Structural and functional modelling of SARS-CoV-2 entry in animal models

Greg N. Brooke[1,2,3] & **Filippo Prischi**[1,2,3]

[1] School of Life Sciences, University of Essex, Colchester, CO4 3SQ, UK

[2]contributed equally

[3]corresponding authors: gbrooke@essex.ac.uk; fprischi@essex.ac.uk.

Abstract

SARS-CoV-2 is the novel coronavirus responsible for the outbreak of COVID-19, a disease that has spread to over 100 countries and, as of the 13 May 2020, has infected over 4 million people. Despite the urgent need to find effective therapeutics, research on SARS-CoV-2 has been affected by a lack of suitable animal models. To facilitate the development of medical approaches and novel treatments, we compared the ACE2 receptor and TMPRSS2 protease usage of the SARS-CoV-2 Spike glycoprotein in human and in a panel of animal models, i.e. guinea pig, dog, cat, rat, rabbit, ferret and mouse. Here we showed that ACE2, but not TMPRSS2, has a higher level of sequence variability in the Spike protein interaction surface, which greatly influences Spike protein binding mode. Comparison of SARS-CoV and SARS-CoV-2 S proteins bound the ACE2 receptors showed that the SARS-CoV-2 Spike glycoprotein has adapted to bind the human, but not rodents, ACE2 with high affinity. In contrast, we did not detect species-specific adaptation for TMPRSS2. Analysis of binding modes and protein contacts indicates that ferrets are the most suitable model for the study of inhibitory antibodies and small molecules targeting the SARS-CoV-2 Spike protein interaction with ACE2. Since TMPRSS2 is similar across species, our data also suggest that transgenic animal models expressing human ACE2, such as the K18-hACE2 mouse, are also likely to be useful models for studies investigating viral entry.

Introduction

Coronaviruses are a large family of viruses that can cause respiratory diseases in humans. These can be mild, for example the common cold, but some coronaviruses have caused severe respiratory disease outbreaks in recent years. This family of viruses were found to be the cause of the 2002 Severe Acute Respiratory Syndrome (SARS coronavirus, SARS-CoV) and 2012 Middle East Respiratory Syndrome (MERS coronavirus, MERS-CoV) outbreaks [1]. In December 2019, a novel coronavirus (SARS-CoV-2) was identified in Wuhan city of Hubei Province (China) in patients that had developed viral pneumonia, called COVID-19 [2].

SARS-CoV-2 is a positive-sense single-stranded RNA virus, that belongs to the *β-coronaviruses* family along with SARS and MERS [3]. Sequencing of the genome of SARS-

CoV-2 has demonstrated that it is closely related to coronaviruses isolated from bats. Indeed, these analyses have shown that the genome of SARS-CoV-2 has 96.1% sequence similarity with SARSr-Ra-BatCoV-RaTG13, identified in *Rhinolophus affinis* bats captured in Pu'er (China) in 2013 [2, 4]. For this reason, it has been proposed that bats have acted as an ecological reservoir for SARS-CoV-2. However, since humans have limited contact with bats, it is believed that SARS-CoV-2 passed through an intermediate host before spilling over into the human population. It remains unclear as to which species may have acted as the intermediate host. In the case of SARS, bats were also likely to have been the ecological reservoir and farmed Civet cats were suggested to have been the intermediate species, although several studies have disputed this [5].

The SARS-CoV-2 genome contains five genes that code for four structural proteins - *spike* (S), *envelope* (E), *membrane* (M) and *nucleocapsid* (N) - and 16 non-structural proteins [6]. Viral entry into human cells is mediated by an interaction between the S glycoprotein and the human Angiotensin-Converting Enzyme 2 (ACE2) receptor [7]. ACE2 is a metalloprotease that lowers blood pressure by catalysing the hydrolyses of angiotensin II [8]. However, ACE2 enzymatic activity is not related, or needed, for SARS-CoV-2 entry into the host cells. Importantly, cells lacking ACE2 are not susceptible to SARS-CoV-2 infection [9]. Cryo–electron microscopy analysis of human ACE2 bound to the neutral amino acid transporter B0AT1 and the Receptor Binding Domain (RBD) of the SARS-CoV-2 S glycoprotein showed that ACE2-B0AT1 is a complex of two heterodimers that interact with two S protein trimers [10]. Detailed structural data have also shown that the S protein binds ACE2 with high affinity (~15 nM). This is 10-20 times higher than the affinity between the SARS-Cov S protein and ACE2, which likely explains the high infectivity of SARS-CoV-2 [11].

Upon binding of the S protein to ACE2, the S protein goes through a conformational change that exposes a cleavage site between the S1 and the S2 domains, which is cleaved by the Transmembrane Serine Protease 2 (TMPRSS2) [12]. Cleavage exposes the fusion peptide on S2, promoting endocytic entry of the virus [13]. TMPRSS2 is a type II transmembrane serine proteases (TTSPs) which has been found to co-express, co-localise and interact with ACE2 [12]. TMPRSS2 belongs to the trypsin (S1) fold subfamily, which are characterised by a highly conserved catalytic serine protease domain stabilised by three intradomain disulphide bonds. The peptidase S1 domain contains the histidine, aspartate, and serine residues (the catalytic triad) necessary for enzymatic activity [14]. Interestingly, studies on SARS-CoV have shown that the binding of the S protein to ACE2 also induces cleavage of ACE2 by TMPRSS2, and it has been suggested that the SARS-S-induced shedding of ACE2 may increase uptake of viral particles [15].

One of the major challenges for the study of SARS-CoV-2, and for the development of effective COVID-19 vaccines and treatments, is the lack of appropriate animal models.

Multiple animals have been shown to be experimentally susceptible to SARS-CoV (e.g. macaques, cats, ferrets, guinea pigs and civet cats) [16-20]. Similarly, SARS-CoV-2 has been shown to infect multiple animal species. For example, Shi et al. [21] found that SARS-CoV-2 can replicate in dogs, pigs, chicken and ducks, although viral replication in these animals is relatively weak. In contrast, the same group found that the virus can replicate efficiently in ferrets and cats.

It remains to be fully elucidated as to why infection rates differ between species, but structural differences in the viral entry receptors are likely to be important. Here we have used a combination of bioinformatics approaches to compare the binding of the SARS-CoV and SARS-CoV-2 S proteins to ACE2 and the structure of TMPRSS2 in a selected group of animal models; namely mouse, rat, guinea pig, rabbit, ferret, cat and dog. Our results show that ferret represents the most promising animal model for the study of ACE2 inhibitors.

Methods

ACE2 structures preparation and docking

The 3D structure of the SARS-CoV-2 and SARS-CoV RBD, in complex with human ACE2 (hACE2), were retrieved from the RCSB Protein Data Bank (PDB ID 6M17 and 2AJF respectively [10, 22]). FASTA sequences were retrieved from NCBI (Table 1). ACE2 homology models were generated using Swiss Model [23] and the hACE2 structure as a template. GROMACS 2019.3 [24] with AMBER99SB-ILDN force field was used to resolve high energy intramolecular interaction and remove modelling biases before docking simulations. Structures were centred in a cubic box filled with TIP3P water molecules and counter ions. Simulations were run applying periodic boundary conditions. The energy of the system was minimised with 10,000 steps using a steepest descent algorithm and equilibrated by running 100 ps of NVT (using V-rescale temperature coupling with tau-t of 0.1) and 100 ps NPT (applying Berendsen pressure coupling setting a tau-p of 0.5). ACE2 docking simulations with the RBD of the SARS-CoV-2 s glycoprotein were performed using the web server version of HADDOCK [25] (https://haddock.science.uu.nl). The docking simulations were driven using the binding interface derived from the RBD-ACE2 structure (PDB ID 6M17 and 2AJF) using PDBePISA [26]. ACE2-RBD complex structures were compared using PDBePISA and PyMOD2.0 [27]. ACE2 N-glycosylation sites were retrieved from UniProt [28].

TMPRSS2 structures preparation

TMPRSS2 FASTA sequences were retrieved from NCBI (Table 1). The human TMPRSS2 (hTMPRSS2) model was generated using I-TASSER [29]. This structure was then used as a

template for homology modelling using Swiss Model [23]. Structures were compared and analysed using PyMOD2.0.

Results

ACE2 interaction with the SARS-Cov-2 spike protein differs between species

ACE2 is a zinc carboxypeptidase type I transmembrane protein, with an extracellular N-terminal peptidase domain (PD) and a cytosolic C-terminal collectrin-like domain (CLD) (Figure S1). The receptor binding domain (RBD) of the SARS-CoV-2 S protein binds directly to the ACE2 PD, and analysis of the crystal structure of this complex shows that the interaction is mostly driven by polar interactions (Table 2). Of particular interest are the hydrogen bond between ACE2 E35 and S protein RBD Q493, the salt bridge between ACE2 D30 and RBD K417, and the hydrophobic interaction between ACE2 M82 and RBD F486. Interestingly, Q493, K417 and F486 are not conserved between SARS-CoV and SARS-CoV-2, and these differences are linked to the higher affinity of the SARS-CoV-2 S protein for ACE2 [30].

In order to predict if the SARS-CoV-2 S protein binds ACE2 of other animal species, focusing mostly on laboratory model systems, we generated homology models for cavACE2, dogACE2, catACE2, ratACE2, rabACE2, ferACE2, musACE2 (Table1, Figure 1A). Alignment of the ACE2 sequences from these species, revealed a high conservation, with a sequence identity between 77.5 % and 85.2% (Supplemental Figure S1-2). This allowed us to produce reliable ACE2 PD models via homology modelling using the hACE2 as a template. We then ran docking simulations between the ACE2 PD models and the SARS-CoV-2 S protein RBD to generate optimised complexes. A docking simulation using hACE2 as a control was also performed. The presence of a similar network of interactions in the docking output for the hACE2 simulation compared to the one observed in the EM structure was used to validate the approach adopted (Table 2).

Overall the hydrophobic contributions that stabilise the RBD-ACE2 complex are similar in all models, with the ferACE2-RBD having a slightly higher number of hydrophobic contacts (Table 2). Interestingly, M82 in the hACE2 is not conserved across species (Supplemental Figure S1), and only the catACE2 and musACE2 form van der Waals interaction between residue 82 and the RBD F486 (Supplemental Figure S3). However, RBD F486 is in contact, in all complex structures, with a relatively hydrophobic patch formed by the ACE2 residues 28, 79 and 83 (Table 2 and Supplemental Figure S3). Comparison of the surface electrostatic potentials of the ACE2 models identified a similar distribution of charges on the α1 helix, α2 helix and β3-β4 loop across species (Figure 1A).

Differences between the structure of hACE2 and musACE2 have been previously described to explain why SARS-CoV is a mild infection in mice [31]. The most strikingly difference between the hACE2 and musACE2 are the D30 to N30 and K31 to N31 substitutions. This results in the lack of salt bridges and concomitant lower number of H-bonds in the musACE2-SARS-CoV-2 RBD complex (Table 2). Specifically, the salt bridge with the K417 of the RBD seems to be a major driver of the interaction. In fact, similar to the musACE2, the ratACE2 has an Asn in position 30, which prevents formation of a salt bridge with K417 in the RBD. However, E26 in the ratACE2 forms a salt bridge with K417, resulting in a slightly altered complex structure with a shift of 6.5 Å of the RBD over the ratACE2, compared to its relative position in the human complex (Figure 1B). Furthermore, substitution of M82 to N82 in the ratACE2 introduces an N-glycosylation site [32], which may create steric clashes with F486 and N487. Taken together, the differences in binding mode would suggest that mice and rat are unsuitable models for the study of COVID-19. Similarly, the presence of a salt bridge between K35 and E484 in the cavACE2-RBD complex would make guinea pig an unsuitable model for the study of inhibitory antibodies and small molecules targeting the ACE2 - SARS-CoV-2 S protein interaction.

SARS-CoV S protein in complex with ACE2

In order to further validate the approach adopted we carried out docking simulations between SARS-CoV RBD, for which more experimental data are available, and hACE2, cavACE2, dogACE2, catACE2, ratACE2, rabACE2, ferACE2, musACE2. Indeed, comparison of the SARS-CoV and the SARS-CoV-2 RBD in complex with ACE2 shows that the two co-crystal structures are comparable (RMSD 2AJF) and the binding interfaces are similar (Table 3). In line with previously published data, we see that the SARS-CoV S has a smaller interaction surface and a lower number of interactions with ACE2 compared to the SARS-CoV-2 S protein (Table 3) [33].

The mode of binding of SARS-CoV RBD to hACE2 has several differences compared to that of the other ACE2 proteins analysed. While the binding of RBD to hACE2 is driven by polar interactions, similarly to what we observed for SARS-CoV-2, in all other SARS-CoV RBD-ACE2 models there are fewer H-bonds and a concomitant increase in hydrophobic interactions. Importantly, the substitution H34 to Y/L34 introduces a steric interference, which results in a shift of ~3 Å of the 441-456 loop of RBD bound to the dogACE2 and the cavACE2, compared to its relative position in the human complex. Similarly, the substitution of E329 with T/A/Q/K329 prevents the formation of a salt bridge with R426, and in the cavACE2 complex creates a charge repulsion. Overall, this would suggest a lower affinity of the RBD for the cavACE2, dogACE2, catACE2, ratACE2, rabACE2, ferACE2, musACE2, in line with previously published data showing different susceptibility to infection of animal models [34].

Indeed, SARS infection in cats, ferrets, mice, guinea pigs, and rats is weaker and does not replicate the human disease in all its aspects [34]. Taken together our approach may provide a rational for the observed experimental differences of the infection in human and animal models.

TMPRSS2 is highly conserved across species

TMPRSS2 is a type II transmembrane serine protease (TTSPs), with an extracellular region composed of a low-density lipoprotein (LDL) receptor class A domain, a scavenger receptor cysteine-rich (SRCR) domain and a peptidase S1 domain containing the catalytic triad (Supplemental Figure S4). TMPRSS2, similar to other TTSPs, has high affinity towards substrates containing an Arg residue in the P1 position. Indeed, TMPRSS2 recognises the SPRRAR/SVASQS sequence in the SARS-CoV-2 S glycoprotein and cleaves S1 from S2 between residues 685/686 [35].

In order to predict if TMPRSS2 from other animal species can cleave the SARS-CoV-2 S protein, the extracellular domain sequences of hTMPRSS2, cavTMPRSS2, dogTMPRSS2, catTMPRSS2, ratTMPRSS2, rabTMPRSS2, ferTMPRSS2 and musTMPRSS2 were aligned. The alignment revealed high conservation with a sequence identity between 75.11% and 83.97 % (Supplemental Figure S2). We then generated a model for hTMPRSS2 (Figure 2A and S4) using I-TASSER, and the best model had a C-score of -0.52 with a TM-score of 0.65 ± 0.13 and an RMSD of 7.9 ± 4.4 Å. The model has the three conserved disulphide bonds on the peptidase S1 domain, characteristic feature of all TTSPs, between residues C281-C297, C410-C426 and C437-C465 [36]. Disulphide bonds are also present between C113-C126, C120-C139, C133-C148, C172-C231 and C185-C241, which further validates the reliability of the models generated (Figure 2A and S4).

The pocket containing the catalytic triad has a uniform negative charge, which favours electrostatic interactions with the Arg reach peptide of the S protein (Figure 2B). Using this structural information, we identified the residues surrounding the catalytic triad that form a pocket on the head of TMPRSS2. Interestingly, this pocket is identical in all species studied and no substantial differences from the hTMPRSS2 were observed (Figure 2C). Taken together these data suggest that cavTMPRSS2, dogTMPRSS2, catTMPRSS2, ratTMPRSS2, rabTMPRSS2, ferTMPRSS2, musTMPRSS2 can cleave the SARS-CoV-2 S glycoprotein in a similar way to hTMPRSS2.

Discussion

The rate at which new infectious diseases are discovered has dramatically increased in the last 20 years [37]. Most of the new viral infections are caused by viruses belonging to

well-characterised virus families, like the novel coronavirus SARS-CoV-2 [1]. SARS-CoV-2 was identified in December 2019 to be the viral agent causing COVID-19, which has now spread to over 100 countries [1]. As a result, there is a global priority to identify and develop effective vaccines and drugs for the treatment of the disease. However, this effort is in part being hampered by the lack of suitable animal models [21]. Ideal animal models should be infected in a similar way to humans, present comparable symptoms, present a correlation between disease severity and virus titer, have similar histopathologic changes, virus growth kinetics, and comparable levels of mortality [38].

Coronaviruses are characterised by the presence of the S glycoprotein on the viral surface, which confer a unique crown-like morphology to the virion [39]. The S protein mediates both the attachment to the host cell and the fusion with the cell membranes [7]. As such, the S protein is the key element that determines cell tropism and the host range [40]. For cellular entry, the S protein binds to ACE2 and is primed by TMPRSS2, which promotes endocytic entry of the virus [13]. Importantly, ACE2 and TMPRSS2 appear to be widely expressed across mammals. Lung transcriptomic data is not available for all of the species investigated in this study. However, investigation of expression data for *Rattus norvegicus, Mus musculus, Canis lupus familiaris and Oryctolagus cuniculus* demonstrated that ACE2 and TMPRSS2 are detectable in the lungs of these species (Expression Atlas [41] and Bgee [42]). Further, these receptors were also found to be expressed in the lungs of multiple other animal species (e.g. *Bos taurus, Macaca mulatta, Ovis aries and Papio Anubis*).

An increasing body of evidence suggests that the tight binding of the S protein to ACE2 is the reason for the high person-to-person transmission rates and severity associated with this disease [11, 30, 32]. This is most evident when comparing the SARS 2002-2004 pandemic with the SARS-CoV-2 pandemic, with 8,098 cases versus over 4 million cases respectively [30]. Our analysis demonstrated that the SARS-CoV and SARS-CoV-2 S proteins have substantial differences in their ACE2 binding motifs (50% identity), which results in an increased number of contacts between the SARS-CoV-2 S protein and ACE2. This correlates with the observed 10- to 20-fold higher affinity of the SARS-CoV-2 S protein for ACE2, compared to SARS-CoV S protein [11].

It is clear that SARS-CoV-2 infection rates differ between species, and this is linked to the ability of the S protein to exploit ACE2 and TMPRSS2 to enter the host cell [40]. Our results show that the Spike protein recognises ferret and cat ACE2 in a comparable way to human ACE2, in line with data showing high susceptibility of these animals to SARS-CoV-2 infection [21]. In contract, we saw substantial differences in the binding mode of the SARS-CoV and SARS-CoV-2 S protein to rodent (guinea pigs, mice and rats) ACE2. For example, guinea pigs can be infected with SARS-CoV [34] (no data available for SARS-CoV-2) and we report that although SARS-CoV-2 S protein can bind guinea pig ACE2 (cavACE2), the mode

of binding differs to hACE2. In fact, compared to the human protein, the RBD-cavACE2 complex has a reduction of four direct hydrophobic contacts, which reduces protein affinity [43] and a concomitant extra charge reinforced hydrogen bond, which stabilize protein binding [20]. This results in a difference in binding kinetics between the RBD-cavACE2 and the RBD-hACE2. The different types of interaction stabilising the complexes are also likely to affect the ability of Protein-Protein Interaction (PPI) inhibitors to modulate, in a comparable way, the binding of RBD to hACE and cavACE. Therefore, while guinea pigs could be useful models for e.g. toxicity studies, they are unlikely to be suitable for the testing of inhibitory antibodies and small molecules targeting the ACE2 -S protein interaction surface.

Unlike ACE2, TMPRSS2 is highly conserved across the species studied. Modelling demonstrated that the catalytic triad (residues H296, D345 and S441 in hTMPRSS2) forms a negatively charged pocket, which favours electrostatic interactions with the Arg reach peptide of the S protein. This region was found to be identical in all species studied and it therefore appears that differences in infectivity across species is unlikely to be due to structural differences in TMPRSS2. Indeed, previous studies found SARS-CoV to be mildly infective in mice compared to humans and this was suggested to be as a result of differences between the structure of hACE2 and musACE2 [31]. Importantly, transgenic expression of hACE2 in mice (K18-hACE2 Mice) resulted in SARS-CoV becoming a rapidly fatal disease in this model [44]. This supports the findings of this study, which suggests that the high level of TMPRSS2 similarity between species does not appear to affect viral entry, but instead it is the species-specific differences in the structure of ACE2 that affects SARS-CoV and SARS-CoV-2 infectivity. Due to the similarities in TMPRSS2 between species, studies that aim to target this protease have a range of suitable models available. In contrast, this study suggests that the ferret or K18-hACE2 mouse are currently the most suitable models for studies that aim to target the ACE2 – S protein interaction.

Acknowledgements

F.P. received funding from Leverhulme Trust Grant RPG-2018-230. G.B. currently has funding from the Rosetree Trust (M702).

References

1. Peeri, N.C., et al., *The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned?* Int J Epidemiol, 2020.
2. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin.* Nature, 2020. **579**(7798): p. 270-273.
3. Chan, J.F., et al., *Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease.* Clin Microbiol Rev, 2015. **28**(2): p. 465-522.
4. Lau, S.K., et al., *Possible Bat Origin of Severe Acute Respiratory Syndrome Coronavirus 2.* Emerging Infectionus Diseases, 2020.
5. Wang, L.-F. and B.T. Eaton, *Bats, Civets and the Emergence of SARS*, in *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission*, J.E. Childs, J.S. Mackenzie, and J.A. Richt, Editors. 2007, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 325-344.
6. Shang, W., et al., *The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines.* NPJ Vaccines, 2020. **5**: p. 18.
7. Belouzard, S., et al., *Mechanisms of coronavirus cell entry mediated by the viral spike protein.* Viruses, 2012. **4**(6): p. 1011-33.
8. Donoghue, M., et al., *A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9.* Circ Res, 2000. **87**(5): p. E1-9.
9. Li, W., et al., *Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus.* Nature, 2003. **426**(6965): p. 450-4.
10. Yan, R., et al., *Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2.* Science, 2020. **367**(6485): p. 1444-1448.
11. Wrapp, D., et al., *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.* Science, 2020. **367**(6483): p. 1260-1263.
12. Shulla, A., et al., *A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry.* J Virol, 2011. **85**(2): p. 873-82.
13. Ou, X., et al., *Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV.* Nat Commun, 2020. **11**(1): p. 1620.
14. Hooper, J.D., et al., *Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes.* J Biol Chem, 2001. **276**(2): p. 857-60.
15. Heurich, A., et al., *TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein.* J Virol, 2014. **88**(2): p. 1293-307.
16. Fouchier, R.A., et al., *Aetiology: Koch's postulates fulfilled for SARS virus.* Nature, 2003. **423**(6937): p. 240.
17. Liang, L., et al., *Pathology of guinea pigs experimentally infected with a novel reovirus and coronavirus isolated from SARS patients.* DNA Cell Biol, 2005. **24**(8): p. 485-90.
18. Martina, B.E., et al., *Virology: SARS virus infection of cats and ferrets.* Nature, 2003. **425**(6961): p. 915.
19. Wu, D., et al., *Civets are equally susceptible to experimental infection by two different severe acute respiratory syndrome coronavirus isolates.* J Virol, 2005. **79**(4): p. 2620-5.
20. Xu, D., S.L. Lin, and R. Nussinov, *Protein binding versus protein folding: the role of hydrophilic bridges in protein associations.* J Mol Biol, 1997. **265**(1): p. 68-84.

21. Shi, J., et al., *Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2.* Science, 2020.
22. Li, F., et al., *Structure of SARS coronavirus spike receptor-binding domain complexed with receptor.* Science, 2005. **309**(5742): p. 1864-8.
23. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic Acids Res, 2018. **46**(W1): p. W296-W303.
24. Berendsen, H., D. van der Spoel, and R. van Drunen, *GROMACS: a message-passing parallel molecular dynamics implementation.* Computer Physics Communications, 1995. **91**(1-3): p. 43-56.
25. van Zundert, G.C.P., et al., *The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes.* J Mol Biol, 2016. **428**(4): p. 720-725.
26. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.
27. Janson, G., et al., *PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL.* Bioinformatics, 2017. **33**(3): p. 444-446.
28. UniProt, C., *UniProt: a worldwide hub of protein knowledge.* Nucleic Acids Res, 2019. **47**(D1): p. D506-D515.
29. Yang, J. and Y. Zhang, *I-TASSER server: new development for protein structure and function predictions.* Nucleic Acids Res, 2015. **43**(W1): p. W174-81.
30. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.* Cell, 2020. **181**(2): p. 281-292 e6.
31. Li, W., et al., *Efficient replication of severe acute respiratory syndrome coronavirus in mouse cells is limited by murine angiotensin-converting enzyme 2.* J Virol, 2004. **78**(20): p. 11429-33.
32. Li, W., et al., *Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2.* EMBO J, 2005. **24**(8): p. 1634-43.
33. Shang, J., et al., *Structural basis of receptor recognition by SARS-CoV-2.* Nature, 2020.
34. Nagata, N., N. Iwata-Yoshikawa, and F. Taguchi, *Studies of severe acute respiratory syndrome coronavirus pathology in human cases and animal models.* Vet Pathol, 2010. **47**(5): p. 881-92.
35. Xia, S., et al., *Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein.* Cell Mol Immunol, 2020.
36. Wu, Q., *Type II transmembrane serine proteases.* Curr Top Dev Biol, 2003. **54**: p. 167-206.
37. Howard, C.R. and N.F. Fletcher, *Emerging virus diseases: can we ever expect the unexpected?* Emerg Microbes Infect, 2012. **1**(12): p. e46.
38. Gretebeck, L.M. and K. Subbarao, *Animal models for SARS and MERS coronaviruses.* Curr Opin Virol, 2015. **13**: p. 123-9.
39. Li, F., *Structure, Function, and Evolution of Coronavirus Spike Proteins.* Annu Rev Virol, 2016. **3**(1): p. 237-261.
40. Millet, J.K. and G.R. Whittaker, *Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis.* Virus Res, 2015. **202**: p. 120-34.
41. Papatheodorou, I., et al., *Expression Atlas: gene and protein expression across multiple studies and organisms.* Nucleic Acids Res, 2018. **46**(D1): p. D246-D251.
42. Bastian, F., et al., *Bgee: Integrating and Comparing Heterogeneous Transcriptome Data Among Species.*, in *Data Integration in the Life Sciences. DILS 2008. Lecture Notes in Computer Science*, A. Bairoch, S. Cohen-Boulakia, and C. Froidevaux, Editors. 2008, Springer: Berlin, Heidelberg. p. 124-131.

43. Sammond, D.W., et al., *Structure-based protocol for identifying mutations that enhance protein-protein binding affinities.* J Mol Biol, 2007. **371**(5): p. 1392-404.
44. McCray, P.B., Jr., et al., *Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus.* J Virol, 2007. **81**(2): p. 813-21.

Table 1 Summary of species included in the study. Species names and abbreviation list with corresponding NCBI accession codes for ACE2 and TMPRSS2.

Table 2 Residues forming direct interactions in the ACE2 PD – SARS-CoV-2 S protein RBD complexes. Residues forming contacts in the EM structure (PDB ID: 6M17) and in the HADDOCK docking models are listed by their position and by their single-letter identity, with the first residue belonging to ACE2 and the second to the S protein. Interactions were identified using PyMol and PDBePISA.

Figure 1 Structure of the ACE2 interaction surface across species. (A) Electrostatic potential molecular surfaces representation of hACE-RBD, cavACE2, dogACE2, catACE2, ratACE2, rabACE2, ferACE2, musACE2. The cartoon representation of hACE2 (in teal with α1, α2 helix and β3-β4 loop in orange) is used as reference. The RBD binding surface on ACE2 is highlighted in pink. All structures have been superimposed and are in the same orientation. **(B)** The structure of the hACE2-RBD complex is overlapped to the structure of ratACE2-RBD complex. Structures are shown in cartoon, with both ACE2 proteins in grey and α1, α2 helix and β3-β4 loop in orange. The RBD structure of the human complex is shown in green while the rat is in hot pink. ACE2s orientation is the same as in (A).

Table 3 - **List of residues forming direct interactions in the ACE2 PD – SARS-CoV S protein RBD complexes.** Residues forming contacts in the crystal structure (PDB ID: 2AJF) and in the HADDOCK docking models are listed by their position and by their single-letter identity, with the first residue belonging to ACE2 and the second to the S protein. Interactions were identified using PyMol and PDBePISA.

Figure 2 The TMPRSS2 active site is highly conserved among species. (A) Cartoon representation of hTMPRSS2, with the SRCR domain in beige and the Peptidase S1 domain in teal. The catalytic triad (H296, D345 and S441) is shown in orange sticks. **(B)** Electrostatic potential molecular surfaces representation of hTMPRSS2. A short peptide, shown in sticks, has been placed in the Peptidase S1 active site, which is highlighted in pink. The protein is in

the same orientation as in (A). **(C)** Multiple sequences alignment of hTMPRSS2, cavTMPRSS2, dogTMPRSS2, catTMPRSS2, ratTMPRSS2, rabTMPRSS2, ferTMPRSS2, musTMPRSS2. Peptidase S1 active pocket residues have been highlighted in red with the relative consensus sequence. Catalytic triad residues are shown in red in the consensus sequence.

Figure S1 Schematic representation of hACE2. PD domain is shown in green, transmembrane region in purple and the cytoplasmic domain in blue. The S protein binding sites, ADAM17 and TMPRSS2 cleavage sites have been mapped on the model. Multiple sequences alignment was performed using Clustal Omega for hACE2, cavACE2, dogACE2, catACE2, ratACE2, rabACE2, ferACE2, musACE2.

Figure S2 Sequences identity matrixes resulting from the ACE2 and TMPRSS2 multiple sequences alignments. Sequence alignment was performed using Clustal Omega using the HHalign algorithm and Gonnett as transition matrix.

Figure S3 Cartoon representation of the hACE2-RBD complex. RBD F486 side chain (teal sticks) is inserted in a relatively hydrophobic patch formed by the hACE2 residues F28, L79, M82 and Y83, shown as hot pink sticks.

Figure S4 Schematic representation of hTMPRSS2. The cytoplasmic domain is shown in blue, the transmembrane helix is purple, the LDL domain in red, the SRCR domain in orange and the Peptidase S1 in green. The catalytic triad residues H296, D345 and S441 have been mapped on the model. Normalised β-factor and estimated accuracy plots of the I-TASSER hTMPRSS2 model. The estimated local accuracy shows that the LDL (1-150) has relatively higher modelling error while most of other regions are accurate with estimated distance to native smaller than 4 Å.