1	Which animals a	re at risk? Predicting species	s susceptibility to Covid-19
2			
3 4 5	Alexander MR* ^{1,2} , Schoede JA ^{3,8}	er CT* ^{3,4} , Brown JA ^{5,6} , Smart C ³ , Meiler J ^{#3,4,9,10} , Chen W ^{#7} , Ma	D ⁷ , Moth C ^{3,8} , Wikswo JP ^{5,6,7,9} , Capra adhur MS ^{#1,2,7,11}
5		*	
O		autions contributed eq	ually
7	# C	corresponding authors, contrib	uting equally
ð			
9 10 11 12 13 14 15 16 17 18 19 20 21	¹ Department of Medicine, Center (VUMC), Nas Pharmacology, Vanderbilt U Biology, Vanderbilt Univ University, Nashville, TN, U Nashville, TN, USA; ⁶ Vand Vanderbilt University, Nashv Vanderbilt University, Nashv University, Nashville, TN, U Nashville, TN, USA; ¹⁰ Insti Germany; ¹¹ Vand	Division of Cardiovascular Me shville, TN, USA; ² Department Jniversity Medical Center, Nas resity, Nashville, TN, USA; ⁴ De JSA; ⁵ Department of Physics a derbilt Institute for Integrative E ville, TN, USA; ⁷ Department of ashville, TN, USA; ⁸ Departmen JSA; ⁹ Department of Biomedic tute for Drug Discovery, Leipzi erbilt Institute for Infection, Imr	dicine, Vanderbilt University Medical of Medicine, Division of Clinical hville, TN, USA; ³ Center for Structural epartment of Chemistry, Vanderbilt and Astronomy, Vanderbilt University, Biosystems Research and Education, Molecular Physiology and Biophysics, t of Biological Sciences, Vanderbilt al Engineering, Vanderbilt University, g University Medical School, Leipzig, nunology, and Inflammation
22 23		Word Count: 4,515	
24 25 26 27		Address for Corresponde	ence:
28 29 30 31 32 33 34 35 36 37 38 39 40	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

41 Abstract

- 42 In only a few months, the novel coronavirus severe acute respiratory syndrome coronavirus 2
- 43 (SARS-CoV-2) has caused a global pandemic, leaving physicians, scientists, and public health
- 44 officials racing to understand, treat, and contain this zoonotic disease. SARS-CoV-2 has made
- 45 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
- 47 livestock. While there is evidence that certain species, such as cats, are susceptible, the vast
- 48 majority of animal species, including those in close contact with humans, have unknown
- 49 susceptibility. Hence, methods to predict their infection risk are urgently needed. SARS-CoV-2
- 50 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
- 52 species using in-depth sequence and structural analyses of ACE2 and its binding to SARS-
- 53 CoV-2. Our findings have important implications for identification of ACE2 and SARS-CoV-2
- 54 residues for therapeutic targeting and identification of animal species with increased
- 55 susceptibility for infection on which to focus research and protection measures for
- 56 environmental and public health.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the virus responsible for the 58 alobal pandemic of coronavirus disease-2019 (Covid-19) that is impacting millions of lives and 59 the global economy. Covid-19 is a zoonotic infection capable of crossing the species barrier. 60 SARS-CoV-2 is thought to have originated in bats and subsequently transmitted to humans. 61 perhaps through a secondary host.^{1,2} Emerging experimental and observational evidence 62 demonstrates differences in species susceptibility to infection. For example, humans, house 63 cats, tigers, and lions are all susceptible to infection by SARS-CoV-2.³⁻⁶ Golden Syrian hamsters 64 and rhesus monkeys are also capable of being experimentally infected by SARS-CoV-2 and 65 developing Covid-19 pathologies.^{7,8} In contrast, observational and experimental studies with 66 direct intranasal inoculation have demonstrated that chickens, ducks, and mice are not 67 susceptible to SARS-CoV-2 infection.^{5,9-11} Interestingly however, susceptibility is not 68 69 dichotomous. Although ferrets are also susceptible to infection, intranasal inoculation failed to 70 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 71 appear asymptomatic, a minority of experimentally or environmentally exposed dogs exhibited 72 evidence of infection by SARS-CoV-2 PCR or SARS-CoV-2 seroconversion with production of 73 SARS-CoV-2-specific antibodies.^{5,12} While pigs have not demonstrated evidence of infection 74 after intranasal inoculation, overexpression of swine ACE2 in cultured cells supports some 75 degree of viral entry.^{5,9,13} Hence, ferrets, dogs, and pigs are classified as having intermediate 76 77 susceptibility to infection. Despite these findings, the number of animal species tested for 78 susceptibility to infection in experimental or observational studies is very limited. Thus, methods 79 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 80 81 discover animal models for research. Identifying the key residues mediating susceptibility to 82 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 83 84 protein on the surface of the SARS-CoV-2 virus mediates interaction with its receptor.

85 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

89 SARS-CoV-2 spike protein RBD and ACE2 is thus critical for viral cell entry and infection.⁹ The

90 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

92 necessary to enable viral cell entry in HeLa cells *in vitro* and SARS-CoV-2 infection in mouse

- 93 models *in vivo*.^{9,16}
- ACE2 is present in almost all vertebrates, however sequence differences exist that may hold

95 clues to differences in SARS-CoV-2 susceptibility, as has been observed for SARS-CoV.^{17,18}

96 Understanding such differences could provide insight into key structural interactions between

- 97 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

99 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

104 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

106 findings have important implications for identification of ACE2 and SARS-CoV-2 residues for

107 therapeutic targeting and identification of animal species with increased susceptibility for

108 infection on which to focus research and protection efforts.

- 109
- 110

RESULTS

111 Susceptibility does not segregate according to phylogeny and ACE2 sequence similarity

Given experimental evidence for susceptibility of humans, house cats, tigers, lions, rhesus

113 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

116 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.

118 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

120 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

123 hamster (Mesocricetus auratus), than non-susceptible species such as duck (Aythya fuligula) or

124 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

127 phylogenetic relationships nor overall ACE2 protein sequence similarity across species is able

128 to predict susceptibility to SARS-CoV-2 infection.

Sequence alignment identifies ACE2 residues distinguishing susceptible from non susceptible species

131 An alternative approach is to use the experimentally validated differences in infection

132 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,

138 Tyr83, and Gln24, are highly similar in susceptible species and quite different in non-susceptible

139 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 140 GroupSim scores were present at or near the binding interface including His34, Asp30, Thr92, 141 Gln24, Lys31, and Leu79 (Figure 1). We then extended this analysis by focusing on key 142 residues previously demonstrated from prior structural analysis to be important for ACE2 and 143 SARS-CoV-2 RBD interactions (**Table 1**).^{7,22-24} Interestingly, this revealed that key amino acids 144 for the ACE2 and SARS-CoV-2 spike protein interaction were enriched among the top scoring 145 146 GroupSim positions (7 of 35; p<0.0001; Fisher's exact test). Such key residues based on 147 structural analysis being over-represented in amino acid positions that best discriminated susceptible from non-susceptible species suggests that structural interactions between ACE2 148 and SARS-CoV-2 spike protein importantly determine differences in species susceptibility to 149 infection. In addition, these data suggest that certain ACE2 amino acid residues may be 150 particularly important for determining susceptibility, including Leu79, His34, Tyr83, Gln24, 151 Lys31, Asp30, and Glu329. 152

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

- 155 We used homology modeling to identify structural determinants of binding the ACE2 protein
- 156 from species with known differences in susceptibility to SARS-CoV-2 infection. The models
- 157 were based on previously reported crystal structures of the human ACE2 in complex with
- 158 SARS-CoV-2 (PDB: 6LZG and 6M0J).¹⁴ We modeled ACE2 in the presence of the SARS-CoV-2
- 159 RBD to allow backbone adjustment to the binder and refined by redocking of the RBD domain to
- 160 optimize sidechains. Models were selected by overall calculated protein stability of the SARS-
- 161 CoV-2 RBD complex, predicted binding energy between ACE2 and SARS-CoV-2 RBD, and
- similarity (as C α -root mean square deviation [C α -RMSD], **Extended Data Figure 3 and**
- 163 **Extended Data Figure 4**). Based on these models, multiple approaches where undertaken to
- 164 investigate the structural interactions between SARS-CoV-2-RBD and ACE2.
- 165 We evaluated the overall calculated protein stability and predicted binding energy for SARS-
- 166 CoV-2-RBD and ACE2 complexes for each species. We considered the 100 best models for
- 167 each species and evaluated evidence for difference in binding energy or stability between
- 168 susceptible and non-susceptible species. The average mean predicted binding energy and
- 169 calculated protein stability differs across species (Figure 2). Consistent with the lack of
- 170 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
- 172 (Aythya fuligula) shows similarly unfavorable scores, indicating that ACE2 sequence differences
- 173 leading to a lower structural binding ability in these two avian species may explain their lack of
- 174 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
- 176 stability than several species that are susceptible, including the lion (*Panthera leo*), tiger
- 177 (Panthera tigris), and cat (Felis catus). Thus, differences in SARS-CoV-2 and ACE2 complex
- 178 stability have some discriminative power but are not the sole factor in differences in
- 179 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 182 susceptible from non-susceptible species, we performed energetic modeling of residue-residue 183 interactions in the interface of SARS-CoV-2 and ACE2 using Rosetta. Although the overall 184 185 interaction pattern across residues is similar between susceptible, non-susceptible, and intermediate susceptibility species, there are significant differences in the magnitude of residue-186 residue interactions (Figure 3). For example, residue 30 (which is an aspartate in all susceptible 187 species) forms a strong ionic interaction with lysine 417 of SARS-CoV-2 RBD and interacts 188 189 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 190 191 longer present and is not substituted by any other structural rearrangements that might 192 accommodate this change. Mouse (Mus musculus) ACE2 contains an asparagine in position 30 193 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 194 195 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 196 197 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, 198 199 this position exhibits significantly decreased binding energy with residues Asn487 and Tyr489 in 200 SARS-CoV-2 RBD in non-susceptible species (Figure 3). Although ACE2 residue 83 also 201 interacts with SARS-CoV-2 RBD phenylalanine 486, this interaction is unlikely to be significantly 202 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 203 204 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 205 identified through the GroupSim analysis described above to be top residues discriminating 206 207 susceptible from non-susceptible species based on sequence alignment (Extended Data Table 208 1). These results suggest that these amino acid positions of ACE2 may be important mediators 209 of the structural interaction of ACE2 and SARS-CoV-2 RBD and determinants of differences to 210 susceptibility to infection across species.

211 Multistate design reveals ACE2 G354 as determinant of susceptibility

212 It is an evolutionary advantage for SARS-CoV-2 to maintain its ability to infect multiple species.
213 Thus, we hypothesized that the sequence of SARS-CoV-2 RBD is not optimized for a single
214 species but is capable of binding ACE2 of multiple species. Multistate design is a computational
215 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)
- 217 multistate design to test this hypothesis. This method determines how many mutations one
- 218 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

221 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.

- 227 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
- 229 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...

237 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
- 240 evaluating residue-specific interactions based on the native sequence recovery from RECON
- 241 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
- 251 in species susceptibility to SARS-CoV-2 infection.

252 ACE2 glycosylation at N90 and N322 as determinants of susceptibility

253 As a final additional approach to structurally evaluate differences in species susceptibility, we 254 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 255 interactions;²⁷ however, given the challenges in identifying glycans in protein crystal structures, 256 257 glycosylation has received considerably less attention than SARS-CoV-2 RBD and ACE2 258 protein-protein interactions. Naturally occurring glycans as posttranslational modifications are 259 not fully visible in crystal structures. Normally only the first N-actylglucosamine is visible or no 260 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 261

262 surface exposed NXT/S sequen was seen three times in proximity to the binding interface on the ACE2. To understand whether the ACE2 of other species have similar glycosylation 263 patterns, glycosylation was predicted using NetNGlyc 1.0, a neural network for predicting N-264 glycosylation sites, and compared to the glycosylation patterns of human ACE2.²⁸ Residues 53, 265 90, 103, and 322 were identified as glycosylation sites in human ACE2, with 53, 90, and 322 266 demonstrating glycosylation in the crystal structure (PDB: 6M0J and 6LZG)¹⁴ (Table 2). Other 267 susceptible species were quite similar to this pattern, except for position 103, which is only 268 269 predicted to be glycosylated in humans and rhesus macaques. Among known susceptible 270 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 271 and intermediate susceptibility species were non-glycosylated. Interestingly, ACE2 from the 272 non-susceptible mouse, despite not showing significant differences in predicted binding energy 273 274 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 275 species. This suggests a potential mechanism by which mice may be non-susceptible despite 276 277 having similar binding energy and SARS-CoV-2 native sequence recovery to susceptible

278 species.

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

280 Taken together, results of these studies reveal a set of key ACE2 residues important for 281 interaction with SARS-CoV-2 RBD and for which differences help discriminate susceptible from 282 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 283 sequence recovery in interaction with SARS-CoV-2, and positions 90 and 322, which exhibit 284 differences in glycosylation. Using these key residues in aggregate, we developed a SARS-285 CoV-2 susceptibility score based on similarity to the human ACE2 sequence using the 286 BLOSUM62 similarity matrix (Table 3).²⁹ This analysis revealed that experimentally validated 287 non-susceptible species have in fact the lowest susceptibility scores, while species with 288 289 previously demonstrated intermediate susceptibility have intermediate susceptibility scores. Using the lowest score of the susceptible species, 23, as the lower cutoff for susceptibility and 290 the highest score of non-susceptible species, 11, as the upper cutoff for non-susceptibility, we 291 extended these results to species with unknown susceptibility. This revealed high scores in the 292 293 susceptible range for the Chinese horseshoe bat (*Rhinolophus sinicus*), horse (*Equus caballus*), and camels (Camelus dromedarius and Camelus bactrianus) and intermediate susceptibility 294 295 scores for the Malayan pangolin (Manis javanica), cow (Bos taurus), goat (Capra hircus), and 296 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 305 structural interactions with SARS-CoV-2 RBD, leading to differences in species susceptibility to 306 SARS-CoV-2 infection. Our results, combining prior knowledge of experimentally validated 307 differences in species susceptibility with multiple methods of determining effects on ACE2 308 309 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 310 exhibit alterations in binding energy, and positions 90 and 322, which exhibit alterations in 311 glycosylation that likely contribute to differences in interactions at the interface. Taken together, 312 our results provide insight into the molecular determinants of species susceptibility to SARS-313 CoV-2 infection and have important implications for identification of key residues for therapeutic 314 targeting and determining susceptibility of additional species to infection. 315

316 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 317 318 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 319 to SARS-CoV-2 with in-depth structural analyses to determine critical ACE2 residues for 320 infection. In addition, we performed multiple structural analyses, including residue-residue 321 322 interactions, RECON multistate design, and glycosylation analysis, to rigorously determine the 323 structural basis for species differences in ACE2 interaction with SARS-CoV-2 RBD. Prior 324 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 325 group.³⁷ Recent experimental work with direct inoculation, however, has demonstrated that pigs 326 are non-susceptible,⁵ and that house cats and Golden Syrian hamsters are susceptible.^{5,7} We 327 328 identified key residues on which to build a susceptibility score that closely matches experimentally verified in vivo susceptibility, including predicting an intermediate susceptibility of 329 the pig and higher susceptibility of house cats and Golden Syrian hamsters. 330

A key principle revealed by our findings is the importance of using multiple methods for 331 determining the structural basis for differences in ACE2 interaction with SARS-CoV-2 RBD. For 332 example, although calculated binding energy, protein stability, and RECON multistate design of 333 SARS-CoV-2 RBD in complex with duck and chicken ACE2 distinguished non-susceptible 334 chicken and duck ACE2 from susceptible species, mouse ACE2 did not fit the pattern of other 335 336 non-susceptible species. However, analysis of ACE2 protein glycosylation revealed two 337 residues, 90 and 322, for which differences in mouse ACE2 distinguished it from susceptible 338 species. In addition, combining ACE2 sequence alignment, GroupSim calculations, and residue-339 residue interaction modeling identified residues 30 and 83, which are distinctly different in all 340 non-susceptible compared to susceptible species. Differences in these residues in nonsusceptible species result in decreased binding energy with SARS-CoV-2 RBD. Although no 341 342 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 343 344 distinguish susceptible from non-susceptible species, as reflected by the calculated

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

Our findings have important implications for determining infectability of animals with heretofore 347 unknown susceptibility to SARS-CoV-2 infection. Determining such susceptibility is critical to 348 prevent disruption to food supplies, identify optimal animal models for research, aid in the 349 350 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 351 species with unknown susceptibility to date. These data suggest that cows (Bos taurus), 352 Malayan pangolin (Manis javanica), and goats (Capra hircus) have intermediate susceptibility to 353 infection, while Chinese horseshoe bats (Rhinolophus sinicus), horses (Equus caballus), and 354 camels (Camelus dromedarius and Camelus bactrianus) have higher susceptibility. Although 355 the ultimate test is direct exposure of live animals to evaluate infectability and transmissability,^{5,7} 356 357 this is complicated by the need for BSL3 containment and is guite costly and challenging with 358 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 359 360 closely monitored for evidence of Covid-19 infection. The close interaction of these animals with 361 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 362 developed here can also be applied to additional species of interest to help direct resources for 363 364 focused research and protection efforts in the future.

365 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 366 structural interaction and potential for therapeutic targeting. By incorporating differences in 367 species susceptibility into the structural analysis, our findings enhance the potential to identify 368 particularly important residues mediating the ACE2 and SARS-CoV-2 RBD interaction. Indeed, 369 although GroupSim scores were not used in the structural analysis, three of the five key 370 identified residues (30, 83, and 90) from the structural modeling are in the top scoring ACE2 371 372 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 373 important contribution of these residues to differences in susceptibility. Amino acid positions 30 374 and 83 of ACE2 in particular exhibited large differences in residue-residue interaction binding 375 376 energies between susceptible and non-susceptible species. Asp30 on ACE2 interacts with 377 residues Lys417, Phe456, and Tyr473 of SARS-CoV-2 RBD, and ACE2 Tyr83 interacts with 378 Asn487 and Tyr489 of SARS-CoV-2 RBD. These amino acids mark sites of SARS-CoV-2 379 interaction with ACE2 that may be important for development of antibody-based therapies or 380 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 nonbinding 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 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

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 differences in other factors across species may also contribute to differences in susceptibility. 392 This includes differences in ACE2 expression levels⁴⁰ and differences in the protein sequence of 393 TMPRSS2, a protein that contributes to viral and host cell membrane fusion through cleavage of 394 spike protein.^{15,41} With further experimental and observational data on infectability of currently 395 unknown species, the susceptibility score we have developed can also help determine species 396 for which differences in ACE2 protein may not inadequately predict differences in susceptibility. 397 For these species future studies could compare differences in expression levels of ACE2 and/or 398 399 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

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

415 **References**

416 1 Adhikari, S. P. et al. Epidemiology, causes, clinical manifestation and diagnosis, prevention and 417 control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. 418 Infectious diseases of poverty 9, 29, doi:10.1186/s40249-020-00646-x (2020). 419 2 Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the 420 COVID-19 Outbreak. *Current biology : CB* **30**, 1346-1351.e1342, doi:10.1016/j.cub.2020.03.022 421 (2020). 422 3 APHISpress@usda.gov. USDA Statement on the Confirmation of COVID-19 in a Tiger in New 423 York. United States Department of Agriculture Animal and Plant Health Inspection Service. 424 https://www.aphis.usda.gov/aphis/newsroom/news/sa by date/sa-2020/ny-zoo-covid-19. 425 (April 6, 2020). 426 Mallapaty, S. Coronavirus can infect cats - dogs, not so much. Nature, doi:10.1038/d41586-020-4 427 00984-8 (2020). 428 5 Shi, J. et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-429 coronavirus 2. Science, doi:10.1126/science.abb7015 (2020). 430 Halfmann, P. J. et al. Transmission of SARS-CoV-2 in Domestic Cats. New England Journal of 6 431 Medicine, doi:10.1056/NEJMc2013400 (2020). 432 7 Chan, J. F.-W. et al. Simulation of the clinical and pathological manifestations of Coronavirus 433 Disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease pathogenesis 434 and transmissibility. Clinical Infectious Diseases, doi:10.1093/cid/ciaa325 (2020). 435 8 Chandrashekar, A. et al. SARS-CoV-2 infection protects against rechallenge in rhesus macaques. 436 Science, eabc4776, doi:10.1126/science.abc4776 (2020). 9 437 Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. 438 Nature 579, 270-273, doi:10.1038/s41586-020-2012-7 (2020). 439 10 Monteil, V. et al. Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-440 Grade Soluble Human ACE2. Cell, doi:10.1016/j.cell.2020.04.004 (2020). 441 Bao, L. et al. The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. Nature, 11 doi:10.1038/s41586-020-2312-y (2020). 442 443 Sit, T. H. C. et al. Infection of dogs with SARS-CoV-2. Nature, doi:10.1038/s41586-020-2334-5 12 444 (2020). 445 13 Liu, Y. et al. Functional and Genetic Analysis of Viral Receptor ACE2 Orthologs Reveals Broad 446 Potential Host Range of SARS-CoV-2. bioRxiv, 2020.2004.2022.046565, 447 doi:10.1101/2020.04.22.046565 (2020). 448 14 Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 449 receptor. Nature, doi:10.1038/s41586-020-2180-5 (2020). 450 15 Hoffmann, M. et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a 451 Clinically Proven Protease Inhibitor. Cell 181, 271-280.e278, doi:10.1016/j.cell.2020.02.052 452 (2020). 453 16 Bao, L. et al. The Pathogenicity of SARS-CoV-2 in hACE2 Transgenic Mice. bioRxiv, 454 2020.2002.2007.939389, doi:10.1101/2020.02.07.939389 (2020). 455 17 Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor Recognition by the Novel Coronavirus 456 from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J Virol 457 94, doi:10.1128/JVI.00127-20 (2020). 458 Li, F. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral Res 100, 18 459 246-254, doi:10.1016/j.antiviral.2013.08.014 (2013). 460 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). 461

462	20	Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic acids
463		research 47 , W636-W641, doi:10.1093/nar/gkz268 (2019).
464	21	Capra, J. A. & Singh, M. Characterization and prediction of residues determining protein
465		functional specificity. <i>Bioinformatics</i> 24, 1473-1480, doi:10.1093/bioinformatics/btn214 (2008).
466	22	Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2.
467		Science 367 , 1444-1448, doi:10.1126/science.abb2762 (2020).
468	23	Wang, Q. et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell,
469		doi:10.1016/j.cell.2020.03.045 (2020).
470	24	Procko, E. The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS
471		coronavirus 2. <i>bioRxiv</i> , 2020.2003.2016.994236, doi:10.1101/2020.03.16.994236 (2020).
472	25	Sevy, A. M., Jacobs, T. M., Crowe, J. E. & Meiler, J. Design of Protein Multi-specificity Using an
473		Independent Sequence Search Reduces the Barrier to Low Energy Sequences. <i>Plos</i>
474		<i>Computational Biology</i> 11 . doi:ARTN e100430010.1371/iournal.pcbi.1004300 (2015).
475	26	Sevy, A. M. <i>et al.</i> Multistate design of influenza antibodies improves affinity and breadth against
476		seasonal viruses. Proceedings of the National Academy of Sciences of the United States of
477		<i>America</i> 116 . 1597-1602. doi:10.1073/pnas.1806004116 (2019).
478	27	Ferreira, I. G. <i>et al.</i> Glycosylation as a Main Regulator of Growth and Death Factor Receptors
479		Signaling. International journal of molecular sciences 19 , 580, doi:10.3390/iims19020580 (2018).
480	28	Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-
481		translational glycosylation and phosphorylation of proteins from the amino acid sequence.
482		Proteomics 4 1633-1649, doi:10.1002/pmic.200300771 (2004)
483	29	Pearson, W. R. Selecting the Right Similarity-Scoring Matrix. <i>Curr Protoc Bioinformatics</i> 43 , 3, 5, 1-
484	23	3.5.9. doi:10.1002/0471250953 bi0305s43 (2013)
485	30	Melin, A. D., Janiak, M. C., Marrone, F., Arora, P. S. & Higham, J. P. Comparative ACE2 variation
486		and primate COVID-19 risk. <i>bioRxiv</i> . 2020.2004.2009.034967. doi:10.1101/2020.04.09.034967
487		(2020).
488	31	7hai, X. <i>et al.</i> Comparison of SARS-CoV-2 spike protein binding to ACE2 receptors from human.
489	01	pets, farm animals, and putative intermediate hosts. J Virol. doi:10.1128/ivi.00831-20 (2020).
490	32	Luan L. Jin, X. Lu, Y. & Zhang, L. SARS-CoV-2 spike protein favors ACF2 from Bovidae and
491		Cricetidae. Journal of medical virology. doi:10.1002/imv.25817 (2020).
492	33	Liu, 7, et al. Composition and divergence of coronavirus spike proteins and host ACF2 receptors
493		predict potential intermediate hosts of SARS-CoV-2. <i>Journal of medical virology</i> .
494		doi:10.1002/imv.25726 (2020).
495	34	Praharai, M. R. <i>et al.</i> Prediction analysis of SARS-COV-2 entry in Livestock and Wild animals.
496	01	<i>bioRxiv</i> , 2020,2005,2008,084327, doi:10.1101/2020.05.08.084327 (2020).
497	35	Lam, S. <i>et al.</i> SARS-CoV-2 spike protein predicted to form complexes with host receptor protein
498	55	orthologues from a broad range of mammals <i>bioRxiv</i> 2020 2005 2001 072371
499		doi:10.1101/2020.05.01.072371 (2020)
500	36	Wan Y Shang I Graham R Baric R S & Li F Recentor Recognition by the Novel Coronavirus
500	50	from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus Journal
502		of Virology 94 e00127-00120 doi:10.1128/ivi.00127-20.(2020)
502	37	Damas L et al Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural
504	57	Analysis of ACE2 in Vertebrates <i>bioRxiv</i> 2020 2004 2016 045302
505		doi:10.1101/2020.04.16.045302 (2020)
506	38	With D. With T. Littin, O. & Yang, Z. The SARS-CoV-2 outbreak: What we know International
507	50	Journal of Infectious Diseases 94 44-48 doi:https://doi.org/10.1016/i.iiid.2020.03.004 (2020)
507		3001101 0/ 11/001.012/ 10.000/ 10.000/ 10.000/ 10.000/ 10.000/ (2020).004 (2020).

Wang, Y., Liu, M. & Gao, J. Enhanced receptor binding of SARS-CoV-2 through networks of

508

509 510		hydrogen-bonding and hydrophobic interactions. <i>Proceedings of the National Academy of</i>
510	40	Sciences, 202008203, doi:10.1075/pilas.2008203117 (2020).
511	40	Fu, J. <i>et ul.</i> Expressions and significances of the angiotensin-converting enzyme 2 gene, the
512		receptor of SARS-COV-2 for COVID-19. <i>Molecular Biology Reports</i> , doi:10.1007/S11033-020-
513		
514	41	Lukassen, S. et al. SARS-COV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial
515		transient secretory cells. <i>Embo J</i> , e105114, doi:10.15252/embj.20105114 (2020).
516	42	Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript
517		server. Nucleic Acids Research 42 , W320-W324, doi:10.1093/nar/gku316 (2014).
518	43	Dereeper, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic
519		Acids Res 36 , W465-469, doi:10.1093/nar/gkn180 (2008).
520	44	Dereeper, A., Audic, S., Claverie, J. M. & Blanc, G. BLAST-EXPLORER helps you building datasets
521		for phylogenetic analysis. BMC evolutionary biology 10, 8, doi:10.1186/1471-2148-10-8 (2010).
522	45	Sayers, E. W. et al. Database resources of the National Center for Biotechnology Information.
523		Nucleic Acids Res 37 , D5-15, doi:10.1093/nar/gkn741 (2009).
524	46	Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments.
525		Nucleic Acids Research 47, W256-W259, doi:10.1093/nar/gkz239 (2019).
526	47	Nivon, L. G., Moretti, R. & Baker, D. A Pareto-optimal refinement method for protein design
527		scaffolds. PLoS One 8 , e59004, doi:10.1371/journal.pone.0059004 (2013).
528	48	Song, Y. et al. High-resolution comparative modeling with RosettaCM. Structure 21, 1735-1742,
529		doi:10.1016/j.str.2013.08.005 (2013).
530	49	Gray, J. J. et al. Protein-protein docking with simultaneous optimization of rigid-body
531		displacement and side-chain conformations. J Mol Biol 331, 281-299, doi:10.1016/s0022-
532		2836(03)00670-3 (2003).
533	50	Chaudhury, S. <i>et al.</i> Benchmarking and analysis of protein docking performance in Rosetta v3.2.
534		<i>PLoS One</i> 6 , e22477, doi:10.1371/journal.pone.0022477 (2011).
535	51	Sauer, M. F., Sevy, A. M., Crowe, J. E., Jr. & Meiler, J. Multi-state design of flexible proteins
536		predicts sequences optimal for conformational change. PLoS Comput Biol 16, e1007339,
537		doi:10.1371/journal.pcbi.1007339 (2020).
538	52	Schrodinger, LLC. The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.8
539		(2015).
540	53	Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8 (2015).
541	54	Schrodinger, LLC. The JyMOL Molecular Graphics Development Component, Version 1.8 (2015).
542		
543		

544 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	т	F	D	к	н	Е	Е	D	Y	Q	L	L	М	Y	Q	E	N	к	G	D	R	R
Macaca mulatta	Rhesus macaque	s	Q	т	F	D	к	н	Е	Е	D	Y	Q	L	L	М	Y	Q	E	N	к	G	D	R	R
Felis catus	House cat	s	L.	т	F	E	к	н	Е	Е	Е	Y	Q	L	L	т	Y	Q	E	N	к	G	D	R	R
Panthera tigris altaica	Tiger	S	L	т	F	Е	к	н	Е	Е	Е	Y	Q	L	L	т	Y	Q	E	N	к	G	D	R	R
Panthera leo	Lion	S	L.	т	F	E	к	н	Е	Е	Е	Y	Q	L	L	т	Y	Q	E	N	к	G	D	R	R
Mesocricetus auratus	Golden Syrian hamster	S	Q	т	F	D	ĸ	Q	Е	Е	D	Y	Q	L	L.	Ν	Y	Q	E	Ν	к	G	D	R	R
Mus musculus	Mouse	S	N	т	F	N	N	Q	Е	Е	D	Y	Q	L	Т	S	E.	Q	Α	N	н	G	D	R	R
Aythya fuligula	Duck	D	-	М	F	Α	E	v	R	Е	D	Y	E	L	N	N	E.	E	К	N	к	Ν	D	R	R
Gallus gallus	Chicken	D	-	т	F	Α	E	v	R	Е	D	Y	Е	L	N	R	F	E	т	N	к	N	D	R	R
Mustela putorius furo	Ferret	s	L	т	F	Е	к	Y	Е	Е	Е	Y	Q	L	н	т	Y	Е	Q	Ν	к	R	D	R	R
Sus scrofa	Pig	s	L	т	F	Е	к	L	Е	Е	D	Y	Q	L	1	т	Y	Q	Ν	Ν	к	G	D	R	R
Canis lupus familiaris	Dog	s	L	т	F	Е	к	Y	Е	Е	Е	Y	Q	L	L	т	Y	Q	G	N	к	G	D	R	R
Rhinolophus sinicus	Chinese horseshoe bat	s	Е	т	F	D	к	т	к	E	D	н	Q	L	L	Ν	Υ	Е	N	N	к	G	D	R	R
Equus caballus	Horse	s	L	т	F	Е	к	S	Е	Е	Е	н	Q	L	L	т	Y	Q	Е	N	к	G	D	R	R
Bos taurus	Cow	s	Q	т	F	Е	к	н	Е	Е	D	Y	Q	L	M	т	Y	Q	D	Ν	к	G	D	R	R
Manis javanica	Malayan pangolin	s	Е	т	F	Е	к	s	Е	Е	Е	Y	Q	L	1	Ν	Y	Q	Е	N	к	н	D	R	R
Capra hircus	Goat	s	Q	т	F	Е	к	н	Е	Е	D	Y	Q	L	М	т	Y	Q	Ν	Ν	к	G	D	R	R
Ovis aries	Sheep	s	Q	т	F	Е	к	н	Е	Е	D	Y	Q	L	М	т	Y	Q	D	Ν	к	G	D	R	R
Camelus dromedarius	Arabian Camel	s	L	т	F	Е	Е	н	Е	Е	D	Y	Q	L	т	т	Y	Q	D	Ν	к	G	D	R	R
Camelus bactrianus	Bactrian Camel	s	L	т	F	Е	Е	н	Е	Е	D	Y	Q	L	т	т	Y	Q	D	N	к	G	D	R	R

546 Highlighted residues that are most similar in susceptible and different in non-susceptible species as

547 determined by GroupSim (Extended Data Table 1). Susceptible species are in orange, non-susceptible in

548 green, intermediate in blue, and unknown in black/grey. Letters indicate amino acids using single-letter

549 naming.

550

545

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	Ν	T	Т	+	Ν	L	Т	+	Ν	Ġ	S	+	Ν	М	Т	+
Macaca mulatta	Ν	Т	Т	+	Ν	L	Т	+	Ν	G	S	+	Ν	Μ	Т	+
Felis catus	N	Т	Т	+	Ν	т	Т	+	S	G	S	-	Ν	Μ	Т	+
Panthera tirgris altaica	N	Т	Т	+	Ν	Т	Т	+	S	G	S	-	Ν	Μ	Т	+
Panthera leo	Ν	Т	Т	+	Ν	т	Т	+	S	G	S	-	Ν	Μ	Т	+
Mesocricetus auratus	Ν	T	Т	+	Ν	L	Т	+	S	G	S	-	Y	М	Т	-
Mus musculus	Ν	Т	Т	+	Т	Ρ	I.	-	S	G	S	-	н	Μ	Т	-
Aythya fuligula	Ν	Т	Т	+	D	Ρ	L	-	К	G	S	-	Ν	Μ	Т	+
Gallus gallus	Ν	Т	Т	+	D	Α	۷	-	R	G	S	-	Ν	Μ	Т	+
Mustela putorius furo	N	Т	Т	+	D	Ρ	1	-	S	G	S	-	Ν	Μ	Т	+
Sus scrofa	Ν	Т	Т	+	Т	L	1	-	S	Ġ	Т	-	Ν	Μ	Т	+
Canis lupus familiaris	Ν	Т	Т	+	D	S	Т	-	S	G	S	-	Ν	Μ	Т	+
Rhinolophus sinicus	Ν	Т	Ν	-	Ν	V	Т	+	S	G	S	-	Ν	Μ	Т	+
Equus caballus	N	Т	Т	+	Ν	L	Т	+	S	G	S	-	Ν	Μ	Т	+
Bos taurus	Ν	Т	Т	+	Ν	L	Т	+	S	G	Т	-	Υ	Μ	Т	-
Manis javanica	Ν	Т	Т	+	Ν	D	Т	+	S	G	S	-	Κ	Μ	Т	-
Capra hircus	N	1	Т	+	Ν	L	Т	+	S	Ġ	Т	-	Υ	Μ	Т	-
Ovies aries	Ν	Т	Т	+	Ν	L	Т	+	S	G	Т	-	Υ	Μ	Т	-
Camelus dromedarius	Ν	T	Т	+	Ν	۷	Т	+	S	G	А	-	Ν	М	Т	+
Camelus bactrianus	N	1	т	+	N	v	т	+	S	G	Δ	-	N	м	т	+

554

555 Susceptible species are in orange, non-susceptible in green, intermediate in blue, and unknown in black.

+ indicates presence, - indicates absence of glycosylation. glyc=glycosylation. Letters indicate amino

557 acids using single-letter naming.

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	Ν	Ν	G	-
Macaca mulatta (Rhesus Macaque)	D	Υ	Ν	Ν	G	31
Felis catus (House cat)	Е	Υ	Ν	Ν	G	27
Panthera tigris altaica (Tiger)	Е	Υ	Ν	Ν	G	27
Panthera leo (Lion)	Е	Υ	Ν	Ν	G	27
Mesocricetus auratus (Golden Syrian Hamster)	D	Υ	Ν	Y	G	23
Mus musculus (Mouse)	Ν	F	т	н	G	11
Aythya fuligula (Duck)	Α	F	D	Ν	Ν	8
Gallus gallus (Chicken)	Α	F	D	Ν	Ν	8
Mustela putorius furo (Ferret)	Е	Υ	D	Ν	R	14
Sus scrofa (Pig)	Е	Υ	т	Ν	G	21
Canis lupus familiaris (Dog)	Е	Υ	D	Ν	G	22
Rhinolophus sinicus (Chinese horseshoe bat)	D	Y	Ν	Ν	G	31
Equus caballus (Horse)	Е	Y	Ν	Ν	G	27
Bos taurus (Cow)	Е	Y	Ν	Y	G	19
Manis javanica (Malayan pangolin)	Е	Y	Ν	κ	н	13
Capra hircus (Goat)	Е	Y	Ν	Y	G	19
Ovis aries (Sheep)	Е	Y	Ν	Y	G	19
Camelus dromedarius (Arabian Camel)	Е	Y	Ν	Ν	G	27
Camelus bactrianus (Bactrian Camel)	Е	Y	Ν	Ν	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
- in magenta. Higher scores are brighter in color. (B) Close-up view of the interface highlighting
 ACE2 residues with high GroupSim scores. (C) Close-up view after 90 degree rotation from (B)
- 570 ACE2 residues with high GroupSim scores. (**C**) Close-up view after 90 degree rotat 571 demonstrating additional residues at the interface with high GroupSim scores.





575 Figure 2: SARS-CoV-2 RBD has lower predicted binding energy and protein complex

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



Figure 3: Energetic modeling of residue-residue interactions identifies a link between 581 582 ACE2 D30 and Y83 and SARS-CoV-2 susceptibility. Residue-residue interactions are 583 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 584 interactions greater than 0.05 Rosetta Energy Units (REU) or smaller than -0.05 REU. 585 Interactions are presented as mean for all included samples. Residues depicted on the y-axis 586 are all observed amino acid identities for the particular position in its susceptibility group. (A) 587 Per-residue interactions for (A) susceptible species (human, cat, lion, tiger, hamster and rhesus 588 589 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 590 591 non-susceptible species.

592



Figure 4: Binding interactions of ACE2 position 30 differ across species. Close-up of the 595 differences in binding interactions of positions 30 and 34 (magenta) of ACE2 from each species 596 597 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 598 alanine (A) in the avian species (Aythya fuligula and Gallus gallus). Glutamic acid (E) is present 599 at position 30 in pig (Sus scrofa) and Malayan pangolin (Manis javanicus), representing 600 intermediate and unknown susceptible species, respectively. Position 34 is conserved as 601 602 histidine (H) in all susceptible species such as humans, yet has another residue identity in 603 intermediate and non-susceptible species. Species names in orange are susceptible, green are non-susceptible, blue are intermediate susceptibility, and black are unknown. 604



Figure 5: 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

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

with the only exception of a histidine in the mouse ACE2. ACE2 residue 354 is a glycine in the

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

are non-susceptible, and black are unknown.



620

Figure 6: Multistate design reveals SARS-CoV-2 RBD Tyr505 to have low native sequence 621

recovery in non-susceptible duck and chicken. (A) RECON multistate design overview. In 622 the presence of ACE2 from two different species the SARS-CoV-2-RBD interface is redesigned. 623

624 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

625 the interface residues is lower, then many sequence changes are required, indicating that one 626

627 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 628

ACE2. Only residues of the SARS-CoV-2-RBD, which are in the protein-protein interface and 629

show changes are depicted. Tyrosine 505 of SARS-CoV-2 RBD shows low native sequence 630

recovery (black) in non-susceptible duck (Gallus gallus) and chicken (Aythya fuligula). The 631

632 orange box outlines susceptible species, the blue box outlines species with intermediate

633 susceptibility, and the green box outlines non-susceptible species.

634

METHODS

637 ACE2 protein alignment

- 638 Protein sequence accession numbers and corresponding FASTA files from multiple species
- 639 (Extended Data Table 2) were pulled from NCBI using Batch Entrez. In the absence of a
- 640 published sequence and accession number, ACE2 protein sequence for the lion (*Panthera leo*)
- 641 was assembled using TBLASTN (National Center for Biotechnology Information) with tiger
- 642 ACE2 protein sequence as the query (**Extended Data Table 3**). Protein sequences were loaded
- 643 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
- 646 (http://espript.ibcp.fr/ESPript/ESPript/), including annotation of secondary structure based on
- 647 Protein Data Bank (PDB) structure 1r42 of human ACE2.⁴² A treedyn format tree diagram
- 648 representing similarity of ACE2 protein sequence across species was generated using
- 649 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
- 651 Package (PHYLIP) tree.⁴⁵ Final visualization was performed using the interactive Tree of Life
- 652 (iTOL) tree viewer v 5.5.1 (https://itol.embl.de/).⁴⁶

653 **Quantification of amino acid differences in alignment of susceptible and non-susceptible** 654 **species**

- 655 Quantification of amino acid positions in the ACE2 protein alignment that optimally distinguish
- 656 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-
- 658 susceptible species. Values are weighted by the BLOSUM62 similarity matrix to incorporate
- 659 similarity of amino acids properties.²⁹

660 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 661 SARS-CoV-2-RBD with the human ACE2 (PDB-IDs 6LZG and 6M0J).¹⁴ One co-crystal structure 662 (PDB-ID 6VW1) was excluded due to its lower resolution as compared to the aforementioned 663 structures. The target sequences were threaded over the ACE2-SARS-CoV-2-RBD co-crystal 664 structure, which was first relaxed with backbone constraints using RosettaRelax.⁴⁷ A total of 665 1000 homology models were constructed using RosettaCM, and subsequently relaxed with 666 backbone constraints.^{47,48} Of these, 25 models were selected based on the total energy as a 667 668 measure of protein stability, predicted binding energy, and C α -root mean square deviation (C α -669 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 670 torsional sampling.^{49,50} A final ensemble of 100 models was selected based on the total energy 671 as measure of protein stability, predicted binding energy and $C\alpha$ -RMSD to the best scoring 672 model (Extended Data Figure 4). The pairwise binding interaction between SARS-CoV-2 and 673 674 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 675

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

678 Calculation of sequence recovery from Restraint Convergence (RECON) multistate 679 design

680 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
- 682 complex.^{25,26,51} As a control, this was also performed solely using the human SARS-CoV-2-
- 683 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).
- 687 All protocols were executed using Rosetta-3.12 (www.rosettacommons.org). Evaluation was
- 688 performed using the numpy, pandas, matplotlib and seaborn libraries in Python 3.7, PyMOL
- 689 2.7⁵²⁻⁵⁴ and GraphPad Prism version 8.3.0 for Windows (GraphPad Software, San Diego,
- 690 California). Example commands and RosettaScripts protocols can be found in the
- 691 Supplementary Methods.

692 **Prediction of glycosylation sites**

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

701 SARS-Cov-2 susceptibility score calculation

Using identified ACE2 key amino acid positions 30, 83, 90, 322, and 354 in the alignment of 702 703 ACE2 across species, a global susceptibility score was calculated as the sum of the Blosum62 704 scoring matrix substitutions for the amino acid at each position compared to the human ACE2 705 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 706 RStudio. The software takes as input alignment of human ACE2 protein sequence with ACE2 of 707 708 another species of interest and provides a susceptibility score as output. Susceptibility scores of 709 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 710 Methods. 711

712 Statistical analysis

- 713 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.).

716 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

vork was supported in part by the National Institutes of Health under awards F32HL144048-01

720 (MRA), DK117147 (WC), UH3TR002097 and U01TR002383 (JPW and JAB), U19AI117905,

721 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

724 authors.

725 Author Contributions

JAB and JPW conceptualized the study; MRA, CTS, WC, MSM, JM, JPW, JAB, CM, and JAC

727 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.

732 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
- 735 available in Supplemental Methods.

736 Additional Information:

737 Supplementary Information is available for this paper.

738

739

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	GroupSim	
Acid	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.

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

748 alignment.

<u>Common name</u>	<u>Genus species</u>	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

749

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

752

753

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

756 FAALTAAQSTTEELAKTFLEKFNHEAEELSYQSSLASWNYNTNITDENVQKMNEAGAKWS 757 AFYEEQSKLAETYPLAEIHNTTVKRQLQALQQSGSSVLSAEKSQRLNTILNAMSTIYSTG 758 KACNPNNPQECLLLEPGLDDIMENSKDYNERLWAWEGWRAEVGKQLRPLYEEYVALKNEM 759 ARANSYEDYGDYWRGDYEEEWTDGYNYSRSQLIKDVEHTFTQIKPLYQHLHAYVRAKLMD 760 SYPSRISPTGCLPAHLLGDMWGRFWTNLYPLTVPFGQKPNIDVTDAMVNQSWDARRIFKE 761 AEKFFVSVGLPNMTOGFWENSMLTEPGDSOKVVCHPTAWDLGKGDFRIKMCTKVTMDDFL 762 TAHHEMGHIQYDMAYAVQPFLLRNGANEGFHEAVGEIMSLSAATPNHLKTIGLLPPGFSE 763 DSETEINFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKEQWMQKWWEMKREIVGVVEP 764 VPHDETYCDPASLFHVANDYSFIRYYTRTIYOFOFOEALCRIAKHEGPLHKCDISNSSEA 765 GKKLLQMLTLGKSKPWTLALEHVVGEKNMNVTPLLKYFEPLFTWLKEQNRNSFVGWNTDW 766 RPCADQSIKVRISLKSALGDKAYEWNDNEMYLFRSSVAYAMREYFSKVKNQTIPFVEDNV 767 WVSNLKPRISFNFFVTASKNVSDVIPRREVEEAIRMSRSRINDAFRLDDNSLEFLGIQPT 768 LSPPYQPPVTIWLIVFGVVMGVVVVGIVLLIVSGIRNRRKEQSSKK

769

771 Extended Data Figures

	1 10	20000000000	αl	200 00000	772
Homo sapiens Felis catus Panthera tigris altaica Panthera leo	MSSSSWLLLSLVAVTAA MSGSFWLLLSFAALTAA LSFAALTAA FAALTAA	OSTIEEQAKTFLDKFNH OSTTEELAKTFLEKFNH OSTTEELAKTFLEKFNH OSTTEELAKTFLEKFNH	EAEDLFYQSSLASWNY EAEELSYQSSLASWNY EAEELSYQSSLASWNY EAEELSYQSSLASWNY	NTNITEENVQ NTNITDENVQ NTNITDENVQ NTNITDENVQ NTNITDENVQ	773
Mesocricetus auratus Macaca mulatta Mus musculus Aythya fuligula Gallus gallus	MSSSSSWLLLSLVAVTTA MSGSSWLLLSLVAVTAA MSSSSWLLLSLVAVTTA MLAHVLLLCGLSAVVIP MLLHFWLLCGLSAVVIP	2SIIEEQAKTFLDKFNC 2STIEEQAKTFLDKFN 2SLTEENAKTFLNNFNC 2DVTNQ.AKMFLAEFNV DVTOE.AOTFLAEFNV	EAEDLSYQSALASWNY EAEDLFYQSSLASWNY EAEDLSYQSSLASWNY RAEDINYENSLASWDY RAEDISYENSLASWNY	NTNITEENAQ NTNITEENVQ NTNITEENAQ NTNITEETAT NTNITEETAR	774
Mustela putorius furo Sus scrofa Canis lupus familiaris Rhinolophus sinicus	MLGSSWLLLSLAALTAA MSGSFWLLLSLIPVTAA MSGSSWLLLSLAALTAA MSGSSWLLLSLAALTAA	OSTTEDLAKTFLEKFNY OSTTEELAKTFLEKFNI OST.EDLVKTFLEKFNY OSTTEDEAKMFLDKFNT	EAEELSYQNSLASWNY BAEDLAYQSSLASWTI BAEELSYQSSLASWNY KAEDLSHQSSLASWDY	NTNITDENIQ NTNITDENIQ NINITDENVQ NINITDENVQ NTNINDENVQ	775
Equus cabanus Bos Taurus Manis javanica Capra hircus Ovis aries	MSGSSWLLLSLVAVTAA MTGSFWLLLSLVAVTAA MSGSSWLLLSLVAVTAA MTGSFWLLLSLVAVTAA MTGSFWLLLSLVAVTAA	2 STTEDLAK TFLEKFNS 2 STTEEQAKTFLEKFNS 2 STSDEEAKTFLEKFNS 2 STTEEQAKTFLEKFN 5 STTEEQAKTFLEKFNS	EAEELSHQSSLASWSY EAEDLSYQSSLASWSY EAEELSYQSSLASWSY EAEDLSYQSSLASWSY EAEDLSYQSSLASWSY	NTNITDENVQ NTNITDENVQ NTNITDENVQ NTNITDENVQ NTNITDENVQ	776
Camelus dromedarius Camelus bactrianus	MSGSFWLLLSLVAVTAA MSGSFWLLLSLVAVTAA	STTEELAKT <mark>FL</mark> EEFNH STTEELAKTFLEEFNH	E <mark>AE</mark> DLSYÖSS LASW NY E <mark>AE</mark> DLSYÖSS LASW NY	NTNITDENVQ NTNITDENVQ	777
	α2 ΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩ 7 9	ао тт <u>2000</u> 80 90	α3 ηι <u>2222222</u> <u>2222</u> <u>2</u>	α4 00000000000 120	
Homo sapiens Felis catus Panthera tigris altaica Panthera tigris altaica Panthera igo Mesocricetus auratus Musea muista Mustela putorius furo Saus scrota Canis lupus familiarios Faluno (aphus familiarios Falunos caballus Bos Taurus Manis javanica Capre hircus Camelus dromedarius Camelus bactrianus	N NNN A GD KW G A G LKE O S K N NE À GA KW S A G Y E E O S K N NE À GA KW S A G Y E E O S K N NE À GA KW S A G Y E E O S K N NE À GA KW S A G Y E E O S K N NE À GA KW S A G Y E E O S K N S A GA KW S A G Y E E O S K N S A GA KW S A G Y E E O S K N N A GA KW S A G Y E E O S S	TIAOMYP FOE IONLTVK KLAKTYP AE IHNTTVK KLAKTYP AE IHNTVK KLAKTYP AE IHNTVK KLAKTYP AE IHNTVK KLAKNYS OEVONLTVK KLAKNYS OEVONLTVK AKNYS OEVONLTVK KLAKTYP AE IONPILK AKTYP BE IODPILF AKTYP BE IODFUT KLAKTYP BE IONLTVK KLAKTYP BE IONLTVK KLAKTYS BE IONLTVK KLAKTYS BE IONLTVK KLAKTYP BE IONLTVK KLAKTYS BE IONLTVK KLAKTYP BE IONLTVK KTAKTYP BE IONVTK	L0L0AC00NGSSVL5 R0L0AL00SGSSVL5 R0L0AL00SGSSVL5 R0L0AL00SGSSVL5 R0L0AL00SGSSVL5 L0L0AL00SGSSVL5 L0L0AL00SGSSVL5 L0L0AL00SGSSVL5 R0L0AL00SGSSVL5 R0LRAL00SGSSVL5 R0LRAL0HSGSSVL5 R0LRAL0HSGSSVL5 R0LAL0AL0SGSSVL5 R0LAL0HSGSSVL5 R0L0AL0 R0 R0 R0 R0 R0 R0 R0 R0 R0 R0 R0 R0 R0	DKSKRLNTIL DKSORLNTIL DKSORLNTIL DKSORLNTIL DKSORLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLNTIL DKSKRLTTVL	
	00000000 β1	β2 α5	α6	0 0000000	
Homo sapiens Felis catus Panthera tigris altaica Mesocricetus auratus Meccae muiatte Mus musculus Aythya fuligula Galius gallus Mustela putorius furo Sus scrofe Canis lugus familiaris Rhinolophus sinicus Bos Taurus Manis javanica Capra hircus Ovis aries Camelus formedarius Camelus bactrianus	130 NTMSTIYSTCKACNPNN NAMSTIYSTCKACNPNN NAMSTIYSTCKACNPNN NAMSTIYSTCKACNPNN NTMSTIYSTCKACNPNN NTMSTIYSTCKVCNPKN NTMSTIYSTCKVCNPKN NTMSTIYSTCKVCNPKN NTMSTIYSTCKVCNPKN NMSTIYSTCKVCNPKN NAMSTIYSTCKVCNPSN NAMSTIYSTCKVCNPSN NAMSTIYSTCKVCNPSN NKMSTIYSTCKVCNPSN NKMSTIYSTCKVDPIN NKMSTIYSTCKVDPIN NKMSTIYSTCKVDPIN NKMSTIYSTCKVDPIN NKMSTIYSTCKVDPIN NKMSTIYSTCKVDPIN NKMSTIYSTCKVDPIN NKMSTIYSTCKVDPIN NKMSTIYSTCKVDPIN NKMSTIYSTCKVDPIN	140 150 P CE L LL P CI D IMEN P CE L P CI D IMEN P CE L V P CI D IMEN P CE L V P CI D IMEN P CE L V P CI D IMEN P CE C LAL P CI D IMEN P CE C V P CI D IMEN P CE C V P CI D IMEN	160 SLDYND RLWAWE SWR SKDYND RLWAWE SWR SKDYND RLWAWE GWR SKDYND RLWAWE GWR SLDYND RLWAWE GWR SLDYND RLWAWE GWR SLDYND RLWAWE GWR SKDYND RLWAWE GWR SKDYND RLWAWE GWR SKDYND RLWAWE GWR SKDYND RLWAWE GWR SKDYND RLWAWE GWR SKDYND RLWAWE GWR SRDYND RLWAWE GWR	180 E VGKQLRPLY EVGKQLRPLY	
	α7 <u>0000000000000</u> 190	α8 <u>000000</u> 200 210	<u>000000000</u> 229 2	α9 20000000000 230	
Homo sapiens Felis catus Panthera ligni altaica Panthera ligni altaica Muscaca mulatta Mus musculus Aythya fuligula Gallus gallus Authya fuligula Gallus gallus Mustela putorius furo Sus scrofa Cania lupus familiaris Rhinolophus sinicus Equus caballus Bos Taurus Manis javanice Capra hircus Ovis aries Camelus dromedarius Camelus decrinaus	E E Y V V L K N E M RAN . H E E Y V A L K N E M K S K O V N E E Y V A L K N E M K A N . N E E Y V A L K N E M RAN . S E E Y V L K N E M RAN . S E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N E E Y V L K N E M RAN . N	YEDYGDYWRGDYEVNGV YEDYGDYWRGDYEESWI YDYGDYWRGDYEESWI YDYGDYWRGDYEESWI YDYGDYWRGDYESWI YDYGDYWRGDYEASGA YDYGDYWRGDYEASGA YDYGDYWRGDYESWI YDYGDYWRGDYESWI YDYGDYWRGDYESWI YGDYGDYWRGDYESWI YGDYGDYWRGDYESWI YGDYGDYWRGDYESWI YGDYGDYWRGDYEWIGA YGDYGDYWRGDYEWIGA YGDYGDYWRGDYEWIGA YGDYGDYWRGDYEWIGA YGDYGDYWRGDYEWIGA YGDYGDYWRGDYEWIGA	D G Y D Y S R G Q I I E D V E I D G Y N Y S R S Q I I K D V E D G Y N Y S R S Q I K D V E D G Y N Y S R S Q I K D V E D G Y N Y S R S Q I K D V E D G Y D N N R D Q I E D V E D G Y D N N R D Q I E D V E E Y K Y S R D Q I V Q D V S E Y K Y S R D Q I V Q D V S G D Y D Y S R N Q I I D V E G D Y D Y S R N Q I I D V E S G Y D Y S R D Q I M K D V E S G Y D Y S R D Q I M K D V E S G Y D Y S R D Q I M K D V E G D Y D Y S R D Q I M K D V E S G Y D Y S R D Q I M K D V E G D Y D Y S R D Q I M K D V E G D Y D Y S R D Q I M K D V E G D Y D Y S R D Q I M K D V E G D Y D Y S R D Q I M K D V E G D Y D Y S R D Q I M K D V E G D Y D Y S R D Q I M K D V E G D Y D Y S R D Q I M K D V E G D Y D Y S R D Q I M K D V E G D Y D Y S R D Q I M K D V E	ITF E E IKPLYE ITF I O IKPLYE ITF I O IKPLYO ITF I O IKPLYO ITF I O IKPLYO ITF E E IKPLYE ITF E E IKPLYE ITF E O IKPLYE ITF E O IKPLYE ITF I O IKPLYE ITF I O IKPLYE ITF I O IKPLYE ITF E E IKPLYE	

	0000000	00000	μ π.π	η2	η3 α10	
Homo sapiens Felis catus Panthera tigris altaica	240 HLHAYVR HLHAYVR HLHAYVR	250 AKLMNAYPSY. AKLMDTYPSR. AKLMDSYPSR.	260 ISPIGCLPA ISPTGCLPA ISPTGCLPA	270 HLLGDMWGRF HLLGDMWGRF HLLGDMWGRF	280 WTNLYSLTVPF WTNLYPLTVPF WTNLYPLTVPF	290 GOKPNIDVTDAM GOKPNIDVTDAM GOKPNIDVTDAM
Panthera leo Mesocricetus auratus Macaca mulatta Mus musculus Aythya fuligula	LHAYVR QLHAYVR HLHAYVR QLHAYVR	AKLMDSYPSK. TKLMNTYPSY. AKLMNAYPSY. RKLMDTYPSY. HRLEQAYGSQF	ISPTGCLPA ISPTGCLPA ISPTGCLPA ISPTGCLPA ISSTGCLPA	HLLGDMWGRF HLLGDMWGRF HLLGDMWGRF HLLGDMWGRF	WTNLYPLIVPE WTNLYPLIVPE WTNLYSLIVPE WTNLYPLIVPE WTNLYPLIVPY	GOKPNIDVTDAM GOKPNIDVTDAM GOKPNIDVTDAM AOKPNIDVTDAM PAKPNIDVTDAM
Gallus gallus Mustela putorius furo Sus scrofa Canis lupus familiaris Phinolophus sinicus	HLHAYVR HLHAYVR HLHAYVR HLHAYVR	HRLEQVYGSEL AKLMDAYPSR AKLMDAYPSR TKLMDTYPSY AKLMDTYPSY	INPTGCLPA ISPTGCLPA ISPTGCLPA ISPTGCLPA	HLLGDMWGRF HLLGDMWGRF HLLGDMWGRF HLLGDMWGRF	WTNLYNLTVPY WTNLYPLMVPF WTNLYPLTVPF WTNLYPLTVPF	PEKPNIDVTSAM ROKPNIDVTDAM GEKPSIDVTEAM GOKPNIDVTNAM
Equus caballus Bos Taurus Manis javanica Capra hircus	HLHAYVR QLHAYVR HLHAYVR QLHAYVR	AKLMDTYPSH. AKLMHTYPSY. AKLMDNYPSH. AKLMNTYPSY.	INPTGCLPA ISPTGCLPA ISPTGCLPA ISPTGCLPA	HLLGDMWGRF HLLGDMWGRF HLLGDMWGRF HLLGDMWGRF	WTNLYSLTVPF WTNLYSLTVPF WTNLYPLTVPF WTNLYSLTVPF	GOKPNIDVTDAM EHKPSIDVTEKM ROKPNIDVTDAM EHKPSIDVTEKM
Ovis aries Camelus dromedarius Camelus bactrianus	QLHAYVR HLHAYVR HLHAYVR	AKLMDTYPSY. AKLMDVYPSH. <u>AKLM</u> DV <mark>Y</mark> PSH.	ISPTGCLPA ISPT <mark>GCLPA</mark> ISPT <mark>GCLPA</mark>	.HLLGDMWGRF .HLLGDMWGRF .HLLGDMWGRF	WTNLYSLTVPF WTNLYSLTVPF WTNLYSLTVPF	EHKPSIDVTEKM GQKPNIDVTEAM GQKPNIDVTEAM
	η4 <u>220</u> 2 300	αll 0000000000 310	⁰ 320	α12 2020 330	ТТ 340	β4 350 350
Homo sapiens Felis catus Panthera tigris altaica Panthera leo	VDQAWDA VNQSWDA VNQSWDA VNQSWDA	ORIFKEAEKFF RRIFKEAEKFF RRIFKEAEKFF RRIFKEAEKFF	VSVGLPNMT VSVGLPNMT VSVGLPNMT VSVGLPNMT	QGFWENSMLT QGFWENSMLT QGFWENSMLT QGFWENSMLT	DPGNVQKAVCH EPGDSRKVVCH EPGNSQKVVCH EPGDSQKVVCH	PTAWDLGKGDFR PTAWDLGKGDFR PTAWDLGKGDFR PTAWDLGKGDFR
Mesocricetus auratus Macaca mulatta Mus musculus Aythya fuligula Gallus gallus	VNQGWNA VNQAWNA MNQGWDA VQKNWDA AOKNWDA	ERIFKEAEKFF ORIFKEAEKFF ERIFOEAEKFF VKIFKAAEAFF MKIFKTAEAFF	VSVGLPYMT VSVGLPNMT SSIGLYNMT ASIGLYNMT	QGFWENSMLT QGFWENSMLT QGFWANSMLT EGFWKNSMLT EGFWTNSMLT	DPGDDRKVVCH DPGNVQKVVCH EPADGRKVVCH EPTDNRKVVCH EPTDNRKVVCH	PTAWDLGKGDFR PTAWDLGKGDFR PTAWDLGHGDFR PTAWDMGKNDYR PTAWDMGKNDYR
Mustela putorius furo Sus scrofa Canis lupus familiaris Rhinolophus sinicus	VNQSWDA VNQSWDA VNQSWDA LKQGWDA	RRIFEEAETFF IRIFEEAEKFF RKIFKEAEKFF DRIFKEAEKFF	VSVGLPNMT VSIGLPNMT VSVGLPNMT VSVGLPNMT	EGFWQNSMLT QGFWNNSMLT QEFWGNSMLT EGFWNNSMLT	EPGDNRKVVCH EPGDGRKVVCH EPSDSRKVVCH EPGDGRKVVCH	PTAWDLGKRDFR PTAWDLGKGDFR PTAWDLGKGDFR PTAWDLGKGDFR
Equus caballus Bos Taurus Manis javanica Capra hircus Ovis aries	VDQSWDA ENQSWDA VNQTWDA KNQSWDA KNQSWDA	KRIFEEAEKFF ERIFKEAEKFF ERIFKEAEKFF ERIFKEAEKFF	VSVGLPNMI VSISLPYMI VSVGLPKMI VSIGLPYMI VSIGLPYMI	OGFWENSMLT OGFWDNSMLT OTFWENSMLT OGFWNNSMLT OGFWDNSMLT	EPGDGRKVVCH EPGDGRKVVCH EPGDGRKVVCH EPGDGRKVVCH	PTAWDLGKGDFR PTAWDLGKGDFR PTAWDLGKGDFR PTAWDLGKGDFR
Camelus dromedarius Camelus bactrianus	ENQS <mark>WD</mark> A ENQSWDA	KR <mark>IFKEAEK</mark> FF K <mark>RIFKEAEK</mark> FF	V <mark>SIGL</mark> PNMT V <mark>SIGL</mark> PNMT	QG <mark>FWD</mark> NSMLT QGFWDNSMLT	EPGDGRKVVCH EPGDGRKVVCH 2	PTAWDL <mark>GKGDF</mark> R PTAWD <mark>LGKGDF</mark> R
	360	α13 <u>0000000000</u> 370	<u>380</u>	η5 <u>000</u> 390	<u>000000</u> 400	α14 α15 00000000 000 410
Homo sapiens Felis catus Panthera tigris altaica Panthera leo Mesocricetus auratus	ILMCTKV IKMCTKV IKMCTKV IKMCTKV IKMCTKV	TMDDFLTAHHE TMDDFLTAHHE TMDDFLTAHHE TMDDFLTAHHE TMDNFLTAHHE	MGHIQYDMA MGHIQYDMA MGHIQYDMA MGHIQYDMA	YAAQPFLLRN YAVQPFLLRN YAVQPFLLRN YAVQPFLLRN YATOPFLLRN	GANEGFHEAVG GANEGFHEAVG GANEGFHEAVG GANEGFHEAVG GANEGFHEAVG	EIMSLSAATPKH EIMSLSAATPNH EIMSLSAATPNH EIMSLSAATPNH EIMSLSAATPNH
Macaca mulatta Mus musculus Aythya fuligula Gallus gallus	IIMCTKV IKMCTKV IKMCTKV IKMCTKV	TMDDFLTAHHE TMDNFLTAHHE TMDDFLTAHHE TMDDFLTAHHE	MGHIOYDMA MGHIOYDMA MGHIEYDMA MGHI <mark>E</mark> YDMA	YAAQPFLLRN YARQPFLLRN YSVQPFLLRD YSVQPFLLRN	GANEGFHEAVG GANEGFHEAVG GANEGFHEAVG GANEGFHEAVG	EIMSLSAATPKH EIMSLSAATPKH EIMSLSAATPQH EIMSLSAATPQH
Mustela putorius furo Sus scrofa Canis lupus familiaris Rhinolophus sinicus Equus caballus	IKMCTKV IKMCTKV IKMCTKV IKMCTKV	TMDDFLTAHHE TMDDFLTAHHE TMDDFLTAHHE TMEDFLTAHHE TMDDFLTAHHE	MGHIOYDMA MGHIOYDMA MGHIOYDMA MGHIOYDMA	YAEQPFLLRN YAIQPYLLRN YAAQPFLLRN YASQPYLLRN YAYOPYLLRN	GANEGFHEAVG GANEGFHEAVG GANEGFHEAVG GANEGFHEAVG GANEGFHEAVG	EIMSLSAATPNH EIMSLSAATPHY EIMSLSAATPNH EVMSLSVATPKH ETMSLSAATPNH
Bos Taurus Manis javanica Capra hircus Ovis aries	IKMCTKV IKMCTKV IKMCTKV IKMCTKV	TMDDFLTAHHE TMDDFLTAHHE TMDDFLTAHHE TMDDFLTAHHE	MGHIOYDMA MGHIOYDMA MGHIOYDMA MGHI <mark>O</mark> YDMA	YAAQPYLLRN YAMQPYLLRN YATQPYLLRN YATQPYLLRN YATQPYLLRN	GANEGFHEAVG GANEGFHEAVG GANEGFHEAVG GANEGFHEAVG	EIMSLSAATPHY EIMSLSAATPKH EIMSLSAATPHY EIMSLSAATPHY EIMSLSAATPHY
Cameius aromedarius Cameius bactrianus	IRMCTRV IKMCTRV 2	TMDDFLTAHHE TM <mark>DD</mark> FLTAHHE	MGHIQYDMA MGHI <mark>Q</mark> YDMA	YAIQPELLRN Y <u>AIQPELLRN</u>	GANEGFHEAVG GANEGFHEAVG	EIMSLSAATPHY EIMSLSAATPHY
Mana analan	<u>000</u> 420	TT <u>0000</u> 430	α16 00000000 44 0	00 <u>000000</u> 450	α17 000000000000 460	η6 <u>00000000</u> 470
Felis catus Panthera tigris altaica Panthera leo Mesocricetus auratus	LKTIGLL LKTIGLL LKTIGLL LKSIGLL	SPGFSEDSETE PPGFSEDSETE PPGFSEDSETE PSDFQEDNETE	INFLLKQAL INFLLKQAL INFLLKQAL INFLLKQAL	TIVGTLPFTY TIVGTLPFTY TIVGTLPFTY TIVGTLPFTY	MLEKWRWMVFK MLEKWRWMVFK MLEKWRWMVFK MLEKWRWMVFK	GEIPKEOWMOKW GEIPKEOWMOKW GDIPKEOWMOKW
Macaca mulatta Mus musculus Aythya fuligula Gallus gallus Mustela putorius furo	LKSIGLL LKSIGLL LKSLDLL LKSLDLL	SPDFQEDNETE PSDFQEDSETE EPAFQEDEETE EPTFQEDEETE PPDFSEDSETD	INFLLKQAL INFLLKQAL INFLLKQAL INFLLKQAL INFLLKQAL	TIVGTLPFTY TIVGTLPFTY TIVGTMPFTY TIVGTMPFTY TIVGTMPFTY	MLEKWRWMVFK MLEKWRWMVFR MLEKWRWMVFR MLEKWRWMVFR MLEKWRWMVFR	GEIPKDQWMKKW GEIPKEQWMKKW GEITKQEWMKQW GEITKQEWTKRW GEIPKEQWMQKW
Sus scrofa Canis lupus familiaris Rhinolophus sinicus Equus caballus	LKALGLL LKNIGLL LKTMGLL LKAIGLL	PPDFYEDSETE PPSFFEDSETE SPDFREDNETE PPDFYEDSETE	INFLLKQAL INFLLKQAL INFLLKQAL INFLLKQAL	TIVGTLPFTY TIVGTLPFTY NIVGTLPFTY TIVGTLPFTY	MLEKWRWMVFR MLEKWRWMVFR MLEKWRWMVFR MLEKWRWMVFR	GEIPKEOWMOKW GEIPKDOWMKTW GEIPKEEWMKKW GEIPKEEWMKKW
Bos Taurus Manis javanica Capra hircus Ovis aries Camelus dromadarius	LKALGLL LKALGLL LKALGLL LKALGLL	APDFHEDNETE PPDFYEDNETE APDFYEDNETE APDFYEDNETE	INFLLKQAL INFLLKQAL INFLLKQAL INFLLKQAL	TIVGTLPFTY TIVGTLPFTY TIVGTLPFTY TIVGTLPFTY	MLEKWRWMVF MLEKWRWMVFS MLEKWRWMVF MLEKWRWMVF	GEIPKQQWMEKW GEIPKQQWMEKW GEIPKQQWMEKW
Camelus bactrianus	LKALGLL	PADFYEDSETE	INFLLKQAL	TIVGTLPFTY	MLEKWRWMVF	GEIPKEOWMOKW

Homo sapiens Felis catus Panthera tigris altaica Panthera leo Mesocricetus auratus Mus musculus Aythya fuligula Gallus gallus Mustela putorius furo Sus scrofa Canis Iguas familiaris Rhinolophus sinicus Eguus caballus Bos Taurus Manis javanica Capra hircus Ovis aries Camelus dromedarius Camelus bactrianus	a18 0000000 480 WEMKREIV	β6 490 GVVEP VPHDE GVVEP V	η7 SOO YCOPASLF YCOPASLF YCOPASLF YCOPASLF YCOPASLF YCOPASLF YCOPASLF YCOPASLF YCOPASLF YCOPASLF YCOPASLF YCOPASLF YCOPACLF YCOPACLF YCOPACLF YCOPACLF YCOPACLF	α19 Ω 510 Ω HVSNDYSFIR HVANDYSFIR HVANDYSFIR HVANDYSFIR HVSNDYSFIR HVANDYSFIR HVANDYSFIR HVANDYSFIR HVAEDYSFIR HVAEDYSFIR HVAEDYSFIR HVAEDYSFIR HVAEDYSFIR HVAEDYSFIR	220	5 5 5 5 5 5 5 5
Homo sapiens Felis catus Fanthera tigris altaica Panthera tigris altaica Macaca mulatta Mustela putorius furo Sus scrofa Canis lupus familiaris Rhinolophus sinicus Equus caballus Bos Taurus Manis Javanica Capra hircus Ovis aries Camelus dromedarius Camelus bactrianus	N8 540 540 6 HKCDIS 6 HKCDIS 7 HKCDIS 7 HKCDIS 7 HKCDIS 7 HKCDIS 7 HKCDIS 8	C21 0000000 550 NSTEAGXKLL NSSEAGKKLL NSSEAGKKLL NSTEAGOXLL NSTEAGOXLL NSTEAGOXLL NSTEAGOXLL NSTEAGOXLL NSSEAGOXLL	S 6 9 S 6 9 MI RLGKSK MI TLGKSK MI TLGKSK MI TLGKSK MI RLGKSE MI SLGKSE MI SLGKSE	C22 COOCOCO 570 PWTLALEH/VV PWTLALEH/VV PWTLALEH/VV PWTLALEH/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALER/VV PWTLALER/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALEN/VV PWTLALES/VV	GAKINGNYRPIL GEKKMNYTPIL GEKKMNYTPIL GEKKMNYTPIL GARNMOVRPIL GARNMOVRPIL GARNMOVRPIL GARNMOVRPIL GIKYMNYRPIL GVKTMOVRPIL GYKNMOVRPIL GIKTMOVRPIL GIKTMOVRPIL GIKTMOVRPIL GIKTMOVRPIL GIKTMOVRPIL GIKTMOVRPIL GIKTMOVRPIL GIKTMOVRPIL GIKTMOVRPIL	a23 590 SPD S
Homo sapiens Felis catus Panthera tigris altaica Panthera leo Mesocricetus auratus Mus musculus Aythya fuligula Galius galius Mustela putorius furo Sus scrofa Canis Iguas familiaris Rhinolophus sinicus Equus caballus Bos Taurus Manis Javanica Capra hircus Ovis aries Camelus dromedarius Camelus bactrianus	Q 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		620 SIKVRISL SIKVRISL SIKVRISL SIKVRISL SIKVRISL SIKVRISL SIKVRISL SIKVRISL SIKVRISL SIKVRISL SIKVRISL SIKVRISL SIKVRISL SIKVRISL	630 KSALGDKAYE KSALGDKAYE KSALGDKAYE KSALGDNAYE KSALGANAYE KSALGANAYE KSALGEKAYE KSALGEKAYE KSALGEKAYE KSALGENAYE KSALGENAYE KSALGENAYE KSALGENAYE KSALGENAYE KSALGENAYE KSALGENAYE	6449 WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS WDNEMYLPRS WDNEMYLPRS WDSELFLPRS WDSELFLPRS WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS WNDNEMYLPRS	650 SVAYAMR CYFLK SVAYAMR BYFSK SVAYAMR BYFSK SVAYAMR YFSK SVAYAMR YFSK SVAYAMR YFSK SVAYAMR YFSS SIAYAMR YFSS SIAYAMR YFSS SVAYAMR YFSS SVAYAMR YFSL SVAYAMR YFSL SVAYAMR YFLK SVAYAMR YFLK SVAYAMR YFLK SVAYAMR YFLK
Homo sapiens Felis catus Panthera tigris altaica Panthera leo Mesocricetus suratus Macaca mulatta Mus musculus Aythya tuligula Gallus gallus Mustela putorius furo Sus scrofa Canis lupus familiaris Rhinolophus sinicus Equus caballus Bos Taurus Manis javanica Capra hircus Ovis aries Camelus bactrianus	660 VKNQMILS VKNQTIPS VKNQTIPS VKNQTIPS IKNQTVPS IKNQTVPS IKNQTVPS KKQTIPS KKQTIPS KKQTIPS KKQTIPS KKQTIPS EXETVLS VKNQTILS EXETIPS VXNQTILS	670 VEDNVWVSNI VEDNVWVSNI VEDNVWVSNI VEDNVWVSNI VEDNVWVSDI SEEDVRVSDI SEEDVRVSDI SEEDVRVSDI GAUDVRVSDI GAUDVWVSDI GAUNVWVSDI GAUNVWVSDI GEDNVWSDI SEEDVWVSDI SEENVWSDI SEENVWSDI SEENVWSDI SEENVWSDI	E S O F P R I S F NFFF F R I S F NFFF F R I S F NFFY F R I S F NFF F R I S F NFFF F R I S F NFFF F R I S F NFFF F R I S F S F F F F R I S F S S F S F S F F R I S F S F S F S F S F S F S F S F S F S	690 VTASKNVSDV VTASKNVSDV VTASKNVSDV VTASKNVSDV VTASPQNVSDV VTSPQNVSDV VSSPQNVSDV VSSPANNSDI VSSPANNSDI VTSPANNSDI VTSPANNSDI VTSPANNSDI VTSPANNSDI VTSPKNNVSDV VSSPANNSDI VTSPKNNVSDU VTSPKNNVSDU VTSPKNNVSDU	T O O T O O T D E VE KA I R R R I P VE KA I R R R R S R C VE E A I R I P P VE E A I R R S R R R V VE A I R I P R R R V VE A I R V VE A I R R I P R R R V VE A I R V VE A I R R I P R R R V VE A I R R R R R R R R R R R R R R R R R R	710 MSRSRINDAPRL MSRSRINDAPRL MSRSRINDAPRL LSRSRINDAPRL LSRSRINDAPRL ISRSRINDAPRL MSRGRIDDAPRL MSRGRIDAPRL MSRGRIDAPRL MSRGRIDAPRL MSRSRINDAPRL LCRDRINDAPCL LCRDRINDAPCL MSRSRINDAPRL

	720	730	740	750	760	770
Homo sapiens	NDNSLEFLO	IOPTLGPP	OPPVSIWLIV	FGVVMGVIVV	GIVILIET	GIRDRKKKNK
Felis catus	DDNSLEFLO	JOPTLSPP	OPPVTIWLIV	/FGVVMGVVV	GIVLLIVS	GIRNRRKNNO
Panthera tigris altaica	DDNSLEFLG	JOPTLSPP	COPPVTIWLIV	FGV <mark>V</mark> MGVV <mark>V</mark> V	GIVLLIVS	GIRNRRKNNÖ
Panthera leo	DDNSLEFLO	IOPTLSPP	OPPVTIWLIV	FGVVMGVVV	GIVLLIVS	GIRNRRKEOS
Mesocricetus auratus	DDNSLEFLO	INPTLSPP	OPPVTIWLI	FGVVMGIVV	GIIIL	GIKGRKKKNE
Macaca mulatta	NDNSLEFIC	TOTTLAPP	OSPVTTWLIV	FGVVMGVTVA	GIVYLIFT	GTRDRKKKNO.
Mus musculus	NDNSLEELO	THPTLEPP	OPPVTTWLTI	FGVVMALVVV	GITTTTTT	GIKGRKKKNE
Avthva fuliqula	DDNTLEFVO	TVPTLAAP	PPVTTWLTT	FGVVISLVVI	GVIVLIIS	GORDRKKKAKGRE
Gallus gallus	DDNTLEED	TVPTLATP	KPPVTTWTT	FGVVMSLTVT	GVIVITT	GORDKRKKARGRA
Mustela putorius furo	DDNSLEEL	TOPTIEPP	OPPVTTWLTV	TECVVMCVVVV	CTELLTES	TRNPPKNNO
Sus scrofa	DDNTLEFLO	TOPTLOPPI	EPPVTVWIT	FGVVMGLVVV	GIVVLTET	GTRDRRKKO
Canis lunus familiaris	DDNSLEEL	TOPTPOPP	FPDVTTWLTV	IF CVVMCVVVV	GIVLLTES	AT PNPPKNDO
Rhinolophus sinicus	DDNSLEELC	TOPTICPP	OPPUTTWLTU	IT CUVMANUVV	CIVULTIT	GT D D D D P T D O
Fauus caballus	DDNTLEFLO	TOPTLOPP	OPPVTVWLTZ	EGUYMGLUV	GTVVLTAT	CT D C D D KKNO
Ros Taurus	DDNSIFIC	TOPTICEP		TCVVMCVVVT	CTVVI TET	CT DND DV HDND
Manis iavanica	DDNSLEFT	TOPTLOPP	OPPVTTWLT	FGVVMGVVV	GTVVLTET	GTRDRKKKDO
Canra hircus	DDNSLEET	TOPTIPE	VED DVT TWI TT	IF CVVMCVVV	GTUVI TET	CT D D D D D KKNO
Ovicarias	DDNGLERIC	TODUTODD			GIVVLI TOT	
Camalus dromadarius	DDNSLEFLG	TODUTORPE			GIVVLI T	STRUCK KNO
Camelus uromedarius	DDNSLEFLG	TOPTLGPP		I E GI VMGL VVV	GIVVLI I	STRUKERENOU
(:smellic hactrianlic		SIVEL BOPP.		L OH ANOLA AN	GIAART FI	
Cameius bactrianus						
Cameius bactrianus		780	790	800		
Camelus bactrianus	ARSGEN	780 	790	800 99 . GEONTDDV	OTSE	
Camelus bactrianus	ARSGEN	780 1PYA	790 SIDISKGE.NN SVDLSKGE.NN	800 P.GFQNTDDV NP.GFOHADDV	QTSF	
Cameius bactrianus Homo sapiens Felis catus Panthera tioris altaica	ARSGEN ARSEEN ARSEEN	780 	790 Sidiskge.NN Svdlskge.NN	800 NP.GFQNTDDV IP.GFQHADDV NP.GFQHADDV	QTSF QTSF OTSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera leo	ARSGEN ARSEEN ARSEEN	780 4PYA 4PYA 4PYA	790 SIDISKGE.NN SVDLSKGE.NN SVDLSKGE.NN SVDLSKGE.NN	800 9 . GFQNTDDV 9 . GFQHADDV 9 . GFQHADDV	VQTSF VQTSF VQTSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera leo Mesocricetus auratus	ARSGEN ARSEEN ARSEEN 	780 	790 SIDISKGE.NN SVDLSKGE.NN SVDLSKGE.NN SKK	800 19. GFQNTDDV 19. GFQHADDV 19. GFQHADDV 10. GFQHADDV	QTSF QTSF QTSF 	
Cameius bactmanus Homo sapiens Felis catus Panthera tigris altaica Panthera teo Mesocoricetus auratus Mesocor mulata	ARSGEN ARSEEN ARSEEN TKREEN ARSEN	780 4PYA 4PYA 4PYA 4PYD	790 SIDISKGE.NN SVDLSKGE.NN SVDLSKGE.NN SVDIGKGE.SN SVDIGKGE.SN	800 NP.GFQNTDDV NP.GFQHADDV NP.GFQHADDV NA.GFLSNDDA NP.GFQNDDA	QTSF QTSF QTSF 	
Homo sapiens Felis catus Panthera tigris altaica Panthera leo Mesocricetus auratus Mesoca mulatta Muse musculus	ARSGEN ARSEN ARSEN ARSEN ARSEN ARSEN	780 	790 SIDISKGE, NN SVDLSKGE, NN SVDLSKGE, SN SVDIGKGE, SN SIDINKGE, SN SIDINKGE, SN	800 P. GFQNTDV P. GFQHADDV P. GFQHADDV A. GFLSNDDA A. GFLSNDDA A. GFQNTDDV A. GFQNTDDV	QTSF QTSF QTSF QTSF QTSF QTSF QTSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera tigris altaica Panthera leo Mesocricetus auratus Mus musculus Adthar fullouta	ARSGEN . ARSEEN ARSEEN 	780 	799 SIDISKGE.NN SVDISKGE.NN SVDISKGE.NN SVDIGKGE.SN SIDINKGE.NN SMDIGKGE.SN SMDIGKGE.SN	800 IP GFQHTDDV IP GFQHADDV IP GFQHADDV IA GFLSNDDA IP GFQNTDDV IA GFQNSDDA K GFQLSSET	QTSF 'QTSF 'QTSF QTSF QTSF QTSF QTSF	
Homo sapiens Felis catus Panthera tigris eliteica Panthera tigris eliteica Panthera teo Mesocricetus auratus Macaca mulatta Mus musculus Aythya fuligula Gallus nallus	ARSGEN ARSEN ARSEN TKREN TKREN TKREN TKREN EAESNCEN	780 VPYA VPYA VPYA VPYA VPYD VPYD VPYD	790 SIDISKGE N SVDLSKGE NN SVDLSKGE N SVDIGKGE SN SIDINKGE N MDIGKGE SN SMDIGKGE SN SMDIGKGE SN SMDIGKGE SN	800 P. GFQNTDDV P. GFQHADDV P. GFQHADDV A. GFQLSNDDA P. GFQNTDDV A. GFQLSET W. GFELSET K. GFELSET	QTSF QTSF QTSF QTSF QTSF QTSF QTSF QTSF	
Homo sapiens Pelis catus Panthera itoris altaica Panthera leo Mesocricetus auratus Macaca mulatta Mus musculus Aythya fuligula Gallus gallus Mustela qudruis fum	ARSGEN ARSEEN ARSEEN TKREEN ARSEEN TKREEN EAGSNCEAN ARSEEN	789 	790 SIDISKGE, NN SVDLSKGE, NN SVDLSKGE, SN SVDIGKGE, SN SIDINKGE, NN BMDIGKGE, SN DDGK, SN EDGR, SN	800 P. GFQNTDDV P. GFQHADDV P. GFQHADDV A. GFLSNDDA NP. GFQNTDDV A. GFQLSEDT K. GFELSET K. GFELSET K. GFELSET	QTSF QTSF QTSF QTSF QTSF QTSF QTSF QTSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera leo Mesocrioetus auratus Macaca mulatta Mus musculus Aythya fuliguia Gallus gallus Mustela putorius furo Sus serofic	ARSGEN ARSEEN ARSEEN TKREEN ARSEN EAESNCEAN ARSEEN ARSEEN ARSEEN	780 PYA PYA PYA PYA PYA PYA PYA PYA	790 SIDISKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE SN SVDLGKGE SN DDGK SN DDGK SN SVDLSKGE NN	800 IP.GPQNTDDV IP.GPQHADDV IP.GPQHADDV IP.GPQNTDDV IA.GPQNTDDV IA.GPQNSDDA IK.GPEQSET IP.GPQNDDV IS.GPQNDDV	015F 015F 015F 015F 015F 015F 015F 015F	
Homo sapiens Felis catus Panthera tigris sitaica Panthera leo Mesocricetus auratus Macaca mulatta Mus musculus Aythya fuligula Gallus gallus Mustela putorius furo Sus scrofa Casis lugus famillaria	ARSGEN ARSEEN ARSEEN TKREEN ARSEEN ARSEEN EAESNCEAN ARSEEN ARSEEN ARSEEN	780 PYA PYA PYA PYD PYD PYD PYD PYD PYD PYD PYD	790 SIDISKGE, NN SVDLSKGE, NN SVDISKGE, SN SIDINKGE, SN SIDINKGE, SN DIGKGE, SN DIGKGE, SN SVDLSKGE, SN SMDLSKGE, SN	800 P. GFQNTDDV P. GFQHADDV P. GFQHADDV A. GFLSNDDA A. GFLSNDDA A. GFQNDDV A. GFQNDDV VK. GFELSET K. GFEQSEDT VS. GFQNGDDV VS. GFQNGDDV VS. GFQNGDDV	QTSF QTSF QTSF QTSF QTSF QTSF QTSF QTSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera tigris altaica Panthera teo Mescericetus auratus Macaca mulato Museulus Aythya fuligula Gallus gallus Mustel putorius furo Sus scrofa Canis Iupus familiaris Phinolonbus einicus	ARSGEN ARSEN ARSEN TKREN ARSEN ARSEN ARSEN ARSEN ARSEN ARSEN ARSEN ARGEN	780 	790 SIDISKGE NN SVDLSKGE NN SVDLSKGE SN SVDLGKGE SN SUDIKGE SN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN	800 IP.GPONTDDV IP.GPOHADDV IP.GPOHADDV IP.GPONTDDV IA.GPONSDDA IK.GPEQSET IV.GPEQSET IV.GPONSDDI IV.GPONSDDA IV.GPEQSET IV.GPONSDDI IV.GPONSDDA IV.GPONSDDA IV.GPONSDA IV.	0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera tigris altaica Panthera leo Mascoricetus auratus Maccaa mulatta Mus musculus Adhya fuligula Gallus gallus Mustela putorius furo Sus scrofa Canis lupus familiaris Rhinolophus sinicus	ARS GEN ARSEN ARSEN TKREEN EAESNCEN ARSEN EAESNCEN ARSEN ARSEN ARSEN ARSEN ARSEN ARSEN	780 	790 SIDISKGE NN SVDLSKGE NN SVDLSKGE NN SVDIGKGE SN SIDINKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN	800 P. GFQNTDDV P. GFQHADDV P. GFQHADDV A. GFQLSNDDA A. GFQLSEDV A. GFQLSET V. GFEQSET P. GFQNVDDV S. GFQNGDDV V. GFQNGDDV P. GFQNGDDV P. GFQNGDDV P. GFQNGDDV	QTSF QTSF QTSF QTSF QTSF QTSF QTSF QTSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera tigris altaica Panthera teo Mesocricetus auratus Macsca mulatta Mus musculus Aythya fuligula Gallus gallus Mustela putorius furo Sus scrofa Canis lupus familiaris Ruinolophus sinicus Equus caballus Bos Taurus	ARSGEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN ASSEEN ARGEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN	780 	790 SIDISKGE NN SVDLSKGE NN SVDLSKGE SN SVDIGKGE SN SUDIKGE SN SUDIKGE SN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN	BOQ P.GPQNTDDV P.GPQHADDV P.GPQHADDV P.GPQNTDDV NA.GPLSNDDV NA.GPPLSEDT IK.GPPLSET IK.GPPQSET IF.GPQNGDDV P.GPQNGDDV P.GPQNGDDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGDV IF.GPQNGV IF.GPQNGV IF.GPQNGV IF.GPQNGV IF.GPQNGV IF.GPQNGV IF.GPQNGV IF.GPQNV IF.GPV IF.GPQNV IF.GPV IF.	QTSF QTSF QTSF QTSF QTSF QTSF QTSF QTSF	
Homo sapiens Felis catus Panthera tigris altaice Panthera tigris altaice Panthera tigris altaice Mascaca mulatta Muse nusculus Aythya fuligula Gallus gallus Mustela putorius funo Sus scrofa Canis lupus familiaris Rhinolophus sinicus Bos Taurus Masin jangine	ARS GEN ARSEEN ARSEEN TKREEN TKREEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN	780 	790 SIDISKGE N SVDLSKGE NN SVDLSKGE NN SVDIGKGE SN SVDIGKGE SN MDIGKGE SN MDIGKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN	800 P. GFQHADDV P. GFQHADDV P. GFQHADDV A. GFQNDDA A. GFQNDDA K. GFELSEET IP. GFQNGDDV S. GFQNGDDV S. GFQNGDDV P. GFQNGDDV P. GFQNGDDV P. GFQNGDDV P. GFQNGDV P. GFQNDV P.	QTSF QTSF QTSF QTSF QTSF QTSF QTSF QTSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera leo Mesocricetus auratus Maccaca mulatta Mus musculus Aythya fuligula Gallus gallus Mustela putorius furo Sus scrofa Canis lupus famililaris Ruinolophus sinicus Equus caballus Bos Taurus Manis javanica	ARSGEN ARSEEN ARSEEN TKREEN ARSEEN ARSEEN ARSEEN ARSEEN ARGEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN ARSEEN	780 PYA PYA PYA PYA PYD PYA PYD PYA PYA PYA PYA PYA PYA PYA PYA	799 SIDISKGE.NN SVDLSKGE.NN SVDLSKGE.NN SVDIGKGE.SN SVDIGKGE.SN SVDLSKGE.NN SVDLSKGE.NN SVDLSKGE.NN SVDLSKGE.NN SVDLSKGE.NN SVDLSKGE.NN	BOO P. GFONTDDV P. GFOHADDV P. GFOHADDV P. GFOTDV P. GFONTDDV NA. GFLSNDDA WK. GFELSE S. GFONSDDA WK. GFELSE S. GFONGDD P. GFONGDDV P. GFONGDV P. GFONGDDV P. GFONGDDV P. GFONGDDV P. GFONGDDV P. GFONGDV P. GFONGV P. G	0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera tigris altaica Panthera tigris altaica Mascaca mulata Museulus Aythya fuliquia Gallus gallus Mustel putorius furo Suss eso pallus Canis lupus familiaris Rhinolophus sinicus Bos Taurus Manis javanica Capra hircus	ARS GEN ARS EEN ARS EEN	780 PYA PYA	790 SIDISKGE NN SVDISKGE NN SVDISKGE NN SVDISKGE SN SVDIGKGE SN MDIGKGE SN MDIGKGE SN SVDISKGE NN SVDISKGE NN SVDISKGE NN SVDISKGE NN SVDISKGE NN SVDISKGE NN SVDISKGE NN SVDISKGE NN	800 P. GPQNTDDV P. GPQHADDV P. GPQHADDV P. GPQNTDDV A. GPQNSDDA K. GPEQSED V. GPQNSDDV V. GPCQSED V. GPCQSGDDV P. GPQNGDDV P. GPQNGDDV P. GPQNGDDV P. GPQNDDV V. GPQNDV V. G	QTSF QTSF QTSF QTSF QTSF QTSF QTSF QTSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera leo Mesocritetus auratus Muscaca mulatta Mus musculus Aythya fuligula Gallus gallus Mustela putorius furo Sus sorofa Canis lupus familitaris Rhinolophus sinicus Equus caballus Bos Taurus Manis javanica Capra hircus Ovis aries	ARSGEN ARSEEN	780 PYA PYA PYA PYA PYD PYD PYD PYD PYA PYA PYA PYA PYA PYA PYA PYA	799 SIDISKGE NN SVDLSKGE NN SVDLSKGE NN SVDISKGE SN SIDINKGE SN MDIGKGE SN MDIGKGE SN SVDLSKGE NN SVDLSKGE NN SVDLNSKGE NN SVDLNSKGE NN SVDLNSKGE NN SVDLNSKGE NN	BOO P. GFONTDDV P. GFONTDDV P. GFONTDDV A. GFLSNDDV A. GFLSNDDV A. GFEQSET IK. GFEQSET IK. GFEQSET IK. GFEQSEDV IS. GFONGDDV IS. GFONGDDV IS. GFONTDDV IS. GFONTDV IS. GFON	0TSF 0TSF	
Homo sapiens Felis catus Panthera tigris altaica Panthera tigris altaica Panthera tigris altaica Panthera teo Mescaca mulato Mustela putorius furo Gallus gallus Mustela putorius furo Gallus gallus Funolophus sinicus Equus caballus Bos Taurus Manis javanica Capra hircus Carles dromedanise	ARS GEN ARS EEN ARS ARS ARS ARS ARS ARS ARS ARS ARS ARS	780 PYA PYA	790 SIDISKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDIGKGE SN DDGK SN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN SVDLSKGE NN	800 P. GFQNTDDV P. GFQHADDV P. GFQHADDV P. GFQNTDDV A. GFQNTDDV A. GFQNSDDA K. GFEQSET P. GFQNGDDV P. GFQNGDDV P. GFQNGDDV P. GFQNGDDV P. GFQNGDDV S. GFQNTDDV VIS. GFQNTDV VIS. GF	0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF 0TSF	

780

- 781 **Extended Data Figure 1**: ACE2 protein sequence alignment from susceptible (orange), non-
- susceptible (green), intermediate susceptible (blue), and unknown susceptible (black) species.
- 783 MAFFT alignment with visualization using ESPript. Secondary structure elements defined based
- on human ACE2 (PDB 1r42). Spirals represent α or 3₁₀ helices and arrows represent beta
- 785 strands.

787 **A**



790

791 Extended Data Figure 2: Differences in ACE2 protein sequences and phylogenetic

792 **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

795 infection.



796

797 Extended Data Figure 3: Construction of SARS-CoV-2-RBD-ACE2 complex models for

798each species resulted in comparable high-quality models. Models were evaluated for their799 $C\alpha$ -root mean square deviation ($C\alpha$ -RMSD) as a measure of similarity to the best performing800model by predicted binding energy versus calculated protein stability. Models in the lower left801quadrant of the plots show good convergence of calculated protein stability and similarity, and802were thus selected for re-docking of SARS-CoV-2-RBD to the respective ACE2 as in Extended

803 Data Figure 4.



805

806 Extended Data Figure 4: Re-docking of SARS-CoV-2-RBD to ACE2 of different species

resulted in high-quality models. $C\alpha$ -root mean square deviation ($C\alpha$ -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.

- 811 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
- 813 (magenta), as they did not converge on the best model.



- 827 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.
- 830
- 831



Native sequence recovery

832

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.

838

840 Supplemental Methods

```
841
       R code for calculating susceptibility scores on new species
842
843
844
       # This is a Shiny web application. You can run the application by clicking
845
       # the 'Run App' button above or Ctrl-Enter on each line.
846
847
       install.packages(c("protr", "devtools", "ggplot2", "dplyr", "shiny"), dependencies = TRUE)
848
       if (!requireNamespace("BiocManager", quietly = TRUE))
849
850
           install.packages("BiocManager")
       if (!requireNamespace("Biostrings", quietly = TRUE))
851
           BiocManager::install("Biostrings")
852
853
       library(protr)
       library (Biostrings)
854
855
       library(ggplot2)
       library(dplyr)
856
857
       librarv(shinv)
       library("BiocManager")
858
859
       # Define UI for application that draws a histogram
860
       ui <- fluidPage(
861
862
            #Title of page
            titlePanel("SARS-CoV-2 Susceptibility Score Calculator"),
863
           HTML("This calculator requires an input file in FASTA format of human ACE2 amino acid
864
       sequence (accession: NP 001358344.1) aligned with ACE2 of another species of interest. This
865
       alignment can be generated using the MAFFT alignment tool found <\!a
866
       href='https://www.ebi.ac.uk/Tools/msa/mafft/'> here.</a>"),
867
            #Input protein sequence of interest
868
            fileInput(inputId = "file", label="Upload alignment file", accept = ".fasta"),
869
           tags$h2("Susceptibilty Score by Species"),
870
           plotOutput("plot", width = 310, height = 500),
871
            "*Red shaded area represents susceptible species while blue/grey represents non-susceptible",
872
873
           tags$h2("Key Residues Extracted"),
tableOutput("residues"),
874
875
876
877
878
           tags$br(),
           tags$h2("Calculated susceptibility score "),
            tags$h3(textOutput("score")),
            "Compare the above score with that of previously calculated species above"
       )
879
880
881
       # Define server logic required to draw a histogram
882
       server <- function(input, output) {</pre>
883
           output$residues <- renderTable({</pre>
884
885
                req(input$file)
                aligned.seq <- readAAMultipleAlignment(input$file$datapath)</pre>
886
887
                df.dash <- as.data.frame(AAStringSet(aligned.seq))</pre>
                human.dash <- as.character(df.dash["NP 001358344.1 angiotensin-converting enzyme 2
888
       precursor [Homo sapiens]",])
889
                human.dash <- strsplit(human.dash[1], split="")</pre>
890
                aa.positions <- c(30,83,90,322,354)
891
892
                residue.df <- as.data.frame(AAStringSet(aligned.seq, aa.positions[1], aa.positions[1]))
                j=1
893
                h=1
894
                for (x in 1:nchar(df.dash["NP 001358344.1 angiotensin-converting enzyme 2 precursor [Homo
       sapiens]",])) {
895
896
                    if (j %in% aa.positions) {
897
                        residue.df[h] <- cbind(as.data.frame(AAStringSet(aligned.seq, aa.positions[h],
898
899
       aa.positions[h])))
                        h = h+1
900
                        j = j+1
901
902
                    else if (human.dash[[1]][j] == "-") {
903
                        aa.positions = aa.positions + 1
904
                        j= j+1
905
                    }
906
                    else \{j = j + 1\}
907
                }
908
           return (residue.df)
```

```
909
            }, colnames = FALSE, rownames = TRUE)
910
911
            output$score <- renderText({</pre>
912
                 reg(input$file)
913
                 aligned.seq <- readAAMultipleAlignment(input$file$datapath)
<u>914</u>
                 df.dash <- as.data.frame(AAStringSet(aligned.seq))</pre>
915
                 human.dash <- as.character(df.dash["NP 001358344.1 angiotensin-converting enzyme 2
916
        precursor [Homo sapiens]",])
917
                 human.dash <- strsplit(human.dash[1], split="")</pre>
918
919
                 aa.positions <- c(30,83,90,322,354)
                 residue.df <- as.data.frame(AAStringSet(aligned.seq, aa.positions[1], aa.positions[1]))</pre>
920
                 j=1
921
922
                 h=1
                 for (x in 1:nchar(df.dash["NP 001358344.1 angiotensin-converting enzyme 2 precursor [Homo
923
924
        sapiens]",])) {
                     if (j %in% aa.positions) {
924
925
926
927
928
929
                          residue.df[h] <- cbind(as.data.frame(AAStringSet(aligned.seq, aa.positions[h],
        aa.positions[h])))
                          h = h+1
                          j = j+1
930
931
932
933
                     else if (human.dash[[1]][j] == "-") {
                          aa.positions = aa.positions + 1
                          j= j+1
934
935
                     else {j = j + 1}
                 }
936
937
938
939
                 seq1 <- as.data.frame(residue.df[1,])</pre>
                 seq2 <- as.data.frame(residue.df[2,])</pre>
                blosum.matrix <- data.frame(AABLOSUM62)</pre>
                 i <- 0
940
                score <- 0
<u>9</u>41
                 value <- 0
942
                 sus.score <- for(x in 1:5) {</pre>
943
                     i = i+1
944
                     value <- as.numeric(subset(blosum.matrix,</pre>
945
                                                    colnames(blosum.matrix) %in% seq1[,i],
946
                                                    rownames(blosum.matrix) %in% seq2[,i]))
947
                     score = score + value
948
                 }
949
                 print (score)
949
950
951
952
953
954
            })
            species <- c("Felis catus", "Panthera tigris altaica", "Panthera leo",
                           "Mesocriceteus auratus", "Macaca mulatta", "Mus musculus", "Aythya fuligula",
955
                           "Gallus gallus", "Mustela putoriusfuro", "Sus scrofa", "Canis lupus familiaris", "Rhinolophus sinicus", "Equus caballus", "Bos taurus", "Manis javanica",
956
957
                           "Capra hircus", "Ovis aries", "Camelus dromedarius", "Camelus bactrianus")
958
959
960
            species.scores <- c(27,27,27,23,31,11,8,8,14,21,22,31,27,19,13,19,19,27,27)</pre>
            df.species.score <- data.frame("Species" = species, "Score" = species.scores)
            df.species.score <- df.species.score[order(df.species.score$Species),]</pre>
961
             # We also add shading to represent cutoffs for susceptible and non-susceptible species
962
            output$plot <- renderPlot(ggplot(df.species.score, aes(y=Score,x='')) +</pre>
963
964
                                              geom_jitter(aes(color=Species, shape=Species), size = 2.5,
        width=0.5) +
965
966
967
        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) +
968
                                              scale_color_discrete(name="Species")+
969
                                              geom rect(aes(xmin = 0, xmax = Inf, ymin = 23, ymax = Inf,
970
        fill="blue"), alpha = 0.008, show.legend = F) +
971
972
                                              geom rect(aes(xmin = 0, xmax = Inf, ymin = -Inf, ymax = 11,
        fill="red"), alpha = 0.005, show.legend = F) +
973
                                              theme classic() +
974
                                              theme(axis.title.y = element text(size=14),
975
                                                    axis.title.x = element blank(),
976
977
                                                     legend.title = element text(size=18),
                                                    legend.text = element_text(size=12))
978
                                         )
979
        }
```

980 981 982	# Run the application shinyApp(ui = ui, server = server)
983	
984	Homology modeling of ACE2 based on the ACE2-SARS-CoV-2-RBD co-crystal structure
985	using RosettaCM
000	
900	
987	Structure and input preparation
988	For all modeling purposes Rosetta-3.12 was used.
989	Preparation of input structures using RosettaRelax:
990	/rosetta-3.12/main/source/bin/relax.default.linuxgccrelease -s 6m0j.pdb -
991	<pre>database/rosetta-3.12/main/database/ -constrain_relax_to_start_coords -</pre>
992	<pre>out:prefix relaxnstruct 25 -ex1 -ex2 -use_input_sc -ignore_unrecognized_res</pre>
993	
994	Command used for partial thread:
995	/rosetta-3.12/main/source/bin/partial thread.default.linuxgccrelease -
996	in:file:fasta felis RBD 02.fasta -in:file:alignment
997	human_felis_align_for_thread_with_RBD_02.txt -database/rosetta-
998	3.12/main/database -in:file:template_pdb ./threads/6lzg_thread_0001.pdb
999	
1000	Construction of the initial ACE2-SARS-CoV-2-RBD complex with RosettaCM
1001	Command used for executing RosettaCM
1002	
1003	/rosetta-3.12/main/source/bin/rosetta_scripts.default.linuxgccrelease
1004	<pre>@hybridize_RBD.options -\$arrayfile -nstruct 11 -out:prefix CMdatabase</pre>
1005	/rosetta-3.12/main/database/ -out:path:all ./output/ >& logfile.log
1006	
1007	RosettaScripts protocol for RosettaCM
1008	<rosettascripts></rosettascripts>
1009	<taskoperations></taskoperations>
1010	
1011	<scorefxns></scorefxns>
1012	<pre><scorefunction name="stage1" symmetric="0" weights="stage1.wts"></scorefunction></pre>
1013	<pre><reweight scoretype="atom_pair_constraint" weight="1"></reweight> </pre>
1014	<pre></pre> <pre><</pre>
1016	<pre><reweight scoretype="atom pair constraint" weight="0.5"></reweight></pre>
1017	
1018	<scorefunction name="fullatom" symmetric="0" weights="stage3.wts"></scorefunction>
1019	<reweight scoretype="atom_pair_constraint" weight="0.5"></reweight>
1020	
1021	<scorefunction name="ref2015" weights="ref2015_cart.wts"></scorefunction>
1022	
1023	

1024	<filters></filters>
1025	
1026	<movers></movers>
1027	<hybridize <="" name="hybridize" scorefxn="stage1" stage1="" th=""></hybridize>
1028	<pre>stage2 scorefxn="stage2" fa scorefxn="fullatom" batch="1"</pre>
1029	<pre>stage1 increase cycles="1.0" stage2 increase cycles="1.0" linmin only="1"</pre>
1030	disulf file="disulfide.txt">
1031	<pre>- Fragments 3mers="1u19 3.frags" 9mers="1u19 9.frags"/></pre>
1032	<template cst="" file="AUTO" pdb="RBD 6m0j.pdb" weight="1.000"></template>
1033	<detailedcontrols <="" sample_template="1" start_res="599" stop_res="792" th=""></detailedcontrols>
1034	sample_abinitio="0"/>
1035	
1036	
1037	<apply_to_pose></apply_to_pose>
1038	
1039	<protocols></protocols>
1040	<add mover="hybridize"></add>
1041	
1042	
1043	

1044 RosettaCM options:

1045

1046	#i/o
1047	-in:file:fasta felis RBD 02.fasta
1048	-parser:protocol hybridze RBD 02.xml
1049	
1050	# relax options
1051	-relax:minimize bond angles
1052	-relax:minimize bond lengths
1053	-relax:jump move true
1054	-default max cycles 200
1055	-relax:min type lbfgs armijo nonmonotone
1056	-relax:jump move true
1057	-score:weights stage3.wts
1058	-use bicubic interpolation
1059	-hybridize:stage1_probability 1.0
1060	
1061	# reduce memory footprint
1062	-chemical:exclude patches LowerDNA UpperDNA Cterm amidation SpecialRotamer
1063	VirtualBB ShoveBB VirtualDNAPhosphate VirtualNTerm CTermConnect sc orbitals
1064	pro hydroxylated case1 pro hydroxylated case2 ser phosphorylated
1065	thr phosphorylated tyr phosphorylated tyr sulfated lys dimethylated
1066	lys monomethylated lys trimethylated lys acetylated glu carboxylated
1067	cys acetylated tyr diiodinated N acetylated C methylamidated
1068	MethylatedProteinCterm

1069

1070 Stage1 weights:

1071	env	1.0
1072	pair	1.0
1073	cbeta	1.0
1074	cenpack	1.0
1075	hs_pair	2.0

ss pair	2.0	
rsigma	2.0	
sheet	2.0	
vdw	0.2	
rg	2.0	
rama	0.3	
linear d	chainbreak	2.0
atom_pai	lr_constraint	1.0
	<pre>ss_pair rsigma sheet vdw rg rama linear_c atom_pai</pre>	<pre>ss_pair 2.0 rsigma 2.0 sheet 2.0 vdw 0.2 rg 2.0 rama 0.3 linear_chainbreak atom_pair_constraint</pre>

1084

```
1085
```

```
1086 Stage2 weights
```

1087	<pre># stage2 weights for hybridization</pre>
1088	hbond sr bb 2.0
1089	hbond_lr_bb 2.0
1090	rama 0.2
1091	omega 0.2
1092	rg 2.0
1093	vdw 1.0
1094	cen_env_smooth 2.0
1095	cen_pair_smooth 1.0
1096	cbeta_smooth 1.0
1097	cenpack_smooth 1.0
1098	cart_bonded 0.05
1099	atom_pair_constraint 0.5

1100

1101 Stage3 weights

1102 # stage3 fullatom weights for hybridization 1103 METHOD WEIGHTS ref 0.16 1.7 -0.67 -0.81 0.63 -0.17 0.56 0.24 -0.65 -0.1 -1104 0.34 -0.89 0.02 -0.97 -0.98 -0.37 -0.27 0.29 0.91 0.51 1105 fa atr 0.8 1106 fa rep 0.44 1107 fa_sol 0.65 1108 fa intra rep 0.004 1109 fa pair 0.49 1110 #fa plane 0 1111 fa_dun 0.56 1112 ref 1 1113 hbond lr bb 1.17 1114 hbond sr bb 0.585 1115 hbond bb sc 1.17 1116 hbond sc 1.1 1117 0.32 p aa pp 1118 dslf_ss_dst 0.5 1119 dslf_cs_ang 2 1120 dslf_ss_dih 5 1121 dslf ca dih 5 pro_close 1122 1.0 1123 0.2 rama 1124 omega 0.5 1125 atom pair constraint 0.5

1126 coordinate constraint 0.0 1127 cart bonded 0.5 1128 RosettaRelax of ACE2-SARS-CoV-2 complex 1129 1130 Command for RosettaRelax 1131 ../rosetta-3.12/main/source/bin/relax.default.linuxgccrelease -1 \$arrayfile -1132 nstruct 1 -out:prefix relax_ -database ../rosetta-3.12/main/database/ out:path:all ./relax/ -ex1 -ex2 -constrain_relax_to_start_coords >& 1133 1134 logfile.log 1135 1136 Docking of ACE2 and SARS-CoV-2 RBD

1137 Command for RosettaDock

```
1138 ../rosetta-3.12/main/source/bin/rosetta_scripts.default.linuxgccrelease -s
1139 renum_relax_14_S_0003_0001.pdb -nstruct 5 -parser:protocol docking_full.xml -
1140 in:file:native./RBD_template/renum_relax_6m0j_0020.pdb -out:prefix dock_ -
1141 database ../rosetta-3.12/main/database/ -out:path:all ./output/
1142 @docking.options -out:prefix retest_
```

1143

1144 Options for protein-protein docking:

1145 1146	-docking # the docking option group -partners A B # set rigid body docking partners
1147 1148	-dock_pert 1 2 # set coarse perturbation parameters (degrees and angstroms)
1149 1150	<pre>-dock_mcm_trans_magnitude 0.01 # refinement translational perturbation</pre>
1151	<pre>-dock_mcm_rot_magnitude 1.0 # refinement rotational perturbation</pre>
1152 1153 1154	<pre>-run:max_retry_job 10 # if the mover fails, retry 50 times -use_input_sc</pre>
1155 1156	-ex1 # increase rotamer bins to include mean +- 1
1157 1158	-ex2 # increase rotamer bins to include mean +- 2 standard deviations

1159

1160 RosettaScripts xml-protocol for protein-protein docking

1161	<rosettascripts></rosettascripts>
1162	<scorefxns></scorefxns>
1163	<scorefunction name="ref2015" symmetric="0" weights="ref2015.wts"></scorefunction>
1164	
1165	
1166	<taskoperations></taskoperations>
1167	<initializefromcommandline name="ifcl"></initializefromcommandline>
1168	<restricttorepacking name="rtr"></restricttorepacking>

1169 1170	Restrict to residues within a distance and vector cutoff of the
1170	<pre><restricttointerfacevector <="" chain1="" chain2="" name="rtiv" num="2" pre=""></restricttointerfacevector></pre>
1172	CB_dist_cutoff="10.0" nearby_atom_cutoff="5.5" vector_angle_cutoff="75"
1173	<pre>vector_dist_cutoff="9.0" /></pre>
11/4	Fix residues known experimentally to be critical in interaction
1175	<pre>PreventkesiduesFromkepacking name="prirp" residues="11,41,345" /> </pre>
1177	<pre><filters></filters></pre>
1178	
1179	<movers></movers>
1180	MINIMIZATION MOVERS
1181	Single cycle of FastRelax to minimize backbone of docking partners
1182	<pre></pre>
1103	
1184	DOCKING MOVERS
1186	<pre><docking <="" low="score docking low" name="dock low" pre="" score=""></docking></pre>
1187	score_high="ref2015" fullatom="0" local_refine="0" optimize_fold_tree="1"
1188	<pre>conserve_foldtree="0" ignore_default_docking_task="0" design="0"</pre>
1189	task_operations="ifcl" jumps="1"/>
1190	<pre><docking <="" fullatom="1" local_refine="1" name="dock_nign" optimize_fold_tree="1" pre="" score_bigh="ref2015" score_low="score_docking_low"></docking></pre>
1192	conserve foldtree="0" design="0" task operations="ifcl" jumps="1"/>
1193	
1194	<saveandretrievesidechains allsc="0" name="srsc"></saveandretrievesidechains> Speeds the move
1195	from centroid to full atom mode
1196	<pre><interfaceanalyzermover <="" name="interface" pre="" scorefxn="ref2015"></interfaceanalyzermover></pre>
1197	pack_separated="1" pack_input="0" packstat="0" interface="A_B"/>
1198 1100	/MOVED CN
1200	<a dock_low"="" href="https://www.eks/contents/conte</th></tr><tr><th>1201</th><th></apply to pose></th></tr><tr><th>1202</th><th><protocols></th></tr><tr><th>1203</th><th>Run docking protocol</th></tr><tr><th>1204
1205</th><th><Add mover=">
1205	<add mover="dock high"></add>
1207	
1208	Minimize interface
1209	<add mover="minimize_interface"></add>
1210	<add mover="interface"></add>
1211 1212	
1212	CONTROL SCOLETAN- LELZOID //

1215 Options for Rosetta protein-protein docking

1216 # the docking option group -docking 1217 # set rigid body docking partners -partners A_B 1218 -dock pert 12 $\ensuremath{\texttt{\#}}$ set coarse perturbation parameters (degrees and 1219 angstroms) -dock_mcm_trans_magnitude 0.01 1220 # refinement translational 1221 perturbation

```
1222
           -dock mcm rot magnitude 1.0
                                          # refinement rotational perturbation
1223
       -run:max retry job 10
                                     # if the mover fails, retry 50 times
1224
      -use input sc
                                           # add the side chains from the input pdb
1225
      to the rotamer library
                                       # increase rotamer bins to include mean +- 1
1226
      -ex1
1227
      standard deviation
1228
                                       # increase rotamer bins to include mean +- 2
      -ex2
1229
      standard deviations
```

1230

- To obtain a control, the relaxed crystal structures were subjected to interface minimization. 1231
- Calculation of native sequence recovery using the RECON multistate design protocol in Rosetta 1232 1233 Command for running RECON multistate design in Rosetta:

```
1234
      ../rosetta-3.12/main/source/bin/recon.default.linuxqccrelease -s homo.pdb
1235
      panthera.pdb -nstruct 50 -out:prefix "$SLURM ARRAY TASK ID" -database
1236
      ../rosetta-3.12/main/database/ -out:path:all ./panthera/output/ -
1237
      parser:protocol multistate design.xml -ex1 -use input sc >& logfile.log
1238
```

RosettaScripts protocol for RECON multistate design: 1239

1240	<rosettascripts></rosettascripts>
1241	<scorefxns></scorefxns>
1242	<scorefunction name="ref2015_cst" weights="ref2015_cst.wts"></scorefunction>
1243	<reweight scoretype="res_type_constraint" weight="1.0"></reweight>
1244	
1245	<pre><scorefunction name="ref2015" weights="ref2015.wts"></scorefunction></pre>
1246	
1247	
1248	<taskoperations></taskoperations>
1249	Include rotamer options from the command line
1250	<initializefromcommandline name="ifcl"></initializefromcommandline>
1251	
1252	<movers></movers>
1253	Design mover to be used in multistate design
1254	<packrotamersmover <="" name="design" scorefxn="ref2015_cst" th=""></packrotamersmover>
1255	task_operations="ifcl" />
1256	
1257	Create MSDMovers to run multistate design - these different in the
1258	constraint weight, with later rounds
1259	having a higher constraint value
1260	<msdmover <="" constraint_weight="0.5" design_mover="design" name="msd1" th=""></msdmover>
1261	resfiles="consensus.resfile, consensus.resfile" />
1262	<msdmover <="" constraint_weight="1" design_mover="design" name="msd2" th=""></msdmover>
1263	resfiles="consensus.resfile, consensus.resfile"/>
1264	<pre><msdmover <="" constraint_weight="1.5" design_mover="design" name="msd3" pre=""></msdmover></pre>
1265	resfiles="consensus.resfile, consensus.resfile" />
1266	<pre><msdmover <="" constraint_weight="2" design_mover="design" name="msd4" pre=""></msdmover></pre>
1267	restiles="consensus.restile, consensus.restile" />
1268	
1269	FindConsensusSequence is needed at the end of the protocol to find a
12/0	single sequence

1271	that agrees with all target states
1272	<findconsensussequence <="" name="finish" scorefxn="ref2015" th=""></findconsensussequence>
1273	resfiles="consensus.resfile,consensus.resfile" />
1274	
1275	Analyze the resulting interface
1276	<interfaceanalyzermover <="" name="analyze" packstat="0" scorefxn="ref2015" th=""></interfaceanalyzermover>
1277	pack input="0" pack separated="1" fixedchains="A" />
1278	
1279	<filters></filters>
1280	
1281	<apply_to_pose></apply_to_pose>
1282	
1283	<protocols></protocols>
1284	Run four rounds of design
1285	<add mover="msd1"></add>
1286	
1287	<add mover="msd2"></add>
1288	
1289	<add mover="msd3"></add>
1290	
1291	<add mover="msd4"></add>
1202	
1292	Find a consensus sequence for all states
1294	<pre>/Ind a consensus sequence for all states /Idd mover="finish" /></pre>
1205	
1295	Calculate interface metrics for the final correspondence
1290	Calculate interface metrics for the final sequence
1297	Add mover- analyze //
1298	
1299	
1300	<pre></pre>
1301	

```
1302
```

1303 Calculation of native sequence recovery:

```
1304
1305
1305
1306
--resfile ../../../consensus.resfile --multiproc *pdb
1306
./calc_nat_seq_recovery.py --native ../../../output/analysis/nativ.pdb --res
1307
1308
1308
```

1309