How does the novel Coronavirus interact with the human ACE2 enzyme? A thermodynamic answer.

Summary

The SARS-CoV-2 coronavirus pandemic is certainly the most important public health concern today. Until now there are no vaccines or treatments available, despite intensive international efforts. One of the targets for new drugs is the Coronavirus Spike Protein, responsible for its binding and entry into the host cells. The Receptor Binding Domain (RBD) found at the Spike Protein recognizes the human angiotensin-converting enzyme 2 (hACE2). The present *in silico* study discusses structural and thermodynamic aspects of the protein complexes involving the RBD's from the 2002 SARS-CoV and 2019 SARS-CoV-2 with the hACE2. Molecular docking and molecular dynamics simulations of the complexes and isolated proteins were performed, providing insights on their detailed pattern of interactions, and estimating the free energy of binding. The obtained results support previous studies indicating that the chemical affinity of the new SARS-CoV-2 for the hACE2 enzyme virus is much higher than the 2002 SARS-CoV. The herein calculated Gibbs free energy of binding to the hACE2 enzyme is, depending on the technique, from 5.11 kcal/mol to 8.39 kcal/mol more negative in the case of the new coronavirus' RBD. In addition, within each employed technique, this free energy is consistently 61±2% stronger for SARS-CoV-2 than for SARS-CoV. This work presents a chemical reason for the difficulty in treating the SARS-CoV-2 virus using drugs targeting its Spike Protein, as well as helps to explain its infectivity, while defining a minimum free energy of binding for new drugs to be designed against this disease.

Jones de Andrade^{a,b}, Paulo Fernando Bruno Gonçalves^{a,c}, Paulo Augusto Netz^{a,d}

^aTheoretical Chemistry Group, Institute of Chemistry, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9.500, CEP: 91.501-970 Porto Alegre – RS, Brazil

^bjdandrade@iq.ufrgs.br, ^cpaulo@iq.ufrgs.br, ^dnetz@iq.ufrgs.br

Introduction

In November 2002, the zoonotic coronavirus SARS-CoV, the agent to cause the disease now known as SARS (Severe Acute Respiratory Syndrome), led to an outbreak that threatened the world with the risk of a pandemic. A high percentage of the infected people developed severe symptoms (20% to 30% needing mechanical ventilation), but with a late viral peak load and concentrated on the lower respiratory tract, leading to a low infection rate, it was declared finished in July 2003.⁽¹⁾ Recently, in December 2019, a new coronavirus, named SARS-CoV-2, appeared causing a widespread disease now named as COVID19 (Coronavirus Disease 2019). More infectious than the previous virus, however with a lower rate of infected patients developing severe symptoms (20% need any hospital treatment, and only a fraction of those need mechanical respiration), it initially spread almost unnoticed by a much larger number of people, beyond international borders, this time causing an outbreak that rapidly led to a worldwide pandemic that, until 2020, April 30, had already infected more than 4.6 million people and took the lives of more than 312 thousand.^(2,3) Up to this very moment, there are no treatments or vaccines available, leading several governments to instate countrywide quarantines in different levels up to full lock-downs, as well as social distancing as means to try to contain the disease propagation down to bearable levels to each healthy system.⁽⁴⁾

In such circumstances, the scientific community began to simultaneously study the pathogen itself, started more than a hundred different vaccine development projects⁽⁵⁾, searching for re-purposing of different already available medication^(6,7) and began the search and development of new and alternative

drugs.^(8,9) and treatments⁽¹⁰⁾. Any potential drug alternative aiming for success in fighting any pathogen will attack a specific target on it. In case of the SARS-CoV-2, the main drug targets are its Protease⁽¹¹⁾, Replicase⁽¹²⁾ and the Spike^(13,14) Proteins, responsible respectively for the host cell proteins cleavage, production of new viruses' genetic material and for its initial attachment to the human target cells during the virus infectious cycle.

Both the SARS-CoV and SARS-CoV-2 Spike Proteins are composed by two subunits, S1 and S2.⁽¹³⁾ The S1 subunit hosts the receptor-binding domain (RBD), whose receptor-binding motif (RBM)⁽¹³⁾ in its turn interacts with the human angiotensin-converting enzyme 2 (hACE2) of the host cells.⁽¹⁴⁾ This human enzyme is localized at the cell membrane of tissues of several human organs, like lungs, arteries, intestines, liver, kidneys and the upper respiratory tract.⁽¹⁵⁾

Recent experimental studies have already shown a greater chemical affinity of the new virus Spike Protein RBD towards the common ACE2 target molecule in human cells when compared to the previous SARS counterpart⁽¹⁶⁻¹⁸⁾, linking this observation as one possible cause of its much higher observed infectivity. Moreover, X-rays diffraction^(17,18,20) and Cryo-EM^(16,19) determined structures of their bound complexes are also already available, providing their structure as needed by computational studies. A precise knowledge of the Gibbs free energy of binding between the new virus Spike Protein RBD and the hACE2 target molecule is necessary in order to provide a proper minimum thermodynamic goal for any successful rational design of drugs to competitively inhibit the SARS-CoV-2 virus, at least when the Spike Protein is the target. The comparison of this thermodynamic property towards the one from 2002 SARS-CoV is necessary to properly understand their different infectivity and resistance to different drugs, in addition to providing a clearer overview of the biochemical battleground ahead.

In the present *in silico* study the SARS-CoV RBD, SARS-CoV-2 RBD, hACE2 and their binding complexes are discussed. Molecular docking and molecular dynamics simulations of the five systems were performed as detailed in the methods section and in the online content. The results provide an insight about their hydrogen bonding features and free energy of binding and support the previous observations^(16,18) that the chemical affinity of the new SARS-CoV-2 virus is much higher than the previous 2002 SARS-CoV.

Results

Molecular Docking of SARS-CoV and SARS-CoV-2 RBD's in the receptor hACE2

The main docking results are summarized in Fig. 1, that shows the hydrogen bonding pattern (1A), the histograms of conformations (1C and 1E) with clusters built from similarity ranked by free energy of binding (docking scores) and the 3D-view representation of the interactions (1B and 1D). The shown results of four independent docking runs are statistically identical. Both the RBD's show strong interactions, but the SARS-CoV-2 RBD interactions with the hACE2 are much stronger (-13.76 kcal/mol against -8.65 kcal/mol favoring the SARS-CoV-2 RBD, therefore, a difference of 5.11 kcal/mol or 59%). This difference must be taken into account when developing new drugs targeting the SARS-CoV-2 RBD, and explains why effective drugs against SARS-CoV were not effective against SARS-CoV-2, as it has greater thermodynamic affinity with the hACE2 receptor.

Fig. 1. Docking Results: SARS-CoV and SARS-CoV-2 RBD's in the hACE2 Receptor. (A) Mapping representation of the SARS-CoV RBD (top), hACE2 (middle) and SARS-CoV-2 RBD (below) intermolecular hydrogen bonding interactions of the lowest binding energy complex conformation. (B) SARS-CoV RBD (in orange) lowest free energy of binding configuration docked on hACE2 (in light blue), showing the intermolecular hydrogen bonded residue pairs. Depicted are: (from hACE2 α -helix 1: TYR41–THR'486; from α -helix 12: GLU329–ARG'426; and from the coil between the two β -sheets: LYS353–GLY'488). Green regions are hACE2 residues with important hydrophobic interactions with the SARS-CoV RBD chain (from hACE2 α -helix 1: GLN24, THR27, PHE28, LYS31, HIS34, GLY35, ASP38, GLN42 and LEU45; from α -helix 2: MET82; from the coil just out of the α -helix 2: TYR83; from α -helix 12: ASN330; from the coil between the two β -sheets: GLY354; and from the second β -sheet: ASP355 and ARG357) revealed by docking. (C) Histogram of SARS-CoV RBD docked configurations on hACE2 from 4 different random-seed runs, ranked by their free energy of binding in 0.5 kcal/mol bins. (D) SARS-CoV-2 RBD (in orange) lowest free energy of binding docked on hACE2 (in light blue), with the intermolecular hydrogen bonded residue pairs. Depicted are: (from the coil just before the hACE2 α -helix 1: SER19–ALA'475; on the α -helix 1: LYS31–GLN'493, GLU35–GLN'493, ASP38–TYR'449, TYR41–THR'500 and GLN42–GLN'498; from the coil just out of α -helix 2: TYR83–ASN'487; and from the coil between the two β -sheets: LYS353–GLY'496 and LYS353–GLY'502). Green regions are hACE2 residues with important hydrophobic interactions with the SARS-CoV RBD chain (GLN24, THR27, PHE28 and LEU45, all from α -helix 1) revealed by docking. (E) Histogram of SARS-CoV-2 RBD docked configurations on hACE2 from 4 different random-seed runs, ranked by their free energy of binding in 0.5 kcal/mol wide bins. Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Concerning the chemical interactions between each ligand (viruses RBDs) and the receptor (hACE2 enzyme), the virus interaction mainly occurs with three of the enzyme α -helices (1, 2, 12), both β -sheets and a few coils. As expected, the hydrogen bonds are fundamental for an effective connection of the viruses' RBD's and the hACE2. In this metric, the lowest free energy of binding conformation from the 2002 SARS-CoV RBD makes 3 hydrogen bonds to the hACE2 enzyme: TYR41-THR'486 on α -helix 1, GLU329-ARG'426 on α -helix 12 and LYS353-GLY'488 in the coil between the two β -sheets (from now on, the apostrophe indicates the RBD residue in the pair). Furthermore, the hACE2 hydrophobically interacts 14 of its residues (GLN24, THR27, PHE28, LYS31, HIS34, GLU35, ASP38, GLN42 and LEU45 on hACE2 α -helix 1; MET82 on hACE2 α -helix 2 and TYR83 on the coil just out of it; ASN330 on hACE2 α -helix 12; GLY354 on the coil between both hACE2 β -sheets; and ASP355 and ARG357 from the second β -sheet) with the SARS-CoV RBD. As such, the docking results indicates that the SARS-CoV virus' RBD evolved to successfully interact with a considerably large surface of hACE2.

Comparatively, the docking results seem to indicate that the SARS-CoV-2 virus' RBD evolved to completely disregard the helices 2 and 12 of hACE2, as well as its β -sheets, focusing almost exclusively on the hACE2 α -helix 1 to make a total of 9 hydrogen bonds to the human enzyme receptor (LYS31-GLN'493, GLU35-GLN'493, ASP38-TYR'449, TYR41-THR'500 and GLN42-GLN'498 on α -helix 1; SER19-ALA'475 on the coil just before of it; TYR83-ASN'487 on the coil just out of α -helix 2; and both LYS353-GLY'496 and LYS353-GLY'502 from the hairpin between the two β -sheets), while at the same time moving away from many hydrophobic interactions, which appear with only 4 of hACE2 residues (GLN24, THR27, PHE28 and LEU45), all at its α -helix 1. It is relevant that for at least 5 hACE2 residues to which the previous virus only had weak hydrophobic interactions (LYS31, GLU35, ASP38, GLN42 and TYR83), the SARS-CoV-2 binds through stronger hydrogen bonds cannot be overrated. There is also one hACE2 residue (LYS353) pinched by the SARS-CoV-2 RBD's GLY'496 and GLY'502, while the GLN'493 residue is precisely positioned to be pinched back by the hACE2 LYS31 and GLU35. There is no similar pinching behavior observed in the SARS-CoV RBD interactions with hACE2, which may result in its weaker overall interactions. This greater number of well positioned residues of the SARS-CoV-2 RBD contributes to their stronger hydrogen bonds, which is the main reason for its much more favorable and stronger ligand-receptor interaction observed in the molecular docking results reported here, while the 2002 SARS-CoV showed mostly weaker hydrophobic interactions.

Molecular dynamics of SARS-CoV and SARS-CoV-2 complexes with hACE2: Hydrogen Bonding Features

While the docking results show a static picture of the interaction, the MD simulations, which explicitly take into account the explicitly the flexibility of both proteins, show a qualitative similar, yet somewhat different, qualitative picture.

According to the MD results, it can be observed in both complexes the interesting feature of at least three distinct "hydrogen bonding networks" (NET1, NET2 and NET3) between each virus's RBD and the hACE2 enzyme. These networks were defined as involving common hACE2 residues for both viruses RBDs that were interconnected in at least one of the complexes. In addition to these three networks, there is also one arginine residue from each virus RBD (ARG'426 in case of SARS-CoV and ARG'439 in case of SARS-CoV-2) making several simultaneous hydrogen bonds to the same hACE2 glutamate residue (GLU329), and three intermolecular hydrogen bonding residue pairs that are exclusive to each RBD-hACE2 complex. It must be pointed out that the GLU329-ARG'439 hydrogen bonds found for the SARS-CoV-2 RBD are due to the chimera structure used in the present study, as it is not found in the wild enzyme.

The first hydrogen bond network (henceforward “NET1”) identified from MD simulations (Fig. 2), involves the hACE2 residues SER19, GLN24, THR27 and TYR83. SARS-CoV lacks interactions with THR83, while SARS-CoV-2 does not interact with both SER19 and THR27.

The SARS-CoV RBD has a NET1 network of 5 hydrogen bonds connecting 2 of its residues (PRO’462 and ASP’463) to three hACE2 residues (SER19, GLN24 and THR27) in 4 pairs of residues (SER19–PRO’462, THR27–PRO’462, SER19–ASP’463 and GLN24–ASP’463). The ASP’463 carboxylate group usually make 2 simultaneous hydrogen bonds to the SER19 protonated amine hydrogen atoms. Both groups free rotation occasionally not only changes the specific atoms involved in the hydrogen bonding of NET1, but can also make a second hydrogen bond to the GLN24 amine instead of SER19. Also important for this network is the double connection from the PRO’462 main chain oxygen atom towards simultaneously the hydroxyls from both SER19 and THR27 residues.

On the other hand, the SARS-CoV-2 RBD has a NET1 network of 4 hydrogen bonds on average, connecting 3 of its residues (ALA’475, GLY’485 and ASN’487) to two hACE2 residues (GLN24 and TYR83) in 4 pairs of residues (GLN24–ALA’475, GLN24–ASN’487, TYR83–GLY’485 and TYR83–ASN’487). For instance, the hACE2 TYR83 hydrogen bonding, although it has a strong bind by its hydroxyl oxygen atom towards the virus RBD ASN’487 at the main chain amine hydrogen atom, it can also form a second bonding to either the same ASN’487 or from its hydroxyl hydrogen atom towards the GLY’485 residue main chain oxygen atom. The hACE2 GLN24 residue is on the other hand persistently bound by its amide group hydrogen atoms to both the ALA’475 main chain oxygen atom and ANS’487 amide group oxygen atom. The free rotation of these GLN24, ASN’487 and ALA’475 groups is probably responsible for the hydrogen profile shown for them, which does not usually break any connection but also (rarely) leads to the formation of two simultaneous hydrogen bonds with the ASN’487 residue while keeping the one with ALA’475.

Fig. 2. Molecular dynamics results: SARS-CoV and SARS-CoV-2 RBD's hydrogen bond network NET1 with hACE2. (A) Mapping representation of the SARS-CoV RBD (top), hACE2 (middle) and SARS-CoV-2 RBD (below) intermolecular hydrogen bonding interactions of the network NET1 (in blue). (B) SARS-CoV RBD to hACE2 hydrogen bonds presence profiles for hydrogen bond network NET1 (from top: SER19-PRO'462, SER19-ASP'463, GLN24-ASP'463 and THR27-PRO'462). (C) SARS-CoV-2 RBD and hACE2 hydrogen bonds presence profiles for hydrogen bond network NET1 (from top: GLN24-ALA'475, GLN24-ASN'487, TYR83-GLY'485 and TYR83-ASN'487). (D) SARS-CoV RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting the hydrogen bond network NET1 (SER19-PRO'462, SER19-ASP'463, GLN24-ASP'463 and THR27-PRO'462). (E) SARS-CoV-2 RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting the hydrogen bond network NET1 (GLN24-ALA'475, GLN24-ASN'487, TYR83-GLY'485 and TYR83-ASN'487). Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

The second relevant intermolecular hydrogen bond network ("NET2") identified from the MD simulations (Fig. 3), involves the hACE2 residues LYS31, HIS34 and GLU35, and, of those three, the SARS-CoV RBD only interacts with HIS34, while SARS-CoV-2 bonds to all of them.

The SARS-CoV RBD has a NET2 network of 3 hydrogen bonds connecting 2 of its residues (TYR'440 and ASN'479) to a single hACE2 residue (HIS34). The ASN'479 amide group usually makes two simultaneous hydrogen bonds from its oxygen and nitrogen atoms to the hydrogen atom on the iminic nitrogen of the HIS34 imidazole group. This attack is reinforced by the TYR'440 hydroxyl oxygen atom. Despite the freely rotating imidazole group from the TYR'440, both it and the hydroxyl group

from the ASN'479 on the SARS-CoV RBD are kept tight locked in position by extra intramolecular interactions caused by those two residues, thus resulting in the hydrogen bonding presence profiles observed.

The SARS-CoV-2 RBD however presents a NET2 network of, in average, 5 hydrogen bonds connecting 2 of its residues (PHE'490 and GLN'493) to all three NET2 hACE2 residues (LYS31, HIS34 and GLU35) in 4 pairs of residues (LYS31–PHE'490, LYS31–GLN'493, HIS34–GLN'493 and GLU35–GLN'493). This greater number of hydrogen bonds not only favors the SARS-CoV-2 RBD network NET2 over the SARS-CoV network in the classical mechanical context, but probably also when studied by the quantum chemistry approach, as greater stability should arise due to relaxation of charges over the atoms in the network. The HIS34 residue, whose imidazole ring secondary amine hydrogen atom was the focal point of 3 hydrogen bonds from 2 different residues in the SARS-CoV RBD, now have a single constant hydrogen bond of its main chain oxygen atom towards one of the GLN'493 residue amide group hydrogen atoms. The strongest interaction now comes from the two hydrogen bonds between other GLN'493 amide group hydrogen atom and the GLU35 carboxylate oxygen atoms. Also, from the GLN'493 residue amide group, however from its oxygen atom this time, comes the interaction with one of the LYS31 residue protonated amine hydrogen atoms. It is important to notice that both LYS31 and specially the HIS34 residues seems to play important roles to lock the GLN'493 in the correct orientation for the formation of its two hydrogen bonds towards the GLU35 carboxylate group oxygen atoms. Finally, another of the LYS31 protonated amine hydrogen atoms is simultaneously attacked by the PHE'490 main chain oxygen atom. Both LYS31 interactions are, as expected, mostly synchronous in their hydrogen bond presence profiles, and only show some interruptions between 60 ns and 175 ns, most probably due to a temporary displacement of the freely rotating LYS31 alkyl chain.

Fig. 3. Molecular dynamics results: SARS-CoV and SARS-CoV-2 RBD's Hydrogen Bond Network NET2 with hACE2. (A) Mapping representation of the SARS-CoV RBD (top), hACE2 (middle) and SARS-CoV-2 RBD (below) intermolecular hydrogen bonding interactions of the network NET2 (in green). (B) SARS-CoV RBD to hACE2 hydrogen bonds presence profiles for hydrogen bond network NET2 (from top: HIS34–TYR'440 and HIS34–ASN'479). (C) SARS-CoV-2 RBD and hACE2 hydrogen bonds presence profiles for hydrogen bond network NET2 (from top: LYS31–PHE'490, LYS31–GLN'493, HIS34–GLN'493 and GLU35–GLN'493). (D) SARS-CoV RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting the hydrogen bond network NET2 (HIS34–TYR'440 and HIS34–ASN'479). (E) SARS-CoV-2 RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting the hydrogen bond network NET2 (LYS31–PHE'490, LYS31–GLN'493, HIS34–GLN'493 and GLU35–GLN'493). Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

The third important intermolecular hydrogen bonds network (henceforward, “NET3”), identified in the MD simulations (Fig. 4), involves the hACE2 residues TYR41, LYS353 and ASP355. Hydrogen bonds between all of them and several residues from both the SARS-CoV and the SARS-CoV-2 RBDs are observed: however, while in the SARS-CoV-2 they all form a single highly interconnected network, in the SARS-CoV they can be easily divided in three “partial-networks”. This makes this network different from the previous networks, specially the first one. As a result, the overall interaction picture of the NET3 network, at least exclusively in the classical mechanics approach, is more favorable to the SARS-CoV-2 RBD interactions over the SARS-CoV ones.

In the SARS-CoV RBD the NET3 network consists of 7 hydrogen bonds between 6 of its residues (THR'433, GLY'434, TYR'481, GLY'482, THR'486 and GLY'488) and all NET3 hACE2 residues forming 6 residue pairs (TYR41–THR'433, TYR41–GLY'434, LYS353–TYR'481, LYS353–GLY'482, LYS353–GLY'488 and ASP355–THR'486). They can be divided in three “partial networks”: “NET3a” consists of the first two pairs, “NET3b” of the following 3 pairs, and “NET3c” is composed only by the last pair. In NET3a, the TYR41 hydroxyl hydrogen atom makes hydrogen bonds to both the main chain oxygen atom from THR'433 and the main chain nitrogen atom from GLY'434. While not present in the first 100 ns of the MD simulation, the freely rotating TYR41 hydroxyl is later locked in position by those hydrogen bonds, as can be observed from their presence profiles. On the other hand, NET3c partial network presents the THR'486 hydroxyl tight locked in position by two hydrogen bonds between its hydrogen and both carboxylate oxygen atoms from ASP355. Finally, NET3b is centered at the LYS353, which possesses 2 interaction sites. At one interaction site, the LYS353 main chain oxygen atom bonds to the GLY'488 residue at its main chain amine hydrogen atom. On the other interaction site, the LYS353 also bonds simultaneously by two of its protonated amine group hydrogen atoms to the main chain oxygen atoms from both TYR'481 and GLY'482 residues.

In contrast, the SARS-CoV-2 RBD NET3 network is composed by 10-11 hydrogen bonds between 6 of its residues (TYR'495, GLY'496, GLN'498, THR'500, ASN'501 and GLY'502) and all hACE2 residues from this network, forming 9 residue pairs (TYR41–THR'500, TYR41–ASN'501, LYS353–TYR'495, LYS353–GLY'496, LYS353–GLN'498, LYS353–ASN'501, LYS353–GLY'502, ASP355–THR'500 and ASP355–ASN'501). It should be stressed that this is a fully connected network that is not divided in “partial networks” as in the SARS-CoV virus RBD, accounting for a much more intense stabilization. Despite that, this NET3 will be discussed in relation to the findings from the other virus one as much as possible.

As such, similarly to the SARS-CoV RBD, the TYR41 is stabilized by two hydrogen bonds from both the THR'500 and ASN'501 residues. However, instead of the TYR41 hydroxyl hydrogen atom, the interaction is made with hydroxyl oxygen atom. Moreover, instead of the aminoacids main chain atoms, it is the THR'500 hydroxyl hydrogen atom and one of the ASN'501 primary amide hydrogen atoms that are involved in the attack. However, this last interaction appears to weakens past 200 ns of simulation.

Also, in the ASP355 target site, its freely rotating hydroxyl was originally restrained by two simultaneous hydrogen bonds with the SARS-CoV RBD THR'486 carboxylate group oxygen atoms, while in the SARS-CoV-2 RBD it is similarly bonded to the two THR'500 carboxylate group oxygen atoms. However, the ASP355 also forms one additional hydrogen bond with the other ASN'501 amide hydrogen atom.

Finally, the LYS353 have two independent interaction sites. First, its main chain oxygen atom interacts with the SARS-CoV-2 GLY'502 main chain amine hydrogen atom in a mirror image to the SARS-CoV GLY'488 bond. Secondly, one of the LYS353 protonated primary amine hydrogen atoms makes two bonds to the SARS-CoV-2 RBD TYR'495 and GLY'496 main chain oxygen atoms, also mirroring with the SARS-CoV RBD TYR'481 and GLY'482 interactions. However, the LYS353 protonated amine other hydrogen atom makes one more hydrogen bond with the ASN'501 aminoacid at the amide's oxygen atom. Moreover, after about 200ns of simulation it also makes two additional simultaneous hydrogen bonds to the GLN'498 aminoacid, simultaneously at its oxygen and nitrogen atoms from its originally freely rotating amide group, locking it too in position.

Fig. 4. Molecular dynamics results: SARS-CoV and SARS-CoV-2 RBD's Hydrogen Bond Network NET3 with hACE2. (A) Mapping representation of the SARS-CoV RBD (top), hACE2 (middle) and SARS-CoV-2 RBD (below) intermolecular hydrogen bonding interactions of the network NET3 (in dark red). (B) SARS-CoV RBD to hACE2 hydrogen bonds presence profiles for hydrogen bond network NET3 (from top: TYR41–THR'433, TYR41–GLY'434, LYS353–TYR'481, LYS353–GLY'482, LYS353–GLY'488 and ASP355–THR'486). (C) SARS-CoV-2 RBD and hACE2 hydrogen bonds presence profiles for hydrogen bond network NET3 (from top: TYR41–THR'500, TYR41–ASN'501, LYS353–TYR'495, LYS353–GLY'496, LYS353–GLN'498, LYS353–ASN'501, LYS353–GLY'502, ASP355–THR'500 and ASP355–ASN'501). (D) SARS-CoV RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting the hydrogen bond network NET3 (TYR41–THR'433, TYR41–GLY'434, LYS353–TYR'481, LYS353–GLY'482, LYS353–GLY'488 and ASP355–THR'486). GLY'488 is not visible in the figure, and ASN'501 view is heavily obstructed. (E) SARS-CoV-2 RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting the hydrogen bond network NET3 (TYR41–THR'500, TYR41–ASN'501, LYS353–TYR'495, LYS353–GLY'496, LYS353–GLN'498, LYS353–ASN'501, LYS353–GLY'502, ASP355–THR'500 and ASP355–ASN'501). Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Besides these hydrogen bond networks, there are also many hydrogen bonds that do not belong to this pattern. Nevertheless, they were found in the molecular dynamics simulations. It is worth mentioning those that are depicted in Fig. 5, and include for instance the important interactions that the hACE2 GLU329 residue makes to ARG'426 and ARG'439 residues, respectively for SARS-CoV and SARS-CoV-2 RBDs. The importance of these two residue pairs is that the (sparse however strong) hydrogen bonding behavior of these residues is a natural feature on the wild SARS-CoV used fragment, but on the SARS-CoV-2 chimera employed it is an artifact resulting from the ARG'439 inclusion. The hACE2 GLU329 carboxylate and both viruses RBD arginines protonated guanidino group makes several (often up to three) simultaneous hydrogen bonds.

Moreover, there are a few interaction sites that are characteristic and exclusive to each one of the viruses. First there is the SARS-CoV TYR'442 hydroxyl hydrogen interaction with the ASP30 residue carboxylate group oxygen atoms, which after 100ns of simulation forms initially two hydrogen bonds but later shows only one fairly persistent one between 130ns and 230ns. No similar interaction is found on the SARS-CoV-2 RBD complex, but its TYR'505 successfully bonds its hydroxyl hydrogen atom towards the GLU37 carboxylate group two oxygen atoms, forming two very persistent hydrogen bonds that only temporarily fades between 160ns and 240ns. This interaction also has no equivalent observed at the SARS-CoV RBD.

There are also two residue pairs for each RBD complexes that, while related to the hydrogen bonding network NET3, are not part of it in either one of them. For instance, it must be observed that the SARS-CoV RBD NET3 is closely localized to a pair of exclusive intermolecular hydrogen bonded residue pairs, formed by the its TYR'484 and the hACE2 LEU351 and GLY352 residues. These two pairs form very persistent (from 75ns of simulation) single hydrogen bonds, between the virus TYR'484 and the main chain oxygen atoms from both LEU351 and GLY352. The occasional event of a second hydrogen bond being formed with the later residue is mostly due to its well positioned main chain nitrogen atom. On the other hand, the SARS-CoV-2 RBD TYR'449 residue also makes two completely different worth mentioning hydrogen bonds with the hACE2 enzyme, at its ASP38 and GLN42 residues. Although not so close to its RBD NET3 as in the SARS-CoV case, it does form an ephemeral hydrogen bonding connection between GLN42 and GLN'498, specially around 150 ns of simulation. Aside from this connection, the TYR'449 hydroxyl group hydrogen atom does form two very persistent hydrogen bonds towards both the ASP38 carboxylate group oxygen atoms. Simultaneously, the TYR'449 hydroxyl group oxygen atom makes another single hydrogen bond with the GLN42 amine hydrogen atoms.

Fig. 5. Molecular dynamics results: SARS-CoV and SARS-CoV-2 RBD's Hydrogen Bonds not belonging to the discussed Networks with hACE2. (A) Mapping representation of the SARS-CoV RBD (top), hACE2 (middle) and SARS-CoV-2 RBD (below) intermolecular hydrogen bonding interactions involving the hACE2 enzyme residues GLU329 (in black), ASP30 (in orange), both LEU351 and GLY352 (in light cyan), GLU37 (in pink), both ASP38 and GLN42 (in dark magenta), as well as the ephemeral interconnection with NET3 (dashed yellow line). (B) SARS-CoV RBD to hACE2 hydrogen bonds presence profiles for hACE2 enzyme complex with SARS-CoV RBD GLU329-ARG'439 (in black), ASP30-TYR'442 (in orange) and both LEU351-TYR'484 and GLY352-TYR'484 (in light cyan) pairs of residues. (C) SARS-CoV-2 RBD and hACE2 hydrogen bonds presence profiles for hACE2 enzyme complex with SARS-CoV-2 RBD GLU329-ARG'439 (in black), GLU37-TYR'505 (in pink) and both ASP38-TYR'449 and GLN42-TYR'449 (in dark magenta), as well as the ephemeral connection with NET3 GLN42-GLN498 (in yellow), pairs of residues. (D) SARS-CoV RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting hACE2 enzyme complex hydrogen bonds with SARS-CoV RBD of the GLU329-ARG'426, ASP30-TYR'442, LEU351-TYR'484 and GLY352-TYR'484 pairs of residues. (E) SARS-CoV-2 RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting hACE2 enzyme complex hydrogen bonds with SARS-CoV-2 RBD GLU329-ARG'439, GLU37-TYR'505, ASP38-TYR'449 and GLN42-TYR'449 pairs of residues (the GLN42-GLN'498 interactions are now shown). Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

The hydrogen bonds play a major role for the chemical systems stability, and it is believed that the presence and extension of the networks reported are decisive for the stabilization of each complex, as they would lead to a major charge distribution, which in turn will better distribute its energy and

increase the entropic contribution to free energy. Studies of non-covalent interactions and frontier orbitals to elucidate this problem are being carried out and will be the subject to future publications.

Binding free energy estimates for SARS-CoV and SARS-CoV-2 Complexes with hACE2

The docking scores, reported in Fig. 1, are estimates of the free energy of binding from an empirical free energy force field^(21,22) – within limitations of the molecular docking methodology – where the receptor and (in this case) also the ligand are kept rigid, using the experimentally determined structures⁽¹⁶⁾. For the SARS-CoV RBD docked in the hACE2 receptor, the free energy of binding of -8.65 kcal/mol was obtained. A much more negative value of -13.76 kcal/mol was obtained for the SARS-CoV-2 RBD. The two systems differ by 5.11 kcal/mol, or a 59.1% favoring the SARS-CoV-2 binding over the SARS-CoV.

Based on averages over the entire molecular dynamics trajectories using the AMBER 03⁽²⁴⁾ force field with full molecular flexibility, the LIE – Linear Interaction Energy⁽²³⁾ – free energy of binding calculations resulted in -13.48 kcal/mol and -21.87 kcal/mol respectively to SARS-CoV and SARS-CoV-2 viruses RBDs to the hACE2 enzyme, a difference of 8.39 kcal/mol. While all these values strongly differ from the molecular docking calculation results, such value is 62,2% more negative for the SARS-CoV-2, showing consistency with the docking estimates.

The two above reported different techniques employed (Molecular Docking scores based on an empirical free energy force field^(21,22) and LIE⁽²³⁾ estimated free energies based on MD trajectories) to evaluate the free energy of binding between each virus' RBD and the hACE2 enzyme yields different absolute values and even absolute differences between them within each technique. That is expected, since each one is based on different configurations (from X-ray crystallography^(17,20) studies and MD simulations themselves) and interaction parameters (from Autodock 4.2⁽²²⁾ and AMBER 03⁽²⁴⁾ force fields). However, the observed consistency in the relative differences within the techniques cannot be understated and provides a rational explanation for the known difficulties faced in combating the SARS-CoV-2 virus, at least when targeting the Spike Protein with the existing drugs.

Conclusions

From figure 1A to 5A, it can be seen that any of the residues involved in the intermolecular hydrogen bonding pattern on molecular docking are maintained in the results of molecular dynamics simulation, with the inclusion of other residues. Also, the number of hydrogen bonds identified for either technique in the SARS-CoV and SARS-CoV-2 comparison favored the later in both cases: respectively, 3 and 9 for molecular docking, and 16 and 21 in molecular dynamics. All this accounts for the greater affinity⁽¹⁶⁻¹⁸⁾ that exists between the hACE2 and SARS-CoV-2 when compared to the interaction with SARS-CoV, as well as to a large rearrangement of the complex and additional stabilization in solution when compared to the crystal structure. However, additional information comes from comparing the three main hydrogen bonds networks identified in each complex. The respective SARS-CoV and SARS-CoV-2 RBDs interacts with the hACE2 by 5 and 4 hydrogen bonds in NET1, 3 and 5 in NET2 and finally 7 and 10-11 in NET3. Whereas in the previous virus the NET3 network is divided into 3 partial networks with 2 to 3 hydrogen bonds each, this means that an effective drug targeting the SARS-CoV RBD can probably be expected to bind to its NET1 residues, while a potential drug against SARS-CoV2 targeting its RBD should be more likely targeting its NET3 (or the residues of each network in hACE2).

In addition, when taking into account the specific residue pairs involved in each complex hydrogen bonds, few changes are observed. For instance, in SARS-CoV RBD complex, from the 3 hydrogen bonds observed in molecular docking to the 16 in molecular dynamics, two (GLU329–ARG'426 and

LYS353–GLY'488) are maintained and 1 (TYR41–THR'486) is broken. Also, from the SARS-CoV-2 RBD complex from the docking 9 hydrogen bonds to the molecular dynamics 21, seven were maintained (LYS31–GLN'493, GLU35–GLN'493, ASP38–TYR'449, TYR41–THR'500, TYR83–ASN'487, LYS353–GLY'496 and LYS353–GLY'502), one is broken (SER19–ALA475) and a last one (GLN42–GLN'498) is reported as the ephemeral hydrogen bond between two other pairs of residues (forming three hydrogen bonds) and NET3.

It was expected that the rigid Molecular Docking configured using the X-ray crystallography^(17,20) obtained structures would show large agreement with those experimental results. For most of the hydrogen bond interactions involved residues, that was true. The Molecular Docking identified that three hACE2 residues (TYR41, GLU329 and LYS353) would effectively form hydrogen bonds towards three SARS-CoV residues (ARG'426, THR486 and GLY488). Also, 8 hACE2 residues (SER19, LYS31, GLU35, ASP38, TYR41, GLN42, TYR83 and LYS353) were prone to hydrogen bond formation towards 8 SARS-CoV2 RBD residues (TYR'449, ALA'475, ASN'487, GLN'493, GLY'496, GLN'498, THR'500 and GLY'502).

Shang *et al.*⁽¹⁷⁾ and Lan *et al.*⁽¹⁸⁾ reported X-ray diffraction studies comparing the structures of both viruses RBD complexes with hACE2. Regarding the SARS-CoV complex, the first identified 27 residue pairs in close contact to each other, while the second identified 12 residue pair connected by hydrogen bonds. Much less (only 3) hydrogen bonds were identified in the docking calculation, however all them were observed in both experimental measurements.

On the other hand, the molecular dynamics simulation identified a greater number of hydrogen bonds (16), but some differences were observed when compared to the experiments. In common with Shang *et al.*⁽¹⁷⁾, the molecular dynamics simulations have only 6 out of its 16 hydrogen bonded residue pairs: HIS34–TYR'440, HIS34–ASN'479, GLU329–ARG'426, LYS353–GLY'482, LYS353–GLY'488 and ASP355–THR'486. However, in comparison with Lan *et al.*⁽¹⁸⁾, the molecular dynamics results show only two pairs of residues connected by hydrogen bonds: GLU329–ARG'426 and LYS353–GLY'488.

Experimental measurements from Lan *et al.*⁽¹⁸⁾ identified 10 extra residue pairs (GLN24–ASN'473, GLU37–TYR'491, ASP38–TYR'436, TYR41–THR'486, TYR41–THR'487, GLN42–TYR'436, TYR83–ASN'473, TYR83–TYR'475, GLN325–ARG'426, and ASN330–THR'486) not observed in the simulation, which in turn yielded 14 extra residue pairs (SER19–PRO'462, SER19–ASN'463, GLN24–ASP'463, THR27–PRO'462, ASP30–TYR'442, HIS34–TYR'440, HIS34–ASN'479, TYR41–THR'433, TYR41–GLY'434, LEU351–TYR'484, GLY352–TYR'484, LYS353–TYR'481, LYS353–GLY'482 and ASP355–THR'486) not found in the experiment.

Given the larger amount of hydrogen bonds identified by Shang *et al.*⁽¹⁷⁾, a larger number of extra residue pairs was found: 21. Also, 10 hydrogen bonded pairs identified in the molecular dynamics simulations (SER19–PRO'462, SER19–ASN'463, GLN24–ASP'463, THR27–PRO'462, ASP30–TYR'442, TYR41–THR'433, TYR41–GLY'434, LEU351–TYR'484, GLY352–TYR'484 and LYS353–TYR'481) were not observed in Shang *et al.*⁽¹⁷⁾ X-ray diffraction measurements.

The closest comparisons were obtained for the SARS-CoV-2 RBD complex with hACE2. Comparing with the molecular docking results, Shang *et al.*⁽¹⁷⁾ observed 33 contact residues directly involved in the complex formation, in contrast to 13 hydrogen bonds identified by Lan *et al.*⁽¹⁸⁾ and 9 from the docking calculations. All nine pairs connected by hydrogen bonds identified by the docking calculations were

found by Shang *et al.*⁽¹⁷⁾, but four (SER19–ALA'475, LYS31–GLN'493, GLN42–GLN'498 and LYS353–GLY'496) were not reported by Lan *et al.*⁽¹⁸⁾.

The molecular dynamics simulation identified 21 residue pairs forming hydrogen bonds, and when compared with the results presented by Shang *et al.*⁽¹⁷⁾, only 7 out of the 21 molecular dynamics identified pairs (GLN24–ALA'475, LYS31–PHE'490, HIS34–GLN'493, TYR83–GLY'485, LYS353–TYR'495, LYS353–GLN'498 and ASP355–ASN'501) were not observed in the diffraction measurements. Moreover, at least two have close relatives found in the experiment: GLN24–GLY'476 and TYR83–PHE'486.

However, in comparison with Lan *et al.*⁽¹⁸⁾, only 9 out of 21 molecular dynamics residue pairs (GLN24–ASN'487, GLU35–GLN'493, GLU37–TYR'505, ASP38–TYR'449, TYR41–THR'500, TYR41–ASN'501, GLN42–TYR'449, TYR83–ASN'487 and LYS353–GLY'502) were found on both. Lan *et al.*⁽¹⁸⁾ have also identified 4 additional hydrogen bonded residue pairs not observed in molecular dynamics: ASP30–LYS'417, GLN42–GLY'446, TYR83–TYR'489 and ARG393–TYR'505.

Additionally, the results here reported indicates that the Gibbs Free Energy of binding is about ~61% more negative for the current virus' RBD than the previous one (differences ranging from 5.11 kcal/mol to 8.39 kcal/mol accordingly to molecular docking and molecular dynamics LIE⁽²³⁾ techniques, respectively). Considering a thermodynamic constant $K \sim \exp(-\Delta G/RT)$, this relative increment yields a relative 4.91–5.06 fold increase favoring the complexed form over the free proteins for the SARS-CoV-2 RBD in comparison to the SARS-CoV one. This is in line with independently reported experimental surface plasmon resonance equilibrium measurement of the dissociation constants (K_D) for both complexes, whose ratios range from 4.17⁽¹⁶⁾ and 4.19⁽¹⁷⁾ to 6.76⁽¹⁸⁾ for the wild RBD, up to 7.97⁽¹⁷⁾ for the chimeric RBD used. Respectively, those experimental ratios indicate 42.7%⁽¹⁶⁾, 43.2%⁽¹⁷⁾, 91.1%⁽¹⁸⁾ and 107.6%⁽¹⁷⁾ more negative free energies of binding. This extra Gibbs free energy amount poses itself as a challenge for the rational design of drugs, while also most probably screens out several potential drugs. This provides a possible chemical reason for the difficulty of treatment of the SARS-CoV-2 virus with compounds targeting its Spike Protein and for its high infectiousness, while also setting a minimum target free energy of binding for new compounds to be rationally designed against this disease.

All methods used in this work shown that the SARS-CoV-2 virus RBD is better adapted to bind to the hACE2 enzyme than in case of SARS-CoV, in line with experimental data already published⁽¹⁶⁻¹⁸⁾. Concerning the relative free energy of binding, both rigid ligand and receptor Molecular Docking calculations and flexible molecules MD simulations were performed and an increment of ~61% of the relative free energy of binding for the SARS-CoV-2 in comparison to the SARS-CoV was consistently observed. The relative increase in free energy of binding of SARS-CoV-2 compared to SARS-CoV is 61±2%. This observation, in line with experimental measurements of affinity already reported⁽¹⁶⁻¹⁸⁾, proves the consistency of the methodology used in the present study and also provides a clear thermodynamic goal for the development of rationally designed drugs, targeting the SARS-CoV-2 Spike Protein RBD region or the host cell hACE2 enzyme. While it strongly raises the bar on what must be accomplished in terms of the Spike protein targeted drugs to overcome the COVID19 pandemic challenge, it sets a consistent parameter that must be crossed independently of the technique used to rationally develop them. Any designed inhibitor (especially a competitive one) for the treatment of COVID-19 will need to overcome this interaction energy to become effective against this virus, otherwise the virus-hACE2 binding will continue to be thermodynamically favored. Molecules whose free binding energy is only marginally more negative than the 2002 SARS-CoV virus-hACE2 complex formation will not be effective. These results therefore provide both the possibility of drug screening

and rational design drugs the necessary knowledge for the their development, a minimum condition of success.

References

- (1) Paules, C. I.; Marston, H. D.; Fauci, A. S. Coronavirus Infections – More Than Just the Common Cold, *Journal of the American Medical Association* 323, 707-708 (2020).
- (2) <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses>, accessed in may 18th, 2020.
- (3) <https://covid19.who.int>, accessed in may 18th, 2020.
- (4) Gibney, E. Whose coronavirus strategy worked best? Scientists junt most effective policies, *Nature* 581, 15-16 (2020).
- (5) Le, T. T. *et al.* The COVID-19 vaccine development landscape, *Nature Reviews – Drug Discovery* 19, 305-306 (2020).
- (6) COVID-19 Clinical Research Coalition “Global Coalition to accelerate COVID-19 clinical research in resource limited settings”, *The Lancet* 395, 1322-1325 (2020).
- (7) Mori, M. *et al.* A deadly spillover: SARS-CoV-2 outbreak, *Expert Opinion on Therapeutical Patents* 2020, yet unpublished
- (8) Dai, W. *et al.* Structure based desing of antiviral drug candidates targeting the SARS-CoV-2 main protease, *Science* 2020, yet unpublished
- (9) Sheahan, T. P. *et al.* An orally bioavaialable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cells culture and multiple coronavirus in mice, *Science Translational Medicine* 12, 1-15 (2020).
- (10) Cao, Y. *et al.* Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients’ B cells, *Cell* 2020, yet unpublished. X
- (11) Jin, Z. *et al.* Structure of M^{PRO} from COVID-19 virus and discovery of its inhibitors, *Nature* 2020, yet unpublished.
- (12) <https://www.biorxiv.org/content/10.1101/2020.03.28.013920v1>, accessed in may 18th, 2020. H
- (13) Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science* 367, 1260-1263 (2020).
- (14) Luan, J. *et al.* Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection, *Biochemical and Biophysical Research Communications* 526, 165-169 (2020).
- (15) <https://www.proteinatlas.org/ENSG00000130234-ACE2/tissue>, accessed in may 18th, 2020. Y
- (16) Walls, A. C. *et al.* Structure, Function and Antigenicity of the SARS-CoV-2 Spike Glycoprotein, *Cell* 180, 281-292 (2020).
- (17) Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2, *Nature* 581, 221-224 (2020).
- (18) Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, *Nature* 581, 215-220 (2020).
- (19) Yan, R. *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, *Science* 367, 1444-1448 (2020).
- (20) Li, F. *et al.* Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor, *Science* 309, 1864-1868 (2005).
- (21) Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated Docking with Selected Receptor Flexibility, *Journal of Computational Chemistry* 30, 2785-2791 (2009).
- (22) Huey, R. *et al.* A Semiempirical Free Energy Force Field with Carge-Based Dessolvation, *Journal of Computational Chemistry* 28, 1145-1152 (2006).
- (23) Hanson, T., Marelus, J. and Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods, *Journal of Computer-Aided Molecular Design* 12, 27-35 (1998).

- (24) Duan, Y. *et al.* A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations, *Journal of Computational Chemistry* **24**, 1999-2012 (2003).
- (25) Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* **1-2**, 19-25 (2015).
- (26) Jorgensen, W. L. *et al.* Comparison of simple potential functions for simulating liquid water, *Journal of Chemical Physics* **79**, 926-93 (1983).
- (27) Tironi, I. G. *et al.* A generalized reaction field method for molecular dynamics simulations, *Journal of Chemical Physics* **102**, 5451-5459 (1995).
- (28) Bussi, G., Donadio, D. and Parrinello, M. Canonic sampling through velocity rescaling, *Journal of Chemical Physics* **126**, 014101 1-7 (2007).
- (29) Parrinello, M. and Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method *Journal of Applied Physics* **52**, 7182-7190 (1981).
- (30) Vosmeer, C. R. *et al.* Improving the iterative Linear Interaction Energy approach using automated recognition of configurational transitions, *Journal of Molecular Modeling* **26**, 31 (2016).

Acknowledgments

The authors would like to acknowledge CNPq and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 for the financial support, and the High-Performance Computing Center (NPAD) at UFRN for the computational resources provided. We would also like to express our deepest gratitude towards Engr. Danilo Chaves de Sousa Ichihara and Engr. Haroldo Watson Teodósio da Silva, whose hard and dedicated work to, even in the midst of a pandemic, put the NPAD supercomputer back online was absolutely fundamental for the completion of the present study in time: for them two, our most sincere acknowledgments.

Authors Contributions:

All three author contributed equally to the present study.

Competing Interests:

The authors declare no competing interests.

Materials and Correspondence:

The supplementary information includes all 5 simulated systems Root Mean Square Deviations (RMSDs) graphs, full MD intermolecular hydrogen bonding mapping, and larger versions of rendered structural hydrogen bond features already present in the main text.

All correspondence and request for materials should be addressed to either one of the authors: JdA (jdandrade@iq.ufrgs.br), PFBG (paulo@iq.ufrgs.br), PAN (netz@iq.ufrgs.br).

Methods

Initial Structures

The coordinates of the chimeric SARS-CoV-2 RBD complexed with the hACE2⁽¹⁷⁾ were obtained from the Protein Data Bank (PDB ID 6VW1). This structure is a chimeric RBD, which uses a core from the SARS-CoV RBD and the RBD from SARS-CoV-2. The SARS-CoV RBD complexed with hACE2^(17,20) was also obtained from the Protein Data Bank (PDB ID 2AJF). From these PDB files, the coordinates of the hACE2, SARS-CoV RBD and SARS-CoV-2 RBD were separated, the missing atoms were manually corrected, obtaining files which were processed using AutoDockTools 4.2⁽²¹⁾ and GROMACS 2018⁽²⁵⁾ for further use in docking and molecular dynamics.

Molecular docking

The SARS-CoV and the SARS-CoV-2 RBDs were initially modeled as rigid ligands in their docking towards the hACE2 also modeled as a rigid receptor using a specially compiled version of Autodock Suite 4.2⁽²¹⁾ to support larger ligands. Autodock was chosen among other alternatives due to both the thin manual control available for the potential grid settings and for its use of a “semiempirical free energy force field to predict the binding free energies”^(21,22) as scoring function. In both cases grid boxes were generated from manually edited input parameter files to have 510 points in each Cartesian direction, spaced by 0.144 Å and 0.132 Å for SARS-CoV and SARS-CoV-2 respectively. These grid boxes had different grid spacings due to the different ligand sizes and relative receptor orientation on each complex, and were centered in the region of the hACE2 where the interactions are known to occur, according to experimental results^(11-13,16-20). Lamarckian genetic algorithms were employed with a population size of 150 individuals evolving through 27,000 generations (where only the one best individual surviving to the next generation), with 0.02 and 0.8 mutation and crossover rates respectively and a maximum of 2.5 million energy evaluations. Four independent molecular docking runs were performed for each ligand-receptor complex, each one with different random-seeds. The best 100 conformations were clustered by structural similarity and ranked relative to its (free) energy of binding in bins of 0.5 kcal/mol, and the most negative energy conformation analyzed in terms of the most important interactions, with special focus on the hydrogen bonds formed between ligand and receptor.

Molecular dynamics

The structures of hACE2, SARS-CoV RBD, SARS-CoV-2 RBD and the two hACE2-RBD complexes were solvated in TIP3P water⁽²⁶⁾ in cubic boxes under periodic boundary conditions and minimum image convention, with van der Waals interactions cut off at 1 nm and electrostatic interactions modeled with the Generalized Reaction Field⁽²⁷⁾ method. AMBER 03⁽²⁴⁾ force field was employed in MM optimization and MD stages of the modeling. The initial systems were energetically optimized with the Steepest-Descent algorithm through parameters using GROMACS 2018⁽²⁵⁾ simulation software package. Afterwards, the boxes underwent successive phases of molecular dynamics (MD) simulations employing the Leap-Frog algorithm with a 2 fs time-step. Their initial equilibration consisted of 1 ns of position restrained NVT simulation using the V-Rescale⁽²⁸⁾ thermostat with randomly generated initial velocities, followed by 1 ns of position restrained proteins NpT simulation using the Parrinello-Rahman⁽²⁹⁾ barostat and the V-Rescale thermostat and another 1 ns of fully flexible NpT simulation performed using the same algorithms as the previous step. The following acquisition stage consisted of 3 successive 100 ns runs recording the configuration each 1 ps. The obtained trajectories were analyzed to identify and account for the presence of the intermolecular interactions in the complexes through time. The proteins Root Mean Square Deviations (RMSDs, provided in the supplementary materials) were monitored to ensure the system was successfully equilibrated. Energies were stored at each picosecond for LIE⁽²³⁾ binding free energy calculations, as described below.

Free energy of binding

Free energy of binding was calculated using two different approaches: first, from the molecular docking⁽²¹⁾ calculations, based on previously published X-ray crystallography^(17,20) structures. Secondly, employing the LIE⁽²³⁾ technique with some of the latest parameters published⁽³⁰⁾.

The LIE approach⁽²³⁾ estimates the ligand binding affinity (measured as free energy) to a given receptor as a weighted average considering the differences on the van der Waals and electrostatic energies between the ligand bound to the receptor and the free ligand, as described by equation (1) below:

$$\Delta G_{bind} = \alpha \left(\langle V_{l-s}^{vdW} \rangle_{bound} - \langle V_{l-s}^{vdW} \rangle_{free} \right) + \beta \left(\langle V_{l-s}^{elect} \rangle_{bound} - \langle V_{l-s}^{elect} \rangle_{free} \right) \quad (1)$$

In the context of molecular dynamics, the averages in the equation (1) above are calculated considering all interactions (van der Waals or electrostatic) involving the ligand with its surroundings, in a simulation of the ligand bound to the receptor (complex), minus the same interaction in a simulation of the free ligand (without receptor). We used the parameters $\alpha = 0.448$ and $\beta = 0.090$, as proposed by Vosmeer *et al*⁽³⁰⁾.

Supplementary Information

Fig. S1. Root Mean Square Deviations (RMSDs) of each systems proteins from the initial structure for 300ns of MD simulation.

Fig. S2. Full mapping representation of the SARS-CoV RBD (top), hACE2 (middle) and SARS-CoV-2 RBD (below) intermolecular hydrogen bonding interactions networks NET1 (in blue), NET2 (in green), NET3 (in dark red) and involving the hACE2 enzyme residues GLU329 (in black), ASP30 (in orange), both LEU351 and GLY352 (in light cyan), GLU37 (in pink), both ASP38 and GLN42 (in dark magenta), as well as the ephemeral interconnection with NET3 (dashed yellow line) from MD.

Fig. S3. SARS-CoV RBD (in orange) lowest free energy of binding configuration docked on hACE2 (in light blue), showing the intermolecular hydrogen bonded residue pairs. Depicted are: (from hACE2 α -helix 1: TYR41–THR'486; from α -helix 12: GLU329–ARG'426; and from the coil between the two β -sheets: LYS353–GLY'488). Green regions are hACE2 residues with important hydrophobic interactions with the SARS-CoV RBD chain (from hACE2 α -helix 1: GLN24, THR27, PHE28, LYS31, HIS34, GLY35, ASP38, GLN42 and LEU45; from α -helix 2: MET82; from the coil just out of the α -helix 2: TYR83; from α -helix 12: ASN330; from the coil between the two β -sheets: GLY354; and from the second β -sheet: ASP355 and ARG357) revealed by docking. Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Fig. S4. SARS-CoV-2 RBD (in orange) lowest free energy of binding docked on hACE2 (in light blue), with the intermolecular hydrogen bonded residue pairs. Depicted are: (from the coil just before the hACE2 α -helix 1: SER19–ALA'475; on the α -helix 1: LYS31–GLN'493, GLU35–GLN'493, ASP38–TYR'449, TYR41–THR'500 and GLN42–GLN'498; from the coil just out of α -helix 2: TYR83–ASN'487; and from the coil between the two β -sheets: LYS353–GLY'496 and LYS353–GLY'502). Green regions are hACE2 residues with important hydrophobic interactions with the SARS-CoV RBD chain (GLN24, THR27, PHE28 and LEU45, all from α -helix 1) revealed by docking. Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Fig. S5. SARS-CoV RBD to hACE2 hydrogen bonds presence profiles for hydrogen bond network NET1 (from top: SER19–PRO'462, SER19–ASP'463, GLN24–ASP'463 and THR27–PRO'462). Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Fig. S6. SARS-CoV RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting the hydrogen bond network NET1 (SER19–PRO'462, SER19–ASP'463, GLN24–ASP'463 and THR27–PRO'462). Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Fig. S7. SARS-CoV RBD to hACE2 hydrogen bonds presence profiles for hydrogen bond network NET2 (from top: HIS34–TYR'440 and HIS34–ASN'479). Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Fig. S8. SARS-CoV RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting the hydrogen bond network NET2 (HIS34–TYR'440 and HIS34–ASN'479). Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Fig. S9. SARS-CoV RBD to hACE2 hydrogen bonds presence profiles for hydrogen bond network NET3 (from top: TYR41–THR'433, TYR41–GLY'434, LYS353–TYR'481, LYS353–GLY'482, LYS353–GLY'488 and ASP355–THR'486). Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Fig. S10. SARS-CoV RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting the hydrogen bond network NET3 (TYR41–THR'433, TYR41–GLY'434, LYS353–TYR'481, LYS353–GLY'482, LYS353–GLY'488 and ASP355–THR'486). GLY'488 is not visible in the figure, and ASN'501 view is heavily obstructed. Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Fig. S11. SARS-CoV RBD to hACE2 hydrogen bonds presence profiles for hACE2 enzyme complex with SARS-CoV RBD GLU329–ARG'439 (in black), ASP30–TYR'442 (in orange) and both LEU351–TYR'484 and GLY352–TYR'484 (in light cyan) pairs of residues. Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.

Fig. S12. SARS-CoV RBD (in orange) complexed with hACE2 (in light blue) snapshot depicting hACE2 enzyme complex hydrogen bonds with SARS-CoV RBD of the GLU329–ARG'426, ASP30–TYR'442, LEU351–TYR'484 and GLY352–TYR'484 pairs of residues. Carbon atoms are grey, hydrogens are white, oxygens are red, nitrogens are blue and hydrogen-bonds are represented in yellow dashes.