Biophysical characterization of the SARS-CoV2 spike protein binding with the ACE2 receptor explains increased COVID-19 pathogenesis

Ratul Chowdhury[1] and Costas D. Maranas[2,*]

[1]Laboratory of Systems Pharmacology, Harvard Medical School, Boston. MA 02115

[2]Department of Chemical Engineering, The Pennsylvania State University, University Park. PA 16802

[*]*Corresponding author* | Email: costas@psu.edu

Abstract

SARS-CoV2 is a novel highly virulent pathogen which gains entry to human cells by binding with the cell surface receptor – angiotensin converting enzyme (ACE2). We contrasted the binding interactions between human ACE2 and coronavirus spike protein receptor binding domains (RBDs) from (i) SARS-CoV2, (ii) the related but less virulent OC43 (Singapore COVID-19 strain), and (iii) the 2002 epidemic-causing SARS-CoV1. We find that the RBDs of the spike protein of SARS-CoV2 are highly optimized to achieve the strongest possible binding interaction with ACE2 which is consistent with its enhanced pathogenicity. SARS-CoV2 RBDs form the most stable complex with ACE2 (64.91%, and 28.07% higher binding scores than SARS-CoV1 and OC43, respectively) while occupying the greatest number of residues in the ATR1 binding site. In fact, the spike protein RBDs from SARS-CoV2 out-compete the angiotensin 2 receptor type I (ATR1) which is the native binding partner of ACE2 by 64% in terms of binding affinity (quantified using the Rosetta scoring function). They accomplish this through strong electrostatic attachments with every fourth residue on the N-terminus alpha-helix (starting from Ser19 to Asn53) as the turn of the helix makes these residues solvent accessible. These results offer a computational explanation for the increased pathogenicity of SARS-CoV2 and allude to therapeutic modalities by identifying and rank-ordering the ACE2 residues involved in binding with the virus.

Main

The causative agent of coronavirus disease 2019 (COVID-19) was identified in January 2020 to be a novel β-coronovirus of the same subgenus as SARS-CoV. In comparison with the 2002 SARS-CoV1 outbreak and the COVID-19 Singapore strain (OC43), the SARS-CoV2 strain has caused the greatest number of infections and fatalities and an effective antiviral treatment and vaccine remains elusive. It has been reported that first step to viral entry is characterized by an association between the viral spike RBD and human ACE2 protein[1]. This study first assesses the molecular interactions between the three spike

RBDs with the human ACE2 complex to explain the difference in the degree of pathogenesis for the three strains. We provide a comparative analysis of the most important RBD residues from the three spike proteins that drive the binding with ACE2. We computationally assess that SARS-CoV2 outcompetes the human ATR1 surface receptor protein to preferentially bind the human ACE2. The role of the highly enhanced binding of RBDs from the spike protein of SARS-CoV2 in explaining pathogenicity has received so far limited attention in the COVID-19 scientific reports.

There has been a recent communication by Wan *et al.*[2] to explain interactions between ACE2 and SARS-CoV1 and SARS-CoV2 RBDs, but the study was limited to (a) using a homology modeled structure of SARS-CoV2 RBD, (b) only exploring five residues from the spike RBDs, (c) not exploring the case of the OC43 strain, and (d) not analyzing the effect of spike-ACE2 binding on ATR1-ACE2 complex. To this end, we have used the experimentally confirmed atomic scale maps (cryo-EM structures) of not only the SARS-CoV2 RBD in complex with ACE2, but also that of OC43 (which however, is not complexed with ACE2). We generated the OC43-ACE2 complex using protein-protein docking (see Methods).

Analysis of human ACE2 in complex with spike RBDs from the three different coronavirus strains

Rosetta-based energy minimization of the ACE2-RBD complexes with RBDs from SARS-CoV1, SARS-CoV2, and OC43 reveals that SARS-CoV2 exhibits the strongest binding (613.02 Rosetta energy units). SARS-CoV1 and OC43 binding scores with ACE2 are 215.159 and 440.94 Rosetta energy units, respectively. In an uninfected human cell, ATR1 receptor binds to ACE2 to form a receptor complex, whereas at the onset of infection, the coronavirus presents the RBD of its spike protein to human ACE2 protein for an electrostatically-driven association between the two. Our results indicate that ACE2 can bind with either human ATR1 or the viral spike as the binding domains overlap. ACE2 can form hydrophobic and strong electrostatic (including pi-pi, and cation-pi) interactions with the binding domain of ATR1. The viral spike RBDs mimics these interactions to gain entry via strong non-covalent attachment.

To understand the role of the inter-residue interaction network formed during viral entry, we first constructed a contact map depicting all such interactions for the spike-binding interface of unbound ACE2 (see **Figure 1**). We then computed the transitions in this contact map for the ACE2-RBD complex for each one of the three cases (i.e., SARS-CoV2, OC83, and SARS-CoV1). We observe that SARS-CoV2 co-opts the original contact map to form the strongest possible ACE2-RBD interface (see **Figure 1**).

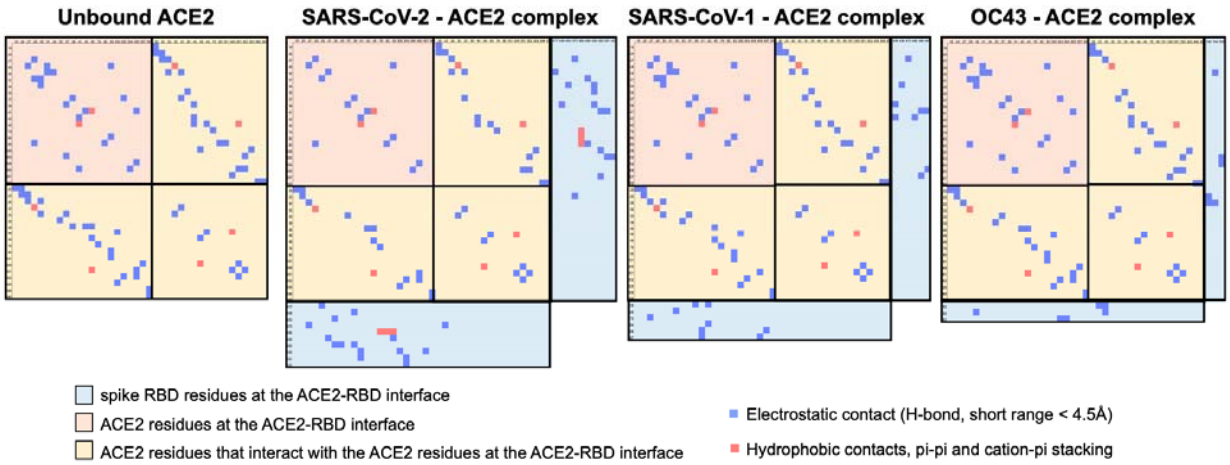

Figure 1. Contact map of ACE2 receptor in unbound state and when bound with the viral spike protein RBDs from SARS-CoV2, SARS-CoV1, and OC43. The red regions show stabilizing contacts between residues of the RBD binding motif of ACE2. The yellow non-diagonal regions show contacts between the RBD binding motif of ACE2 and other neighboring residues of ACE2. The yellow diagonal region shows contacts within these neighboring residues. Finally, the blue padded regions show contacts at the ACE2-RBD binding interface. The width of these blue regions determines the number of RBD residues that form strong electrostatic or hydrophobic contacts with ACE2. Long-range electrostatic interactions (4.5Å< but >6.0Å) between the spike and ACE2 are not shown.

We observe that SARS-CoV2 forms the greatest number of effective ACE2 contacts (16 electrostatic and three hydrophobic) using ten RBD residues at the ACE2 binding interface (see blue region in **Figure 1**). In comparison, the spike protein RBD for SARS-CoV1 and OC43 are only able to form nine (six) strong electrostatic contacts using five (three) RBD residues, respectively. This does not imply that SARS-CoV1 and OC43 only present five (three) residues, respectively to bind to ACE2. The other interface residues either form weak electrostatic contacts or are simply non-interacting. **Table 1** lists the strong pairwise interactions between the RBDs and ACE2 along with the corresponding distances. Therefore, it appears that SARS-CoV2 reforms the original contact map with ACE2 by leveraging 35% (9 out of 26) of self-stabilizing contacts around the spike-binding domain to form 19 new complex-stabilizing contacts. SARS-CoV1 and OC83 show weaker attachments as they are each able to co-opt only 15% (4 out of 26) contacts.

In silico alanine scanning to identify spike residues most important for ACE2 binding

Each one of the ACE2 binding residues from the three viral spike RBDs was computationally mutated to alanine (one at a time) and the resultant ACE2-RBD complexes were energy minimized and scored using Rosetta force field. This procedure assesses how important is the identity of the native residues by defaulting them to alanine and observing whether this significantly affects binding. In detail, the percent

loss of ACE2 binding upon an alanine mutation was used as a proxy score for assessing the importance a particular RBD residue in binding and subsequent pathogenesis. The results from the alanine scan study (see **Figure 2**) reveal that 88% (15 out of 17) of the ACE2 interface residues of SARS-CoV2 are important for complex formation. Even a single alanine mutation in any of these reduces the binding score by ~60%. These results imply that the SARS-CoV2 RBD of the spike protein are highly optimized for binding with ACE2. We note that positions Lys417 and Gly502 have the strongest impact on binding (75% and 83% reduction upon mutation to Ala, respectively). This is because they help establish one strong electrostatic contact with Asp30, and three with Gln325, Lys353, and Gly354 (as listed in **Table 1**). Our results also identify the same three residues Phe486, Gln493, and Asn501 to be important for ACE2 binding as proposed by Wan *et al.*[2]. In their study, they used a homology modeled SARS-CoV2 RBD instead of experimentally resolved atomic coordinates for the description of the protein binding event. We calculated that Phe486, Gln493, and Asn501 establish three new contacts each (see **Table 1** for interaction partners and inter-residue distances), and their mutation to Ala (even for only one of them) leads to loss of ACE2 binding by more than ~62.5%.

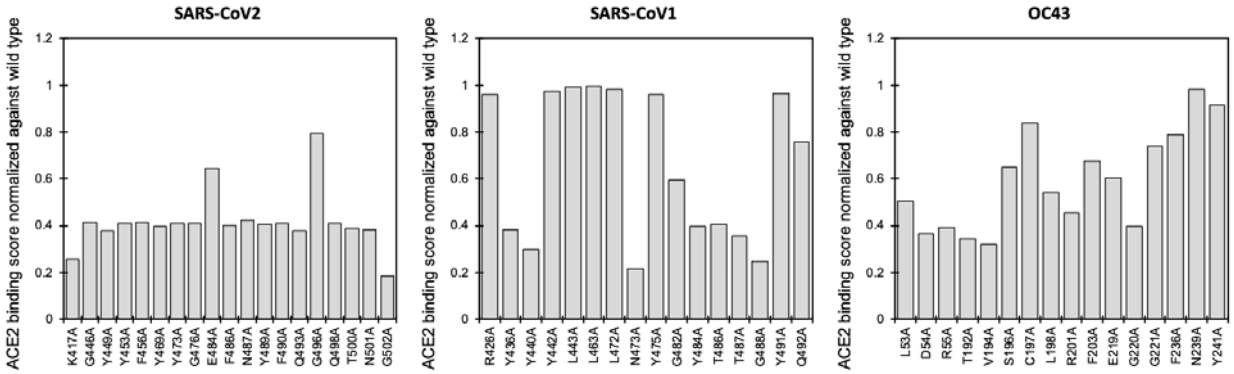

Figure 2. Alanine scan on ACE2 binding residues of spike RBDs of SARS-CoV2, SARS-CoV1, and OC43 coronavirus. Bars represent the residual ACE2 binding score upon alanine mutation at the indicated site normalized with respect to binding score prior to mutation.

The alanine scan of the spike protein RBD of SARS-CoV1, on the other hand shows less significant penalty to the binding score upon mutation to alanine. Only six residues (Tyr436, Tyr440, Asn473, Tyr484, Thr486, and Thr487) seen to be involved in strong electrostatic coupling with ACE2 residues (indicated in **Table 1**). In addition, Gly488Ala and Gly482Ala mutations lead to 41% and 72% loss of ACE2 binding, respectively. In summary, alanine scans indicate that SARS-CoV2 has the highest number of "effectively" interacting residues at the ACE2 binding interface whereas the SARS-CoV1 spike forms only a few strong ACE2 connectors with a large number of "idle" interface residues (43% - 7 out of 16) which do not affect ACE2 binding upon mutation to alanine. OC83 seems to be between the two with a

maximum number of weak electrostatic interactors (43% - 7 out of 16) and only two idle residues at the interface (i.e., residues Asn239 and Tyr241).

Table 1. List of strong contacts between the spike RBDs from (SARS-CoV2, SARS-CoV1, and OC43) and human. ACE2

Strain	RBD residue	ACE2 residue	Distance (Å)
SARS-CoV2	K417	D30	2.9
	Y449	D38, Q42	3.2/3.7
	Y453	H34	3.3
	Y473	T27	3.9
	F486	M82/L79/Q24	3.1/3.4/3.6
	Y489	K31	2.6
	Q493	K31/H34/E35	2.9/3.5/4.1
	T500	Y41	3.6
	N501	Y41/K353/G354	3.2/3.4/3.6
	G502	Q325/K353/G354	3.1/3.3/3.5
SARS-CoV1	Y436	D38	3.2
	Y440	H34	3.5
	473	Q24	3.3
	Y484	D38/Q42/K353	3.4/3.7/3.2
	T486	Y41/G354	3.4/4.0
	T487	K353	3.5
OC43	T192	E22/E24/D30	4.2/3.7/3.1
	V194	T92	3.6
	N202	H34/D38	3.1/3.6

Presence of tyrosines and glycines in ACE2 binding domains of these spike proteins

Both the SARS-CoV1 and SARS-CoV2 ACE2 binding domains are enriched in tyrosines. As many as 25% (4 out of 16 residues) and 26.3% (5 out of 19 residues) are tyrosines, respectively. In contrast, OC83 has only one tyrosine (among 16 residues). We have not explored the phylogenetic basis for the presence of tyrosines but they do seem to be important for conferring high binding affinity between SARS-CoV2 and ACE2 as alluded to by the alanine scan (see **Figure 2**). In contrast, the tyrosines in SARS-CoV1 (Tyr442, Tyr475, and Tyr491) are mostly idle. We use **Figure 3** to explain two representative cases of interface tyrosines from both SARS-CoV2 (Tyr473 and Tyr489) and SARS-CoV1 (Tyr442 and Tyr475) RBDs.

The SARS-CoV2 tyrosine backbones, even though on a loop, are mutually stabilized by hydrogen bonding and the side chains are locked in place by a pi-pi aromatic interactions between the phenyl rings. This enables both of these tyrosine side-chains to form a strong electrostatic contact with Thr27 side chain of ACE2. It is thus unsurprising that both mutation Y473A and Y489A result in a similar (58.5% and 59.1%, respectively - as shown in **Figure 2**) reduction in binding with ACE2. In contrast, in the energy

minimized complex of SARS-CoV1 RBD and ACE2 both Tyr442 and Tyr475 (**Figure 3**) only contribute to internal stability of the spike by forming strong electrostatic contacts with Trp476 backbone O-atom, and Asn473 side chain N-atom, and hence are not available (or are too far > 6.0Å) for binding with the neighboring ACE2 residues.

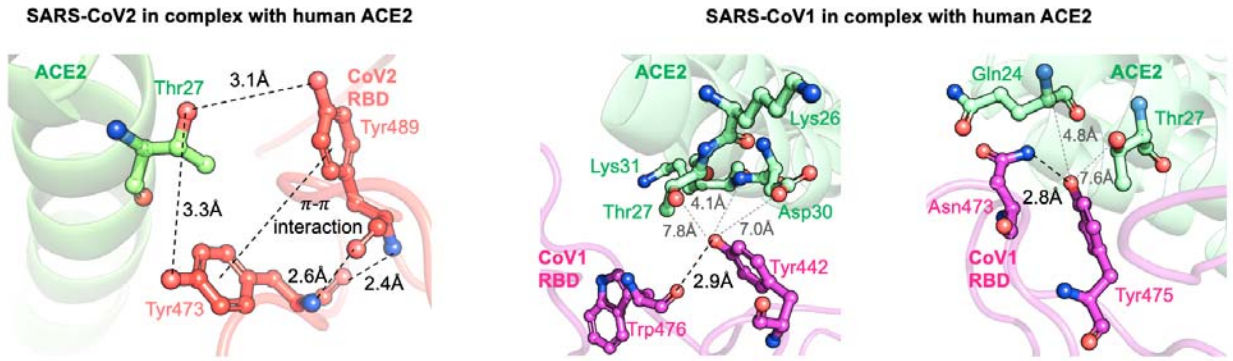

Figure 3. The role of tyrosines in SARS-CoV2 RBD is to form strong contacts with ACE2 residues, while in SARS-CoV1, they are primarily responsible for forming stabilizing contacts within the spike and are hence unavailable for ACE2 binding.

Next, we focus on the role of glycine residues in all three spike RBDs which form important electrostatic contacts with ACE2. They lead to more than 55% loss of binding (on average) upon mutation to alanine. We choose to study in detail one such representative glycine from each spike RBD – Gly502 from SARS-CoV2, Gly488 from SARS-CoV1, and Gly220 from OC83.

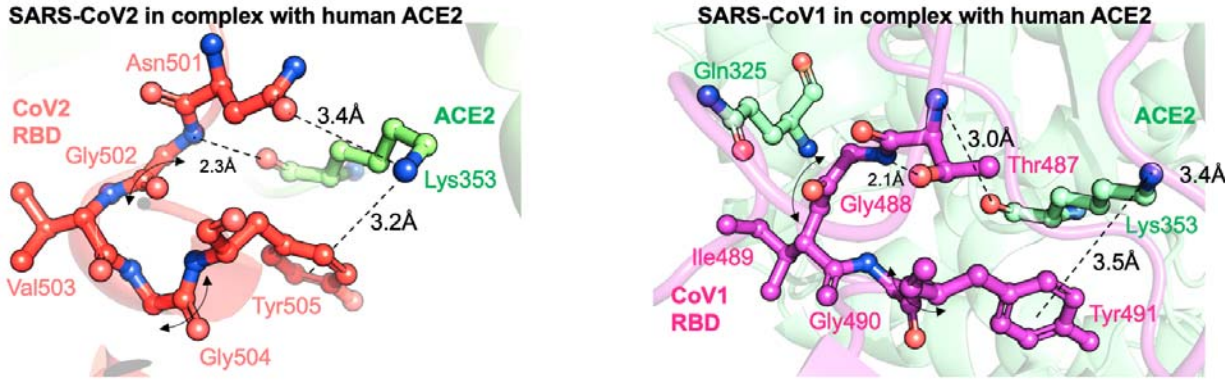

Figure 3. The role of glycines in both SARS-CoV2 RBD is to provide a XGxGY motif for capturing ACE2 Lys353 using a strong electrostatic and a cation- interaction. Here, 'X' is a short-chain polar residue, and 'x' is a short chain hydrophobic residue. The glycines hinge the residue 'X' about the Tyr residue for a non-covalent binding with Lys353.

Glycine residues seem to play a very crucial, though indirect, role in effecting a strong capture of the ACE2 protein. Interestingly, in both SARS-CoV variants, the capture mechanism of the ACE2 residue Lys353 by glycine residues in the spike protein is the same. Atomic coordinates of both these complexes were independently, and experimentally confirmed by Song *et al.*[3] in 2018 and Wang *et al.* in 2020 (manuscript unpublished but structure deposited at - www.rcsb.org/structure/6lzg). In both cases the SARS-CoV spike RBD uses a combination of a cation- and a strong electrostatic interaction to bind with Lys353. The electrostatic interaction is mediated by Asn501 in SARS-CoV2 and Thr487 in SARS-CoV1. Two glycine residues and a short hydrophobic group together allows the Asn501 (and Thr487 for SARS-CoV1) to be within strong electrostatic reach of Lys353 while ensuring a cation- interaction between Tyr505 (and Tyr491 for SARS-CoV1) and Lys353. Mutation Y505A (or Y491A for SARS-CoV1) has no effect on ACE2 binding. However, mutation to alanine of either Gly502 or Asn501 (or their corresponding residues in SARS-CoV1) leads to >70% loss of ACE2 binding. Thus, we recover a strong functional motif, which is conserved between the two SARS-CoV strains.

Sequence alignment of the three RBD sequences reveal a conserved motif in all three sequences (see **Figure 4**). Further inspection of the sequence alignment reveals (see **Figure 4**) that all of the three other conserved binding motifs for ACE2 hotspot residues (Thr27, Asp38, and Gln42) have a key glycine (and rarely alanine) which presumably offers a hinge to orient a neighboring polar residue to bind to a specific ACE2 residue.

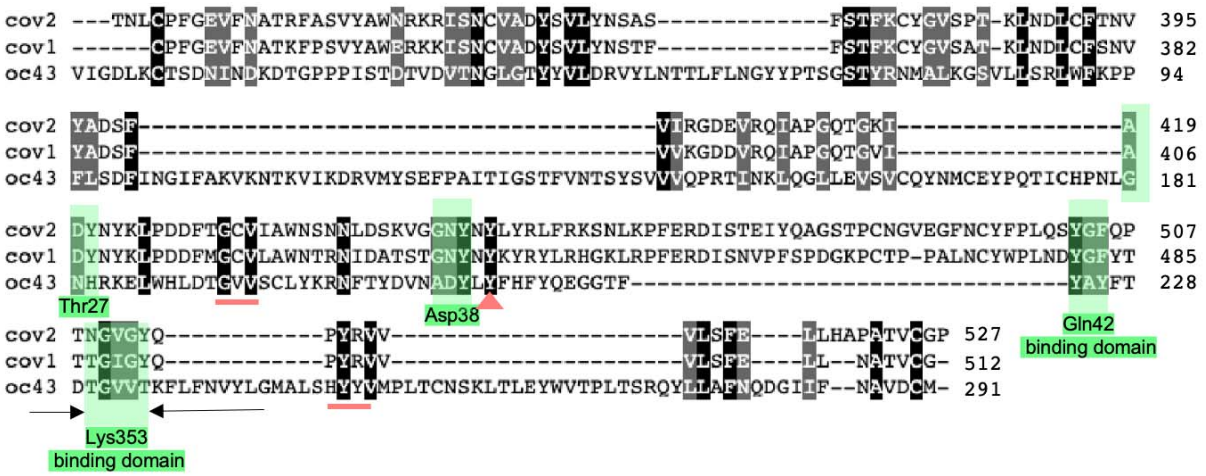

Figure 4. Sequence alignment of SARS-CoV2, SARS-CoV1, and OC43 RBDs indicate conserved motifs for binding specific ACE2 residues indicated in green. The red motifs are conserved and near to the binding interface but do not play a role in effecting a strong ACE2 bind.

Competitive ACE2 binding of spike RBDs and angiotensin receptor (ATR1)

We reported on the differences between three viral ACE2-RBD complexes and how they exploit the contact map of RBD-binding domain of human ACE2 receptor to facilitate binding. In an uninfected cell, through the action of the renin angiotensin system (RAS), ACE2 forms a complex with the angiotensin 2 receptor type I (ATR1)[4]. Due to the lack of an experimentally resolved structure for the ACE2-ATR1 complex, we used protein-protein docking and Rosetta-binding score screening to identify the most stable configuration of the complex. Analysis of the ACE2-ATR1 binding interface reveals 41 ACE2 residues and 26 ATR1 residues at the interface connected by five strong electrostatic contacts and several long range weak electrostatic contacts. We find that eleven SARS-CoV2 RBD binding residues of ACE2 are shared by the ATR1 binding region. Moreover, SARS-CoV2 RBD binds ACE2 with ~64% better binding score than ATR1. In contrast, SARS-CoV1 show ~2.1% worse binding score with ACE2 compared to the ACE2-ATR1 complex while sharing only nine residues with the ATR1 binding interface of ACE2 (see **Figure 5**). These numbers indicate that SARS-CoV2 (and OC43) outcompete the ATR1-ACE2 complex thus facilitating the formation of the ACE2-spike complex whereas SARS-CoV1 cannot. This is in line with their respective viral infection rates.

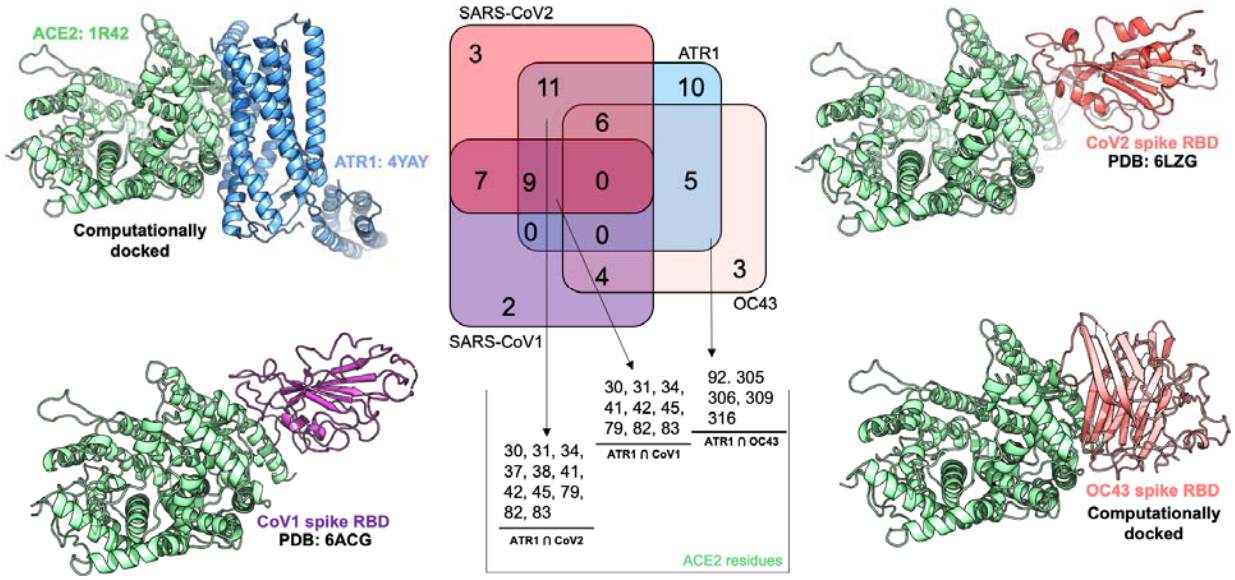

Figure 5. ACE2 complexes with ATR1, SARS-CoV1, SARS-CoV2, and OC43 spike RBDs along with the number of shared ACE2 residues (Venn diagram) at their respective binding regions is shown. Residue positions that are shared between ATR1 and the three spike RBDs (SARS-CoV1, SARS-CoV2, and OC43) have been listed.

We computationally explored the potential margin of improvement for the binding affinity of SARS-CoV2 with ACE2 using the IPRO[5] protein design software. We allowed all 21 contacting residues of the RBD of the spike protein to simultaneously mutate. We run two separate design trajectories (see

supplementary information) and, in both cases the best design achieved an approximately 25% improvement in binding affinity using the Rosetta scoring function. This improvement is less than the difference between the calculated binding scores of SARS-CoV1 and SARS-CoV2 implying that SARS-CoV2 has already achieved over 50% of the theoretically possible binding affinity gain with ACE2 compared to SARS-CoV1. Interestingly, the network of glycine residues in SARS-CoV2 is conserved in all redesigned RBDs.

Summary

We believe that the computational results presented herein are responsive to the recent call to contrast the two SARS-CoV and OC43 strains heralded in the recent press release (March 20, 2020) by the Scripps Research Institute[6]. We also provide key biophysical insights to explain the differences in pathogenicity of SARS-CoV2 in comparison to OC43 and SARS-CoV1. Multiple lines of computational evidence indicate that the spike RBDs gain entry into the human cell through electrostatic attachment with every fourth residue on the N-terminal alpha-helix (starting from Ser19 to Asn53) as the turn of the helix makes these residues solvent accessible.

Methods

We have used experimentally determined coordinates of SARS-CoV1 and SARS-CoV2 in complex with ACE2 (PDB accessions: 6ACG[3] and 6LZG - www.rcsb.org/structure/6lzg, respectively). OC43 RBD was isolated from the crystal structure of the whole OC43 spike (PDB: 6OHW[7]). Similarly, unbound ATR1 structure (PDB: 4YAY[8]) was also separately downloaded and docked against ACE2 using protein-protein docking scripts from Z-DOCK 3.0[9]. ZDOCK uses pairwise shape-complementarity, electrostatics, and implicit solvation terms in scoring the docked poses. Implicit solvation treats the water as a dielectric continuum. The rotational sampling interval was set to 10°. Clustering of the docked poses were done at an 8 Å cutoff. Subsequently, PyRosetta[10] scripts were written to rank and identify the most stable complexes from each cluster which were then energy-minimized and re-ranked. Finally, the complex which ranked high in stability and binding scores was chosen as the model. An alanine scan was again performed using PyRosetta scripts, where the computational models of the alanine variants were first generated, energy minimized, and ACE2 binding scores computed. The ACE2 interface definitions for each binding partner (RBDs and ATR1) were obtained by feeding the energy minimized protein-protein complexes through the *find_contacts* module of OptMAVEn-2.0[11].

Author Contributions

RC, and CDM conceived, designed, and wrote the study.

Acknowledgement

RC thanks Debolina Sarkar for advice on the renin angiotensin system and also editing the paper.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* (2020). doi:10.1016/j.cell.2020.02.052
2. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J. Virol.* (2020). doi:10.1128/jvi.00127-20
3. Song, W., Gui, M., Wang, X. & Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog.* (2018). doi:10.1371/journal.ppat.1007236
4. Tikellis, C. & Thomas, M. C. Angiotensin-converting enzyme 2 (ACE2) is a key modulator of the renin angiotensin system in health and disease. *International Journal of Peptides* (2012). doi:10.1155/2012/256294
5. Pantazes, R. J., Grisewood, M. J., Li, T., Gifford, N. P. & Maranas, C. D. The Iterative Protein Redesign and Optimization (IPRO) suite of programs. *J. Comput. Chem.* (2015). doi:10.1002/jcc.23796
6. Ed Yong, T. A. *Why the Coronavirus Has Been So Successful*.
7. Alejandra Tortorici, M. *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* (2019). doi:10.1038/s41594-019-0233-y
8. Towler, P. *et al.* ACE2 X-Ray Structures Reveal a Large Hinge-bending Motion Important for Inhibitor Binding and Catalysis. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M311191200
9. Mintseris, J. *et al.* Integrating statistical pair potentials into protein complex prediction. *Proteins Struct. Funct. Genet.* (2007). doi:10.1002/prot.21502
10. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq007
11. Chowdhury, R., Allan, M. F. & Maranas, C. D. OptMAVEn-2.0: De novo Design of Variable Antibody Regions Against Targeted Antigen Epitopes. *Antibodies* (2018). doi:10.3390/antib7030023