HHS Public Access

Author manuscript

FASEB J. Author manuscript; available in PMC 2021 December 01.

Published in final edited form as:

FASEB J. 2020 December ; 34(12): 15946–15960. doi:10.1096/fj.202001808R.

Predicting susceptibility to SARS-CoV-2 infection based on structural differences in ACE2 across species

Matthew R. Alexander[1,2,*], **Clara T. Schoeder**[3,4,*], **Jacquelyn A. Brown**[5,6], **Charles D. Smart**[7], **Chris Moth**[3,8], **John P. Wikswo**[5,6,7,9], **John A. Capra**[3,8], **Jens Meiler**[3,4,9,10,†], **Wenbiao Chen**[7,†], **Meena S. Madhur**[1,2,7,11,†]

[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, Nashville, TN, USA

Abstract

Correspondence Wenbiao Chen, Department of Molecular, Physiology and Biophysics, Vanderbilt University, 735B Light Hall, 2215 Garland Ave, Nashville, TN 37232, USA. wenbiao.chen@vanderbilt.edu, Meena S. Madhur, Department of Medicine, Vanderbilt University, 2215 Garland Avenue, P415D MRB IV, Nashville, TN 37232, USA. meena.madhur@vanderbilt.edu.

†Jens Meiler, Wenbiao Chen, and Meena S. Madhur contributed equally to this work.

*Matthew R. Alexander and Clara T. Schoeder contributed equally to this work.

AUTHOR CONTRIBUTIONS

J. A. Brown and J. P. Wikswo conceptualized the study; M. R. Alexander, C. T. Schoeder, W. Chen, M. S. Madhur, J. Meiler, J. P. Wikswo, J. A. Brown, C. Moth, and J. A. Capra made substantial contributions to experimental design; M. R. Alexander and C. T. Schoeder conducted the majority of the experiments with help from W. Chen, J. Meiler, C. D. Smart, J. A. Capra, J. A. Brown, and J. P. Wikswo; M. R. Alexander, C. T. Schoeder, J. A. Capra, W. Chen, C. Moth, and J. Meiler significantly contributed to the data acquisition and interpretation; M. R. Alexander and C. T. Schoeder drafted the manuscript with W. Chen, M. S. Madhur, J. Meiler, J. P. Wikswo, and J. A. Capra contributing critical revisions. All authors approved the final version.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest in connection with this article.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of the global pandemic of coronavirus disease-2019 (COVID-19). SARS-CoV-2 is a zoonotic disease, 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. Certain species, such as domestic cats and tigers, are susceptible to SARSCoV-2 infection, while other species such as mice and chickens are not. Most animal species, including those in close contact with humans, have unknown susceptibility. Hence, methods to predict the infection risk of animal species 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 integrate species differences in susceptibility with multiple in-depth structural analyses to identify key ACE2 amino acid positions including 30, 83, 90, 322, and 354 that distinguish susceptible from resistant species. Using differences in these residues across species, we developed a susceptibility score that predicts an elevated risk of SARS-CoV-2 infection for multiple species including horses and camels. 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 transmissibility of this virus among humans. Taken together, our findings define potential ACE2 and SARS-CoV-2 residues for therapeutic targeting and identification of animal species on which to focus research and protection measures for environmental and public health.

Keywords

angiotensin-converting enzyme 2; COVID-19; protein structural elements; severe acute respiratory syndrome coronavirus 2

1 | INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARSCoV-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, tigers and lions are susceptible as evidenced by the presence of respiratory symptoms and PCR confirmation of SARS-CoV-2 infection.[3,4] Golden Syrian hamsters, house cats, and rhesus macaques can also be experimentally infected by SARS-CoV-2 and develop COVID-19 pathologies including respiratory symptoms and alveolar barrier dysfunction in the lung.[5–8] 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.[7,9–11] Interestingly, however, susceptibility is not dichotomous. Although ferrets are susceptible to infection, intranasal or intratracheal inoculation resulted in either no or low levels of viral RNA in the lower respiratory tract along with limited clinical symptoms and no alveolar/capillary barrier dysfunction in the lung, as opposed to that observed with rhesus macaques, house cats, and Syrian hamsters.[7,12,13] In addition, although dogs failed to exhibit the 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 the

production of SARS-CoV-2-specific antibodies.[7,14] While pigs have not demonstrated evidence of infection after intranasal inoculation, overexpression of swine ACE2 in cultured cells supports some degree of viral entry.[7,9,15] 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.[16] Following spike protein cleavage, the fusion of the viral and host cell membranes occurs to enable viral entry into the cell.[17] 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.[9,11]

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.[18,19] Understanding such differences could provide insight into key structural interactions between ACE2 and SARS-CoV-2 RBD important for infection, and permit the 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 SARSCoV-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 among humans. Our findings have important implications for the 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.

2 | MATERIALS AND METHODS

2.1 | ACE2 protein alignment

Protein sequence accession numbers and corresponding FASTA files from multiple species (Table S1) were pulled from NCBI using Batch Entrez. In the absence of a published sequence and accession number, the 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 (Table S2). 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] The resulting alignment was uploaded to ESPript 3.0 to generate a graphical version of the alignment (http://espript.ibcp.fr/ESPript/ESPript/), including the annotation of the secondary structure based on the Protein Data Bank (PDB) structure 1r42 of human ACE2.[21] A TreeDyn format tree diagram representing the similarity of ACE2 protein sequence across species was generated using phylogeny.fr (https://www.phylogeny.fr/).[22,23] NCBI Taxonomy Browser was used to generate a taxonomic tree of phylogenetic relationships among species as a Phylogeny Inference Package (PHYLIP) tree.[24] Final visualization was performed using the interactive Tree of Life (iTOL) tree viewer v 5.5.1 (https://itol.embl.de/).[25]

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

Quantification of amino acid positions in the ACE2 protein alignment that optimally distinguish susceptible vs non-susceptible species was performed using GroupSim.[26] 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 the similarity of amino acids properties.[27]

2.3 | Homology modeling of ACE2-SARSCoV-2 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).[16] 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-SARSCoV-2-RBD co-crystal structure, which was first relaxed with backbone constraints using RosettaRelax.[28] A total of 1000 homology models were constructed using RosettaCM, and subsequently energetically relaxed with backbone constraints.[28,29] 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 (Figure S1). The SARS-CoV2-RBD-ACE2 complex was optimized using a rigid-body docking with limited degrees for rotational and torsional sampling.[30,31] A final ensemble of 100 models was selected based on the total energy as the measure of protein stability, predicted binding energy, and Cα-RMSD to the best scoring model (Figure S2). 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.

2.4 | 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-2RBD-ACE2 complex.[32–34] 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 (Figure S3).

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,[35–37] 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.

2.5 | Prediction of glycosylation sites

The NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) was used to predict glycosylation sites.[38] Based on 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.

2.6 | 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.[27] 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 the 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 and the graphic user interface implementation is available at https://meilerlab.shinyapps.io/RShinyApps/.

2.7 | Statistical analysis

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

3 | RESULTS

3.1 | Susceptibility does not segregate according to phylogeny and ACE2 sequence similarity

Given experimental evidence for the 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,[3–5,7,9–11,39,40] we performed protein sequence alignment of ACE2 from these organisms using MAFFT (Figure S4).[20] We

also included species with intermediate susceptibility, including dogs, pigs, and ferrets,[7,9,13,14] 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 (Figure S5). 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*).[9,11] 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.[9,11] 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.

3.2 | 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 vs 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.[26] 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 (Table S3). 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 (Figure 2).[5,41-43] 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; $P < .0001$; Fisher's exact test). Such key residues based on the 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.

3.3 | 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).[16] We modeled ACE2 in the presence of the

SARS-CoV-2 RBD to allow backbone adjustment to the binder and refined by the 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 [CαRMSD], Figure S1 and S2). Based on these models, multiple approaches were 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 lowest energy models for each species and evaluated the evidence for the 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 3). 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.

3.4 | 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 4). 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 environments available in the non-susceptible species chicken, duck, and mouse as compared to susceptible species, including humans (Figure 5). This analysis also identifies residue 83 of ACE2 as having differential energetic interactions across species. Residue 83 is a tyrosine in susceptible species and phenylalanine in non-susceptible species (Figure 2). 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 4). 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 6A). 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 (Table S3). 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.

3.5 | Multistate design reveals ACE2 G354 as a 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. This method determines how many mutations one protein requires to acquire an affinity for multiple targets at once.[32,33]

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 in 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 7A).

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 in 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 7B). 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 (Figure S6). 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 (Figure 2). Tyrosine 505 also interacts less strongly with glycine 354, which is occupied by asparagine in the avian species (chicken and duck) (Figure 2 and Figure 6B). 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.

3.6 | 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[44]; 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*-acetylglucosamine 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 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.[38] 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)[16] (Figure 8). 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.

3.7 | 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 (Figure 9).[27] 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 the 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 the implementation of this algorithm as a graphical user interface is available in Supplemental Methods and the graphical user interface is available online at https://meilerlab.shinyapps.io/RShinyApps/.

4 | 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 the identification of key residues for therapeutic targeting and determining the 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 in the SARS-CoV-2 and ACE2 interface.[5,45–51] 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 SARSCoV-2 RBD. Prior studies of ACE2 sequence alignment with limited

structural modeling have suggested that pigs are susceptible to infection,[52] and that hamsters and house cats are in an intermediate-risk group.[53] Recent experimental work with direct inoculation, however, has demonstrated that pigs are non-susceptible,[7] 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 the 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 the identification of potential animal reservoirs that can propagate transmission.[54] 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 transmissibility,[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 the determination of their susceptibility to 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[29] 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 the 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 virus among humans. This finding is also consistent with recent results that compared SARS-CoV and SARS-CoV-2 and determined that a number of differences in the SARS-CoV-2 RBD have made it a much more potent binder to human ACE2 through the introduction of numerous hydrogen bonding and hydrophobic networks.[55] In addition, although several mutations in the SARS-CoV-2 spike protein have been identified, the vast majority of these mutations are outside of the receptor-binding domain (RBD) that interfaces with ACE2.[56–59] The potential significance of these mutations remains under investigation, though few, including the now dominant D614G mutation, have shown a clear association with increased severity of illness.[58] Our findings that the receptor-binding domain of SARS-CoV-2 spike protein is nearly optimal for binding human ACE2 suggests that mutations at this site are unlikely to improve binding, and thus these strains may be negatively selected from the virus populations infecting humans. One rare mutation, G476S, has been identified within the RBD of the spike protein.[59] However, this residue is found at the edge of the binding interface, and in our RECON multistate analysis exhibited little variance in binding ACE2 across species, indicating that this mutation is unlikely to differentially alter binding to ACE2 across the breadth of species tested. Thus, although continued surveillance for additional mutations is warranted, currently known mutations do not appear to have altered the RBD of the spike protein in a manner expected to alter ACE2 binding or cross-species susceptibility.

Although ACE2 and SARS-CoV-2 RBD interactions are critical to SARS-CoV-2 infection,[9–11] differences in other factors across species may also contribute to differences in susceptibility. This includes differences in ACE2 expression levels[60] and differences in the

protein sequence of TMPRSS2, a protein that contributes to viral and host cell membrane fusion through cleavage of spike protein.[17,61] With further experimental and observational data on the infectability of currently unknown species, the susceptibility score we have developed can also help determine species for which differences in ACE2 protein may not inadequately predict differences in susceptibility. For these species, future studies could compare differences in expression levels of ACE2 and/ or differences in TMPRSS2 structure. These structural comparisons of TMPRSS2, however, will require elucidation of the protein crystal structure, which is not yet available.

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

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

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.

Funding information

National Institutes of Health, Grant/Award Number: F32HL144048-01, DK117147, UH3TR002097, U01TR002383, U19AI117905, U01AI150739, R01AI141661, R35GM127087 and DP2HL137166; American Heart Association, Grant/Award Number: 20PRE35080177 and EIA34480023

Abbreviations:

ACE2	angiotensin-converting enzyme 2
Cα-RMSD	Cα-root mean square deviation
COVID-19	coronavirus disease-2019
Itol	interactive tree of life
PDB	protein data bank

PHYLIP	phylogeny inference package
RBM	receptor-binding motif
RBD	receptor-binding domain
RECON	restraint convergence
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2

REFERENCES

1. Adhikari SP, Meng S, Wu YJ, et al. Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. Infect Dis Poverty. 2020;9:29. [PubMed: 32183901]
2. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. CB. 2020;30:1346–1351. e1342. [PubMed: 32197085]
3. APHISpress@usda.gov. (April 6, 2020). 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
4. APHISpress@usda.gov. (August 13, 2020). Confirmed cases of SARS-CoV-2 in animals in the United States. United States Department of Agriculture Animal and Plant Health Inspection Service. https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/SA_One_Health/sars-cov-2-animals-us
5. Chan JF, Zhang AJ, Yuan S, 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. 2020. 10.1093/cid/ciaa325
6. Chandrashekar A, Liu J, Martinot AJ, et al. SARS-CoV-2 infection protects against rechallenge in rhesus macaques. Science. 2020;eabc4776.
7. Shi J, Wen Z, Zhong G, et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. Science. 2020;368:1016–1020. [PubMed: 32269068]
8. Halfmann PJ, Hatta M, Chiba S, et al. Transmission of SARS-CoV-2 in domestic cats. N Engl J Med. 2020;383:592–594. [PubMed: 32402157]
9. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579:270–273. [PubMed: 32015507]
10. Monteil V, Kwon H, Prado P, et al. Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. Cell. 2020;181:905–913.e907. [PubMed: 32333836]
11. Bao L, Deng W, Huang B, et al. The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. Nature. 2020;583:830–833. [PubMed: 32380511]
12. Cleary SJ, Pitchford SC, Amison RT, et al. Animal models of mechanisms of SARS-CoV-2 infection and COVID-19 pathology. Br J Pharmacol. 2020. 10.1111/bph.15143
13. Kim YI, Kim SG, Kim SM, et al. Infection and rapid transmission of SARS-CoV-2 in ferrets. Cell Host Microbe. 2020;27:704–709.e702. [PubMed: 32259477]
14. Sit THC, Brackman CJ, Ip SM, et al. Infection of dogs with SARS-CoV-2. Nature. 2020. 10.1038/s41586-020-2334-5
15. Liu Y, Hu G, Wang Y, et al. Functional and genetic analy-sis of viral receptor ACE2 orthologs reveals a broad potential host range of SARS-CoV-2. bioRxiv. 2020. 10.1101/2020.04.22.046565
16. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;581(7807):215–220. 10.1038/s41586-020-2180-5 [PubMed: 32225176]
17. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020;181:271–280.e278. [PubMed: 32142651]

18. 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. 10.1128/JVI.00127-20
19. Li F. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral Res. 2013;100:246–254. [PubMed: 23994189]
20. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;47:W636–W641. [PubMed: 30976793]
21. Robert X, Gouet P. Deciphering key features in protein struc-tures with the new ENDscript server. Nucleic Acids Res. 2014;42:W320–W324. [PubMed: 24753421]
22. Dereeper A, Guignon V, Blanc G, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008;36:W465–469. [PubMed: 18424797]
23. Dereeper A, Audic S, Claverie JM, Blanc G. Blast-explorer helps you building datasets for phylogenetic analysis. BMC Evol Biol. 2010;10:8. [PubMed: 20067610]
24. Sayers EW, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2009;37:D5–15. [PubMed: 18940862]
25. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47:W256–W259. [PubMed: 30931475]
26. Capra JA, Singh M. Characterization and prediction of resi-dues determining protein functional specificity. Bioinformatics. 2008;24:1473–1480. [PubMed: 18450811]
27. Pearson WR. Selecting the right similarity-scoring matrix. Curr Protoc Bioinformatics. 2013;43(1):1–3. [PubMed: 26270170]
28. Nivon LG, Moretti R, Baker D. A Pareto-optimal refinement method for protein design scaffolds. PLoS One. 2013;8:e59004. 10.1371/journal.pone.0059004
29. Song Y, DiMaio F, Wang RY, et al. High-resolution comparative modeling with RosettaCM. Structure. 2013;21:1735–1742. [PubMed: 24035711]
30. Gray JJ, Moughon S, Wang C, et al. Protein-protein docking with simultaneous optimization of rigid-body displacement and sidechain conformations. J Mol Biol. 2003;331:281–299. [PubMed: 12875852]
31. Chaudhury S, Berrondo M, Weitzner BD, Muthu P, Bergman H, Gray JJ. Benchmarking and analysis of protein docking performance in Rosetta v3.2. PLoS One. 2011;6:e22477.
32. Sevy AM, Jacobs TM, Crowe JE, Meiler J. Design of protein multi-specificity using an independent sequence search reduces the barrier to low energy sequences. PLoS Comput Biol. 2015;11:e1004300.
33. Sevy AM, Wu NC, Gilchuk IM, et al. Multistate design of influenza antibodies improves affinity and breadth against seasonal viruses. Proc Natl Acad Sci USA. 2019;116:1597–1602. [PubMed: 30642961]
34. Sauer MF, Sevy AM, Crowe JE Jr, Meiler J. Multi-state design of flexible proteins predicts sequences optimal for conformational change. PLoS Comput Biol. 2020;16:e1007339.
35. Schrodinger LLC. The AxPyMOL molecular graphics plugin for microsoft powerpoint. Version 1.8. 2015.
36. Schrodinger LLC. The PyMOL molecular graphics system. Version 1.8. 2015.
37. Schrodinger LLC. The JyMOL molecular graphics development component. Version 1.8. 2015.
38. 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. 2004;4:1633–1649. [PubMed: 15174133]
39. Yu P, Qi F, Xu Y, et al. Age-related rhesus macaque models of COVID-19. Animal Models Exp Med. 2020;3:93–97.
40. Mallapaty S. Coronavirus can infect cats—Dogs, not so much. Nature. 2020. 10.1038/d41586-020-00984-8
41. 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. [PubMed: 32132184]
42. Wang Q, Zhang Y, Wu L, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. Cell. 2020;181:894–904.e899. [PubMed: 32275855]

43. Procko E. The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2. bioRxiv. 2020. 10.1101/2020.03.16.994236
44. Ferreira IG, Pucci M, Venturi G, Malagolini N, Chiricolo M, Dall'Olio F. Glycosylation as a main regulator of growth and death factor receptors signaling. Int J Mol Sci. 2018;19:580.
45. Melin AD, Janiak MC, Marrone F, Arora PS, Higham JP. Comparative ACE2 variation and primate COVID-19 risk. bioRxiv. 2020.
46. Zhai X, Sun J, Yan Z, et al. Comparison of severe acute respiratory syndrome coronavirus 2 spike protein binding to ACE2 receptors from human, pets, farm animals, and putative intermediate hosts. J Virol. 2020; 94. 10.1128/JVI.00831-20
47. Luan J, Jin X, Lu Y, Zhang L. SARS-CoV-2 spike protein favors ACE2 from Bovidae and Cricetidae. J Med Virol. 2020. 10.1002/jmv.25817
48. Liu Z, Xiao X, Wei X, et al. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. J Med Virol. 2020;92(6):595–601. 10.1002/jmv.25726 [PubMed: 32100877]
49. Praharaj MR, Garg P, Khan RIN, et al. Prediction analysis of SARS-COV-2 entry in Livestock and Wild animals. bioRxiv. 2020. 10.1101/2020.05.08.084327
50. Lam S, Bordin N, Waman V, et al. SARS-CoV-2 spike protein predicted to form complexes with host receptor protein orthologues from a broad range of mammals. bioRxiv. 2020. 10.1101/2020.05.01.072371
51. Rodrigues JP, Barrera-Vilarmau S, Teixeira JM, Seckel E, Kastritis P, Levitt M. Insights on cross-species transmission of SARSCoV-2 from structural modeling. bioRxiv. 2020. 10.1101/2020.06.05.136861
52. 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:e00127–00120. [PubMed: 31996437]
53. Damas J, Hughes GM, Keough KC, et al. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. Proc Natl Acad Sci. 2020;117(36):22311–22322. 10.1073/pnas.2010146117 [PubMed: 32826334]
54. Wu D, Wu T, Liu Q, Yang Z. The SARS-CoV-2 outbreak: what we know. Int J Infect Dis. 2020;94:44–48. [PubMed: 32171952]
55. Wang Y, Liu M, Gao J. Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. Proc Natl Acad Sci. 2020;117:13967–13974. [PubMed: 32503918]
56. Saha P, Banerjee AK, Tripathi PP, Srivastava AK, Ray U. A virus that has gone viral: amino acid mutation in S protein of Indian isolate of Coronavirus COVID-19 might impact receptor binding, and thus, infectivity. Biosci Rep. 2020;40. 10.1042/BSR20201312
57. Ou J, Zhou Z, Dai R, et al. Emergence of RBD mutations in circulating SARS-CoV-2 strains enhancing the structural stability and human ACE2 receptor affinity of the spike protein. bioRxiv. 2020.
58. Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. Cell. 2020. 10.1016/j.cell.2020.06.043
59. Korber B, Fischer W, Gnanakaran S, et al. Spike mutation pipeline reveals the emergence of a more transmissible form of SARSCoV-2. bioRxiv. 2020. 10.1101/2020.04.29.069054
60. Fu J, Zhou B, Zhang L, et al. Expressions and significances of the angiotensin-converting enzyme 2 gene, the receptor of SARSCoV-2 for COVID-19. Mol Biol Rep. 2020;47:4383–4392. [PubMed: 32410141]
61. Lukassen S, Chua RL, Trefzer T, et al. SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. Embo J. 2020;e105114.

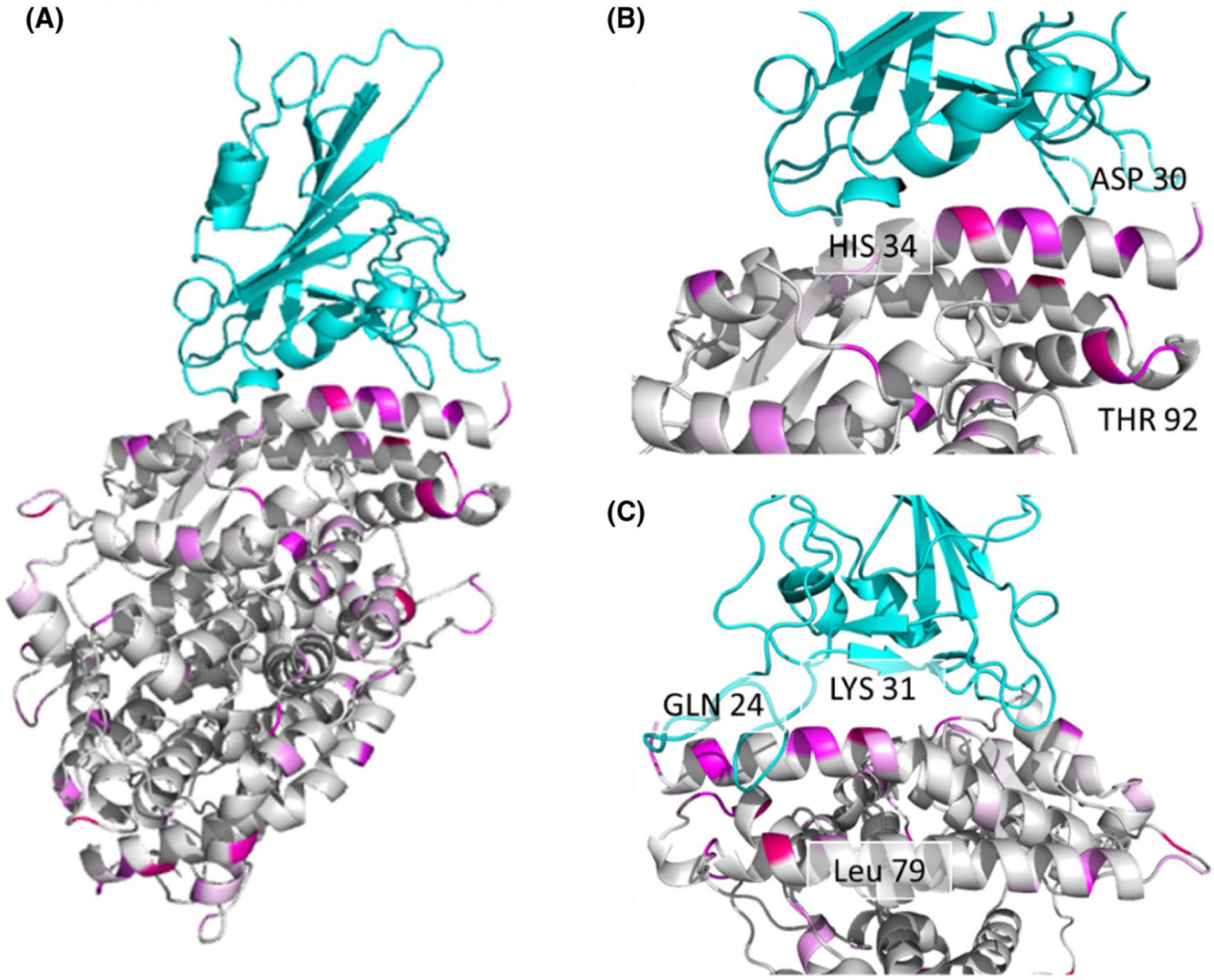

FIGURE 1.

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

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

FIGURE 2.

Twenty-four key residues for SARS-CoV-2 RBD and ACE2 interactions. Highlighted residues are most similar in susceptible and different in non-susceptible species as determined by GroupSim (Table S1). 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

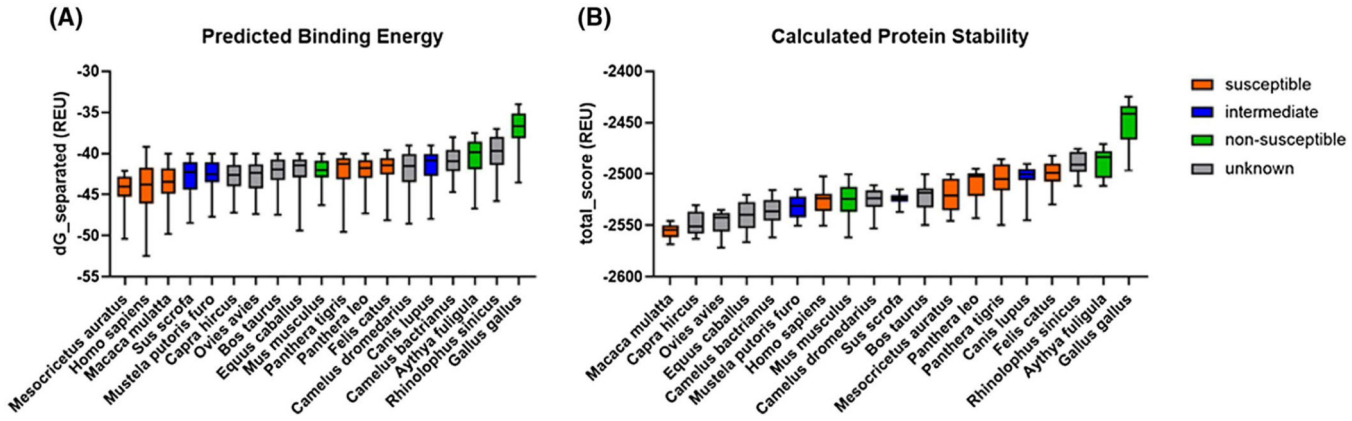

FIGURE 3.

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

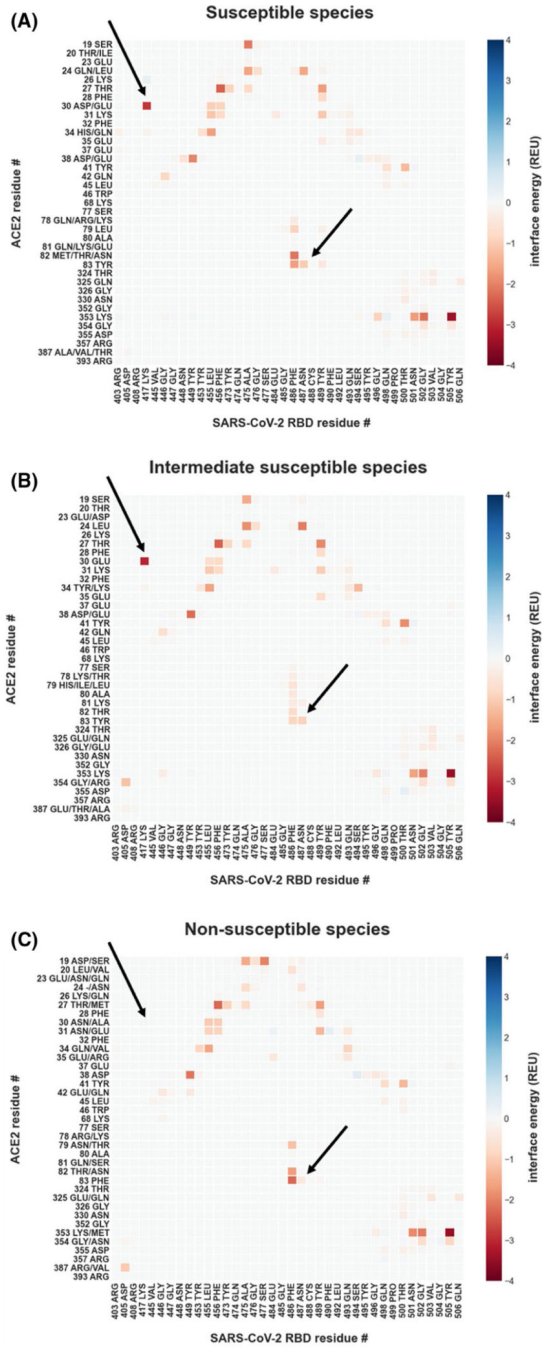

FIGURE 4.

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 SARSCoV-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 the 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

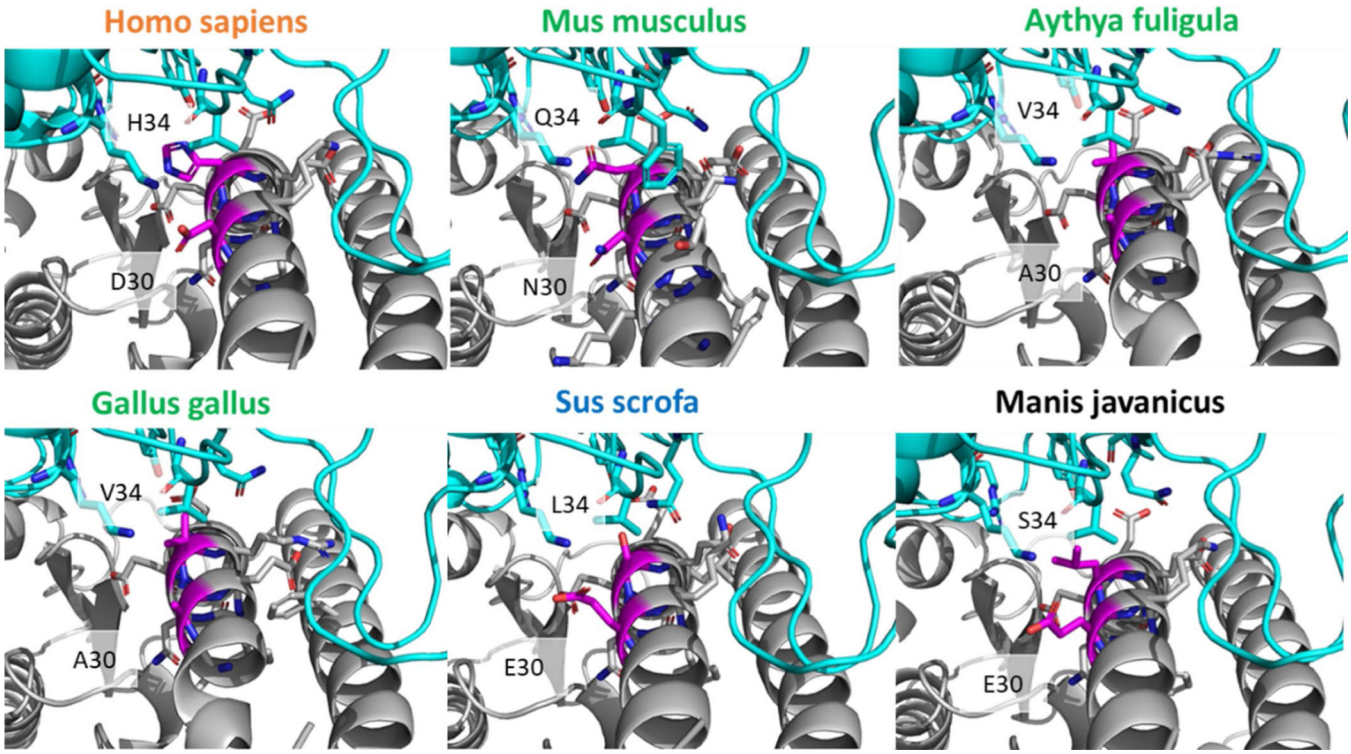

FIGURE 5.

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 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 javanica), 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

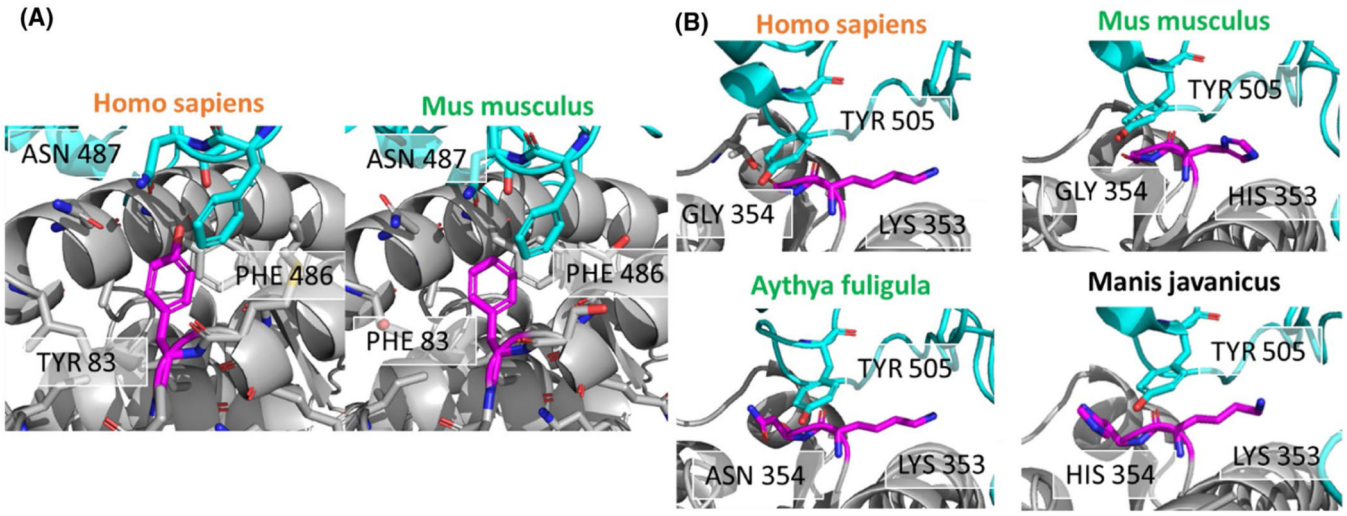

FIGURE 6.

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

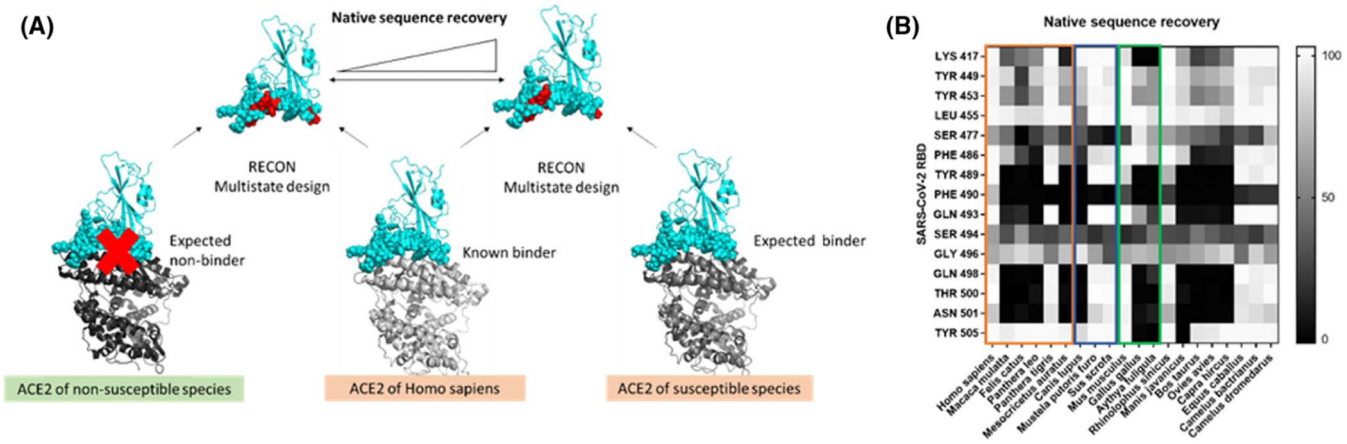

FIGURE 7.

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

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	+

FIGURE 8.

Predicted glycosylation profiles for ACE2 amino acid positions 53, 90, 103, and 322. Susceptible species are in orange, non-susceptible in green, intermediate in blue, and unknown in black. + indicates presence,—indicates the absence of glycosylation. glyc = glycosylation. Letters indicate amino acids using single-letter naming

	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

FIGURE 9.

Key residues of aligned ACE2 proteins with calculated SARS-CoV-2 susceptibility score for each species. Susceptible (orange), non-susceptible (green), intermediate (blue), and unknown (black/grey) species are indicated