 Ferreira, I. G. *et al.* Glycosylation as a Main Regulator of Growth and Death Factor Receptors Signaling. *International journal of molecular sciences* **19**, 580, doi:10.3390/ijms19020580 (2018).

28 Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633-1649, doi:10.1002/pmic.200300771 (2004).

29 Pearson, W. R. Selecting the Right Similarity-Scoring Matrix. *Curr Protoc Bioinformatics* **43**, 3.5.1-3.5.9, doi:10.1002/0471250953.bi0305s43 (2013).

30 Melin, A. D., Janiak, M. C., Marrone, F., Arora, P. S. & Higham, J. P. Comparative ACE2 variation and primate COVID-19 risk. *bioRxiv*, 2020.2004.2009.034967, doi:10.1101/2020.04.09.034967 (2020).

31 Zhai, X. *et al.* Comparison of SARS-CoV-2 spike protein binding to ACE2 receptors from human, pets, farm animals, and putative intermediate hosts. *J Virol*, doi:10.1128/jvi.00831-20 (2020).

32 Luan, J., Jin, X., Lu, Y. & Zhang, L. SARS-CoV-2 spike protein favors ACE2 from Bovidae and Cricetidae. *Journal of medical virology*, doi:10.1002/jmv.25817 (2020).

33 Liu, Z. *et al.* Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *Journal of medical virology*, doi:10.1002/jmv.25726 (2020).

34 Praharaj, M. R. *et al.* Prediction analysis of SARS-COV-2 entry in Livestock and Wild animals. *bioRxiv*, 2020.2005.2008.084327, doi:10.1101/2020.05.08.084327 (2020).

35 Lam, S. *et al.* SARS-CoV-2 spike protein predicted to form complexes with host receptor protein orthologues from a broad range of mammals. *bioRxiv*, 2020.2005.2001.072371, doi:10.1101/2020.05.01.072371 (2020).

36 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. *Journal of Virology* **94**, e00127-00120, doi:10.1128/jvi.00127-20 (2020).

37 Damas, J. *et al.* Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural Analysis of ACE2 in Vertebrates. *bioRxiv*, 2020.2004.2016.045302, doi:10.1101/2020.04.16.045302 (2020).

38 Wu, D., Wu, T., Liu, Q. & Yang, Z. The SARS-CoV-2 outbreak: What we know. *International Journal of Infectious Diseases* **94**, 44-48, doi:https://doi.org/10.1016/j.ijid.2020.03.004 (2020).

39 Wang, Y., Liu, M. & Gao, J. Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. *Proceedings of the National Academy of Sciences*, 202008209, doi:10.1073/pnas.2008209117 (2020).

40 Fu, J. *et al.* Expressions and significances of the angiotensin-converting enzyme 2 gene, the receptor of SARS-CoV-2 for COVID-19. *Molecular Biology Reports*, doi:10.1007/s11033-020-05478-4 (2020).

41 Lukassen, S. *et al.* SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *Embo j*, e105114, doi:10.15252/embj.20105114 (2020).

42 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320-W324, doi:10.1093/nar/gku316 (2014).

43 Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**, W465-469, doi:10.1093/nar/gkn180 (2008).

44 Dereeper, A., Audic, S., Claverie, J. M. & Blanc, G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC evolutionary biology* **10**, 8, doi:10.1186/1471-2148-10-8 (2010).

45 Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **37**, D5-15, doi:10.1093/nar/gkn741 (2009).

46 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**, W256-W259, doi:10.1093/nar/gkz239 (2019).

47 Nivon, L. G., Moretti, R. & Baker, D. A Pareto-optimal refinement method for protein design scaffolds. *PLoS One* **8**, e59004, doi:10.1371/journal.pone.0059004 (2013).

48 Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735-1742, doi:10.1016/j.str.2013.08.005 (2013).

49 Gray, J. J. *et al.* Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* **331**, 281-299, doi:10.1016/s0022-2836(03)00670-3 (2003).

50 Chaudhury, S. *et al.* Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS One* **6**, e22477, doi:10.1371/journal.pone.0022477 (2011).

51 Sauer, M. F., Sevy, A. M., Crowe, J. E., Jr. & Meiler, J. Multi-state design of flexible proteins predicts sequences optimal for conformational change. *PLoS Comput Biol* **16**, e1007339, doi:10.1371/journal.pcbi.1007339 (2020).

52 Schrodinger, LLC. *The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.8* (2015).

53 Schrodinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* (2015).

54 Schrodinger, LLC. *The JyMOL Molecular Graphics Development Component, Version 1.8* (2015).

Table 1: Twenty-four key residues for SARS-CoV-2 RBD and ACE2 interactions

Genus species	Common name	19	24	27	28	30	31	34	35	37	38	41	42	45	79	82	83	325	329	330	353	354	355	357	393
Homo sapiens	*Human*	S	Q	T	F	D	K	H	E	E	D	Y	Q	L	L	M	Y	Q	E	N	K	G	D	R	R
Macaca mulatta	*Rhesus macaque*	S	Q	T	F	D	K	H	E	E	D	Y	Q	L	L	M	Y	Q	E	N	K	G	D	R	R
Felis catus	*House cat*	S	L	T	F	E	K	H	E	E	E	Y	Q	L	L	T	Y	Q	E	N	K	G	D	R	R
Panthera tigris altaica	*Tiger*	S	L	T	F	E	K	H	E	E	E	Y	Q	L	L	T	Y	Q	E	N	K	G	D	R	R
Panthera leo	*Lion*	S	L	T	F	E	K	H	E	E	E	Y	Q	L	L	T	Y	Q	E	N	K	G	D	R	R
Mesocricetus auratus	*Golden Syrian hamster*	S	Q	T	F	D	K	Q	E	E	D	Y	Q	L	L	N	Y	Q	E	N	K	G	D	R	R
Mus musculus	*Mouse*	S	N	T	F	N	N	Q	E	E	D	Y	Q	L	T	S	F	Q	A	N	H	G	D	R	R
Aythya fuligula	*Duck*	D	-	M	F	A	E	V	R	E	D	Y	E	L	N	N	F	E	K	N	K	N	D	R	R
Gallus gallus	*Chicken*	D	-	T	F	A	E	V	R	E	D	Y	E	L	N	R	F	E	T	N	K	N	D	R	R
Mustela putorius furo	*Ferret*	S	L	T	F	E	K	Y	E	E	E	Y	Q	L	H	T	Y	E	Q	N	K	R	D	R	R
Sus scrofa	*Pig*	S	L	T	F	E	K	L	E	E	D	Y	Q	L	I	T	Y	Q	N	N	K	G	D	R	R
Canis lupus familiaris	*Dog*	S	L	T	F	E	K	Y	E	E	E	Y	Q	L	L	T	Y	Q	G	N	K	G	D	R	R
Rhinolophus sinicus	*Chinese horseshoe bat*	S	E	T	F	D	K	T	K	E	D	H	Q	L	L	N	Y	E	N	N	K	G	D	R	R
Equus caballus	*Horse*	S	L	T	F	E	K	S	E	E	E	H	Q	L	L	T	Y	Q	E	N	K	G	D	R	R
Bos taurus	*Cow*	S	Q	T	F	E	K	H	E	E	D	Y	Q	L	M	T	Y	Q	D	N	K	G	D	R	R
Manis javanica	*Malayan pangolin*	S	E	T	F	E	K	S	E	E	E	Y	Q	L	I	N	Y	Q	E	N	K	H	D	R	R
Capra hircus	*Goat*	S	Q	T	F	E	K	H	E	E	D	Y	Q	L	M	T	Y	Q	N	N	K	G	D	R	R
Ovis aries	*Sheep*	S	Q	T	F	E	K	H	E	E	D	Y	Q	L	M	T	Y	Q	D	N	K	G	D	R	R
Camelus dromedarius	*Arabian Camel*	S	L	T	F	E	E	H	E	E	D	Y	Q	L	T	T	Y	Q	D	N	K	G	D	R	R
Camelus bactrianus	*Bactrian Camel*	S	L	T	F	E	E	H	E	E	D	Y	Q	L	T	T	Y	Q	D	N	K	G	D	R	R

Highlighted residues that are most similar in susceptible and different in non-susceptible species as determined by GroupSim (Extended Data Table 1). Susceptible species are in orange, non-susceptible in green, intermediate in blue, and unknown in black/grey. Letters indicate amino acids using single-letter naming.

552 **Table 2: Predicted glycosylation profiles for ACE2 amino acid positions 53, 90, 103 and**
553 **322**

Species	53	54	55	glyc.	90	91	92	glyc.	103	104	105	glyc.	322	323	324	glyc.
Homo sapiens	N	I	T	+	N	L	T	+	N	G	S	+	N	M	T	+
Macaca mulatta	N	I	T	+	N	L	T	+	N	G	S	+	N	M	T	+
Felis catus	N	I	T	+	N	T	T	+	S	G	S	-	N	M	T	+
Panthera tirgris altaica	N	I	T	+	N	T	T	+	S	G	S	-	N	M	T	+
Panthera leo	N	I	T	+	N	T	T	+	S	G	S	-	N	M	T	+
Mesocricetus auratus	N	I	T	+	N	L	T	+	S	G	S	-	Y	M	T	-
Mus musculus	N	I	T	+	T	P	I	-	S	G	S	-	H	M	T	-
Aythya fuligula	N	I	T	+	D	P	L	-	K	G	S	-	N	M	T	+
Gallus gallus	N	I	T	+	D	A	V	-	R	G	S	-	N	M	T	+
Mustela putorius furo	N	I	T	+	D	P	I	-	S	G	S	-	N	M	T	+
Sus scrofa	N	I	T	+	T	L	I	-	S	G	T	-	N	M	T	+
Canis lupus familiaris	N	I	T	+	D	S	T	-	S	G	S	-	N	M	T	+
Rhinolophus sinicus	N	I	N	-	N	V	T	+	S	G	S	-	N	M	T	+
Equus caballus	N	I	T	+	N	L	T	+	S	G	S	-	N	M	T	+
Bos taurus	N	I	T	+	N	L	T	+	S	G	T	-	Y	M	T	-
Manis javanica	N	I	T	+	N	D	T	+	S	G	S	-	K	M	T	-
Capra hircus	N	I	T	+	N	L	T	+	S	G	T	-	Y	M	T	-
Ovies aries	N	I	T	+	N	L	T	+	S	G	T	-	Y	M	T	-
Camelus dromedarius	N	I	T	+	N	V	T	+	S	G	A	-	N	M	T	+
Camelus bactrianus	N	I	T	+	N	V	T	+	S	G	A	-	N	M	T	+

554

555 Susceptible species are in orange, non-susceptible in green, intermediate in blue, and unknown in black.
556 + indicates presence, - indicates absence of glycosylation. glyc=glycosylation. Letters indicate amino
557 acids using single-letter naming.

558

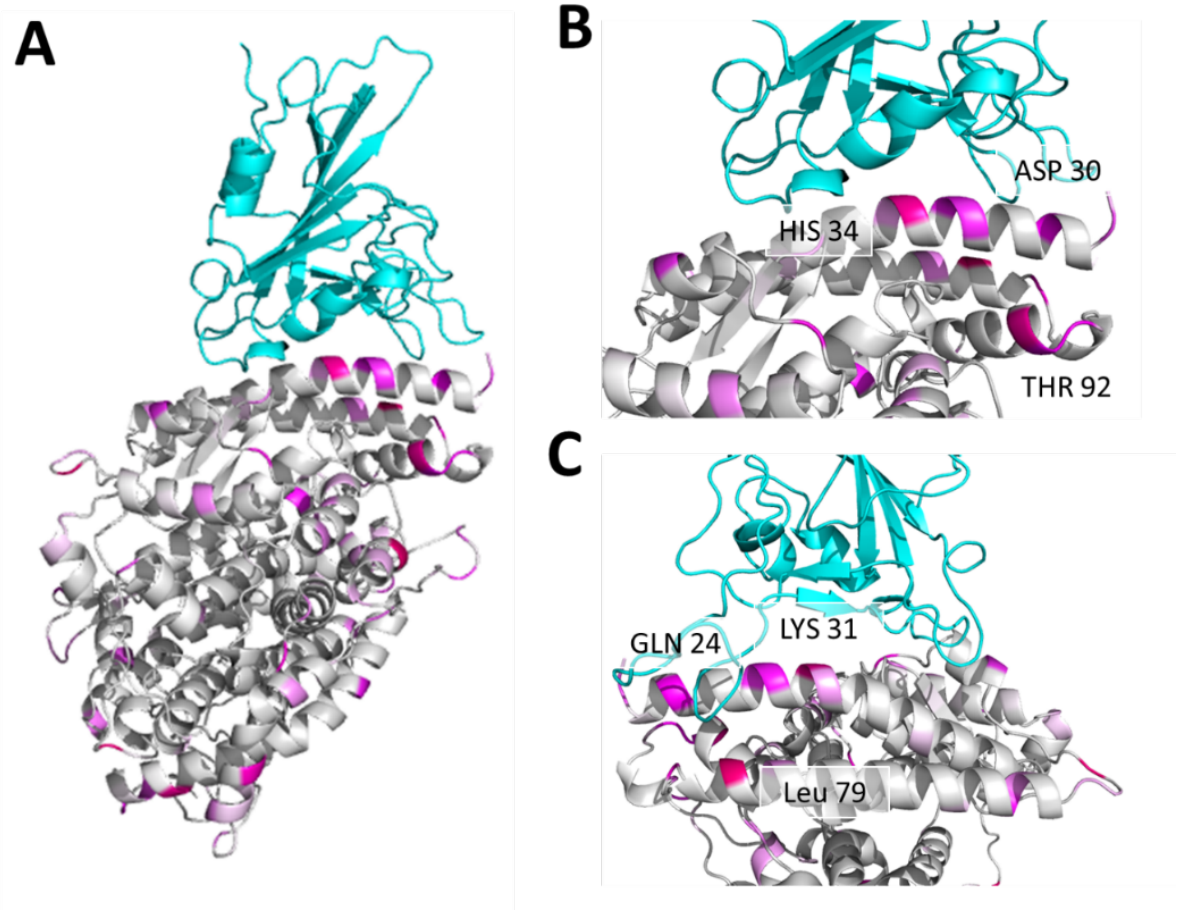
559 **Table 3: Key residues of aligned ACE2 proteins with calculated SARS-CoV-2**
560 **susceptibility score for each species**

	30	83	90	322	354	Susceptibility score
Homo sapiens (Human)	D	Y	N	N	G	-
Macaca mulatta (Rhesus Macaque)	D	Y	N	N	G	31
Felis catus (House cat)	E	Y	N	N	G	27
Panthera tigris altaica (Tiger)	E	Y	N	N	G	27
Panthera leo (Lion)	E	Y	N	N	G	27
Mesocricetus auratus (Golden Syrian Hamster)	D	Y	N	Y	G	23
Mus musculus (Mouse)	N	F	T	H	G	11
Aythya fuligula (Duck)	A	F	D	N	N	8
Gallus gallus (Chicken)	A	F	D	N	N	8
Mustela putorius furo (Ferret)	E	Y	D	N	R	14
Sus scrofa (Pig)	E	Y	T	N	G	21
Canis lupus familiaris (Dog)	E	Y	D	N	G	22
Rhinolophus sinicus (Chinese horseshoe bat)	D	Y	N	N	G	31
Equus caballus (Horse)	E	Y	N	N	G	27
Bos taurus (Cow)	E	Y	N	Y	G	19
Manis javanica (Malayan pangolin)	E	Y	N	K	H	13
Capra hircus (Goat)	E	Y	N	Y	G	19
Ovis aries (Sheep)	E	Y	N	Y	G	19
Camelus dromedarius (Arabian Camel)	E	Y	N	N	G	27
Camelus bactrianus (Bactrian Camel)	E	Y	N	N	G	27

561

562 Susceptible (orange), non-susceptible (green), intermediate (blue), and unknown (black/grey) species are
563 indicated.

564 **Figures**



565

566 **Figure 1: Multiple residues with high GroupSim scores are present at the interaction**
567 **interface of the SARS-CoV-2 RBD and ACE2 complex.** (**A**) SARS-CoV-2 RBD (top) and
568 human ACE2 (bottom) complex shown as a ribbon diagram with GroupSim scores color coded
569 in magenta. Higher scores are brighter in color. (**B**) Close-up view of the interface highlighting
570 ACE2 residues with high GroupSim scores. (**C**) Close-up view after 90 degree rotation from (B)
571 demonstrating additional residues at the interface with high GroupSim scores.

572

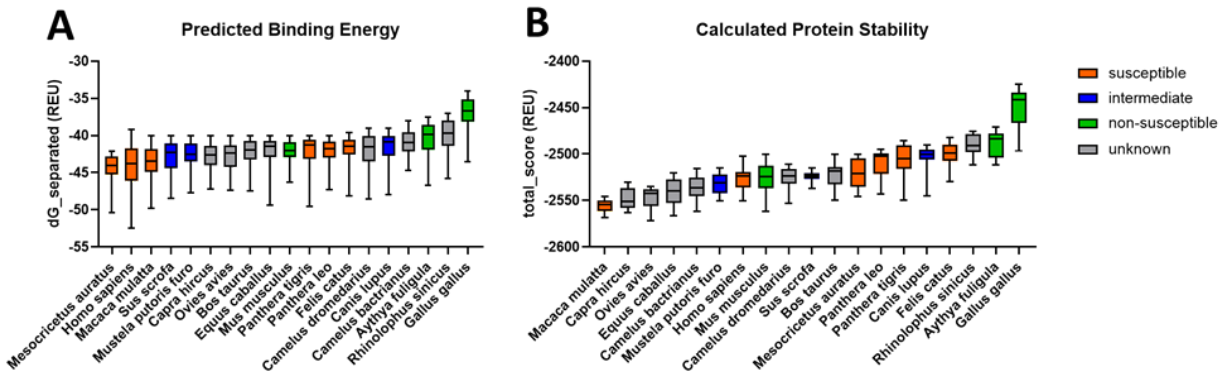


Figure 2: SARS-CoV-2 RBD has lower predicted binding energy and protein complex stability for ACE2 from non-susceptible avian species. (**A**) Predicted binding energy as calculated with Rosetta and (**B**) protein complex stability of SARS-CoV-2 RBD and ACE2 of various species predicted by Rosetta.

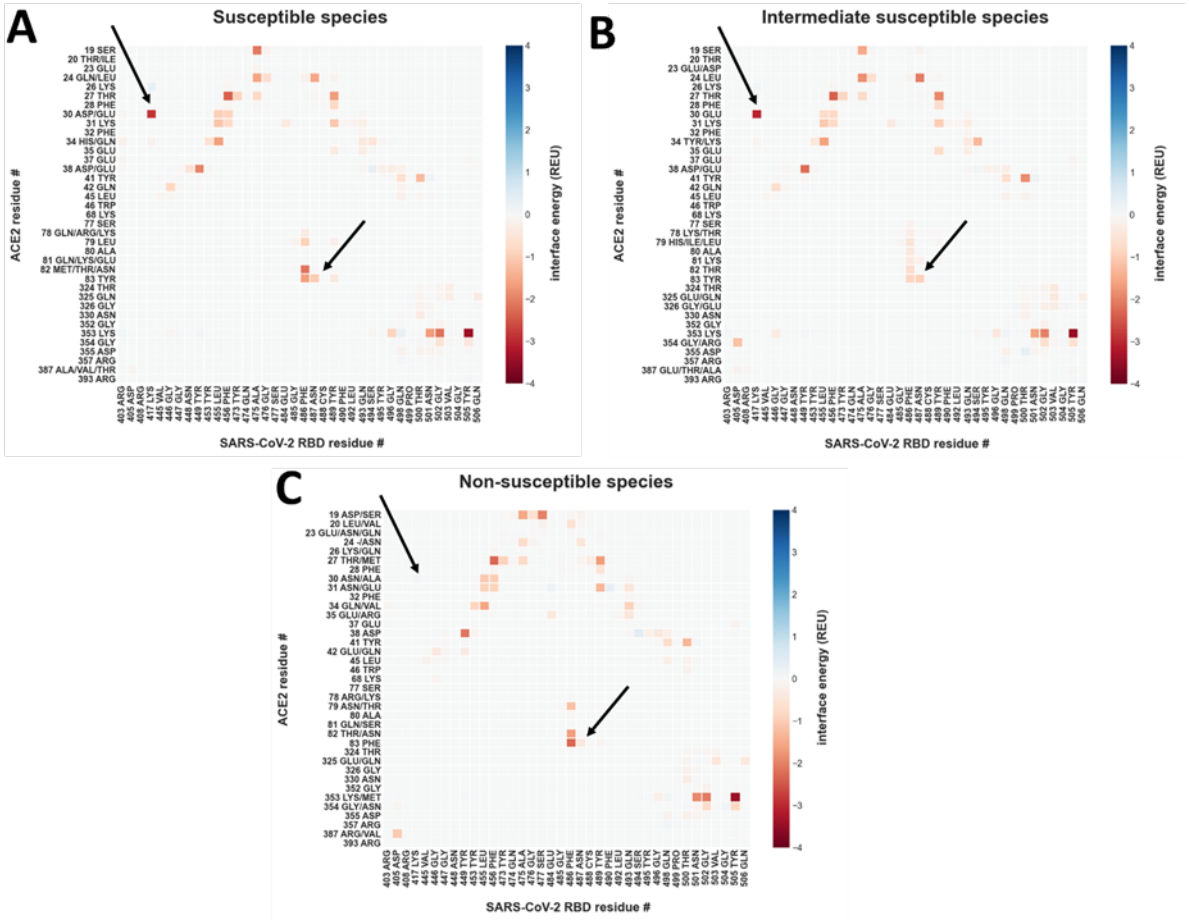


Figure 3: Energetic modeling of residue-residue interactions identifies a link between ACE2 D30 and Y83 and SARS-CoV-2 susceptibility. Residue-residue interactions are calculated with Rosetta, using the co-crystal structure of the human ACE2 in complex with the SARS-CoV-2-RBD (PDB: 6LZG and 6M0J) after backbone-constrained relaxation for all interactions greater than 0.05 Rosetta Energy Units (REU) or smaller than -0.05 REU . Interactions are presented as mean for all included samples. Residues depicted on the y-axis are all observed amino acid identities for the particular position in its susceptibility group. (**A**) Per-residue interactions for (A) susceptible species (human, cat, lion, tiger, hamster and rhesus macaque), (**B**) intermediate susceptibility species (pig, dog and ferret), and (**C**) non-susceptible species (duck, mouse, and chicken). The arrows point to interactions that are not observed in non-susceptible species.

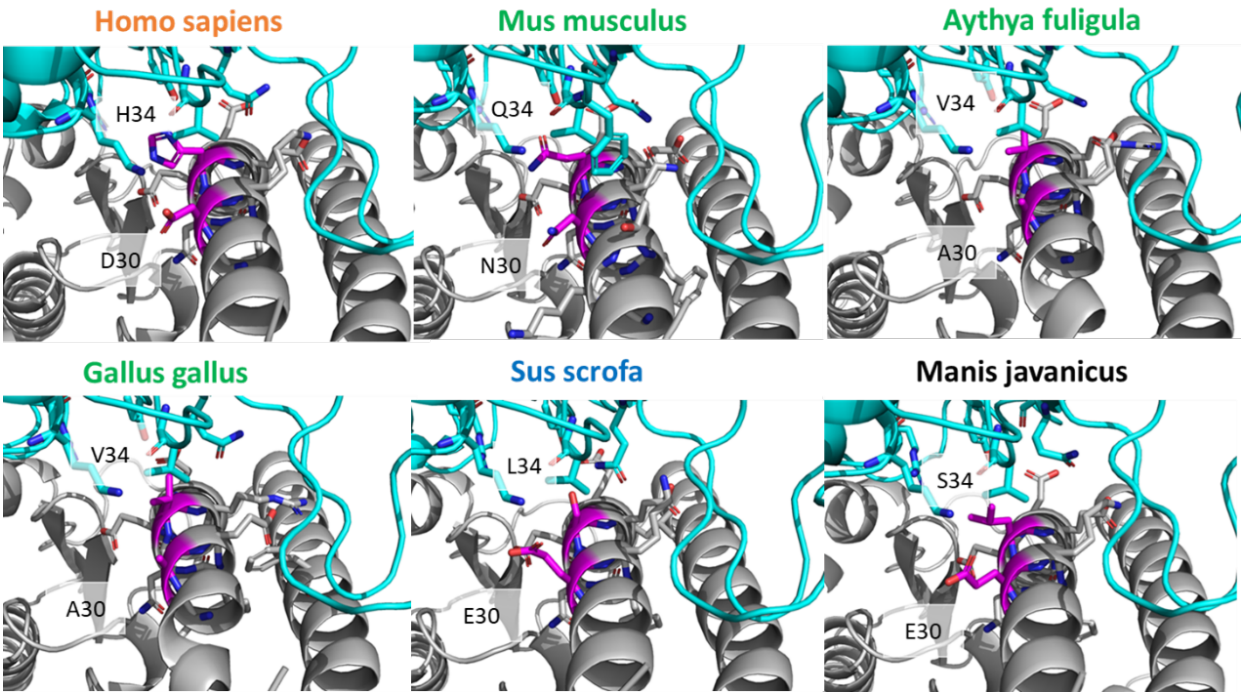
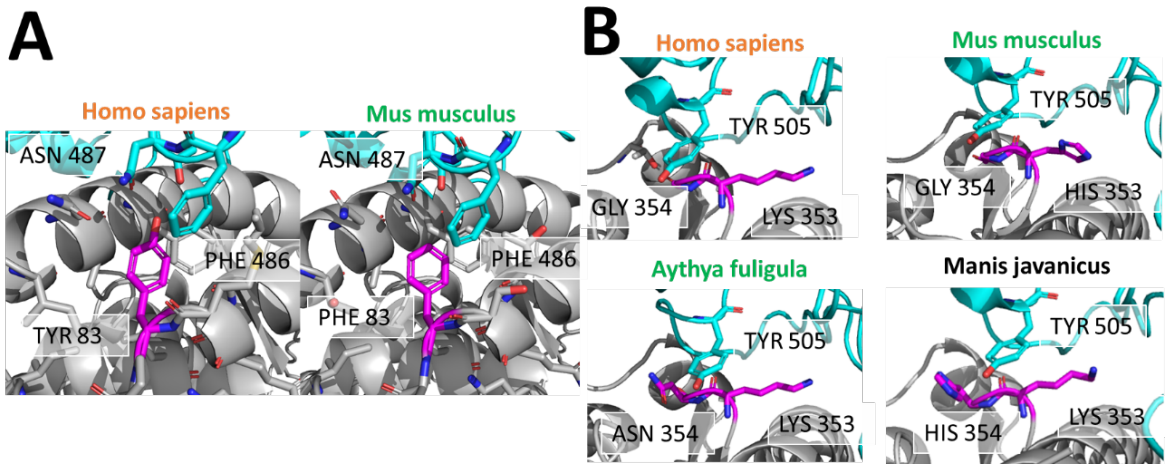


Figure 4: Binding interactions of ACE2 position 30 differ across species. Close-up of the differences in binding interactions of positions 30 and 34 (magenta) of ACE2 from each species with the SARS-CoV-2 RBD. Position 30 is occupied by an aspartic acid (D) in susceptible humans (*Homo sapiens*), is an asparagine (N) in non-susceptible mice (*Mus musculus*), and an alanine (A) in the avian species (*Aythya fuligula* and *Gallus gallus*). Glutamic acid (E) is present at position 30 in pig (*Sus scrofa*) and Malayan pangolin (*Manis javanicus*), representing intermediate and unknown susceptible species, respectively. Position 34 is conserved as histidine (H) in all susceptible species such as humans, yet has another residue identity in intermediate and non-susceptible species. Species names in orange are susceptible, green are non-susceptible, blue are intermediate susceptibility, and black are unknown.



606

607

608 **Figure 5: Binding interactions of ACE2 positions 83 and 354 differ across susceptible**
609 **and non-susceptible species.** (**A**) ACE2 position 83 (magenta) is a tyrosine in the human
610 susceptible species (left) and phenylalanine in the non-susceptible mouse species (right).
611 Tyrosine 83 of human ACE2 interacts with asparagine 87 of SARS-CoV-2 RBD, probably via a
612 hydrogen bond. Phenylalanine in mouse ACE2 cannot interact with asparagine 487 due to the
613 lack of a hydrogen bond donor. (**B**) Interactions of tyrosine r505 of the SARS-CoV-2-RBD (cyan)
614 with ACE2 residues 353 and residue 354 (magenta). ACE2 residue 353 is conserved as lysine
615 with the only exception of a histidine in the mouse ACE2. ACE2 residue 354 is a glycine in the
616 susceptible species (human), but an asparagine in non-susceptible duck and chicken, and a
617 histidine in pangolin (unknown susceptibility). Species names in orange are susceptible, green
618 are non-susceptible, and black are unknown.

619

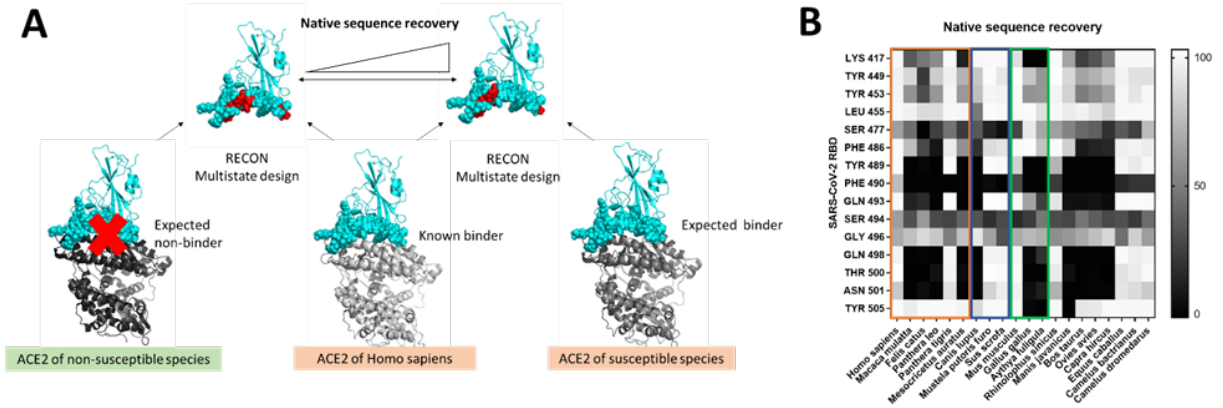


Figure 6: Multistate design reveals SARS-CoV-2 RBD Tyr505 to have low native sequence recovery in non-susceptible duck and chicken. (**A**) RECON multistate design overview. In the presence of ACE2 from two different species the SARS-CoV-2-RBD interface is redesigned. When two true binders are redesigned they should require few sequence changes, thus resulting in a higher native sequence recovery. In contrast, if the native sequence recovery for the interface residues is lower, then many sequence changes are required, indicating that one of the ACE2 proteins is a non-binder. (**B**) Residue-specific native sequence recovery as determined from RECON multistate design against the SARS-CoV-2-RBD complex with human ACE2. Only residues of the SARS-CoV-2-RBD, which are in the protein-protein interface and show changes are depicted. Tyrosine 505 of SARS-CoV-2 RBD shows low native sequence recovery (black) in non-susceptible duck (*Gallus gallus*) and chicken (*Aythya fuligula*). The orange box outlines susceptible species, the blue box outlines species with intermediate susceptibility, and the green box outlines non-susceptible species.

METHODS

ACE2 protein alignment

Protein sequence accession numbers and corresponding FASTA files from multiple species (**Extended Data Table 2**) were pulled from NCBI using Batch Entrez. In the absence of a published sequence and accession number, ACE2 protein sequence for the lion (*Panthera leo*) was assembled using TBLASTN (National Center for Biotechnology Information) with tiger ACE2 protein sequence as the query (**Extended Data Table 3**). Protein sequences were loaded into EMBL-EBI web interface implementation of MAFFT for multiple sequence alignment using default settings (https://www.ebi.ac.uk/Tools/msa/mafft/).[20] Resulting alignment was uploaded to ESPript 3.0 to generate a graphical version of the alignment (http://espript.ibcp.fr/ESPript/ESPript/), including annotation of secondary structure based on Protein Data Bank (PDB) structure 1r42 of human ACE2.[42] A treedyn format tree diagram representing similarity of ACE2 protein sequence across species was generated using phylogeny.fr (https://www.phylogeny.fr/).[43,44] NCBI Taxonomy Browser was used to generate a taxonomic tree of phylogenetic relationships amongst species as a Phylogeny Inference Package (PHYLIP) tree.[45] Final visualization was performed using the interactive Tree of Life (iTOL) tree viewer v 5.5.1 (https://itol.embl.de/).[46]

Quantification of amino acid differences in alignment of susceptible and non-susceptible species

Quantification of amino acid positions in the ACE2 protein alignment that optimally distinguish susceptible versus non-susceptible species was performed using GroupSim.[21] Values from 0 to 1 were obtained with 1 assigned to the position that best stratifies susceptible and non-susceptible species. Values are weighted by the BLOSUM62 similarity matrix to incorporate similarity of amino acids properties.[29]

Homology modeling of ACE2-SARS-CoV2 co-crystal structures using RosettaCM

ACE2 of human and non-human species was modeled based on two co-crystal structures of SARS-CoV-2-RBD with the human ACE2 (PDB-IDs 6LZG and 6M0J).[14] One co-crystal structure (PDB-ID 6VW1) was excluded due to its lower resolution as compared to the aforementioned structures. The target sequences were threaded over the ACE2-SARS-CoV-2-RBD co-crystal structure, which was first relaxed with backbone constraints using RosettaRelax.[47] A total of 1000 homology models were constructed using RosettaCM, and subsequently relaxed with backbone constraints.[47,48] Of these, 25 models were selected based on the total energy as a measure of protein stability, predicted binding energy, and Cα-root mean square deviation (Cα-RMSD) to the best scoring model (**Extended Data Figure 3**). The SARS-CoV-2-RBD-ACE2 complex was optimized using a rigid-body docking with limited degrees for rotational and torsional sampling.[49,50] A final ensemble of 100 models was selected based on the total energy as measure of protein stability, predicted binding energy and Cα-RMSD to the best scoring model (**Extended Data Figure 4**). The pairwise binding interaction between SARS-CoV-2 and ACE2 was evaluated by retrieving the decomposed Rosetta scores for each residue. The protocol was tested by modeling the human ACE2 in complex with SARS-CoV-2-RBD, and

evaluating the recovery of predicted binding energy, total energy, and residue-residue interactions in the interface.

Calculation of sequence recovery from Restraint Convergence (RECON) multistate design

RECON multistate design was carried out as reported previously for each susceptible, non-susceptible, intermediate, and unknown species against the human SARS-CoV-2-RBD-ACE2 complex.[25,26,51] As a control, this was also performed solely using the human SARS-CoV-2-RBD-ACE2 complex. A total of 5000 models were sampled and trajectories with final models that scored lower than -2400 REU were evaluated. The native sequence recovery was calculated for each pairwise experiment and also for the control run for the SARS-CoV-2-RBD complex with the human ACE2 (**Extended Data Figure 6**).

All protocols were executed using Rosetta-3.12 (www.rosettacommons.org). Evaluation was performed using the numpy, pandas, matplotlib and seaborn libraries in Python 3.7, PyMOL 2.7[52-54] and GraphPad Prism version 8.3.0 for Windows (GraphPad Software, San Diego, California). Example commands and RosettaScripts protocols can be found in the Supplementary Methods.

Prediction of glycosylation sites

The NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) was used to predict glycosylation sites.[28] Based from the observation that asparagine in positions 53, 90, and 322 carried glycosylation in the crystal structures PDB: 6LZG and 6M0J, and scored with high confidence from NetNGlyc 1.0, these were selected as reliably glycosylated. Position 103 was included, as it was strongly predicted to be glycosylated by NetNGlyc 1.0, although no glycosylation was observed in the crystal structures. Furthermore, it was evaluated whether the NxT/S sequons were surface accessible and in proximity to the ACE2-SARs-CoV-2-RBD binding interface.

SARS-Cov-2 susceptibility score calculation

Using identified ACE2 key amino acid positions 30, 83, 90, 322, and 354 in the alignment of ACE2 across species, a global susceptibility score was calculated as the sum of the Blosum62 scoring matrix substitutions for the amino acid at each position compared to the human ACE2 sequence.[29] This was calculated for each species, with higher scores suggesting greater susceptibility. An R implementation of this susceptibility score algorithm was also developed in RStudio. The software takes as input alignment of human ACE2 protein sequence with ACE2 of another species of interest and provides a susceptibility score as output. Susceptibility scores of species examined in this manuscript are also graphically demonstrated as reference. Code for implementing this algorithm in R as a graphical user interface is available in Supplemental Methods.

Statistical analysis

713 Contingency testing was performed with Fisher's exact test as a two-sided comparison and
714 alpha equal to 0.05 using GraphPad Prism version 8.2.1 (GraphPad Software, Inc.).

715

Acknowledgements

The authors would like to thank Erkan Karakas for useful suggestions at the beginning of the project and Melissa Farrow for helpful suggestions in the conceptualization of the study. This work was supported in part by the National Institutes of Health under awards F32HL144048-01 (MRA), DK117147 (WC), UH3TR002097 and U01TR002383 (JPW and JAB), U19AI117905, U01AI150739, and R01AI141661 (JM and CTS), R35GM127087 (CM and JAC), and DP2HL137166 (MSM). The work was also supported by the American Heart Association 20PRE35080177 (CDS) and EIA34480023 (MSM). The views expressed are solely those of the authors.

Author Contributions

JAB and JPW conceptualized the study; MRA, CTS, WC, MSM, JM, JPW, JAB, CM, and JAC made substantial contributions to experimental design; MRA and CTS conducted the majority of the experiments with help from WC, JM, CDS, JAC, JAB, and JPW; MRA, CTS, JAC, WC, CM, and JM significantly contributed to the data acquisition and interpretation; MRA and CTS drafted the manuscript with WC, MSM, JM, JPW, and JAC contributing critical revisions. All authors approved the final version.

Data and Code Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request. R code for the susceptibility score algorithm is available in Supplemental Methods.

Additional Information:

Supplementary Information is available for this paper.

741 **Extended Data Tables**

742 **Extended Data Table 1: GroupSim scores of aligned ACE2 amino acid positions with**
743 **greatest differences between susceptible and non-susceptible species.**

Amino Acid	GroupSim score
682	1.000
79	0.661
568	0.643
337	0.643
286	0.643
246	0.554
34	0.548
92	0.536
751	0.518
593	0.494
641	0.464
83	0.464
536	0.458
709	0.446
321	0.446
637	0.429
90	0.411
24	0.406
689	0.405
91	0.404
253	0.393
139	0.393
782	0.381
728	0.381
113	0.381
228	0.380
752	0.375
59	0.375
31	0.375
653	0.371
30	0.361
765	0.357
675	0.357
329	0.357
214	0.357

744 Bold amino acids are structurally predicted or experimentally validated to mediate ACE2 interactions with SARS-CoV-
745 2 spike protein. Amino acid numbers correspond to *Homo sapiens* ACE2.

746

Extended Data Table 2: Species and accession numbers used for ACE2 protein sequence alignment.

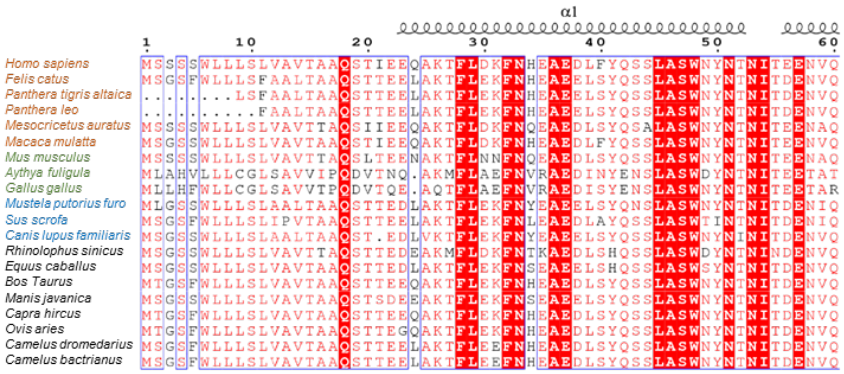
Common name	Genus species	Accession number
Human	*Homo sapiens*	NP_001358344.1
House cat	*Felis catus*	XP_023104564.1
Tiger	*Panthera tigris altaica*	XP_007090142.1
Lion	*Panthera leo*	No accession number
Golden Syrian hamster	*Mesocricetus auratus*	XP_005074266.1
Rhesus macaque	*Macaca mulatta*	NP_001129168.1
Mouse	*Mus musculus*	NP_081562.2
Duck	*Aythya fuligula*	XP_032058386.1
Chicken	*Gallus gallus*	XP_416822.2
Ferret	*Mustela putorius furo*	NP_001297119.1
Pig	*Sus scrofa*	NP_001116542.1
Dog	*Canis lupus familiaris*	NP_001158732.1
Chinese horseshoe bat	*Rhinolophus sinicus*	AGZ48803.1
Horse	*Equus caballus*	XP_001490241.1
Cow	*Bos taurus*	XP_005228485.1
Malayan pangolin	*Manis javanica*	XP_017505746.1
Goat	*Capra hircus*	NP_001277036.1
Sheep	*Ovis aries*	XP_011961657.1
Arabian camel	*Camelus dromedarius*	XP_010991717.1
Bactrian camel	*Camelus bactrianus*	XP_010966303.1

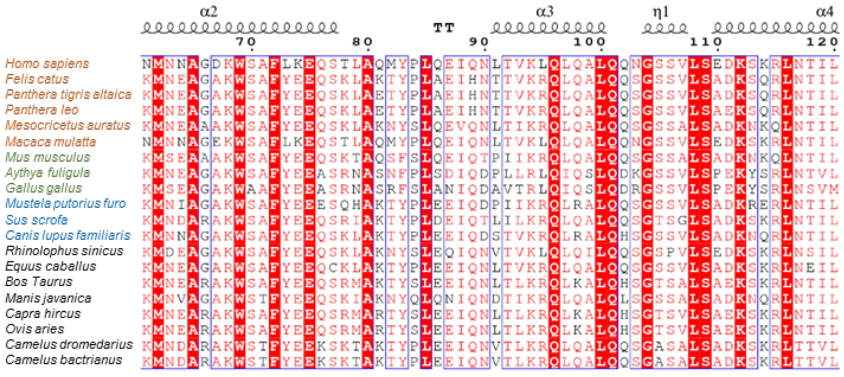
Susceptible (red), non-susceptible (green), intermediate (blue), and unknown (black) species are indicated.

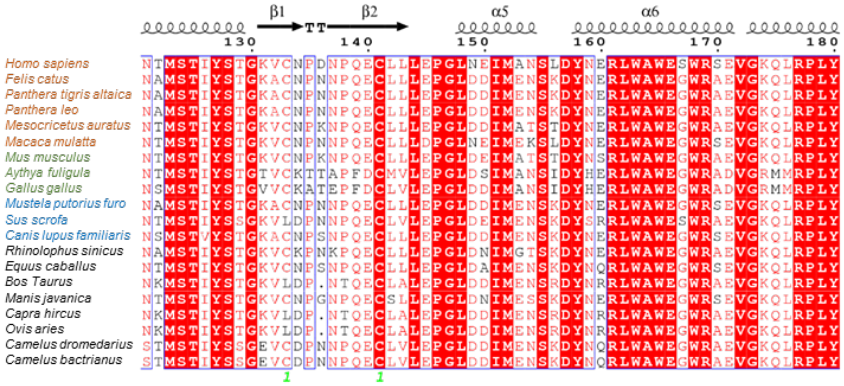
Extended Data Table 3: *Panthera leo* ACE2 protein sequence.

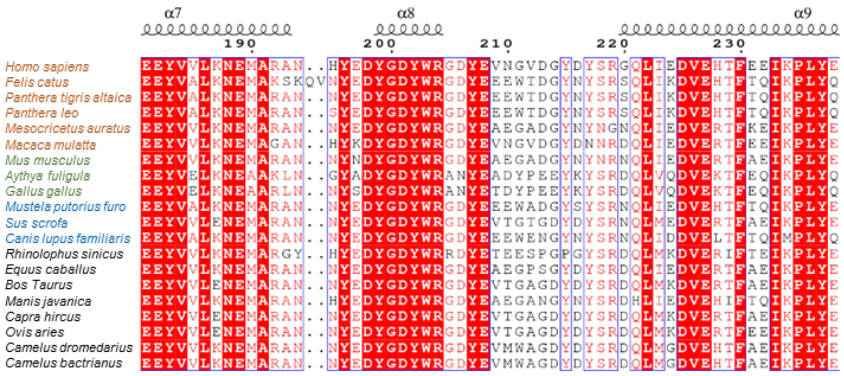
FAALTAAQSTTEELAKTFLEKFNHEAEELSYQSSLASWNYNTNITDENVQKMNEAGAKWS
AFYEEQSKLAETYPLAEIHNTTVKRQLQALQQSGSSVLSAEKSQRLNTILNAMSTIYSTG
KACNPNNPQECLLLEPGLDDIMENSKDYNERLWAWEGWRAEVGKQLRPLYEEYVALKNEM
ARANSYEDYGDYWRGDYEEEWTDGYNYSRSQLIKDVEHTFTQIKPLYQHLHAYVRAKLMD
SYPSRISPTGCLPAHLLGDMWGRFWTNLYPLTVPFGQKPNIDVTDAMVNQSWDARRIFKE
AEKFFVSVGLPNMTQGFWENSMLTEPGDSQKVVCHPTAWDLGKGDFRIKMCTKVTMDDFL
TAHHEMGHIQYDMAYAVQPFLLRNGANEGFHEAVGEIMSLSAATPNHLKTIGLLPPGFSE
DSETEINFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKEQWMQKWWEMKREIVGVVEP
VPHDETYCDPASLFHVANDYSFIRYYTRTIYQFQFQEALCRIAKHEGPLHKCDISNSSEA
GKKLLQMLTLGKSKPWTLALEHVVGEKNMNVTPLLKYFEPLFTWLKEQNRNSFVGWNTDW
RPCADQSIKVRISLKSALGDKAYEWNDNEMYLFRSSVAYAMREYFSKVKNQTIPFVEDNV
WVSNLKPRISFNFFVTASKNVSDVIPRREVEEAIRMSRSRINDAFRLDDNSLEFLGIQPT
LSPPYQPPVTIWLIVFGVVMGVVVVGIVLLIVSGIRNRRKEQSSKK

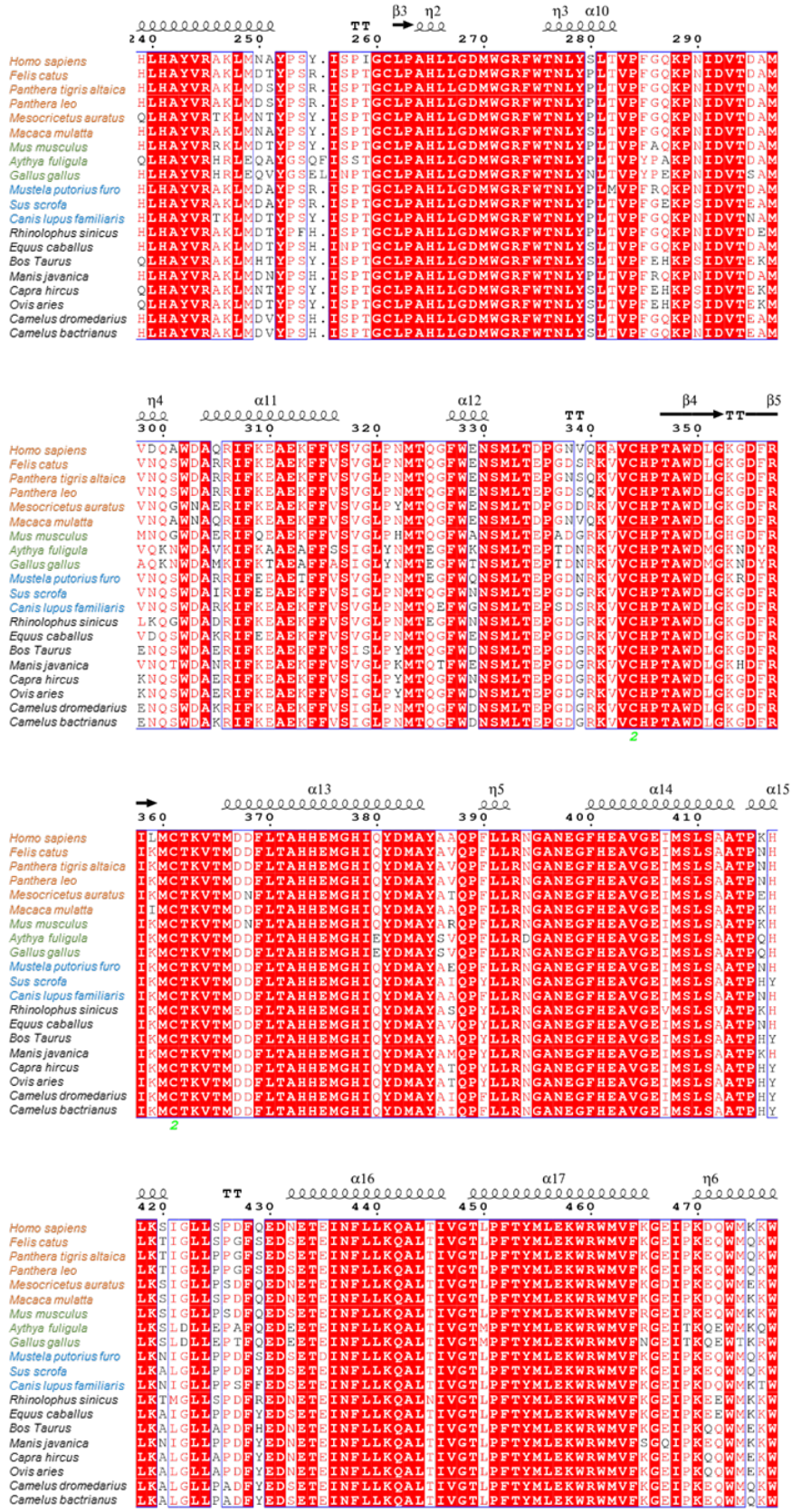
Extended Data Figures



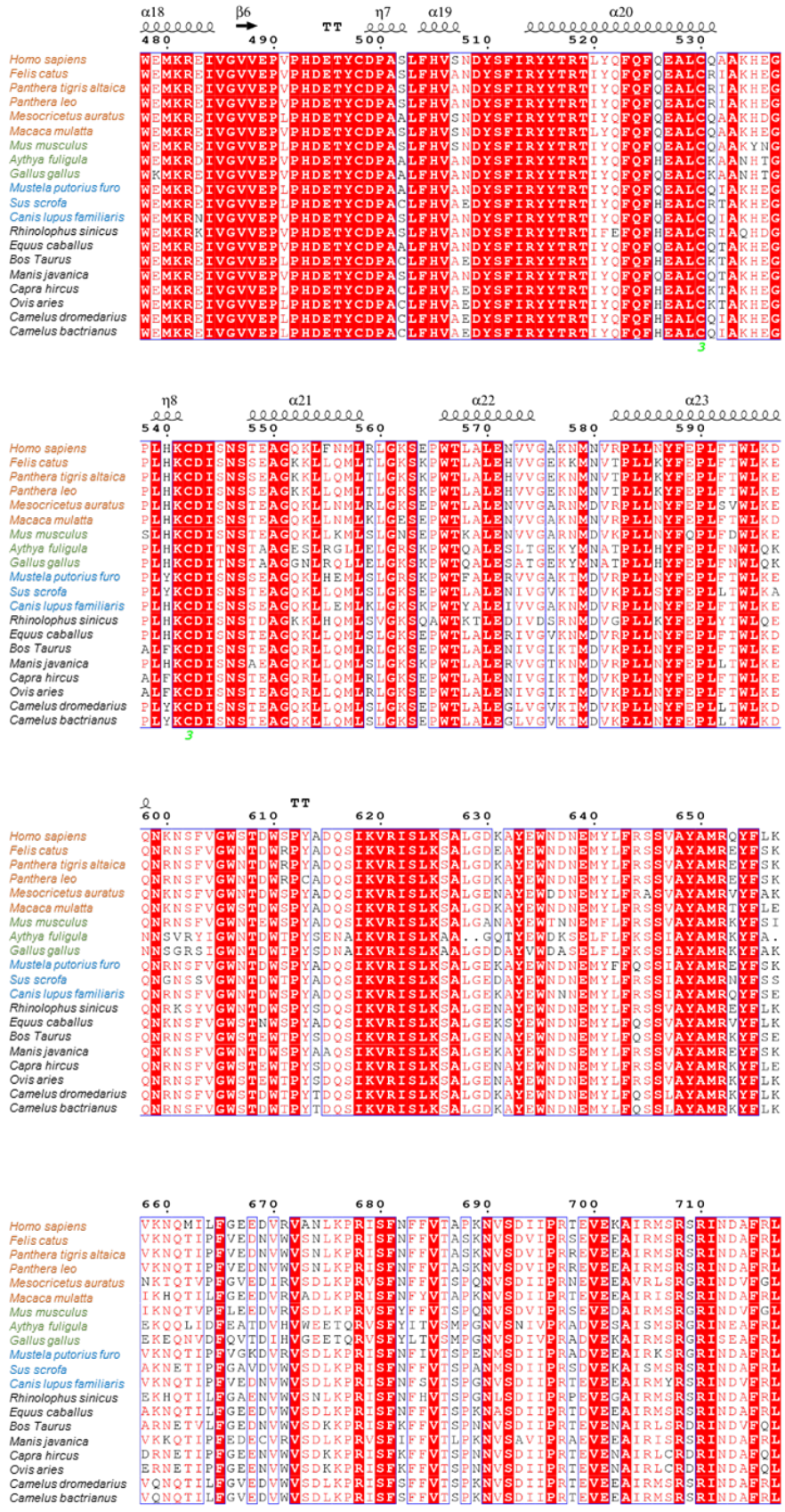


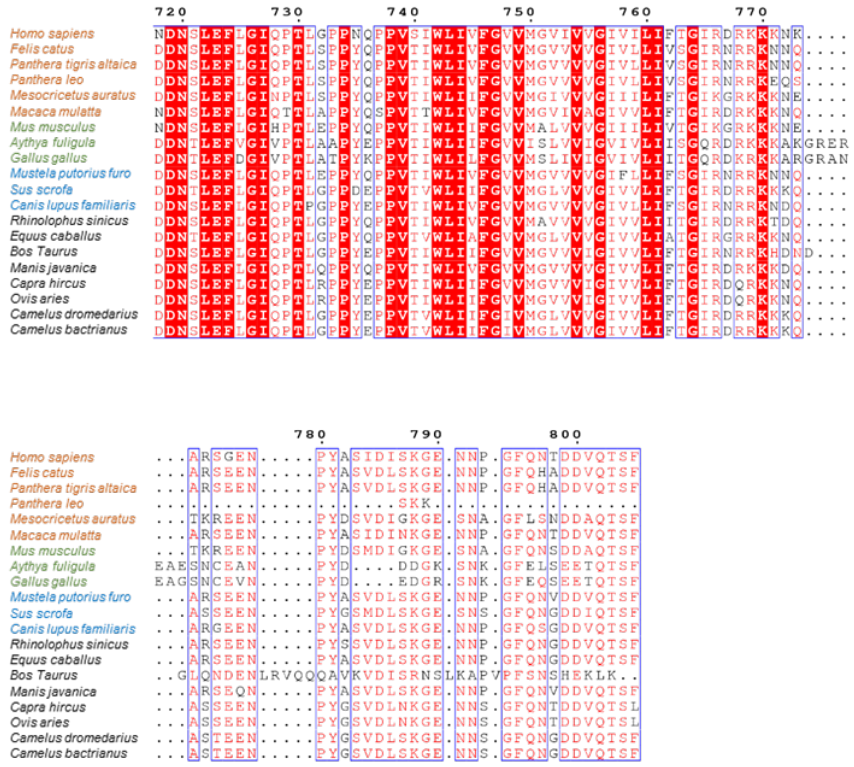






778



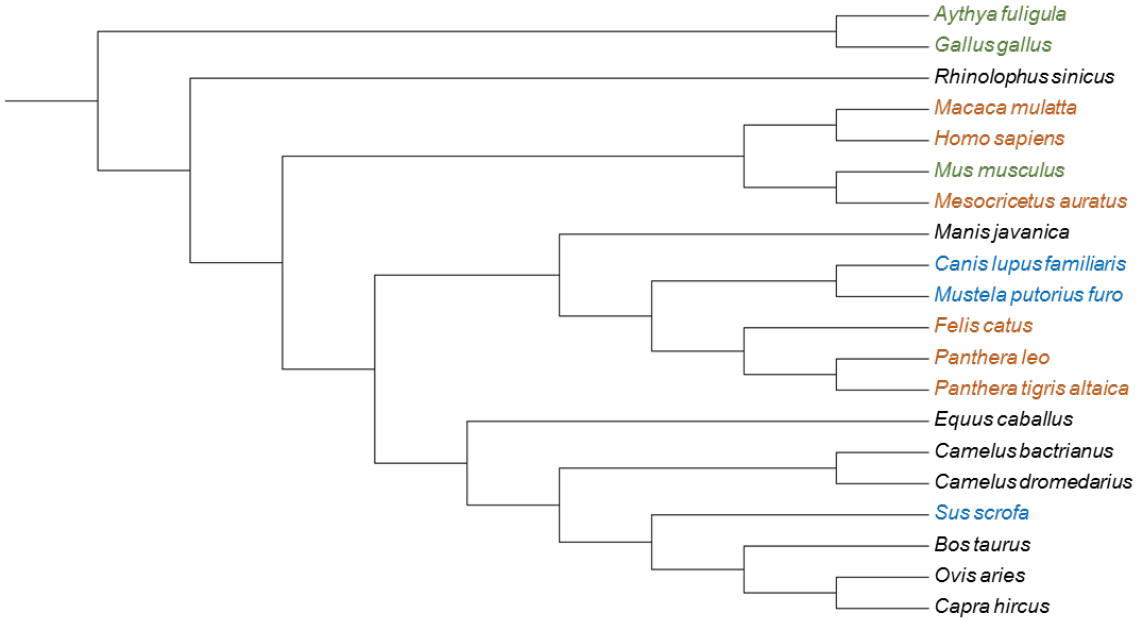


780

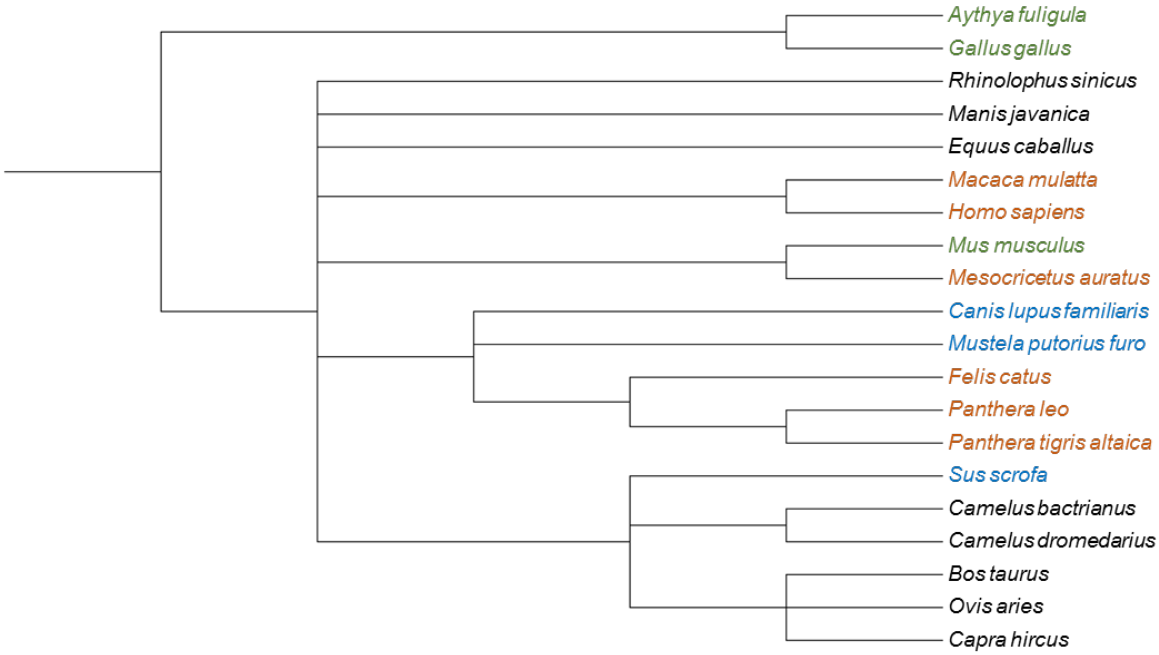
781 **Extended Data Figure 1**: ACE2 protein sequence alignment from susceptible (orange), non-
782 susceptible (green), intermediate susceptible (blue), and unknown susceptible (black) species.
783 MAFFT alignment with visualization using ESPript. Secondary structure elements defined based
784 on human ACE2 (PDB 1r42). Spirals represent α or 3_{10} helices and arrows represent beta
785 strands.

786

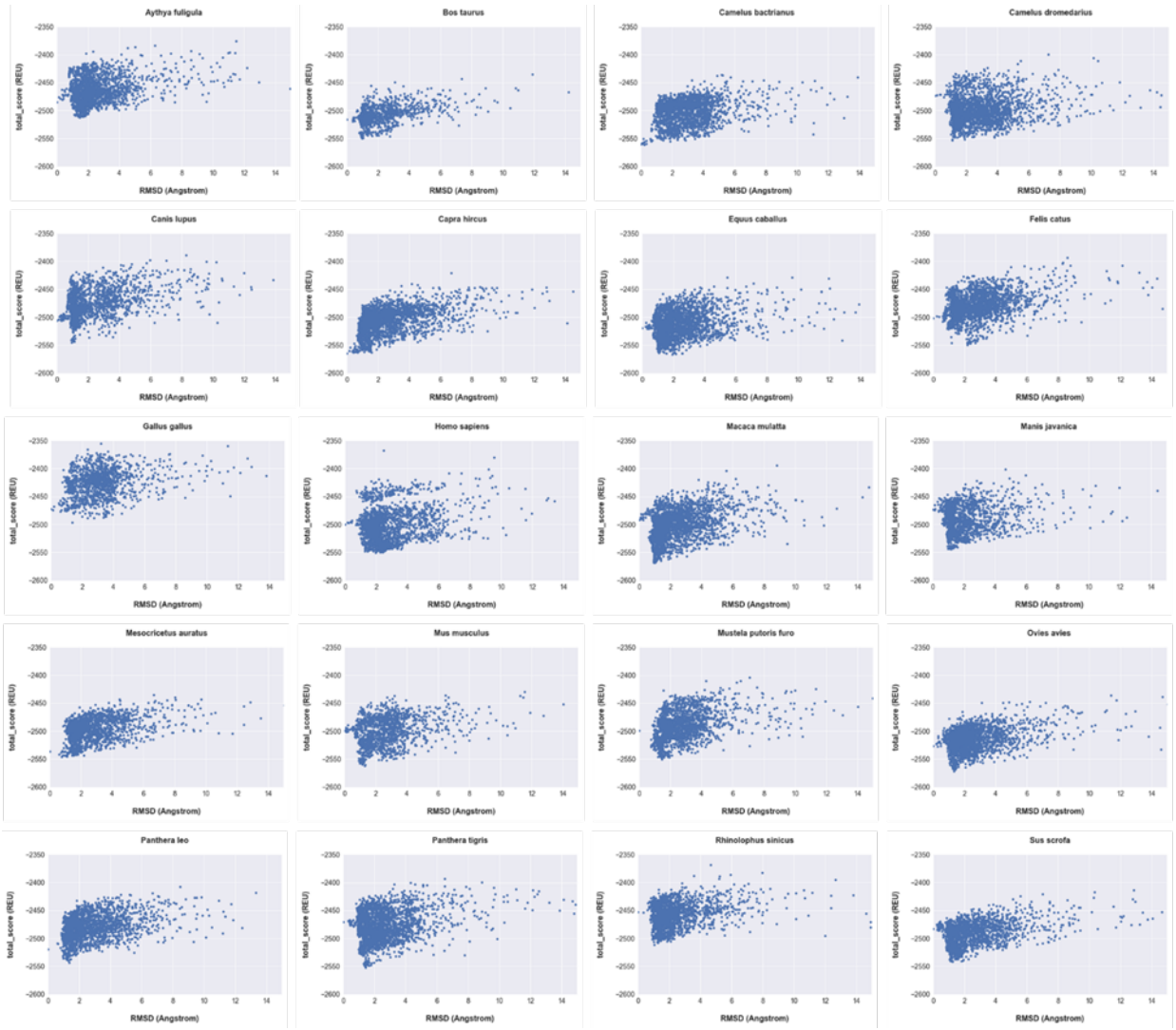
A



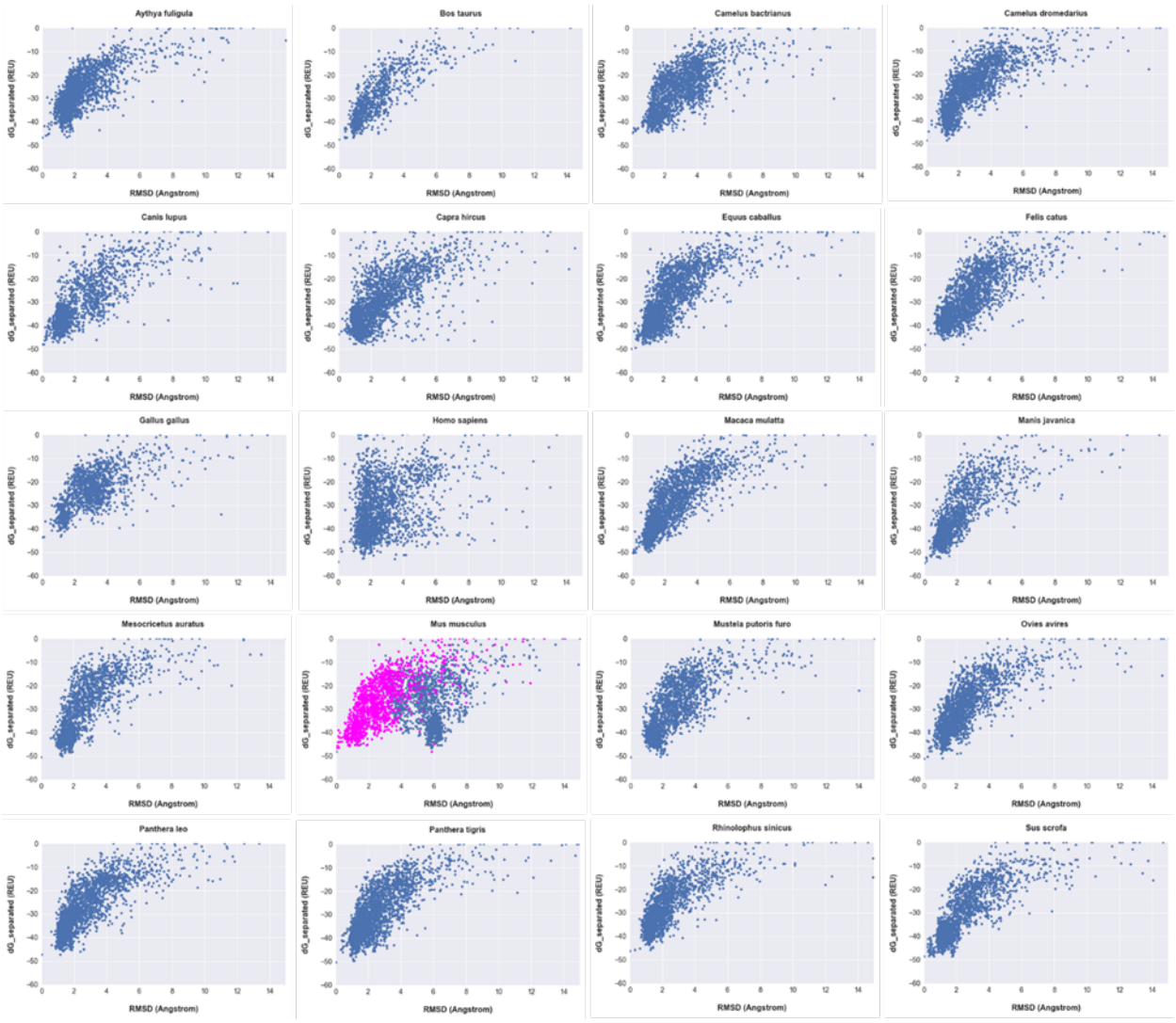
B



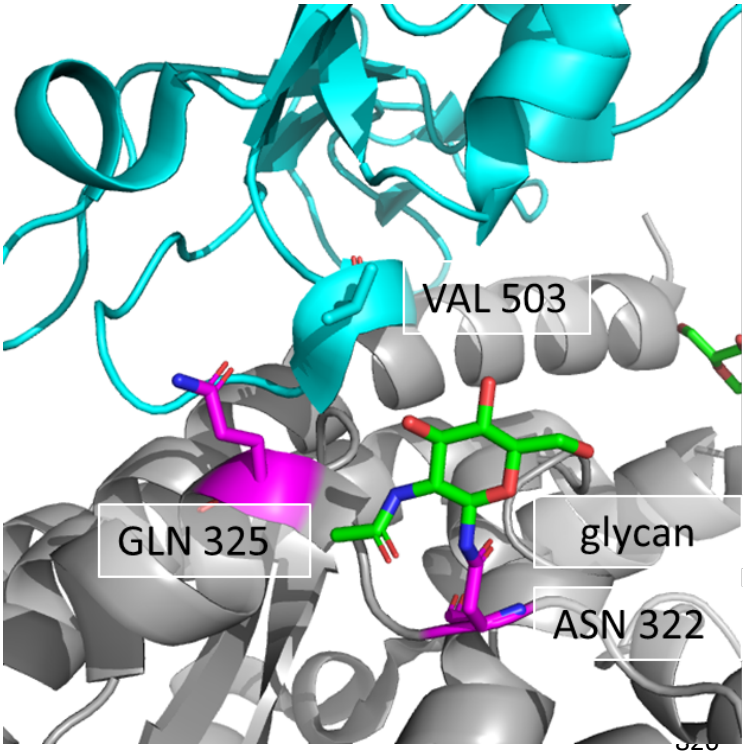
Extended Data Figure 2: **Differences in ACE2 protein sequences and phylogenetic relationships are similar across species.** (A) Dendrogram of ACE2 protein sequence comparisons and (B) phylogenetic relationships of susceptible (orange), non-susceptible (green), intermediate susceptibility (blue), and unknown susceptibility (black) to SARS-CoV-2 infection.



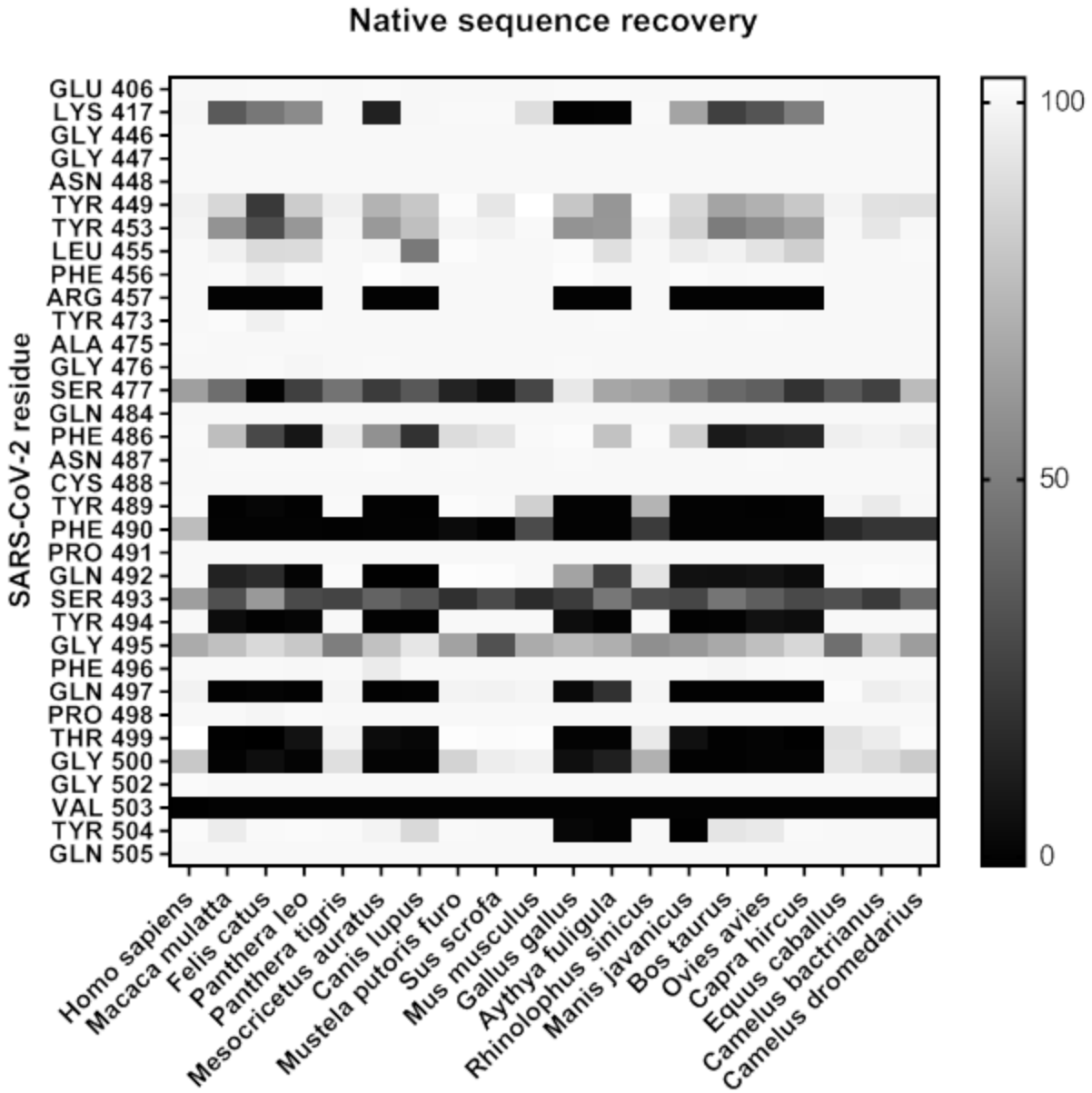
Extended Data Figure 3: Construction of SARS-CoV-2-RBD-ACE2 complex models for each species resulted in comparable high-quality models. Models were evaluated for their Cα-root mean square deviation (Cα-RMSD) as a measure of similarity to the best performing model by predicted binding energy versus calculated protein stability. Models in the lower left quadrant of the plots show good convergence of calculated protein stability and similarity, and were thus selected for re-docking of SARS-CoV-2-RBD to the respective ACE2 as in Extended Data Figure 4.



Extended Data Figure 4: Re-docking of SARS-CoV-2-RBD to ACE2 of different species resulted in high-quality models. Cα-root mean square deviation (Cα-RMSD) were calculated against the best performing model and plotted versus predicted binding energy (dG_separated) after redocking of the SARS-CoV-2-RBD for all SARS-CoV-2-RBD-ACE2 co-complexes. This measure describes the similarity of the models compared to their predicted binding energy. Models from the lowest left corner represent the highest quality models and where chosen for further analysis. The models for *Mus musculus* were recalculated to the second-best model (magenta), as they did not converge on the best model.



Extended Data Figure 5: Proximity of Val503 of the SARS-CoV-2 RBD to an ACE2 glycan at Asn322. At the interface of SARS-CoV-2-RBD (cyan) and ACE2 (grey), Val503 is in close proximity to the *N*-acetylglucosamine (green) at Asn322 as seen in PDB: 6lzg.



Extended Data Figure 6: Native sequence recovery as determined by RECON multistate design allowing the mutation and optimization of SARS-CoV-2-RBD interface residues in the presence of the complex with human ACE2. This figure demonstrates all designable residues, including residues that did not change during multistate design or changed for all species in the same way.

Supplemental Methods

R code for calculating susceptibility scores on new species

```
#
# This is a Shiny web application. You can run the application by clicking
# the 'Run App' button above or Ctrl-Enter on each line.
#
install.packages(c("protr", "devtools", "ggplot2", "dplyr", "shiny"), dependencies = TRUE)
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
if (!requireNamespace("Biostrings", quietly = TRUE))
    BiocManager::install("Biostrings")
library(protr)
library(Biostrings)
library(ggplot2)
library(dplyr)
library(shiny)
library("BiocManager")

# Define UI for application that draws a histogram
ui <- fluidPage(
    #Title of page
    titlePanel("SARS-CoV-2 Susceptibility Score Calculator"),
    HTML("<p>This calculator requires an input file in FASTA format of human ACE2 amino acid
sequence (accession: NP_001358344.1) aligned with ACE2 of another species of interest. This
alignment can be generated using the MAFFT alignment tool found <a
href='https://www.ebi.ac.uk/Tools/msa/mafft/'> here.</a>"),
    #Input protein sequence of interest
    fileInput(inputId = "file", label="Upload alignment file", accept = ".fasta"),
    tags$h2("Susceptibilty Score by Species"),
    plotOutput("plot", width = 310, height = 500),
    "*Red shaded area represents susceptible species while blue/grey represents non-susceptible",
    tags$h2("Key Residues Extracted"),
    tableOutput("residues"),
    tags$br(),
    tags$h2("Calculated susceptibility score "),
    tags$h3(textOutput("score")),
    "Compare the above score with that of previously calculated species above"
)


# Define server logic required to draw a histogram
server <- function(input, output) {
    output$residues <- renderTable({
        req(input$file)
        aligned.seq <- readAAMultipleAlignment(input$file$datapath)
        df.dash <- as.data.frame(AAStringSet(aligned.seq))
        human.dash <- as.character(df.dash["NP_001358344.1 angiotensin-converting enzyme 2
precursor [Homo sapiens]",])
        human.dash <- strsplit(human.dash[1], split="")
        aa.positions <- c(30,83,90,322,354)
        residue.df <- as.data.frame(AAStringSet(aligned.seq, aa.positions[1], aa.positions[1]))
        j=1
        h=1
        for (x in 1:nchar(df.dash["NP_001358344.1 angiotensin-converting enzyme 2 precursor [Homo
sapiens]",])) {
            if (j %in% aa.positions){
                residue.df[h] <- cbind(as.data.frame(AAStringSet(aligned.seq, aa.positions[h],
aa.positions[h])))
                h = h+1
                j = j+1
            }
            else if (human.dash[[1]][j] == "-") {
                aa.positions = aa.positions + 1
                j= j+1
            }
            else {j = j + 1}
        }
    return(residue.df)
```

```
    }, colnames = FALSE, rownames = TRUE)

    output$score <- renderText({
        req(input$file)
        aligned.seq <- readAAMultipleAlignment(input$file$datapath)
        df.dash <- as.data.frame(AAStringSet(aligned.seq))
        human.dash <- as.character(df.dash["NP_001358344.1 angiotensin-converting enzyme 2
precursor [Homo sapiens]",])
        human.dash <- strsplit(human.dash[1], split="")
        aa.positions <- c(30,83,90,322,354)
        residue.df <- as.data.frame(AAStringSet(aligned.seq, aa.positions[1], aa.positions[1]))
        j=1
        h=1
        for (x in 1:nchar(df.dash["NP_001358344.1 angiotensin-converting enzyme 2 precursor [Homo
sapiens]",])) {
            if (j %in% aa.positions){
                residue.df[h] <- cbind(as.data.frame(AAStringSet(aligned.seq, aa.positions[h],
aa.positions[h])))
                h = h+1
                j = j+1
            }
            else if (human.dash[[1]][j] == "-") {
                aa.positions = aa.positions + 1
                j= j+1
            }
            else {j = j + 1}
        }
        seq1 <- as.data.frame(residue.df[1,])
        seq2 <- as.data.frame(residue.df[2,])
        blosum.matrix <- data.frame(AABLOSUM62)
        i <- 0
        score <- 0
        value <- 0
        sus.score <- for(x in 1:5) {
            i = i+1
            value <- as.numeric(subset(blosum.matrix,
                                       colnames(blosum.matrix) %in% seq1[,i],
                                       rownames(blosum.matrix) %in% seq2[,i]))
            score = score + value
        }
        print(score)
    })


    species <- c("Felis catus", "Panthera tigris altaica", "Panthera leo",
                 "Mesocriceteus auratus", "Macaca mulatta", "Mus musculus", "Aythya fuligula",
                 "Gallus gallus", "Mustela putoriusfuro", "Sus scrofa", "Canis lupus familiaris",
                 "Rhinolophus sinicus", "Equus caballus", "Bos taurus", "Manis javanica",
                 "Capra hircus", "Ovis aries", "Camelus dromedarius", "Camelus bactrianus")
    species.scores <- c(27,27,27,23,31,11,8,8,14,21,22,31,27,19,13,19,19,27,27)
    df.species.score <- data.frame("Species" = species, "Score" = species.scores)
    df.species.score <- df.species.score[order(df.species.score$Species),]
    # We also add shading to represent cutoffs for susceptible and non-susceptible species
    output$plot <- renderPlot(ggplot(df.species.score, aes(y=Score,x='')) +
                                  geom_jitter(aes(color=Species, shape=Species),  size = 2.5,
width=0.5) +

scale_shape_manual(values=c(15,16,17,18,15,16,17,18,15,16,17,18,15,16,17,18,15,16,17),
name="Species", labels=df.species.score$Species) +
                                  scale_color_discrete(name="Species")+
                                  geom_rect(aes(xmin = 0, xmax = Inf, ymin = 23, ymax = Inf,
fill="blue"), alpha = 0.008, show.legend = F) +
                                  geom_rect(aes(xmin = 0, xmax = Inf, ymin = -Inf, ymax = 11,
fill="red"), alpha = 0.005, show.legend = F)+
                                  theme_classic() +
                                  theme(axis.title.y = element_text(size=14),
                                        axis.title.x = element_blank(),
                                        legend.title = element_text(size=18),
                                        legend.text = element_text(size=12))
                              )
}
```

```
# Run the application
shinyApp(ui = ui, server = server)
```

Homology modeling of ACE2 based on the ACE2-SARS-CoV-2-RBD co-crystal structure using RosettaCM

Structure and input preparation

For all modeling purposes Rosetta-3.12 was used.

Preparation of input structures using RosettaRelax:

```
../rosetta-3.12/main/source/bin/relax.default.linuxgccrelease -s 6m0j.pdb -
database ../rosetta-3.12/main/database/ -constrain_relax_to_start_coords -
out:prefix relax_ -nstruct 25 -ex1 -ex2 -use_input_sc -ignore_unrecognized_res
```

Command used for partial thread:

```
../rosetta-3.12/main/source/bin/partial_thread.default.linuxgccrelease -
in:file:fasta felis_RBD_02.fasta -in:file:alignment
human_felis_align_for_thread_with_RBD_02.txt -database ../rosetta-
3.12/main/database -in:file:template_pdb ./threads/6lzg_thread_0001.pdb
```

Construction of the initial ACE2-SARS-CoV-2-RBD complex with RosettaCM

Command used for executing RosettaCM

```
../rosetta-3.12/main/source/bin/rosetta_scripts.default.linuxgccrelease
@hybridize_RBD.options -$arrayfile -nstruct 11 -out:prefix CM_ -database
../rosetta-3.12/main/database/ -out:path:all ./output/ >& logfile.log
```

RosettaScripts protocol for RosettaCM

```
<ROSETTASCRIPTS>
    <TASKOPERATIONS>
    </TASKOPERATIONS>
    <SCOREFXNS>
        <ScoreFunction name="stage1" weights="stage1.wts" symmetric="0">
            <Reweight scoretype="atom_pair_constraint" weight="1"/>
        </ScoreFunction>
        <ScoreFunction name="stage2" weights="stage2.wts" symmetric="0">
            <Reweight scoretype="atom_pair_constraint" weight="0.5"/>
        </ScoreFunction>
        <ScoreFunction name="fullatom" weights="stage3.wts" symmetric="0">
            <Reweight scoretype="atom_pair_constraint" weight="0.5"/>
        </ScoreFunction>
        <ScoreFunction name="ref2015" weights="ref2015_cart.wts" >
        </ScoreFunction>
    </SCOREFXNS>
```

```
    <FILTERS>
    </FILTERS>
    <MOVERS>
    <Hybridize name="hybridize" stage1_scorefxn="stage1"
stage2_scorefxn="stage2" fa_scorefxn="fullatom" batch="1"
stage1_increase_cycles="1.0" stage2_increase_cycles="1.0" linmin_only="1"
disulf_file="disulfide.txt">
            Fragments 3mers="1u19_3.frags" 9mers="1u19_9.frags"/>
            <Template pdb="RBD_6m0j.pdb" cst_file="AUTO" weight="1.000" />
        <DetailedControls start_res="599" stop_res="792" sample_template="1"
sample_abinitio="0"/>
    </Hybridize>
    </MOVERS>
    <APPLY_TO_POSE>
    </APPLY_TO_POSE>
    <PROTOCOLS>
        <Add mover="hybridize"/>
    </PROTOCOLS>
</ROSETTASCRIPTS>
```

RosettaCM options:

```
#i/o
-in:file:fasta felis_RBD_02.fasta
-parser:protocol hybridze_RBD_02.xml

# relax options
-relax:minimize_bond_angles
-relax:minimize_bond_lengths
-relax:jump_move true
-default_max_cycles 200
-relax:min_type lbfgs_armijo_nonmonotone
-relax:jump_move true
-score:weights stage3.wts
-use_bicubic_interpolation
-hybridize:stage1_probability 1.0

# reduce memory footprint
-chemical:exclude_patches LowerDNA  UpperDNA Cterm_amidation SpecialRotamer
VirtualBB ShoveBB VirtualDNAPhosphate VirtualNTerm CTermConnect sc_orbitals
pro_hydroxylated_case1 pro_hydroxylated_case2 ser_phosphorylated
thr_phosphorylated  tyr_phosphorylated tyr_sulfated lys_dimethylated
lys_monomethylated  lys_trimethylated lys_acetylated glu_carboxylated
cys_acetylated tyr_diiodinated N_acetylated C_methylamidated
MethylatedProteinCterm
```

Stage1 weights:

```
env     1.0
pair    1.0
cbeta   1.0
cenpack 1.0
hs_pair 2.0
```

```
ss_pair 2.0
rsigma  2.0
sheet   2.0
vdw     0.2
rg      2.0
rama    0.3
linear_chainbreak    2.0
atom_pair_constraint 1.0
```

Stage2 weights

```
# stage2 weights for hybridization
hbond_sr_bb 2.0
hbond_lr_bb 2.0
rama        0.2
omega       0.2
rg          2.0
vdw         1.0
cen_env_smooth  2.0
cen_pair_smooth 1.0
cbeta_smooth    1.0
cenpack_smooth  1.0
cart_bonded     0.05
atom_pair_constraint 0.5
```

Stage3 weights

```
# stage3 fullatom weights for hybridization
METHOD_WEIGHTS ref  0.16 1.7 -0.67 -0.81 0.63 -0.17 0.56 0.24 -0.65 -0.1 -
0.34 -0.89 0.02 -0.97 -0.98 -0.37 -0.27 0.29 0.91 0.51
fa_atr  0.8
fa_rep  0.44
fa_sol  0.65
fa_intra_rep 0.004
fa_pair 0.49
#fa_plane 0
fa_dun  0.56
ref     1
hbond_lr_bb 1.17
hbond_sr_bb 0.585
hbond_bb_sc 1.17
hbond_sc    1.1
p_aa_pp     0.32
dslf_ss_dst 0.5
dslf_cs_ang 2
dslf_ss_dih 5
dslf_ca_dih 5
pro_close   1.0
rama    0.2
omega   0.5
atom_pair_constraint    0.5
```

```
coordinate_constraint   0.0
cart_bonded     0.5
```

RosettaRelax of ACE2-SARS-CoV-2 complex

Command for RosettaRelax

```
../rosetta-3.12/main/source/bin/relax.default.linuxgccrelease -l $arrayfile -
nstruct 1 -out:prefix relax_ -database ../rosetta-3.12/main/database/ -
out:path:all ./relax/ -ex1 -ex2 -constrain_relax_to_start_coords >&
logfile.log
```

Docking of ACE2 and SARS-CoV-2 RBD

Command for RosettaDock

```
../rosetta-3.12/main/source/bin/rosetta_scripts.default.linuxgccrelease -s
renum_relax_14_S_0003_0001.pdb -nstruct 5 -parser:protocol docking_full.xml -
in:file:native ./RBD_template/renum_relax_6m0j_0020.pdb -out:prefix dock_ -
database ../rosetta-3.12/main/database/ -out:path:all ./output/
@docking.options -out:prefix retest_
```

Options for protein-protein docking:

```
-docking                          # the docking option group
    -partners A_B                     # set rigid body docking partners
    -dock_pert 1 2          # set coarse perturbation parameters (degrees and
angstroms)
    -dock_mcm_trans_magnitude 0.01     # refinement translational
perturbation
    -dock_mcm_rot_magnitude 1.0     # refinement rotational perturbation
-run:max_retry_job 10          # if the mover fails, retry 50 times
-use_input_sc                         # add the side chains from the input pdb
to the rotamer library
-ex1                              # increase rotamer bins to include mean +- 1
standard deviation
-ex2                              # increase rotamer bins to include mean +- 2
standard deviations
```

RosettaScripts xml-protocol for protein-protein docking

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
         <ScoreFunction name="ref2015" weights="ref2015.wts" symmetric="0">
      </ScoreFunction>
    </SCOREFXNS>
    <TASKOPERATIONS>
        <InitializeFromCommandline name="ifcl"/>
        <RestrictToRepacking name="rtr" />
```

```
        Restrict to residues within a distance and vector cutoff of the
protein-protein interface
        <RestrictToInterfaceVector name="rtiv" chain1_num="1" chain2_num="2"
CB_dist_cutoff="10.0" nearby_atom_cutoff="5.5" vector_angle_cutoff="75"
vector_dist_cutoff="9.0" />
        Fix residues known experimentally to be critical in interaction
        PreventResiduesFromRepacking name="prfrp" residues="11,41,345" />
    </TASKOPERATIONS>
    <FILTERS>
    </FILTERS>
    <MOVERS>
        MINIMIZATION MOVERS
        Single cycle of FastRelax to minimize backbone of docking partners
        <FastRelax name="minimize_interface" scorefxn="ref2015" repeats="1"
task_operations="ifcl,rtr,rtiv" />

        DOCKING MOVERS
        <Docking name="dock_low" score_low="score_docking_low"
score_high="ref2015" fullatom="0" local_refine="0" optimize_fold_tree="1"
conserve_foldtree="0" ignore_default_docking_task="0" design="0"
task_operations="ifcl" jumps="1"/>
        <Docking name="dock_high" score_low="score_docking_low"
score_high="ref2015" fullatom="1" local_refine="1" optimize_fold_tree="1"
conserve_foldtree="0" design="0" task_operations="ifcl" jumps="1"/>

        <SaveAndRetrieveSidechains name="srsc" allsc="0" /> Speeds the move
from centroid to full atom mode
        <InterfaceAnalyzerMover name="interface" scorefxn="ref2015"
pack_separated="1" pack_input="0" packstat="0" interface="A_B"/>

    </MOVERS>
    <APPLY_TO_POSE>
    </APPLY_TO_POSE>
    <PROTOCOLS>
        Run docking protocol
        <Add mover="dock_low"/>
        <Add mover="srsc" />
        <Add mover="dock_high" />

        Minimize interface
        <Add mover="minimize_interface" />
        <Add mover="interface"/>
    </PROTOCOLS>
    <OUTPUT scorefxn="ref2015" />
</ROSETTASCRIPTS>
```

Options for Rosetta protein-protein docking

```
-docking                          # the docking option group
    -partners A_B                    # set rigid body docking partners
    -dock_pert 1 2          # set coarse perturbation parameters (degrees and
angstroms)
    -dock_mcm_trans_magnitude 0.01     # refinement translational
perturbation
```

```
    -dock_mcm_rot_magnitude 1.0     # refinement rotational perturbation
-run:max_retry_job 10          # if the mover fails, retry 50 times
-use_input_sc                       # add the side chains from the input pdb
to the rotamer library
-ex1                             # increase rotamer bins to include mean +- 1
standard deviation
-ex2                             # increase rotamer bins to include mean +- 2
standard deviations
```

To obtain a control, the relaxed crystal structures were subjected to interface minimization.

Calculation of native sequence recovery using the RECON multistate design protocol in Rosetta

Command for running RECON multistate design in Rosetta:

```
../rosetta-3.12/main/source/bin/recon.default.linuxgccrelease -s homo.pdb
panthera.pdb -nstruct 50 -out:prefix "$SLURM_ARRAY_TASK_ID"_ -database
../rosetta-3.12/main/database/ -out:path:all ./panthera/output/ -
parser:protocol multistate_design.xml -ex1 -use_input_sc >& logfile.log
```

RosettaScripts protocol for RECON multistate design:

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <ScoreFunction name="ref2015_cst" weights="ref2015_cst.wts" >
            <Reweight scoretype="res_type_constraint" weight="1.0" />
        </ScoreFunction>
        <ScoreFunction name="ref2015" weights="ref2015.wts" >
        </ScoreFunction>
    </SCOREFXNS>
    <TASKOPERATIONS>
        Include rotamer options from the command line
        <InitializeFromCommandline name="ifcl" />
    </TASKOPERATIONS>
    <MOVERS>
        Design mover to be used in multistate design
        <PackRotamersMover name="design" scorefxn="ref2015_cst"
task_operations="ifcl" />

        Create MSDMovers to run multistate design - these different in the
constraint weight, with later rounds
        having a higher constraint value
        <MSDMover name="msd1" design_mover="design" constraint_weight="0.5"
resfiles="consensus.resfile,consensus.resfile" />
        <MSDMover name="msd2" design_mover="design" constraint_weight="1"
resfiles="consensus.resfile,consensus.resfile"/>
        <MSDMover name="msd3" design_mover="design" constraint_weight="1.5"
resfiles="consensus.resfile,consensus.resfile" />
        <MSDMover name="msd4" design_mover="design" constraint_weight="2"
resfiles="consensus.resfile,consensus.resfile" />

        FindConsensusSequence is needed at the end of the protocol to find a
single sequence
```

```
        that agrees with all target states
        <FindConsensusSequence name="finish" scorefxn="ref2015"
resfiles="consensus.resfile,consensus.resfile" />

        Analyze the resulting interface
        <InterfaceAnalyzerMover name="analyze" scorefxn="ref2015" packstat="0"
pack_input="0" pack_separated="1" fixedchains="A" />
    </MOVERS>
    <FILTERS>
    </FILTERS>
    <APPLY_TO_POSE>
    </APPLY_TO_POSE>
    <PROTOCOLS>
        Run four rounds of design
        <Add mover="msd1" />

        <Add mover="msd2" />

        <Add mover="msd3" />

        <Add mover="msd4" />

        Find a consensus sequence for all states
        <Add mover="finish" />

        Calculate interface metrics for the final sequence
        <Add mover="analyze" />

    </PROTOCOLS>
    <OUTPUT scorefxn="ref2015" />
</ROSETTASCRIPTS>
```

Calculation of native sequence recovery:

```
./design_analysis.py --native ../../../output/analysis/nativ.pdb --format eps
--resfile ../../../consensus.resfile --multiproc *pdb
./calc_nat_seq_recovery.py --native ../../../output/analysis/nativ.pdb --res
../../../consensus.resfile out.tab
```