Interaction Energies of the Human ACE2 Molecular Recognition by SARS-CoV-2

Joao Batista Junior*

Independent Researcher

Abstract

The *Coronaviridae* family of viruses includes hundreds of viruses common in many different animal species and humans. Seven coronaviruses (CoVs) are known to cause disease in humans. Four of them show low pathogenicity and are endemic in humans and the other three CoV are particularly dangerous and highly pathogenic viruses, which underwent genetic changes rendering them able to jump the species barriers from animal host to humans and also to spread efficiently among humans. SARS-CoV-2 is the seventh coronavirus known to infect humans. The S protein mediates attachment and viral and host cell membrane fusion. The receptor-binding domains (RBDs) are regions in S protein responsible for receptor recognition. Human angiotensin-converting enzyme 2 (ACE2) is recognized by HCoV-NL63, SARS-CoV and SARS-CoV-2 as their functional receptor.

Interaction energy analysis were performed to unveil how precisely SARS-CoV-2 interacts with ACE2 by identifying which amino acid residues are responsible for the interactions across S protein-ACE2 interfaces and how they contribute to the strength, stability and specificity of S protein interactions.

Interaction energies acting on molecular recognition of ACE2 by HCoV-NL63, SARS-CoV and SARS-CoV-2 conduced to a naturally evolved RBD with different combinations of amino acids, providing SARS-CoV-2 binding interface more interacting residue pairs, more hydrogen bonds, increased number of residues engaged in hydrogen bonding, allowing for better distribution of hydrogen bond per residue in interface than SARS-CoV or HCoV-NL63, includes salt bridge, and adds new van der Waals contacts into the network.

Residues across the SARS-CoV and SARS-CoV -2 homologous sequences have been chosen to be remarkably evolutionary conserved in the regions mediating binding of these viruses because of their dominant hydrogen bonding contribution to binding stability to ACE2. SARS-CoV-2 achieves higher binding affinity than SARS-CoV and HCoV-NL63 to human ACE2 molecular recognition primarily by combining its richer interaction network and higher binding stability.

This study presents a comprehensive and quantitative analysis of interaction energies of the human ACE2 molecular recognition by CoVs that may contribute to further understand the higher infectivity and transmissibility of SARS-CoV-2 compared to SARS-CoV and HCoV-NL63, furthermore, this could help explain why SARS-CoV-2 has an enhanced ability for pathogenicity.

Keywords: Interaction energy • SARS-CoV-2 • Spike Glycoprotein • ACE2 • Molecular Recognition

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes an infectious disease named coronavirus disease 2019 (COVID-19). Since December in late 2019, when SARS-CoV-2 first made an appearance as a novel emerging coronavirus in Wuhan, China, it is infecting people and spreading easily, silently and rapidly from person-to-person worldwide. On March 11, 2020, the COVID-19 outbreak was characterized as a pandemic by the World Health Organization (WHO).

The *Coronaviridae* family of viruses includes hundreds of viruses, which are common in many different animal species, including wild animals (bats, civets, raccoons, pangolins) [1,2], domestic and peridomestic animals (cats, cattle, horses, pigs, goats, camels) [3], and humans [4,5]. This virus family consists of two subfamilies, Coronavirinae and *Torovirinae* (members of this subfamily are known to not cause human infection). The subfamily Coronavirinae comprises four genera called *Alpha coronavirus, Beta coronavirus, Gamma coronavirus, and Delta coronavirus* [6].

Seven coronaviruses (CoV) are known to cause disease in humans. Four of them show low pathogenicity and are endemic in humans: human

coronavirus-229E (HCoV-229E) and human coronavirus-NL63 (HCoV-NL63), both from genus *Alpha coronavirus* (α-CoV); and human coronavirus-OC43 (HCoV-OC43) and human coronavirus- HKU1 (HCoV-HKU1), from genus *Beta coronavirus* (β-CoVs). Those CoV are an important cause of upper respiratory tract infections and have most frequently been associated with mild symptoms as those observed in the common cold [7]. The other three CoV are particularly dangerous and highly pathogenic viruses and all belonging to genus β-CoVs, included in the species *Severe acute respiratory syndrome-related coronavirus* (SARS-related CoV) [6], which underwent genetic changes, through mutation and/or recombination, rendering them able to jump the species barriers from animal host to humans and also to spread efficiently among humans [8,9] and cause much more severe respiratory infections, sometimes fatal, such as the outbreaks observed early in this century when severe acute respiratory syndrome coronavirus (SARS-CoV) emerged, in 2002, causing severe acute respiratory syndrome (SARS); Middle East respiratory syndrome coronavirus (MERS-CoV), in 2012, causing Middle East respiratory syndrome (MERS), and the current SARS-CoV-2, the seventh coronavirus known to infect humans, causing COVID-19 pandemic. All three SARS-related CoV posing a severe threat to public health.

Corresponding Author: *João Batista Júnior, Independent Researcher; E-mail: profjbatistajr@gmail.com; ORCID iD: orcid.org/0000-0002-8972-8131*

The CoVs have a RNA genome, a positive sense, single-stranded genome, ranging from 26 to 32 kilobases (kb) in length, the largest genomes for RNA viruses, which encompasses a region encoding an RNA-dependent RNA polymerase, a region with coding sequences of genes which encode each one of the four main structural proteins required to produce a structurally complete viral particle: the spike (S) protein, envelope (E) protein, membrane (M) protein, the nucleocapsid(N), and a region representing several nonstructural proteins [10,11]. Some CoVs species may also possess a gene encoding the structural protein hemagglutinin-esterase (HE) [12].

Since CoVs have their genomic material surrounded by a lipid bilayer membrane and that genomic material needs to be transported through the barriers imposed by the host cell membranes, the S protein protruding from the viral surface mediates cell attachment and membrane fusion processes between the viral and target cell membranes. The S protein is a transmembrane glycoprotein that forms homotrimers. Each monomeric unit of S protein basically consists of three segments: an ectodomain, a transmembrane anchor and a short intracellular tail. The ectodomain comprises two functional subunits (S1 subunit and S2 subunit) used for invading host cells. S1 subunit is responsible for binding receptors and S2 subunit that contains the fusion machinery is responsible for viral and cellular membrane fusion. The S1 subunit N-terminal moiety comprises domain A. The S1 subunit C-terminal folds as three spatially distinct β-rich domains, termed domain B, C and D [13].

The regions responsible for receptor recognition in S protein are only found in domains A or B within S1 subunit, the receptor-binding domains (RBDs), and domain A or B is used in receptor recognition or attachment process specifically according to CoV species and their receptor specificities. A distinct location of S1 subunit domain A of HCoV-OC43 and HCoV-HKU1, both β-CoVs, mediates the binding of these viruses to the receptor 9-O-acetyl-sialic acid (9-O-Ac-Sia), which is terminally linked to oligosaccharides decorating glycoproteins and gangliosides, at the host cell surface [14]. HCoV-229E, an α-CoV, requires the zinc metalloprotease human aminopeptidase N as a receptor for entry into target cells and uses three receptor-binding loops of RBD present in S1 subunit domain B to bind aminopeptidase N [15,16]. MERS-CoV, a β-CoV, recognizes dipeptidyl peptidase 4 (DPP4) as its functional receptor by binding via its S1 subunit domain B. While MERS-CoV S1 subunit domain A selectively binds to sialoglycoconjugates on cell-surface which can serve as an attachment factor for support biding of S1 subunit domain B [17-19].

HCoV-NL63, a prevalent human respiratory virus, uses S1 subunit domain B, its RBD, to recognize angiotensin-converting enzyme 2 (ACE2) as its receptor for infection of target cells (Figure 1A). HCoV-NL63 is the only α-CoV known to use ACE2 as its receptor [20,21]. SARS-CoV and SARS-CoV-2, both β-CoV, also recognize host receptor ACE2 as its functional receptor and uses their S1 subunit domain B, their RBD, to attache the virion directly with ACE2 (Figures 1B and 1C) [22-24].
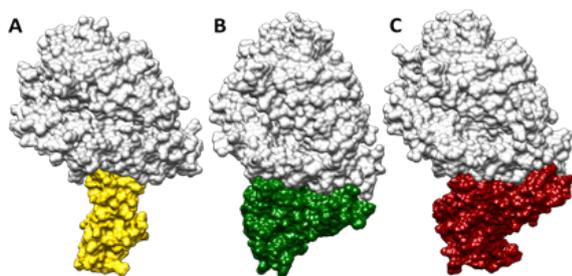

Figure 1. Recognition of human ACE2 by (A) HCoV-NL63, (B) SARS-CoV, and (C) SARS-CoV-2 RBDs. Surface representations of human ACE2 (in gray), HCoV-NL63 RBD (in yellow), SARS-CoV RBD (in green) and SARS-CoV-2 RBD (in red).

S protein is cleaved at the boundary between the S1 and S2 subunits in many CoVs, and those subunits remain non-covalently bound in the prefusion conformation [25]. The S protein of SARS-CoV-2 has a functional polybasic (furin) cleavage site at the S1–S2 boundary through the insertion of 12 nucleotides, which additionally led to the predicted acquisition of three O-linked glycans around the site [26]. After binding of RBD in S1 subunit of S protein on the virion to the ACE2 receptor on the target cell, the heptad repeat 1 (HR1) and heptad repeat 2 (HR2) domains in its S_2 subunit of S protein interact with each other to form a six-helix bundle (6-HB) fusion core, bringing viral and cellular membranes into close proximity for fusion and infection [27].

In addition to being a cellular entry receptor for HCoV-NL63, SARS-CoV and SARS-CoV-2, ACE2 has its own unique functions. Human ACE2 is a glycoprotein, type 1 transmembrane metallopeptidase, expressed and active in most tissues, with remarkable expression observed on lung alveolar epithelial cells, enterocytes of the small intestine, and vascular endothelial cells and arterial smooth muscle cells [28]. ACE2 has an ectodomain containing its single zinc coordinating catalytic site on the cell surface. It functions as a carboxypeptidase and acts as regulatory components of the renin-angiotensin system (RAS), one of the most important hormonal systems in the physiological regulation of blood pressure and fluid balance. ACE2 hydrolyzes the C-terminal dipeptide of Angiotensin II (Ang II), a very powerful vasoconstrictor and the main active peptide of RAS, to convert it into Angiotensin 1-7 (Ang 1–7), a vasodilator. By regulating local levels of Ang II and Ang 1–7, in the cardiovascular system in particular, ACE2 has the importance in maintaining the balance of the RAS activation.

According to various studies, CoVs have existed early in the natural environments [29-31] and they have been present since 1966 in the human history [4]. CoVs have thus had plenty of time to adapt to their environments and to have given rise to numerous versions gaining ability to evolve to new restricted host, where there is less competition from other virus or life-forms.

In this study we unveil how precisely SARS-CoV-2 interacts with its functional host receptor by identifying which amino acid residues are responsible for the interactions across S protein-ACE2 interfaces and by detecting specific atoms from those amino acids and how they contribute to the strength, stability and specificity of S protein interactions.

Methods

To descript the atom-atom interactions across the interfaces of the S protein-ACE2 molecular complex, we selected from Protein Data Bank (PDB) experimental crystal structure for each CoV S protein RBD structure in complex with the human ACE2 receptor (for HCoV-NL63 PDB ID Code 3KBH, resolution 3.31 Å, [21]; for SARS-CoV PDB ID Code 2AJF, resolution 2.90 Å, [23]; and for SARS-CoV-2 PDB ID Code 6M0J, resolution 2.45 Å, [24]). We isolated each chain composing the molecular complex found in the crystal structures, removed water, ions, and all carbohydrates molecules bound to structure. After that, hydrogen atoms were added to the chains, followed by charges addition using AMBER force field. Next we performed interaction energy calculations [32] using parameters derived from AMBER parm99 molecular mechanical force fields for organic and biological molecules [33], in a solvent environment, to identify the key amino acid residues within CoV S protein RBD-ACE2 interfaces, with a maximum distance threshold of 4.00 Å, which are significantly contributing to the stability of that interaction.

The interatomic contact surface and interface areas were determined by calculating the S protein RBD and ACE2 complexed surfaces, the S protein RBD and ACE2 uncomplexed surfaces, and the buried surfaces for each unit in the complex [34].

Multiple sequence alignment of CoVs S protein sequences were computed using a progressive alignment construction method [35] for identifying residue conservation or residue changes in all sequences of the S protein RBDs.

Results and Discussion

The computational structural analysis of interactions between HCoV-NL63 S protein and human ACE2 highlighted 15 amino acid residues (Ser496, His586, Tyr498, Trp585, Pro536, Gly494, Gly537, Gly495, Gly534, Val499, Cys500, Ser539, Cys497, Ser535, Ser540) of the HCoV-NL63 S protein in the region mediating binding in RBD, the receptor-binding motif (RBM) (Figure 2A), and 20 surface residues of ACE2 (Figure 3A), resulting in 33 interacting residue pairs. The residues involved in this interaction cover a surface area of 753,90 Å2 in the RBM of HCoV-NL63 RBD and a surface area of 648,10 Å2 in ACE2, which form an interface of 700.99 Å2 and encompass 4 hydrogen bonds and 84 non-bonded contacts, and no salt bridges.
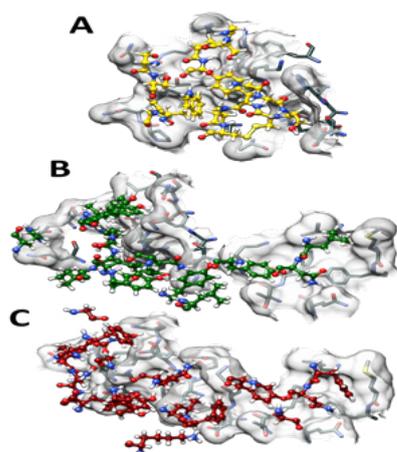

Figure 2. Amino acid residues interaction networks across the interfaces of the CoVs S protein-ACE2 molecular complexes. Binding interfaces for (A) HCoV-NL63, (B) SARS-CoV, and (C) SARS-CoV-2. Interface illustrated in transparent gray; amino acid residues in the RBMs shown in ball-and-stick model, carbon atoms in the RBMs of HCoV-NL63, SARS-CoV, SARS-CoV-2 colored in yellow, green, and dark red, respectively; nitrogen in blue, oxygen in red, hydrogen on heteroatoms in white; interfacial amino acid residues of the ACE2 shown in stick model, carbon atoms in gray.
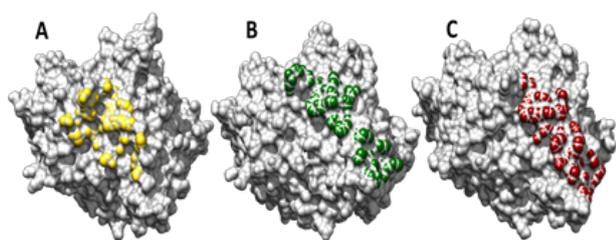

Figure 3. Amino acid residues in human ACE2 surface critical for molecular recognition by CoVs. Surface representations of human ACE2 in gray. Residues recognized by (A) HCoV-NL63 RBD are highlighted in yellow, by (B) SARS-CoV in green, and by (C) SARS-CoV-2 in dark red.

SARS-CoV S protein and ACE2 employ a large fraction of their surfaces, 870,90 Å2 in the S protein RBD and 831,00 Å2 in ACE2, for creating a binding interface of 850.96 Å2 upon complex formation, which counts on 16 interacting residues (Gly488, Tyr484, Tyr491, Asn473, Thr486, Tyr436, Tyr475, Arg426, Tyr440, Asn479, Tyr442, Gly482, Thr487, Leu472, Ile489, Leu443) in RBM of SARS-CoV RBD (Figure 2B) and 20 residues of ACE2 (Figure 3B) responsible for forming 34 interacting residue pairs and building 9 hydrogen bonds and 106 non-bonded contacts, and 1 salt bridge across the binding interface.

Surface and interface areas very similar in size to those areas involved in SARS-CoV S protein-ACE2 interactions were identified in the SARS-CoV-2 S protein-ACE2 complex interactions analysis, which revealed surfaces areas of 867,26 Å2 in the S protein RBD and 824.34 Å2 in ACE2 with a binding interface of 845.80 Å2, encompass 17 (Gly502, Gln498, Tyr505, Asn487, Thr500, Tyr449, Tyr489, Tyr453, Gln493, Leu455, Gly496, Asn501, Phe486, Phe456, Lys417, Gly446, Ala475) interacting residues in RBM of SARS-CoV-2 RBD (Figure 2C)and 20 residues of ACE2 (Figure 3C) and build a network of 35 interacting residue pairs that were found to form 12 hydrogen bonds and 109 non-bonded contacts, and 1 salt bridge across surface area.

Interaction energies analysis was applied to identity amino acid residues of the S protein RBDs important for interaction with ACE2 and their effect on the stability and formation of the S protein RBD-ACE2 complex. The interaction energy for SARS-CoV-2 RBD was higher than for SARS-CoVRBD, and for both the SARS-CoV-2 and the SARS-CoV RBDs was nearly double as for the HCoV-NL63 RBD, interaction energy of -1876.68 kj/mol, -1772.26 kj/mol and -956.59 kj/mol, respectively, with lower values corresponding to higher energy, and thus higher interaction energies.

Since those interaction energies are a composition of the individual interaction energy contributions of each amino acid residue in the interacting surfaces of S protein with ACE2, those energies reflect most precisely the behavior of each individual amino acid residue in the RBMs of S protein RBD-ACE2 complexes. In this computational analysis, the interaction energy for an amino acid residue represents the sum of all stabilizing interaction energies (≤ -0.01 kj/mol) and all destabilizing interaction energies (≥ 0.01 kj/mol) acting on this specific amino acid residue in the complex. The interface interaction energy, meaning the sum of all interface stabilizing interaction energies (≤ -0.01 kj/mol) and all interface destabilizing interaction energies (≥ 0.01 kj/mol) resulted exclusively from the interactions with amino acid residues in the ACE2 surface that act on the interacting amino acid residues in the RBMs surface. The interface interaction energy between SARS-CoV and ACE2 interacting surfaces was -477.82 kj/mol, similar but slightly higher than that between SARS-CoV-2 and ACE2 surfaces, -456.16 kj/mol. While HCoV-NL63-ACE2 interface revealed much lower interface interaction energies of -287.90 kj/mol.

The calculated surface interaction energy pointed to a higher stabilized surface on SARS-CoV-2, with a surface interaction energy of -1418.35 kj/mol, followed by SARS-CoV with -1291.95 kj/mol, and HCoV-NL63, which exhibited a surface interaction energy of -666.82 95 kj/mol, less than half of the energy observed for SARS-CoV-2, and nearly half of the energy for SARS-CoV.

The higher interaction energy obtained for SARS-CoV-2 RBD and its higher stabilized surface, compared to SARS-CoV and HCoV-NL63 RBDs, contribute for higher binding stability and a decreased interface residue fluctuations relative to the SARS-CoV–ACE2 complex, as performed in others calculations, and adopt a stable binding mode [36]

The decomposition of the interaction energy at the level of residue energy contribution for each one of the 15 identified ACE2-interfacing residues in RBM of HCoV-NL63 RBD identified Trp585, His586 and Ser535 as key residues that contribute with the highest interaction energy in the HCoV-NL63-ACE2 complex, -199.52, -114.55, and -96.78 –kj/mol, respectively, which represents 42.95% of the interaction energy in the RBM. In addition, those three amino acids contribute with -41.82, -46.38, and -48.88 kj/mol, respectively, across the interface HCoV-NL63-ACE2.

The multiple amino acid sequence alignments for RBD of the SARS-CoV-2, SARS-CoV and HCoV-NL63, showed that a sequence identity shared by SARS-CoV-2 and SARS-CoV is 74.26%, SARS-CoV-2 and HCoV-NL63 is 17.31%, and SARS-CoV and HCoV-NL63 is 19.34%, which indicated high amino acid sequence conservation in RBD between SARS-CoV-2 and SARS-CoV, and low sequence conservation when compared to RBD of HCoV-NL63. The structure-based sequence alignments of the RBDs demonstrated high structural similarity for SARS-CoV-2 and SARS-

CoV, and when one of those RBDs was taken as the reference structure, they had no structural similarities with HCoV-NL63 RBD. Those results are in agreement with the fact that S1 subunit RBD in the S protein is the most variable part of the coronavirus genome [37]. The S1 subunit of NL63-CoV, SARS-CoV, and SARS-CoV-2, as one might infer, may have undergone divergent evolution from a common ancestor into different structures and then convergent evolution to structures sharing ACE2-binding topologies [38].

Whereas there was low sequence conservation for RBD of HCoV-NL63 when it was compared to the others analyzed RDBs, the higher amino acid sequence conservation and structural similarities shared by SARS-CoV-2 and SARS-CoV between their RBDs allowed a more direct comparison of the importance of the amino acids in specific positions in the RBM of those viruses and their structural and functional role in molecular recognition and formation of the complex with the ACE2.

SARS-CoV-2 compared to SARS-CoV, related to the number of ACE2-interfacing residues in their RBMs, showed that these two CoVs are 47.1% conserved, as 8 interfacing residues are identical in their RBMs, and no deletion or insertion was observed within RBMs. Based on SARS-CoV, those conserved residues and their assigned contributions (in kj/mol) to the total interaction energy are Gly488 (-31.21), Tyr491 (-132.93), Asn473 (-142.27), Thr486 (-107.00), Tyr436 (-95.53), Tyr475 (-132.85), Tyr440 (-164.83), and Gly482 (-52.01), corresponding to residues Gly502 (-64.35), Tyr505 (-146.17), Asn487 (-141.47), Thr500 (-98.72), Tyr449 (-86.28), Tyr489 (-152.29), Tyr453 (-152.84), and Gly496 (-48.71) in SARS-CoV-2, respectively. A remarkable observation is that almost all hydrogen bonds observed in the SARS-CoV-ACE2 binding interface are related to conserved residues in RBM, except for one hydrogen bond formed by SARS-CoV Arg426 and ACE2 Glu329, in the form of a strong saltbridge.

The corresponding conserved residues in SARS-CoV-2 play the same role in the intermolecular interactions with ACE2, justifying once more the role of these conserved residues in their RBMs and the importance of hydrogen bond for interactions across the interfaces of these two viruses with ACE2, and maybe that is why those residues have been chosen to be evolutionary conserved across the SARS-CoV and SARS-CoV-2 homologous sequences.

Four residues in HCoV-NL63 RBD (Tyr498, Ser540, Ser535 and His586) form four hydrogen bonds in the interface with ACE2. The pairs of those hydrogen-bonded residues are Tyr498-Glu37, Ser540-Thr324, Ser535-Lys353, and His586-Asn322. Six conserved residues in SARS-CoV RBM concentrate nine hydrogen bond interactions between ACE2, which forms the following hydrogen-bonded pairs Gly488-Lys353, Tyr491-Glu37, Asn473-Gln24, Asn473-Tyr83, Thr486-Tyr41, Thr486-Asn330, Tyr436-Asp38, Tyr436-Gln42, and Tyr475-Tyr83.

The mutation of SARS-CoV interfacing residues Asn479 and Thr487, and non-interfacing residue Thr433, to their corresponding residues Gln493, Asn501, and Gly 446, in SARS-CoV-2 RBD, add three new hydrogen bonds in the SARS-CoV-2-ACE2 interface.

Thus, 10 residues in SARS-CoV-2 engage in hydrogen bonding to other residues in ACE2 and form 12 hydrogen bonded pairs (Gly502-Lys353, Tyr505-Glu37, Asn487-Gln24, Asn487-Tyr83, Thr500-Tyr41, Tyr449-Asp38, Tyr449-Gln42, Tyr489-Tyr83, Gln493-Glu35, Gly496-Lys353, Asn501-Tyr41, and Gly446-Gln42) across the interface area.

Comparing hydrogen bonded residues in HCoV-NL63, SARS-CoV and SARS-CoV-2 RBMs one can notice that not only there is a progressive augment of the number hydrogen bond interactions, but also that there is an increased number of residues engaged in hydrogen bonding.

SARS-CoV-2 not only shows larger number of hydrogen bonds but also a better distributed hydrogen bonds per residue in the interface than SARS-CoV or HCoV-NL63.

This more favorable hydrogen-bonding arrangement in SARS-CoV-2 is an important contributor to the stability of the SARS-CoV-2-ACE2 complex,

and, by decreasing the number of hydrogen bond donor or acceptor unpaired residues at ACE2 binding surface, it contributes to achieve more specificity to ACE2 molecular recognition.

SARS-CoV Arg426 contributes with second highest interaction energy in the SARS-CoV RBD-ACE2 complex, -161.26 kj/mol, and its salt-bridged bond with ACE2 Glu329 contributes as stabilizing interaction of -7.25 kj/mol. In SARS-CoV-2, residue Arg426 becomes asparagine (Asn439), but Asn439 does not conserve the same role as Arg426, because besides being in RBD of SARS-CoV-2, Asn439 is out of its RBM. Other residue that has a similar behavior when mutated is the residue SARS-CoV Ile489 that becomes valine (Val503) and does not play an interface interacting role in SARS-CoV-2 RBM. Ile489 has a discrete role on SARS-CoV interface, its contribution is higher as a surface stabilizing (-63.48 kj/mol) than an interface binding (-7.24 kj/mol) residue.

Residues Tyr484, Asn479, Tyr442, Thr487, Leu472, Leu443, all identified as ACE2-interfacing residues in SARS-CoV RBD, become Gln498, Gln493, Leu455, Asn501, Phe486, Phe456 in SARS-CoV-2 RBD, respectively. In addition, three more residues in SARS-CoV RBD, VAL404, Thr433, and Pro462, which are not directly involved with ACE2-binding, become Lys417, Gly446, and Ala475 in SARS-CoV-2 RBD, respectively. These mutations enabled new characteristics at specific positions on the surface regions of SARS-CoV-2 that are responsible for binding to the ACE2, and all of those mutated residues kept or gained a role for receptor interaction or surface stabilization.

For instance, the residue mutation at the Val404 position in SARS-CoV to a protonatable Lys417 in SARS-CoV-2 contributes to the binding to ACE2 and to the structure stability of the complex mainly through electrostatic interactions. Valine side-chain has a short and apolar isopropyl group, while lysine side chain has a different length and polarity than valine. Because of that the Asp30 on ACE2 surface has access to the protonatable Lys417 on SARS-CoV-2 surface and form a salt-bridged. Thus, Lys417 incorporates strong interaction energy of -148.49 kj/mol for the binding between SARS-CoV-2 and the receptor.

Some residues mutations preserved their interaction energy contribution at their structural correlated position in viral RBD-receptor complexes, for example, SARS-CoV Thr487 and its corresponding residue SARS-CoV-2 Asn501 have almost identical interaction energy contributions of -147.69 kj/mol and -148.24 kj/mol, respectively, in their complexes. However, the same behavior is not observed when comparing the interaction energy balance between the interface and surface. SARS-CoV Thr487 contributes with -66,25kj/mol for the interface binding energy and with -81,26 kj/mol for the surface interaction energy. While those respectively contributions from SARS-CoV-2 Asn501 are -38,77kj/mol and -109,28 kj/mol.

Both residues SARS-CoV Thr487 and SARS-CoV-2 Asn501 exploit Lys353 as their strongest intermolecular interaction partner in ACE2, forming stabilizing interactions that account for the following energies -26,15 kj/mol, and -26,83 kj/mol, respectively. The interaction energy analysis shows that Lys353 on the surface of ACE2 is a remarkable interacting residue, suggesting a pivotal role of Lys353 for virus–receptor interactions, which is agreement with other studies [21].

There are two more residues in SARS-CoV and other two in SARS-CoV-2 that form binding pairs with Lys353, those residues and their bind energies are: Tyr491 -33.29 kj/mol, and Gly482 -8.00 kj/mol, in RBM of SARS-CoV; and Tyr505 -31.32, and Gly496 -7.48, in RBM of SARS-CoV-2.

Lys353 of ACE2 has also been chosen by 4 residues from the RBM of HCoV-NL63 as their first binding pair interaction, Ser535-Lys353 -27.69 kj/mol (was the highest among the residue pairs), Gly537-Lys353 -10.76 kj/mol, Gly495-Lys353 -4.37 kj/mol, and Gly534 -2,21 kj/mol. Also, as a second binding pair interaction, Gly494-Lys353 -1.74 kj/mol, and a third binding pair Tyr498-Lys353 -5.91 kj/mol, Pro536-Lys353 -5.57 kj/mol, and Cys500-Lys353 -1.56 kj/mol.

The amino acid interaction network by non-polar residues and van

der Waals contacts represents the van der Waals interaction energy contributions of amino acids located at the RBD–ACE2 interface. In HCoV-NL63 RBD-ACE2 interface this interaction network encompasses 84 van der Waals contacts. While SARS-CoV and SARS-CoV-2 RBDs-ACE2 interfaces involve 106 and 109 van der Waals contacts, respectively.

Mutations of SARS-CoV residue Leu472 (-36.00; that is the interaction energy in kj/mol for the residue in the complex) to SARS-CoV-2 Phe486 (-56.12); and SARS-CoV residue Leu443 (-99.81) to SARS-CoV-2 residue Phe456 (-127.25) contribute to the total interaction energy and to enhance the structural stability of the complex, thus to an improved RBD-ACE2 binding for SARS-CoV-2 mainly through van der Walls interactions. Leucine residues (Leu472 and Leu443) have aliphatic hydrophobic isobutyl group as side chain, while phenylalanine residues (Phe486 and Phe456) have aromatic hydrophobic benzyl side chain. Besides the distinct structural aspects of isobutyl and benzyl groups in their respective residues, they seem to present individually similar hydrophobicity. However, those mutations add new van der Waals interactions in the interface because they extend the number of residue contact pairs in ACE2, allowing them to enhance the number of new interatomic contacts.

SARS-CoV residue Leu472 has in ACE2 three contact pairs, whose interface stabilizing interaction energies contribution is less than -0.1 kj/mol, as indicated in parentheses: Leu79 (-6.65), Met82 (-6.49), and Tyr83 (-2.28); while SARS-CoV-2 residue Phe486 has six contact pairs in ACE2: Tyr83 (-11.73), Leu79 (-9.43), Met82 (-9.01), Gln24 (-2.71), Glu75 (-1.76), and Ala80 (-1.27). SARS-CoV residue Leu443 forms one contact pair with Thr27 (-4.53) in ACE2; and SARS-CoV-2 residue Phe456 also has Thr27 (-9.84) in ACE2 as contact pair, but accounts on residues Lys31 (-5.88), Asp30 (-4.23), and Phe28 (-1.19) in ACE2 to form three new contact pairs in the interface. Those mutantions from a smaller leucine to a bigger phenylalanine lead to richer packing and considerable gain in van der Waals contribution to the binding stability.

These results present comprehensive and quantitative descriptions about how the interaction energy acting on molecular recognition of ACE2 by HCoV-NL63, SARS-CoV and SARS-CoV-2 conduced to a naturally evolved RBD with different combinations of amino acids, which provides a SARS-CoV-2 binding interface, with more interacting residue pairs, forms more hydrogen bonds with increased number of residues engaged in hydrogen bonding, allowing for better distribution of hydrogen bond per residue in interface than SARS-CoV or HCoV-NL63, also includes a salt bridge in the interface, and adds new van der Waals contacts into the network.

This richer interaction network and its associated interaction energy might be essential for maintaining ideal RBD stability and high binding affinity to the ACE2 receptors, which is supported by another study, in which was found that ACE2-binding affinity of the RBD in S1 subunit of SARS-CoV-2 is 10- to 20-fold higher than that of SARS-CoV [39].

Conclusion

Receptor recognition represents an important function in the process of virus adaptation to new hosts upon cross-species transmission of distinct viruses. An evolutionary and natural selected recognition mode can lead to new dominant genotypes. Since viral-receptor recognition relies on interfacial interaction energies, one can make the simplifying assumption that a dominant viral genotype is intrinsically linked to the interaction energies of the receptor recognition.

Sugar receptors have been serving to CoVs attach and entry to host cells for a long evolutionary time. Since protein receptors in general have advantages over sugar receptors by providing higher affinity interactions for viral attachment, natural selection events and evolution allowed CoVs to search for high-affinity protein receptors.

SARS-CoV-2 and SARS-CoV share high amino acid sequence conservation between their RBDs and low sequence conservation when compared to RBD of HCoV-NL63. Also, there is high structural similarity

for SARS-CoV-2 and SARS-CoV RBDs, and no structural similarities with HCoV-NL63 RBD is observed. The S1 subunit of NL63-CoV, SARS-CoV, and SARS-CoV-2 may have undergone divergent evolution from a common ancestor into different structures and then convergent evolution to structures sharing ACE2-binding topologies. The amino acid interactions at the binding interface of HCoV-NL63, SARS-CoV and SARS-CoV-2 have progressively evolved in search of a stable binding network of residue–residue contacts to the human receptor ACE2. SARS-CoV-2 is a result of evolutionary optimized binding mode to the human receptor ACE2.

The interatomic interactions across the interfaces of the CoVs RBD-ACE2 molecular complexes show different combinations of amino acids. Residues across the SARS-CoV and SARS-CoV-2 homologous sequences have been chosen to be remarkably evolutionary conserved in the RBMs of this virus because of their dominant hydrogen bonding contribution to binding stability to ACE2 upon complex formation.

Furthermore, some residue mutations add new hydrogen bonds across the SARS-CoV-2-ACE2 interface and engage more residues in hydrogen bonding network enabling SARS-CoV-2 to have a more favorable hydrogen-bonding arrangement in the interface than SARS-CoV or HCoV-NL63, which contributes to enhance SARS-CoV-2-ACE2 complex complementarity and helps SARS-CoV-2 to achieve more specificity to ACE2 molecular recognition. Other mutations add new van der Waals interactions in the interface because they extend the number of residue contact pairs in ACE2, allowing them to enhance the number of new interatomic contacts, which leads to richer packing and considerable gain in van der Waals contribution to the binding stability. SARS-CoV-2 achieves higher binding affinity than SARS-CoV and HCoV-NL63 to human receptor ACE2 not simply because of its enhanced number of interface interactions, but primarily by combining its interface interaction network optimization and the higher binding stability given by its RBD optimized interaction energies in human ACE2 molecular recognition.

Future research on this topic should yield significant new knowledge, however in the present context of receptor recognition mechanism those optimized interaction energies, higher binding stability, higher binding affinity would enable SARS-CoV-2, upon ACE2 binding, to initiate receptor-mediated signaling pathway resulting in its internalization into cell and in triggering a series of molecular and cellular mechanisms through which other signals are integrated during a productive infection causing a different disease outcome.

This comprehensive and quantitative analysis of interaction energies of the human ACE2 molecular recognition by CoVs may contribute to further understand the higher infectivity and transmissibility of SARS-CoV-2 compared to SARS-CoV and HCoV-NL63, furthermore, this could help explain why SARS-CoV-2 has an enhanced ability for pathogenicity.

References

1. Song, Dong Huai, Tu Chang-Chun, Zhang Guo-Wei and Wang Sheng-Yue et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci USA* (2005) 102: 2430-2435.

2. Zhang, Tao , Wu, Qunfu, Zhang Zhigang. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr Biol* (2020) 30:1346-1351.e2.

3. Azhar, Esam I, El-Kafrawy, Sherif A,. Farraj Suha A and Hassanet Ahmed M el. Al-Saeed MS, Hashem AM, Madani TA. Evidence for camel-to-human transmission of MERS coronavirus. *N Engl J Med* (2014) 370: 2499-2505.

4. Hamre, Dorothy, Procknow John . A new virus isolated from the human respiratory tract. *ProcSocExpBiol Med* (1966) 121: 190-193.

5. van der, Hoek L, Pyrc K, Jebbink MF and Vermeulen-Oost W et al. Identification of a new human coronavirus. *Nat Med* (2004) 10: 368-373.

6. Gorbalenya, AE, Baker SC, Baric RS and de Groot RJ et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* (2020) 5: 536-544.

7. Gaunt, ER, Hardie A, Claas EC and Simmonds P et al. Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *J Clin Microbiol* (2010) 48: 2940-2947.

8. Corman, Victor Max, Ithete, Ndapewa Laudika, Richards Leigh Rosanne and Schoeman M Corrie et al. Rooting the phylogenetic tree of middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. *J Virol* (2014) 88:11297-11303.

9. Guan, Y, Zheng BJ, He YQ and Liu XL et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* (2003) 302: 276-278.

10. Marra, Marco A, Jones Steven J M, Astell Caroline R and Holt Robert A et al. The genome sequence of the SARS-associated coronavirus. *Science* (2003) 300:1399-1404.

11. Rota, Paul A, Oberste M Steven, Monroe Stephan S and Nix W Allan et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* (2003) 300: 1394-1399.

12. Vlasak, R, Luytjes W, Spaan W and Palese P. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proc Natl Acad Sci USA* (1988) 85: 4526–4529.

13. Walls, AC, Tortorici MA, Bosch BJ and Frenz B et al. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* (2016) 53: 114-117.

14. Hulswit, Ruben JG, Lang, Yifei, Bakkers Mark JG and Zeshi Li et al. Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *ProcNatlAcadSci USA* (2019) 116: 2681-2690.

15. Yeager, CL, Ashmun RA, Williams RK and Cardellichio CB et al. Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* (1992) 357: 420-422.

16. Wong, Alan HM, Tomlinson Aidan CA, Zhou Dongxia and Satkunarajah Malathy et al. Receptor-binding loops in alphacoronavirus adaptation and evolution. *Nat Commun*. (2017) 23:1735.

17. Raj, Stalin V, Mou Huihui, Smits Saskia L and Dekkers Dick HW et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*. (2013) 495: 251-254.

18. Wang, Nianshuang, Shi Xuanling, Jiang Liwei and Zhang Senyan et al. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res* (2013) 23: 986-993.

19. Li, Wentao, Hulswit Ruben JG, Widjaja Ivy and Raj V. Stalin et al. Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *ProcNatlAcadSci USA* (2017) 114: E8508-E8517.

20. Hofmann, Heike, Pyrc Krzysztof, Hoek Lia van der and Geier Martina et al. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *ProcNatlAcadSci USA* (2005) 102: 7988-7993.

21. Wu, Kailang, Li Weikai, Peng Guiqing and Li Fang. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *ProcNatlAcadSci USA* (2009) 106: 19970-19974.

22. Li, Wenhui, Moore Michael J, Vasilieva Natalya and Sui Jianhua et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* (2003) 426: 450-454.

23. Li, Fang, Li Wenhui, Farzan Michael and Harrison Stephen C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* (2005) 309: 1864-1868.

24. Lan, Jun, Ge Jiwan, Yu Jinfang and Shan Sisi et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* (2020) 581: 215-220.

25. Jan, Bosch Berend, Zee Ruurd van der, Haan Cornelis AM de, and Rottier Peter JM. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* (2003) 77: 8801–8811.

26. Walls Alexandra C, JunPark Young, Tortorici M. Alejandra and AbigailWall et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* (2020) 181:281-292.

27. Bosch, Berend Jan, Martina Byron EE, Zee Ruurd van der and Lepault Jean et al. Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *ProcNatlAcadSci USA* (2004) 101: 8455-8460.

28. Hamming, I, Timens W, Bulthuis ML and Lely AT et al. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis.*J Pathol* (2004) 203: 631-637.

29. Estola, T. Coronaviruses, a New Group of Animal RNA Viruses. *Avian Diseases*. (1970) 14:330–336.

30. Vijgen, Leen, Keyaerts Els, Moës Elien and Thoelen Inge et al. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J Virol* (2005) 79: 1595-1604.

31. Pyrc, Krzysztof, Dijkman Ronald, Deng Lea and Jebbink Maarten F. et al. Mosaic structure of human coronavirus NL63, one thousand years of evolution. *J Mol Biol* (2006) 364: 964-973.

32. Bendova, Biedermannova L, Hobza P, Vondrasek J. Identifying stabilizing key residues in proteins using interresidue interaction energy matrix. *Proteins*. (2008) 72: 402–413.

33. Wang Junmei, Cieplak Piotr, Kollman Peter A.. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem* (2000) 21: 1049–1074.

34. Ribeiro, Judemir, Vera Carlos Ríos and Schüller Andreas. Calculation of accurate interatomic contact surface areas for the quantitative analysis of non-bonded molecular interactions. *Bioinformatics* (2019) 35: 3499-3501.

35. Sievers, Fabian, Wilm Andreas, Dineen David and Gibson Toby J et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *MolSyst Biol*. (2011) 7:539.

36. Brielle Esther S, Duhovny Dina Schneidman, Linial Michal. The SARS-CoV-2 Exerts a Distinctive Strategy for Interacting with the ACE2 Human Receptor. *Viruses* (2020)12: 497.

37. Zhou Peng, Yang Xing-Lou, Wang Xian-Guang and Ben Hu et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* (2020) 579: 270-273.

38. Li Fang. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J. Virol* (2012) 86: 2856-2858.

39. Wrapp, Daniel, Wang Nianshuang, Corbett Kizzmekia S and Goldsmith Jory A et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* (2020) 367: 1260-1263.