Can ACE2 receptor polymorphism predicts species susceptibility to SARS-CoV-2?

Christian A. Devaux (christian.devaux@mediterranee-infection.com)
Aix-Marseille Université, IRD, APHM, MEPHI, IHU–Méditerranée Infection, Marseille, France

Lucile Pinault
Aix-Marseille Université, IRD, APHM, MEPHI, IHU–Méditerranée Infection, Marseille, France

Ikram Omar Osman
Aix-Marseille Université, IRD, APHM, MEPHI, IHU–Méditerranée Infection, Marseille, France

Didier Raoult
Aix-Marseille Université, IRD, APHM, MEPHI, IHU–Méditerranée Infection, Marseille, France

Keywords: COVID-19, SARS-CoV-2, Coronavirus, ACE2, multi-sequence alignment, 3-D structure analysis, electrostatic potential surface generation, angiotensin I converting enzyme 2

Posted Date: April 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-25753/v1

Version of Record: A version of this preprint was published at Frontiers in Public Health on February 10th, 2021. See the published version at https://doi.org/10.3389/fpubh.2020.608765.

Abstract

A novel severe acute respiratory syndrome coronavirus named SARS-CoV-2, emerged in China in December 2019 and recently spread worldwide causing more than 144,987 deaths from COVID-19 disease. Similar to SARS-CoV that was related to a bat-borne coronavirus, this new virus originated from *Rhinolophus affinis* bats. Because bats usually transmit their coronaviruses to intermediate animal hosts that in turn represent a source of virus able to cross the species barrier to finally infect humans, the identification of an intermediate animal reservoir was recently the subject of intense researches. At the very beginning it was claimed that a reptile (*Ophiophagus hannah*) was the intermediate host. This hypothesis was quickly eliminated to be replaced by the pangolin (*Manis javanica*) hypothesis. Yet, several other animal species were recently reported as possible intermediate hosts in between bats and humans. To determine which intermediate animal host could have infected the index case patient with SARS-CoV-2, we used multi-sequence alignment, 3-D structure analysis and electrostatic potential surface generation of the angiotensin I converting enzyme 2 (ACE2) that serves as cellular receptor for SARS-CoV-2. We report evidence that such in *silico* investigation is a powerful tool that may help identification of potential SARS-CoV-2 susceptible species and that positions K31and Y41 in the a1 ridge, N82 and N90 in the loop and a3 and K353 in loop and b5 play a major role for SARS-CoV-2 binding to ACE2.

Introduction

The recent emergence of SARS-CoV-2, responsible for a respiratory disease named COVID- 19 (Coronavirus Disease-2019), threatens public health [1,2]. SARS-CoV-2 is responsible for respiratory infections, frequently asymptomatic but sometimes responsible of pneumonia in its most severe form with a fatality rate estimated about 1%-2.5% [3]. The case fatality rate increase with age and the existence of underlying diseases [4]. Recently, Fang and colleagues [5] reported that the most distinctive comorbidities in patients who died from COVID-19 are hypertension, coronary heart diseases, cerebrovascular diseases and diabetes. This virus is spreading very rapidly worldwide and WHO has declared COVID-19 as pandemic.

SARS-CoV-2 is the 7th human coronavirus (HCoV) reported to date. Previously, the first HCoV described back in the 1960s were the HCoV-229E (*Alphacoronavirus*) and HCoV- OC43 (*Betacoronavirus* lineage 2a), two agents of common winter cold. In 2003, the coronaviruses gained in notoriety with the emergence in Asia of SARS-CoV (*Betacoronavirus* lineage 2b), proven responsible for a severe acute respiratory syndrome (SRAS) in humans with case fatality rate of 9.6% [6]. Within the next couple of years, emerged HCoV-NL63 (*Alphacoronavirus* lineage 1b) and HCoV-HKU1 (*Betacoronavirus* lineage 2a). The HCoV-HKU1 was discovered in Hong Kong. The case fatality rate of the four HCoV responsible for common winter cold was estimated to be 0.5% to 1.5% [7-9]. In contrast to SARS-CoV and HCoV-HKU1 that emerged in South-East Asia suggesting that this region is probably a hotspot for coronaviruses emergence, the Middle East Respiratory Syndrome (MERS), caused by the MERS-CoV (*Betacoronavirus* lineage 2c), was reported in Saudi Arabia in 2012. This epidemic was characterized by an extremely high

case fatality rate of 34.7% [10]. The last coronavirus known to infect humans, SARS-CoV-2 (*Betacoronavirus* lineage 2b), emerged in China in 2019 and shows 79.5% nucleotide identity with SARS-CoV-2.

Based on knowledge of virus-receptor interaction [11], we revisited here the predicted binding properties between the viral Spike (S) protein of SARS-CoV-2 and its human receptor, the angiotensin I converting enzyme 2 (ACE2) peptidase, using *in silico* analysis based on alignment of receptor protein sequences from different species and structural modeling of ACE2 receptors. We found a very good match between the *in silico* predictions of virus tropism and the species already known to be possible intermediates in between bats and humans for transmission of SARS-CoV-2. In addition we identify a possible critical position in ACE2 as being N90.

Material And Methods

ACE2 protein sequence. ACE2 protein sequences from the NCBI reference sequence database: *Rousettus leschenaultii* (GenBank: ADJ19219.1); *Rousettus leschenaultii* (GenBank: BAF50705.1); *Rousettus aegyptiacus* (NCBI Reference Sequence: XP_015974412.1); *Pteropus alecto* (NCBI Reference Sequence: XP_006911709.1); *Pteropus vampyrus* (NCBI Reference Sequence: XP_011361275.1); *Phyllostomus discolor* (NCBI Reference Sequence: XP_028378317.1); *Desmodus rotundus* (NCBI Reference Sequence: XP_024425698.1); *Miniopterus natalensis* (NCBI Reference Sequence: XP_016058453.1); *Pipistrellus abramus* (GenBank: ACT66266.1); *Eptesicus fuscus* (NCBI Reference Sequence: XP_008153150.1); *Myotis davidii* (NCBI Reference Sequence: XP_015426918.1); *Myotis lucifugus* (NCBI Reference Sequence: XP_023609438.1); *Myotis brandtii* (NCBI Reference Sequence: XP_014399780.1); *Hipposideros armiger* (NCBI Reference Sequence: XP_019522936.1); *Rhinolophus ferrumequinum* (GenBank: ADN93470.1); *Rhinolophus pearsonii* (GenBank: ABU54053.1); *Rhinolophus sinicus* (GenBank: AGZ48803.1); *Rhinolophus pusillus* (GenBank: ADN93477.1); *Rhinolophus macrotis* (GenBank: ADN93471.1); *Homo sapiens* (GenBank: BAB40370.1); *Macaca mulatta* (NCBI Reference Sequence: NP_001129168.1), *Paguma larvata* (GenBank: AAX63775.1); *Felis catus* (GenBank: AAX59005.1); *Mustela putorius furo* (NCBI Reference Sequence: NP_001297119.1); *Sus scrofa domestic* (GenBank: ASK12083.1); *Sus scrofa* (NCBI Reference Sequence: NP_001116542.1), *Rhinolophus sinicus* (GenBank: AGZ48803.1), *Manis javanica* (NCBI Reference Sequence: XP_017505752.1), *Mus musculus* (NCBI Reference Sequence: NP_081562.2), *Rattus rattus* (NCBI Reference Sequence: XP_032746145.1), *Gallus gallus* (GenBank: QEQ50331.1), *Pelodiscus sinensis* (NCBI Reference Sequence: XP_006122891.1), *Xenopus tropicalis* (NCBI Reference Sequence: XP_002938293.2), *Ophiophagus hannah* (GenBank: ETE61880.1).

Comparison of sequences. All ACE2 sequences were compared using the Clustal Omega multiple sequence alignment (EMBL-EBI bioinformatic tool; Copyright © EMBL 2020) [12].

3-D analysis and electrostatic potential surface generation. The 3-D structure of ACE2 was retrieved according to the published data (PDB : 6M1D). Three amino acids segments (30-41, 82-93 and 353-358)

from *Rhinolophus sinicus*, *Mus musculus* and *Xenopus tropicalis*

ACE2 proteins were inserted into a human ACE2 backbone sequence to determine whether or not these substitutions may change the 3-D structure of ACE2. Protein modeling for these chimeric sequences was performed using the Phyre2 server [13]. The PyMOL 1.8.0 software [14] and the Adaptative Poisson-Boltzmann Solver (APBS) tools plugin [15] was used to generate electrostatic potential surfaces of the human ACE2 and its modified chimeric versions. The red color indicates an excess of negative charges near the surface and the blue color arises from a positively charged surface.

Results

The ACE2 receptors in bats

With 1230 species, bats (order *Chiroptera*) represent the second species richness (after rodents) in the mammal world. They inhabiting a multitude of ecological niches and carry a huge number of zoonotic viruses worldwide [16,17]. Coronaviruses (CoV) makes no exception to the rule, they are largely present in bats and can be the source of epizootic and zoonosis 18,19,20. The probability for CoV to cross species barrier is higher in SouthEast Asia where bats

are sold in wild life wet markets. Different species of *Rhinolophus* bats in China carry genetically diverse SARS-like coronaviruses, some of which are direct ancestors of SARS- CoV [21]. Based on our knowledge of coronaviruses circulating in Chinese bats, it is not a surprise that SARS-CoV-2 was also considered to have originated from *Rhinolophus* bats. This turned to be confirmed by elegant results showing SARS-CoV-2 shares 96% identity with BatCoV RaTG13 from *Rhinolophus affinis* [22].

It was demonstrated that the SARS-CoV uses the angiotensin I converting enzyme 2 (ACE2) to enter human cells [23]. SARS-CoV-2 was recently reported to bind to ACE2 expressed on pneumocytes [24,25]. Several recently published papers characterized SARS-CoV-2 entry into target cells through interactions with ACE2 and serine protease TMPRSS2 priming, as well as the 3-dimensional (3-D) structures involved in these interactions between the viral S protein and ACE2 [11,26,27]. Very similar to what was described for SAR-CoV, the SARS-CoV-2 S1 spike binds the peptidase domain (PD) of ACE2 and the cleavage of ACE2 COOH-terminal segment (amino acid 697 to 716) by the transmembrane protease serine 2 (TMPRSS2) enhances the viral entry. The S1 domain of the SARS-CoV spike protein mediates ACE2 receptor binding whereas the S2 domain is a membrane-associated portion that likely undergoes postbinding 3-D structure changes allowing membrane fusion. The viral receptor binding domain (RBD) in S1 was mapped to amino acid residues 318 to 510 [28]. A co-crystal structure of ACE2 to the RBD revealed that amino acids 424 to 494 are involved in direct contact with the first a-helix and K353 and proximal residues at the N-terminus of b-sheet 5 of ACE2 [26]. These authors found that K353 and N90 are essential for infection likely due their effect on the conformation of the a-helix 1 of ACE2. A point mutation L584A in ACE2, markedly attenuated the shedding of the enzyme and favor SARS-CoV entry into target cells [29]. A soluble form of ACE2 lacking the cytoplasmic and transmembrane domain of the molecule was reported

capable of blocking the binding of SARS-CoV spike protein to ACE2 [30]. It is interesting to highlight that HCoV-NL63, SARS-CoV and SARS-CoV-2 spike proteins bind ACE2 indicating that several members of the coronavirus family have developed a preferential tropism for this receptor to enter target cells [31,32] **(Table I).**

Using multiple sequence alignment we first compared the ACE2 sequences of 18 bat species. We found that the variant ACE2 proteins perfectly matched in the phylogenetic tree according to the species of bats (**Figure 1A**). When we studied the multiple sequence alignments of ACE2 from bats and examined the regions predicted by cristallography to be the regions of contact with the Spike of the virus (**Figure 1 B**), we observed significant differences between species. *Rhinolophus* bats appeared to be appropriated candidates for replication of SARS- CoV-2 related viruses, yet a species polymorphism was observed among the *Rhinolophus* (i.e. *Rhinolophus sinicus* with K31, Y41H, N82, N90, K353 and *Rhinolophus ferrumequinum* with K31D, Y41H, N82, N90, K353). The K31D variant may possibly alter the binding of the SARS-CoV-2 spike to the ACE2 from *Rhinolophus ferrumequinum*. The ACE2 sequences from other bat species show increased amino acids substitutions at positions considered to be required for viral spike binding (e.g., *Desmodus rotundus* with K31N, Y41, N82T, N90D, K353N). It is worth noting that the three *Rousettus* and two *Pteropus* ACE2 proteins analyzed in this study were characterized by K31, Y41, N82T (*Rousettus*) or N82A (*Pteropus*), N90D, and K353. We also found that the three ACE2 proteins from *Myotis* bats were characterized by K31N, Y41H, N82T, N90, and K353, suggesting that these species are unlikely to replicate SARS-CoV-2-like ancestor viruses.

The central role played by ACE2 for interspecies virus spread

Unraveling which cellular receptor(s) are used by SARS-CoV-2 for entry should provide insights into viral transmission among species. Before SARS-CoV-2, SARS-like CoV was previously found to circulate in Chinese horseshoe bats and to spread through wild Himalayan palm-civet sold as food in Chinese wild life markets from Guangdong [33-35]. SARS-CoV was also identified in weasels and raccoons in Chinese wet markets [33,36]. With regard to SARS- CoV-2 the question remains of how it got to humans. If the transmission via reptiles was considered for some time, it was quickly eliminated to be replaced by the pangolin (*Manis javanica*) hypothesis [37]. This is an entirely possible intermediate animal reservoir, but transmission to humans could just as easily have taken place via another animal [38]. SARS- CoV-2 was reported to bind ACE2 from Chinese horseshoe bats, civet and swine but not mouse ACE2 [30]. A recent study combining phylogenetic analysis and critical site marking to predict the utilizing capability of ACE2 from 250 vertebrate species by SARS-CoV-2, reported that mammal ACE2 proteins from pangolin, cat, cow, buffalo, goat, sheep and pigeon might bind SARS-CoV-2 [39]. Prediction of host susceptibility to SARS-CoV-2 through investigation of ACE2 proteins polymorphism among species together with focused attention to amino acids expected to play a crucial role in the viral spike binding, provide important clues regarding possible intermediate hosts. A recent study that used this *in silico* approach concludes that SARS-CoV-2 could bind ACE2 from cat, dog, pangolin, and Chinese hamster 40.

We investigated the amino acids substitution in 14 species of mammals, bird, reptile and amphibian, expected to be possible intermediate hosts for SARS-CoV-2 (**Figure 2**). Beside the position K31, Y41, and K353 reported in several studies as playing a major role for SARS-CoV-2 spike binding to ACE2, our multisequence alignment suggested that species carrying an N90 are more likely to be susceptible to SARS-CoV-2 infection (it includes *Homo sapiens*, *Macaca mulatta*, *Felis catus*, *Rhinolophus sinicus*, *Manis javanica*, and *Pelodiscus sinensis*) while others should be less susceptible to infection, except if the virus adapts a second receptor for cellular binding and entry.

Amino acids K31, Y41, N90, and K353 in ACE2 are likely to confer susceptible to SARS- CoV-2

The analysis of 3-D structures of different ACE2 with respect to the amino acids found in the region 30-41, 82-93 and 353-358 was studied after designing a backbone from the *Homo sapiens* ACE2 in which the corresponding regions from *Rhinolophus sinicus*, *Mus musculus*, and *Xenopus tropicalis* species were substituted to that from human. We found (**Figure 3A**) that these substitutions did not change the global 3-D structure of the molecule. However,

when we analyzed the electrostatic potential surface of ACE2, more particularly in the regions 30-41, 82-93 and 353-358, we found that the substitution of those human ACE2 segments by the corresponding regions from *Rhinolophus sinicus*, *Mus musculus*, and *Xenopus tropicalis* species clearly altered the electrostatic pattern of the molecule (**Figure 3B**). Indeed, in the region were amino acids Y41 and K353 are located in the human ACE2, when this region was substituted by sequences from mouse or frog origins, we observed a shift from neutral to basic electrostatic surface whereas the substitution for bat sequence did not change the electrostatic charge. The electrostatic surface was also different when the region containing K31 was substituted by that from bat or frog. These modifications are likely to be sufficient to alter the interaction between SARS-CoV-2 spike and the variant ACE2.

Discussion

Soon after the discovery of SARS-CoV-2, the cell surface exopeptidase ACE2 was found to serve as viral receptor in human and the first investigation of species susceptibility to this new virus demonstrated that SARS-CoV-2 is able to use Chinese horseshoe bat, swine, but not mouse ACE2 to bind host cells [41]. Since this pioneering work, several laboratories have intended to predict the utilizing capability by SARS-CoV-2 of ACE2 from different species using amino acids sequence comparisons aimed at identifying the possible intermediate hosts of SARS-CoV-2. This was made possible after the crystallographic analyzes determining which amino acids of ACE2 are essential for the attachment of the viral spike protein [11].

Our investigation suggests that SARS-CoV-like ancestral coronaviruses have adapted the ACE2 receptor to replicate in bats. However, our analysis also suggests that probably not all bat species support SARS-CoV-like coronavirus replication. According to multisequence alignment, *Rhinolophus* bats appear to be appropriated candidates for replication of SARS- CoV-2 related viruses, yet a species polymorphism is observed among the *Rhinolophus. Rhinolophus sinicus* with K31, Y41H, N82, N90, K353 is a good

candidate for SARS-CoV-2- like virus replication whereas *Rhinolophus ferrumequinum* with K31D, Y41H, N82, N90, K353 can be predicted less susceptible to the virus. ACE2 sequences from other bat species show increased amino acids substitutions at positions considered required for viral spike binding (e.g., *Desmodus rotundus* with K31N, Y41, N82T, N90D, K353N). In species expressing variant ACE2 not suitable for virus binding another surface receptor could serves for viral entry into cells but such viruses will be less likely to cross species barriers using an

ACE2 protein as receptor in an intermediate host species. This can support the hypothesis of a long bat and virus co-evolution with bat species which replicate ACE2-tropic viruses like SARS-CoV and other species which replicate CD26-tropic viruses like MERS-CoV.

In order that a SARS-CoV-like virus leave bats to infect a intermediate host in between bat and human, infected bat must come into contact with an animal expressing an ACE2 receptor adapted to SARS-CoV-like virus binding. In agreement with other studies [38, 39,41], our *in silico* search for intermediate host species able to pass the SARS-CoV-2 to humans supports the hypothesis that species bearing K31 and K353 amino acids are more likely to be susceptible to SARS-CoV-2. For example, ACE2 from *Manis javanica*, *Mustela putorius furo* and *Felis catus,* considered SARS-CoV-2 susceptible species, show K31 and K353 amino acids whereas *Mus musculus* considered SARS-CoV-2 resistant species shows a K31N and K353H variant. A Y41 also seems to be important, yet *Rhinolophus sinicus* ACE2 expresses a Y41H variant. It may account for the requirement of an intermediate host before being able to infect humans. A position not particularly stressed in previous SARS-CoV-2 studies that appear important, is N90. Indeed the N90 that is found in *Homo sapiens* final host and *Rhinolophus sinicus* early host, is also found in *Macaca mulatta*, *Manis javanica*, *Felis catus*, and *Pelodiscus sinensis*, previously described susceptible to SARS-CoV-2 and possible intermediate host whereas N90D or N90T variants are found in the other species studied. This is also consistent with the earlier observation indicating that N90 was important for SARS- CoV binding to ACE2 [32]. However, what is surprising is the sequence of the *Paguma larvata* with K31T, Y41, N82T, N90D, and K353, since palm civet has been considered as the intermediate host for SARS-CoV and suggested to also serve as an intermediate host for SARS-CoV- 2 [40], whereas with the absence of K31, the absence of N82 and N90 (which are expected to be glycosylated thereby favoring interaction with the viral spike), palm civet appears to be an animal unlikely to be infected through ACE2. This discrepancy should be further explored. Obviously, not all the species tested are theoretically susceptible to infection by SARS-CoV-2. The ACE2 protein should express amino acids essential for the viral spike binding and variants ACE2 that lack such amino acids are not likely to allow virus binding and entry. Qiu and colleagues [39] compared the ACE2 sequences from 250 species with a specific focus on T20, K31, Y41, K68, Y83, S218, A246, K353, D355, R357, M383, P426,

T593, N636, A714, R716, and A774 and concluded that SARS-CoV-2 might bind *Manis javanica* (pangolin), *Felis catus* (cat), Bos taurus (cow), *Bubalus bubalus* (buffalo), *Capra hircus* (goat), *Ovis aries* (sheep) and *Columbia livia* (pigeon) ACE2 but not (*Mus musculus*) murine ACE2. They also suggested to paid attention to *Protobothrops mucrosquamatus*

(pallas pit viper) a common snake living in the Hubei Province of China. In their study Luan and colleagues [40], investigated 42 mammalian ACE2 proteins from the wild animal protection list of Hubei Province. In their study the authors focused attention on key amino acids K31, E35, D38, M82, and K353. According to their predictions, they considered that beside humans, the mammals whose ACE2 could bind to the S protein of SARSCoV-2 are bats (*Rhinolophus macrotis*, *Rhinolophus sinicus*, *Rhinolophus pearsonii*, *Pteropus vampyrus*, *Rousettus leschenaultii*,), pangolin (*Manis javanica*), palm civet (*Paguma larvata*), monkeys (*Macaca mulatta*, *Pan troglodytes*, *Pongo abelii, Papio Anubis, Callithrix jacchus*), cat (*Felis catus*), dog (*Canis lupus familiaris*), ferret (*Mustela putorius furo*), pig (*Sus scrofa domesticus*) among others (*Rhinopithecus roxellana*, *Mustela erminea*, *Sus scrofa*, *Equus caballus*, *Bos taurus*, *Ovis aries*, *Oryctolagus cuniculus*, *Vulpes vulpes*, *Phodopus campbelli*, *Mesocricetus auratus*, *Heterocephalus glaber*, *Ictidomys tridecemlineatus*, and *Cricetulus griseus*). The mammals whose ACE2 appeared unable to bind to S protein of SARS-CoV-2 included the *Rhinolophus ferrumequinum* bats, rat (*Rattus norvegicus*), mouse (*Mus musculus*), camel (*Camelus dromedarius*), and others (*Procyon lotor*, *Ornithorhynchus anatinus*, *Loxodonta africana*, *Erinaceus europaeus*, *Nyctereutes procyonoides*, *Suricata suricatta*, *Dipodomys ordii*, and *Cavia porcellus*). They draw particular attention to N82 amino acid in the ACE2 protein. Another study by Liu and colleagues [38], based on prediction of interactions between the S protein of SARS-CoV-2 and ACE2, that investigated monkey (*Gorilla*, *Macaca*), bat (*Rhinolophus sinicus*; *Rhinolophus pearsonii*), pangolin (*Manis javanica*), snake (*Ophiophagus hannah*), turtles (*Chrysemys picta bellii*, *Chelonia mydas* and *Pelodiscus sinensis*), and others (dog, cat, mouse), stressed a possible role as intermediate host animal reservoir for turtles. This study which focused on positions T27, F28, D30, K31, H34, D38, Y41, Q42, M82, E329, K353, G354, D355, and R357, indicated that mouse and dog ACE2 showed multiple substitutions (>5) among the 14 amino acids that retained their attention, an observation in agreement with the relative resistance of these species to infection by SARS-CoV-2. They suggested K31, Y41 and K353 to be key amino acids for viral spike binding.

Although the *in silico* studies have the advantage of being able to quickly investigate the probability of infection of a large number of species, nothing can replace the *in vivo* experimentation. Interestingly golden hamster (*Mesocricetus auratus*) and Chinese hamster (*Cricetulus griseus*) are known as animal models for SARS-CoV [42,43]. Monkeys (*Macaca mulatta*; *Macaca fascicularis* ; *Chlorocebus aethiops*) were also found to be animal models

for SARS-CoV with reports of pneumonitis in infected monkeys [44,45]. Ferret (*Mustela putorius furo*), were also used as animal model for SARS-CoV and showed productive infection [46,47]. Although mouse (*Mus musculus*) ACE2 was considered unable to bind SARS- CoV-2 spike, it was previously reported that young inbred mice supported SARS-CoV viral replication but failed to show clinical sign of disease [48,49]. A recent (not peer-reviewed) study [50], describes the investigation of the *in vivo* susceptibility of animals to replicate SARS-CoV-2. The authors claim that the virus replicated poorly in dogs, pigs, chickens and ducks but efficiently infected ferrets and cats. In addition, these authors found that the virus can be transmitted from cat to cat by respiratory droplets.

If the SARS-CoV-2 like ancestral virus from bats can meet an intermediate host bearing an ACE2 molecule to which the virus spike can bind and if this lead to productive infection of the intermediate host, then another conjunction of events must occur for the virus to pass the species barrier and infect humans. Time span between these events can be very long. Upon contact with the infected intermediate animal host, the SARS-CoV-2 can meet the ACE2 protein at the surface of human lung epithelial cells allowing infection to occur. ACE2 is expressed on both type I and II alveolar epithelial lung cells as well as epithelial cells of oral mucosa, enterocytes of the small intestine, and arterial and venous endothelial cells contributing to the COVID-19 disease [51-53]. In human, ACE2 is a 100kDa type I cell-surface glycoprotein of 805 amino acids. It is characterized by a NH_2-terminal signal peptide of 17 amino acid residues, a peptidase domain (PD) (residues 19-615) with its zinc binding metalloprotease HEXXH motif, a C-terminal Collectrin-like domain (CLD) (residues 616- 768) that includes a ferredoxin-like domain (615-726), a transmembrane region of 22 amino acid residues followed by an intracellular segment of 43 amino acid residues [54,55]. Polymorphism of ACE2 in human populations was recently well documented [56].

The question must be asked why not all coronaviruses capable of infecting humans use the ACE2 receptor to infect target cells. There is probably an initial response with MERS-CoV. Instead of binding ACE2, MERS-CoV bind the dipeptidyl peptidase 4 (DPP4)/CD26), a serine peptidase expressed by T cells [57-59]. After attachment of MRES-CoV spike (S) glycoprotein to DPP4 positive human cells the S protein undergoes proteolytic activation through the cellular serine protease TMPRSS2 and cysteine protease cathepsin L once inside endosomes [60]. It was recently reported that among fourteen characterized mutants forms of DPP4, four polymorphisms (K267E, K267N, A291P and D346-348) strongly reduce the binding and penetration of MERS-CoV into target cells and the viral replication [61]. Indeed, MERS-CoV did not emerge in China from *Rhinolophus* bats, it emerged in Saudi Arabia from *Pipistrellus* bats. it was found that MERS-CoV was closely related to Ty-BatCoV HKU4 borne by *Tylonycteris pachypus* and Pi-BatCoV HKU5 borne by *Pipistrellus abramus* [62,63]. It should be noted that a *Betacoronavirus* (BatCoV-P.davyi49/Mexico/2012) isolated from bats captured in Mexico presented 72% nucleotide identity with MERS-like CoV circulating in *Rousettus* and *Pipistrellus* bats [64]. Although we had available only a single sequence of ACE2 from *Pipistrellus* bat and three sequences of *Rousettus* to compare to other ACE2 from other bats, it should be noted that both *Pipistrellus* and *Rousettus* express a N90D variant. If this position is really critical as we suggest, one might hypothesize that this is perhaps the reason why CD26-tropic coronaviruses would have spread in *Pipistrellus abramus* and *Tylonycteris pachypus* bats, rather than ACE2-tropic coronaviruses. Before transmission to human, the ancestral MESR-like CoV from bats crossed the species barrier to infect camels. Humans were infected through camel contacts [65-67].

In conclusion, our results suggest that species carrying a sequence with K31, Y41, N90, and K353 are likely to be susceptible to infection by SARS-CoV-2 (including *Homo sapiens*, *Macaca mulatta*, *Felis catus*, *Rhinolophus sinicus*, *Manis javanica*, and *Pelodiscus sinensis*) while others should be less susceptible to infection except if the virus adapts a second receptor for cellular binding and entry. The combination of 3-D structure analysis and electrostatic potential surface was quite informative. We found that the substitution of human ACE2 regions 30-41, 82-93 and 353-358 by the corresponding regions

from *Rhinolophus sinicus*, *Mus musculus*, and *Xenopus tropicalis* species did not significantly changed the 3-D structure of ACE2 but modify the electrostatic potential surface of the molecule. These modifications are likely to be sufficient to alter the interaction of SARS-CoV-2 spike with the variants ACE2. The crystal structure analysis of ACE2 suggested the presence of several hinge regions and N-glycosylations [68], including the glycosylation of N90 considered essential for SARS- CoV-2 binding. This may explain why N90 may be very important for infection of host cells. Among the various *in vitro* antiviral activities of chloroquine described to date, it has been suggested that this molecule could prevent the glycosylation of ACE2 [69,70]. We could hypothesize that chloroquine blocks the N-glycosylation at position 90 of the ACE2 sequence, thereby preventing the attachment of SARS-CoV and SARS-CoV-2 spike to the receptor.

Declarations

Author contributions

All authors contributed to conceive the manuscript. CD performed the Clustal omega analysis and LP performed the 3-D analysis. CD designed the figures and IOO worked on the art design of figures. CD wrote the paper. DR obtained the funding and supervised the study. All authors reviewed and approved the final version of the manuscript

Funding

This work was supported by the French Government under the « Investissements d'avenir » (Investments for the Future) program managed by the Agence Nationale de la Recherche (ANR, fr: National Agency for Research), (reference: Méditerranée Infection 10-IAHU-03).

Competing Interests

The authors declare that they have no competing interests.

Ethical Approval

Not required

Acknowledgments

We thank Jean-Marc Rolain, Bernard Davoust and Aurélie Caputo for stimulating discussions

References

1. Guan W, Ni Z, Hu Y, et Clinical Characteristics of Coronavirus Disease 2019 in China. *New Engl J Med* 2020; doi: 10.1056/NEJMoa2002032

2. Zhu N, Zhang D, Wang W, et A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020, **382**:727–33.
3. Huang C, Wang Y, Li X, et Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;**395**(10223):497-506
4. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult in patients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020; Published **online** March 11, 2020; doi:10.1016/S0140-6736(20)30566-3
5. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet* 2020; Published **online** March 11, 2020 https://doi.org/10.1016/S2213-2600(20)30116-8
6. Marra MA, Jones SJ, Astell CR, et The genome sequence of the SARS-associated coronavirus. *Science* 2003; **300**:1399-1404
7. Kiyula PK, Agoti CN, Munywoki PK, et (2018). Human Coronavirus NL63 Molecular Epidemiology and Evolutionary Patterns in Rural Coastal Kenya. *J Inf Dis* 2018*,* **217**(11): 1728–1739
8. Zhang SF, Tuo JL, Huang XB, et Epidemiology characteristics of human coronaviruses in patients with respiratory infection symptoms and phylogenetic analysis of HCoV-OC43 during 2010-2015 in Guangzhou. *PLos One* 2018, **13**(1):e0191789.
9. Choi WI (2019). Comparison of the Clinical Characteristics and Mortality of Adults Infected with Human Coronaviruses 229E and OC43. *Research Square* 2019; doi: 21203/rs.2.15496/v1
10. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New Engl J Med* 2012, **367**:1814-1820
11. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science 2020;* 1126/science.abb2762
12. https://www.ac.uk/Tools/msa/clustalo/.
13. Kelley LA, Mezulis S, Yates CM, Wass MK, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 2015, **10**(6): 845-858
14. https://sourceforge.net/projects/pymol/files/pymol/1.8/
15. https://pymolwiki.org/index.php/APBS
16. Afelt A, Devaux CA, Serra-Cobo J, Frutos R. Bats, bat-borne viruses, and environmental changes. *IntechOpen* 2018, chapter 8:pp. 113-131
17. Wong ACP, Li X, Lau SKP, Woo PCY. Global epidemiology of bat *Viruses* 2019, **11**:174 ; doi:10.3390/v11020174
18. Saif, Animal coronaviruses: what can they teach us about the severe acute respiratory syndrome? Rev Sci Tech 2014, **23**: 643–660.
19. Hu B, Ge X, Wang LF, Shi Z. (2015) Bat origin of human coronaviruses. *Virology J* 2015, **12**:221.
20. Forni D, Cagliani R, Clerici M, Sironi M. Molecular Evolution of Human Coronavirus Genomes *Trend Microbiol* 2017, **25**(1): 35-48

21. Afelt A, Frutos R and Devaux C (2018) Bats, Coronaviruses, and Deforestation: Toward the Emergence of Novel Infectious Diseases? *Microbiol* 2018, **9**:702.
22. Zhou P, Yang XL, Wang XG, et A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; doi:10.1038/s41586-020-2012-7
23. Ge XY, Li JL, Yang XL, et Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 2013;**503**:535-538
24. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. bioRxiv. 2020; doi: https://doi.org/10.1101/2020.01.26.919985
25. Wang PH, Cheng Y. Increasing host cellular receptor—angiotensinconverting enzyme 2 (ACE2) expression by coronavirus may facilitate 2019-nCoV infection. bioRxiv 2020 Feb 27; doi: 1101/2020.02.24.963348
26. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 2020, 181: 1–10
27. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020, **367** (6483): 1260-1263
28. Babcock GJ, Esshaki DJ, Thomas WD , Ambrosino DM. Amino acids 270 to 510 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction with receptor. *J Virol* 2004, **78:**4552-4560
29. Jia HP, Look DC, Tan P, Shi L, Hickey M, Gakhar L, Chappel MC, Wohlford-Lenane C, McGray Jr PB. Ectodomain Shedding of Angiotensin Converting Enzyme 2 in Human Airway Epithelia. *Am J Physiol Lung Cell Mol Physiol* 2009, **297** (1): L84-96
30. Lambert D, Yarski M, Warner FJ, et Tumor Necrosis Factor- Convertase (ADAM17) Mediates Regulated Ectodomain Shedding of the Severe-acute Respiratory Syndrome- Coronavirus (SARS-CoV) Receptor, Angiotensin-converting Enzyme-2 (ACE2). *J. Biol. Chem* 2005, **280**(34):30113-30119
31. Glowacka I, Bertram S, Herzog P, Pfefferle S, Steffen I, Muench MO, Simmons G, Hofmann H, Kuri T, Pöhlmann S. Differential Downregulation of ACE2 by the Spike Proteins of Severe Acute Respiratory Syndrome Coronavirus and Human Coronavirus *J Virol* 2010, **84**(2): 1198–1205
32. Li W, Zhang C, Sui J,et Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* 2005, **24**:1634-1643
33. Guan Y; Zheng BJ; He YQ, et Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 2003, **302**: 276–278.
34. Wang M, Yan M, Xu H, et SARS-CoV infection in a restaurant from palm civet. *Emerg. Infect. Dis* 2005, **11**: 1860–1865
35. Song HD, Tu CC, Zhang GW, et Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci USA* 2005, **102**: 2430– 2435.

36. Li W, Wong SK, Li F, Kuhn JH, Huang IC, Choe H, Farzan M. Animal origins of the severe acute respiratory syndrome coronavirus: Insight from ACE2-S-protein interactions. J Virol 2006, **80**: 4211–4219
37. Liu P, Chen W, Chen JP. Viral metagenomics revealed Sendai virus and coronavirus infection of Malayan pangolins (*Manis javanica*). *Viruses* 2019, **11**(11):979.
38. Liu Z, Xiao X, Wei X, Li J, Yang J, Tan H, Zhu J, Zhang Q, Wu J, Liu Composition and divergence of coronavirus spike proteins and hostACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *J Med Virol* 2020, 1-7.
39. Qiu Y, Zhao YB, Wang Q, Li JY, Zhou ZJ, Liao CH, Ge XY. Predicting the angiotensin converting enzyme 2 (ACE2) utilizing capability as the receptor of SARS-CoV-2. *Microbes and Infection*, 2020. Pre-proof https://doi.org/10.1016/j.micinf.2020.03.003
40. Luan J, Lu Y, Jin X, Zhang L. Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochem Biophys Res Com* 2020; https://doi.org/10.1016/j.bbrc.2020.03.047
41. Zhou P, Yang XL, Wang XG, et A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020, **579**(7798): 270-273.
42. Gretebeck LM, Subbarao Animal models for SARS and MERS coronaviruses. *Curr Opin Virol* 2015, **13**: 123-129
43. Sutton TC, Subbarao Development of animal models against emerging coronaviruses: From SARS to MERS coronavirus. Virology 2015, **479**: 247-258
44. McAuliffe J, Vogel L, Roberts A, Fahle G, Fischer S, Shieh WJ, Butler E, Zaki S, St Claire M, Murphy B. Replication of SARS coronavirus administered into the respiratory tract of African Green, rhesus and cynomolgus monkeys. *Virology* 2004, **330**:8–15
45. Rowe T, Gao G, Hogan RJ, et Macaque model for severe acute respiratory syndrome. *J Virol* 2004;**78**:11401–11404.
46. ter Meulen J, Bakker AB, van den Brink EN, Weverling GJ, Martina BE, Haagmans BL. Human monoclonal antibody as prophylaxis for SARS coronavirus infection in *Lancet* 2004, **363**(9427):2139–2141
47. Chu YK, Ali GD, Jia F, Li Q, Kelvin D, Couch RC. The SARS-CoV ferret model in an infection-challenge study. *Virology* 2008, **374**(1):151–163
48. Glass W.G., Subbarao K., Murphy B., Murphy P.M. Mechanisms of host defense following severe acute respiratory syndrome-coronavirus (SARS-CoV) pulmonary infection of *J Immunol* 2004, **173**:4030–4039
49. Subbarao K, McAuliffe J, Vogel L, et al. Prior infection and passive transfer of neutralizing antibody prevent replication of severe acute respiratory syndrome coronavirus in the respiratory tract of *J Virol* 2004, **78**:3572–3577
50. Shi J, Wen Z, Zhong G et Susceptibility of ferrets, cats, dogs, and different domestic animals to SARS-coronavirus-2. 2020; https://doi.org/10.1101/2020.03.30.015347 doi: bioRxiv preprint

51. Hamming I, Timens W, Bulthuis M, Lely T, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS Coronavirus. *J Pathol* 2004, **203**(2):631-637
52. Sun P, Lu X, Xu C, Sun W, Pan B. Understanding of COVID-19 based on current evidence.*J Med Virol* Published **online** February 25, 2020. doi:10.1002/jmv.25722.
53. Xu H, Zhong L, Deng J, Peng J, Dan H, Zeng X, Li T, Chen Q. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of orla mucosa. Int J Oral Science 2020; 12: 8
54. Donoghue M, Hsieh F, Baronas E, et A novel angiotensin-converting enzyme-related carboxypeptidase. (ACE2) converts angiotensin I to angiotensin 1–9. *Circ Res* 2000; **87**:E1–9.
55. Zhang H, Wada J, Hida K, et Collectrin, a Collecting Duct-specific Transmembrane Glycoprotein, Is a Novel Homolog of ACE2 and Is Developmentally Regulated in Embryonic Kidneys. *J Biol Chem* 2001, **276**(20): 17132-17139
56. Cao Y, Li L, Feng Z, Wan S, Huang P, Sun X, Wen F, Huang X, Ning G, Wang W. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell Discovery* 2020, **6**:11
57. Raj VS, Mou H, Smits SL, et Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 2013, **495**(7440):251–254.
58. Wang N, Shi X, Jiang L, et Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res* 2013, **23**(8):986–993
59. Wagner L, Klemann C, Stephan M, von Hörsten S. Unravelling the immunological roles of dipeptidyl peptidase 4 (DPP4) activity and/or structure homologue (DASH) proteins. *Clin Exp Immunol* 2016, **184**(3): 265–283.
60. Shirato K, Kawase M, Matsuyama S. Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2. *J Virol* 2013, **87**(23):12552–12561.
61. Kleine-Weber H, Schroeder S, Krüger N, et Polymorphisms in dipeptidyl peptidase 4 reduce host cell entry of Middle East respiratory syndrome coronavirus. *Emerg Microbes Infect* 2020, **9**(1): 155–168.
62. Woo, P.C.; Wang, M.; Lau, S.K.; Xu, H.; Poon, R.W.; Guo, R.; Wong, B.H.; Gao, K.; Tsoi, H.W.; Huang, Y.; et Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. *J Virol* 2007, **81**: 1574–1585.
63. Woo PC, Lau SK, Li KS, Tsang AK, Yuen KY. Genetic relatedness of the novel human group C betacoronavirus to Tylonycteris bat coronavirus HKU4 and *Pipistrellus* bat coronavirus HKU5. *Emerg Microb Infect* 2012, **1**: e35
64. Goes LGB, Ruvalcaba SG, Campos AA, et Novel bat coronaviruses, Brazil and Mexico. *Emerg Inf Dis* 2013, **19**(10): 1711-1713
65. Wang Q, Qi J, Yuan Y, et Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* 2014,**16**(3):328-337
66. Sabir JSM, Lam TTY, Ahmed MMM, et Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science*, 2016, **351**(6268): 81- 84.

67. Anthony SJ, Gilardi K, Menachery VD, et Further evidence for bats as the evolutionary source of Middle East respiratory syndrome coronavirus. *MBio* 2017, **8**(2). pii: e00373-17. doi: 100.1128/mBio.00373-17
68. Watermeyer JM, Sewell BT, Scwager SL, Natesh R, Corradi HR, Acharya KR, Sturrock ED. Structure of Testis ACE Glycosylation Mutants and Evidence for Conserved Domain Movement. *Biochemistry* 2006, **45**(42): 12655-12663
69. Vincent MJ, Bergeron E, Benjannet S, et Chloroquine is a potent inhibitor of SARS coronavirus infection and spread. *Virol J* 2005, **2**:69.
70. Devaux CA, Rolain JM, Colson P, Raoult New insights on the antiviral effects of chloroquine against coronavirus: what to expect for COVID-19? *Int J Antimicrob Agents* 2020; doi: 10.1016/j.ijantimicag.2020.105938

Table

Table I: Viral tropism

Receptor	Virus	Primary Site of Disease	Other Organs
ACE2	SARS-CoV-1	Lower respiratory tract	Multi-organ failure
	SARS-CoV-2	Lower respiratory tract	Multi-organ failure
	HCoV-NL63	Upper respiratory tract	-
DPP4/CD26	MERS-CoV	Lower respiratory tract	Myocardistis, renal failure
CD13	HCoV-229E	Upper respiratory tract	Gastrointestinal
HLA Class I	HCoV-OC43	Upper respiratory tract	Gastrointestinal
	HCoV-HKU1	Upper respiratory tract	Gastrointestinal

Figures

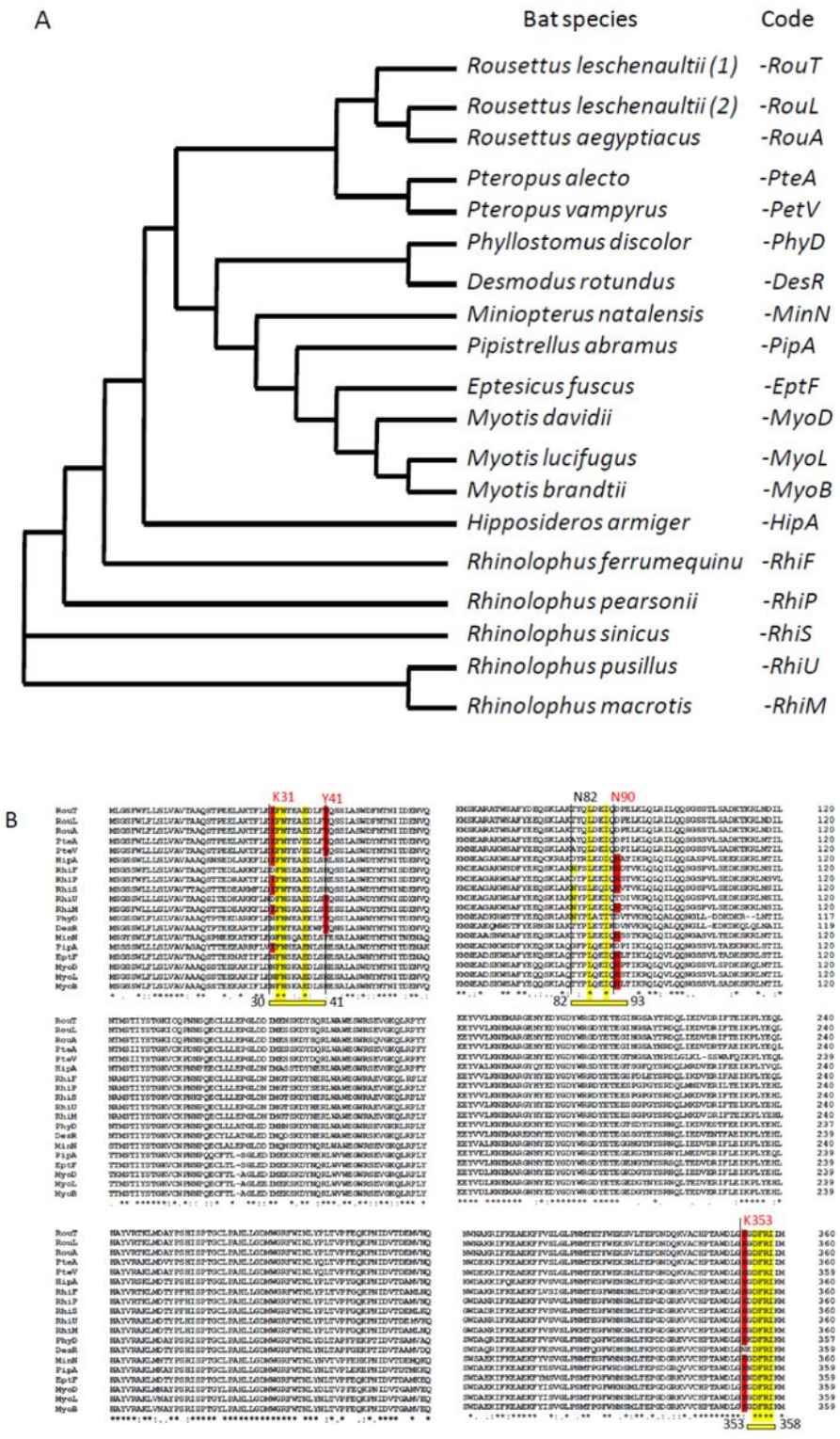

Figure 1

Bat ACE2 sequences alignment. The ACE2 protein sequences from 18 species of bats were obtained from the NCBI reference sequence database: Rousettus leschenaultii; Rousettus aegyptiacus; Pteropus alecto; Pteropus vampyrus; Phyllostomus discolor; Desmodus rotundus; Miniopterus natalensis; Pipistrellus abramus; Eptesicus fuscus; Myotis davidii; Myotis lucifugus; Myotis brandtii; Hipposideros armiger; Rhinolophus ferrumequinum; Rhinolophus pearsonii; Rhinolophus sinicus; Rhinolophus pusillus;

Rhinolophus macrotis. Clustal Omega multiple sequence alignment (EMBL-EBI bioinformatic tool; Copyright © EMBL 2020) was used to compare the ACE2 protein sequences of these mammals considered at the origin of human coronaviruses. 1A: Phylogenetic tree of bat ACE2 sequences built using the Clustal Omega multiple sequence alignment program. The short code is used in figure 1B. 1B. Sequences alignment of bat ACE2 N-terminal (amino acids 1 to 360 of 805) protein sequences. Some of the amino acids important for viral tropism are in red (previous studies showed that residues 31, 41, and regions 82-84 and 353-357 are important for viral spike binding). Within the regions considered important for the interaction with the spike of SARS-CoV-2, the conserved amino acids are in yellow.

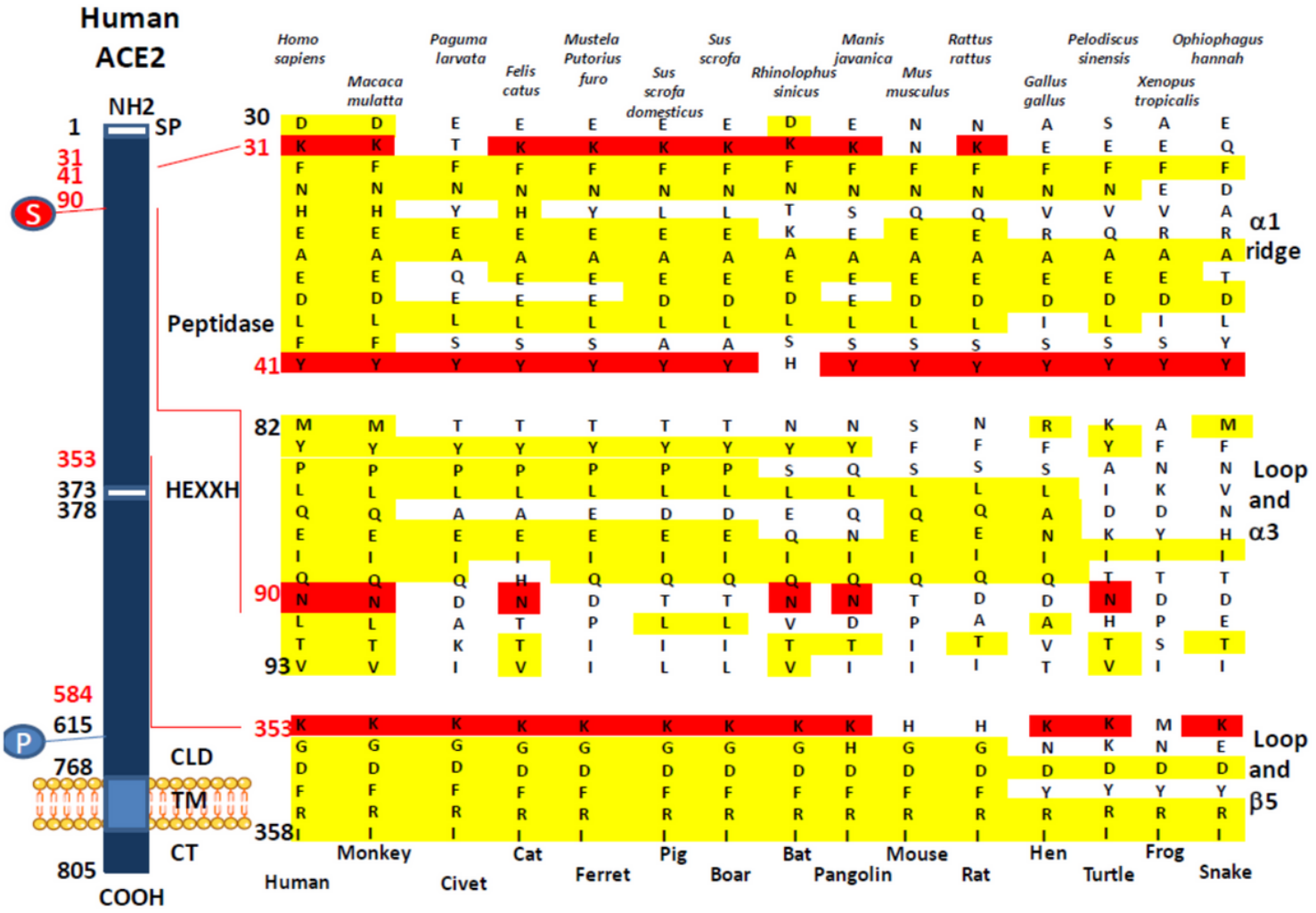

Figure 2

Comparison of the Homo sapiens ACE2 protein sequence and sequences from different mammals, bird, amphibian, and reptile. A schematic representation of the cell surface human ACE2 molecule and its major domains is drawn on left side of the figure. The amino acids position are in black. Some of the amino acids considered important for viral tropism are in red. S=Sugar; P=Phosphorylation. A comparison of ACE2 sequences from 15 different species using Clustal Omega multiple sequence alignment, is shown in the right side of the figure. All sequences were numbered according to amino acids position on the Homo sapiens ACE2 protein. All sequences were obtained from the NCBI reference sequence database. They include Macaca mulatta (monkey), Paguma larvata (palm civet), Felis catus

(cat), Mustela putorius furo (ferret), Sus scrofa domestic (pig), Sus scrofa (boar), Rhinolophus sinicus (bat), Manis javanica (pangolin), Mus musculus (mouse), Rattus rattus (rat), Gallus gallus (hen), Pelodiscus sinensis (turtle), Xenopus tropicalis (frog), Ophiophagus hannah (snake)

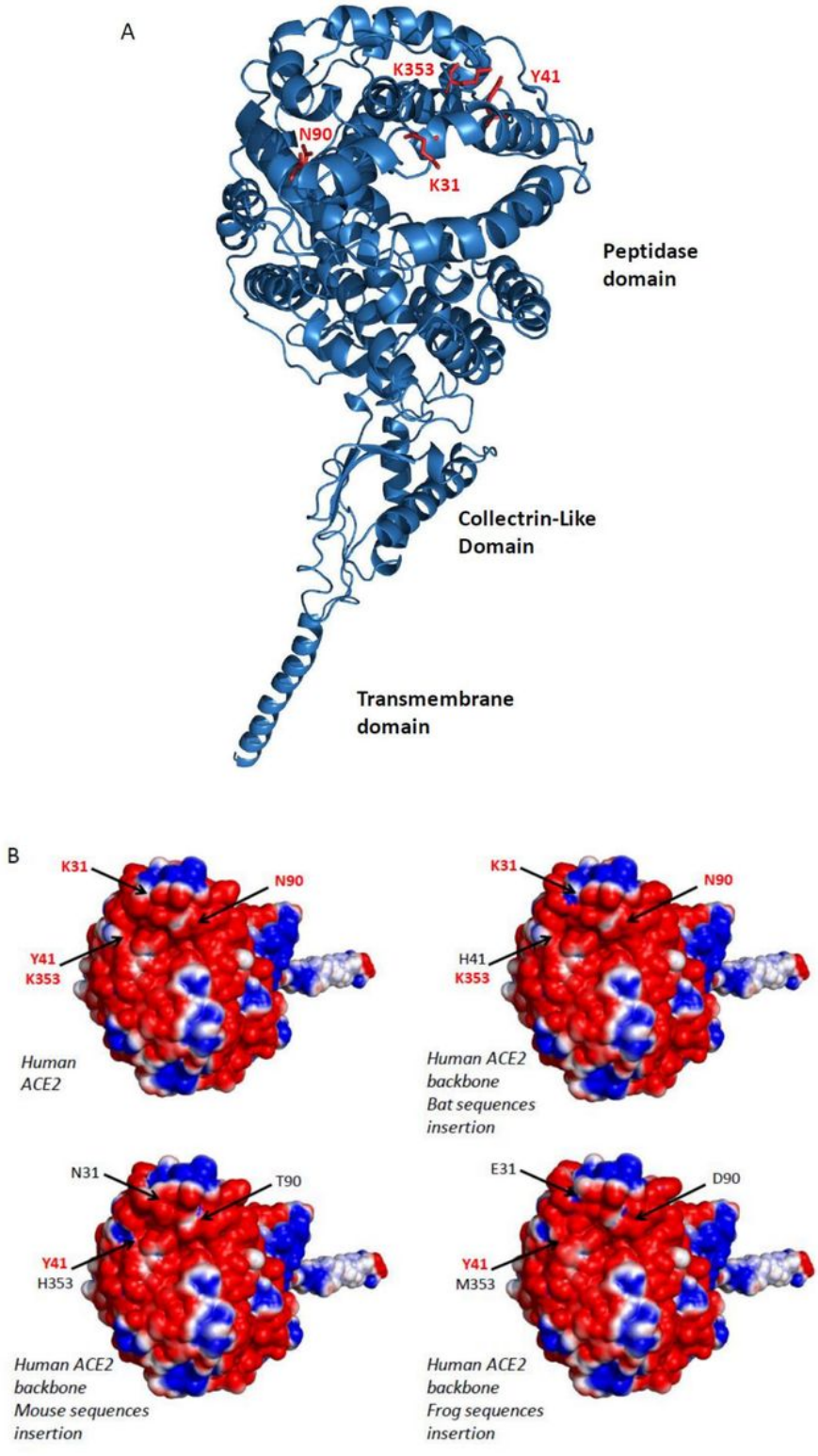

Figure 3

3-D model of ACE2. 3A. Variants amino acids segment (30-41, 82-93 and 353-358) from Rhinolophus sinicus, Mus musculus, and Xenopus tropicalis species were superimposed on the Homo sapiens ACE2 3-

D structure using the Phyre2 server. This model lacks the cytoplasmic tail of ACE2. 3B An electrostatic potential surface was generated using the PyMOL 1.8.0 software along with the APBS tool plugin. Upper left panel is a model of Homo sapiens ACE2 extracellular region with its electrostatic potential distribution (red=acidic; white=neutral; blue=basic). The upper right and the two lower images represent simulation in which the α1 ridge, loop and α3, and loop and α5 (see figure 2 B) sequences from Rhinolophus sinicus (bat), Mus musculus (mouse), and Xenopus tropicalis (frog) were substituted to the corresponding human sequences in an Homo sapiens ACE2 backbone. The locations of amino acids 31, 41, 90 and 353 are indicated by arrows.