bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

In silico study of the spike protein from SARS-CoV-2 interaction with ACE2: similarity with SARS-CoV, hot-spot analysis and effect of the receptor polymorphism

Houcemeddine Othman[1], Zied Bouslama [2], Jean-Tristan Brandenburg [1], Jorge da Rocha[1], Yosr Hamdi[3] Kais Ghedira[4] Najet-Srairi Abid[5] Scott Hazelhurst[1,6]

[1] Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.
[2] Laboratory of veterinary epidemiology and microbiology LR16IPT03. Institut Pasteur of Tunis. University of Tunis El Manar, Tunis, Tunisia.
[3] Laboratory of Biomedical Genomics and Oncogenetics, LR16IPT05, Pasteur Institute of Tunis, University of Tunis El Manar, Tunis, Tunisia.
[4] Laboratory of Bioinformatics, Biomathematics and Biostatistics - LR16IPT09, Pasteur Institute of Tunis, University Tunis El Manar, Tunis, Tunisia.
[5] Université de Tunis El Manar, Institut Pasteur de Tunis, LR11IPT08 Venins et Biomolécules Thérapeutiques, 1002, Tunis, Tunisia.
[6] School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa.

* houcemoo@gmail.com

Abstract

The spread of the COVID-19 caused by the SARS-CoV-2 outbreak has been growing since its first identification in December 2019. The publishing of the first SARS-CoV-2 genome made a valuable source of data to study the details about its phylogeny, evolution, and interaction with the host. Protein-protein binding assays have confirmed that Angiotensin-converting enzyme 2 (ACE2) is more likely to be the cell receptor via which the virus invades the host cell. In the present work, we provide an insight into the interaction of the viral spike Receptor Binding Domain (RBD) from different coronavirus isolates with host ACE2 protein. We used homology-based protein-protein

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

docking, binding energy estimation, and decomposition, aiming to predict both qualitative and quantitative aspects of the interaction. Using *in silico* structural modelling, we brought additional evidence that the interface segment of the spike protein RBD might be acquired by SARS-CoV-2 via a complex evolutionary process rather than mutation accumulation. We also highlighted the relevance of Q493 and P499 amino acid residues of SARS-CoV-2 RBD for binding to hACE2 and maintaining the stability of the interface. Finally, we studied the impact of eight different variants located at the interaction surface of ACE2, on the complex formation with SARS-CoV-2 RBD. We found that none of them is likely to disrupt the interaction with the viral RBD of SARS-CoV-2.

key words: COVID-19, ACE2, viral spike Receptor Binding Domain, homology-based protein-protein docking, variants.

Introduction

The coronavirus SARS-CoV-2 (previously known as nCoV-19) has been associated with the recent epidemic of acute respiratory distress syndrome [9], [2]. Recent studies have suggested that the virus binds to the ACE2 receptor on the surface of the host cell using the spike protein and explored the binary interaction of these two partners [4, 12, 28]. In this work, we focused our analysis on the interface residues to get insight into four main subjects: (1) The architecture of the interface of the spike protein and whether its evolution in many isolates supports an increase in affinity toward the ACE2 receptor; (2) How the affinity of SARS-COV-2-RBD and SARS-CoV-RBD toward different ACE2 homologous proteins from different species is dictated by a divergent interface sequences (3); A comparison of the interaction hotspots between SARS-CoV and SARS-CoV-2; and finally, (4) whether any of the studied ACE2 variants may show a different binding property compared to the reference allele. To tackle these questions we used multiscale modelling approaches combined to sequence and phylogeny analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Materials and Methods

Sequences and data retrieval

Full genome sequences of 10 Coronaviruses isolates were retrieved from NCBI GeneBank corresponding to the following accessions: AY485277.1 (SARS coronavirus Sino1-11), FJ882957.1 (SARS coronavirus MA15), MG772933.1 (Bat SARS-like coronavirus isolate bat-SL-CoVZC45), MG772934.1 (Bat SARS-like coronavirus isolate bat-SL-CoVZXC21), DQ412043.1 (Bat SARS coronavirus Rm1), AY304488.1 (SARS coronavirus SZ16), AY395003.1 (SARS coronavirus ZS-C), KT444582.1 (SARS-like coronavirus WIV16), MN996532.1 (Bat coronavirus RaTG13) in addition to Wuhan seafood market pneumonia virus commonly known as SARS-CoV-2 (accession MN908947.3).

The sequences of the surface glycoprotein were extracted from the Coding Segment (CDS) translation feature from each genome annotation or by locally aligning the protein from SARS-CoV-2 with all possible ORFs from the translated genomes. ACE2 orthologous sequences from Human (Uniprot sequence Q9BYF1), Masked palm civet (NCBI protein AAX63775.1 from *Paguma larvata*), Chinese rufous horseshoe bat (NCBI protein AGZ48803.1 from *Rhinolophus sinicus*), King cobra snake (NCBI protein ETE61880.1 from *Ophiophagus hannah*), chicken (NCBI protein XP_416822.2, *Gallus gallus*), domestic dog (NCBI protein XP_005641049.1, *Canis lupus familiaris*), Wild boar (NCBI protein XP_020935033.1, *Sus scrofa*) and Brown rat (NCBI protein NP_001012006.1 *Rattus norvegicus*) were also computed and retrieved.

Human variants of the ACE2 gene were collected from the gnomAD database [7]. Only variants that map to the protein coding region and belonging to the interface of interaction with the RBD of the spike protein were retained for further analyses.

Sequence analysis and phylogeny tree calculation

MAFFT was used to align the whole genome sequences and the protein sequences of viral RBDs [8]. For the genome comparison, we selected the best site model based on lowest Bayesian Information Criterion calculated

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

using model selection tool implemented in MEGA 6 software [21]. The General Time Reversible (GTR) model was chosen as the best fitting model for nucleotide substitution with discrete Gamma distribution (+G) with 5 rate categories. For the RBD sequences, the best substitution model for ML calculation was selected using a model selection tool implemented on MEGA 6 software based on the lowest Bayesian Information criterion (BIC) score. Therefore, the WAG model [24] using a discrete Gamma distribution (+G) with 5 rate categories has been selected.

Phylogenetic trees were generated using a maximum likelihood (ML) method in MEGA 6. The consistency of the topology, for the RBD sequences, was assessed using a bootstrap method with 1000 replicates. The resulting phylogenetic tree was edited with iTOL [13].

Homology based protein-protein docking and binding energy estimation

The co-crystal structure of the spike protein of SARS coronavirus complexed to human human-civet chimeric receptor ACE2 was solved at 3 Åof resolution (PDB code 3SCL). We used this structure as a template to build the complex of spike protein from different virus isolates with the human ACE2 protein (Uniprot sequence Q9BYF1). The template sequences of the ligand (spike protein) and the receptor (ACE2) were aligned locally with the target sequences using the program Water from the EMBOSS package [16]. Modeller version 9.22 [18] was then used to predict the complex model of each spike protein with the ACE2 using a slow refining protocol. For each model, we generated ten conformers from which we selected the model with the best DOPE score [19].

To calculate the binding energy we used the PRODIGY server [27]. The Calculation of contribution of each amino acid in a protein partner was computed with MM-GBSA method implemented in the HawkDock server [23]. Different 3D structures of hACE2, each comprising one of identified variants, were modeled using the BuildModel module of FoldX5 [3]. Because it is more adapted to predict the effect of punctual variations of amino acids, we used DynaMut at this stage of analysis [17].

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Flexibility analysis

We ran a protocol to simulate the spike RBD fluctuation of SARS-CoV-2 and SARS-Cov using the standalone program CABS-flex (version 0.9.14) [11]. Three replicates of the simulation with different seeds were conducted using a temperature value of 1.4 (dimensionless value related to the physical temperature). The protein backbone was kept fully flexible and the number of the Monte carlo cycles was set to 100.

Results

Sequence and phylogeny analysis

Phylogenetic analysis of the different RBD sequences revealed two well supported clades. Clade 2 includes SARS-CoV-2, RatG13, SZ16, ZS-C, WIV16, MA15, and SARS-CoV-Sino1-11 isolates (Figure 1A). SARS-CoV-2 and RatG13 sequences are the closest to the common ancestor of this clade. Clade 1 includes Rm1 isolate, Bat-SL-CoVZC45 and Bat-SL-CoVZXC21. These three isolates are closely related to SARS-CoV-2 as revealed by the phylogenetic tree constructed from the entire genome (Figure 1A). The exact tree topology is reproduced when we used only the RBD segment corresponding to the interface residues with hACE2. This is a linear sequence spanning from residue N481 to N501 in SARS-CoV-2.

Multiple sequence alignment showed that the interface segment of SARS-CoV-2 shares higher similarity to sequences from clade 2 (Figure 1B). However, we noticed that S494, Q498 and P499 are exclusively similar to their equivalent amino acids in sequences from clade 1. SARS-CoV-2 interface sequence is closely related to Bat-CovRaTG13 sequence, isolated from a *Rhinolophus affinis* bat.

Prediction of the RBD/hACE2 complex structure

To investigate whether the interface of the spike protein isolates evolves by increasing the affinity toward the ACE2 receptor in the final host, we predicted the interaction models of the envelope anchored spike protein (SP) from several clinically relevant Coronavirus isolates with the human receptor

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

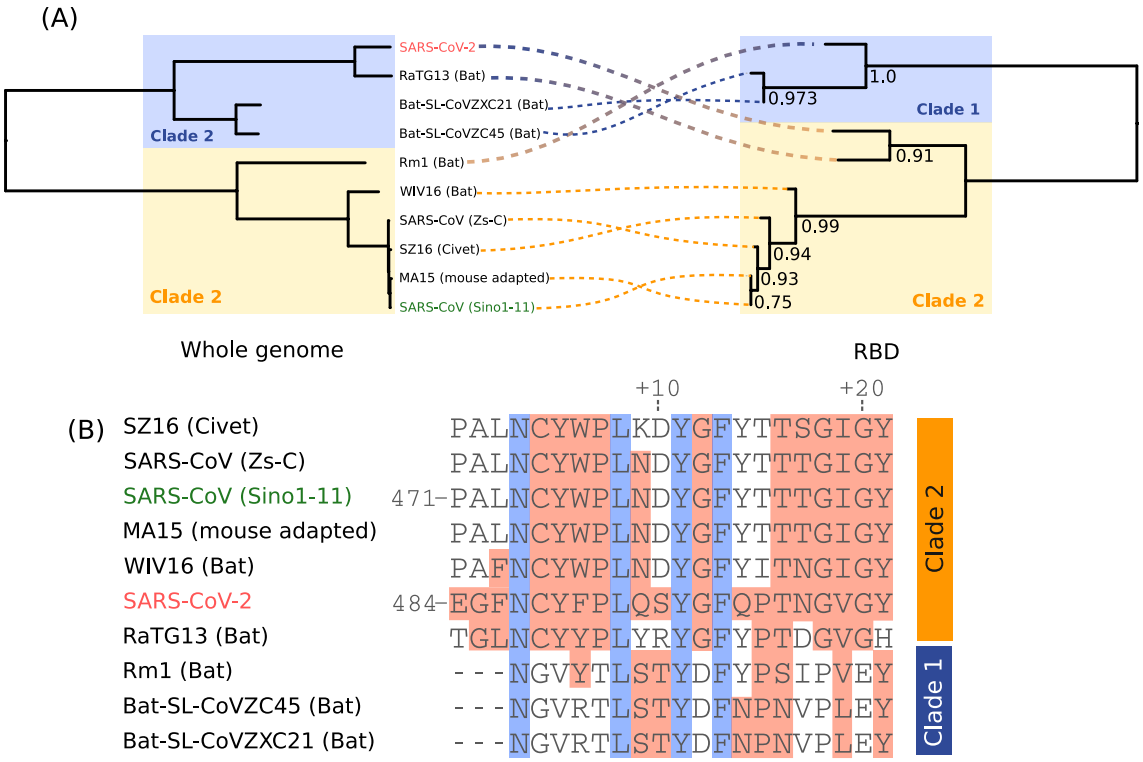

Figure 1. Phylogeny and sequence analysis based on full genome and RBDs from the different isolates included in this study. (A) Phylogeny trees are opposed to each other to show the clade discrepancies and discontinuous lines shows the equivalent taxon between each tree. (B) Multiple sequence alignment of the interface residues of RBD. Blocks in red indicate the residues with similar biochemical properties to the positions in SARS-CoV-2. Conserved residues are colored in blue.

ACE2 (hACE2). The construction of the complex applies a comparative-based approach that uses a template structure in which both partners (ligand and receptor) are closely related to those in the target system respectively. In our study, we only modeled the interaction of the RBD which was shown to be implicated in the physical interaction with ACE2 receptor (Figure 2A). The lowest sequence identity of the modeled spike proteins as well as those of any of the orthologous ACE2 sequences (Human, civet, bat, pig, rat, chicken and snake) do not fall below 63% compared to their respective templates. At such values of sequence identities between the equivalent partners of the receptor or the ligand, it is expected that the template and the target complexes share the same binding mode [1, 10].

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Analysis of energy Interaction of hACE2 with other virus

We calculated the binding energy of RBD from different virus isolates interacting with hACE2 using the PRODIGY method (Figure 2b). The binding energies are converted to dissociation constant estimations (Kd). SP proteins from bat-SL-CoVZC45, bat-SL-CoVZXC21 and Rm1-Cov show the highest values (least favorable) which are all above 50 nM. All the other estimations fall below 18 nM. The interaction between hACE2 and the RBDs of SARS-CoV-2 isolate (Whuhan-Hu-1) and the SARS-CoV-Sino1-11 show Kd values of 5.1 and 18 nM, respectively.

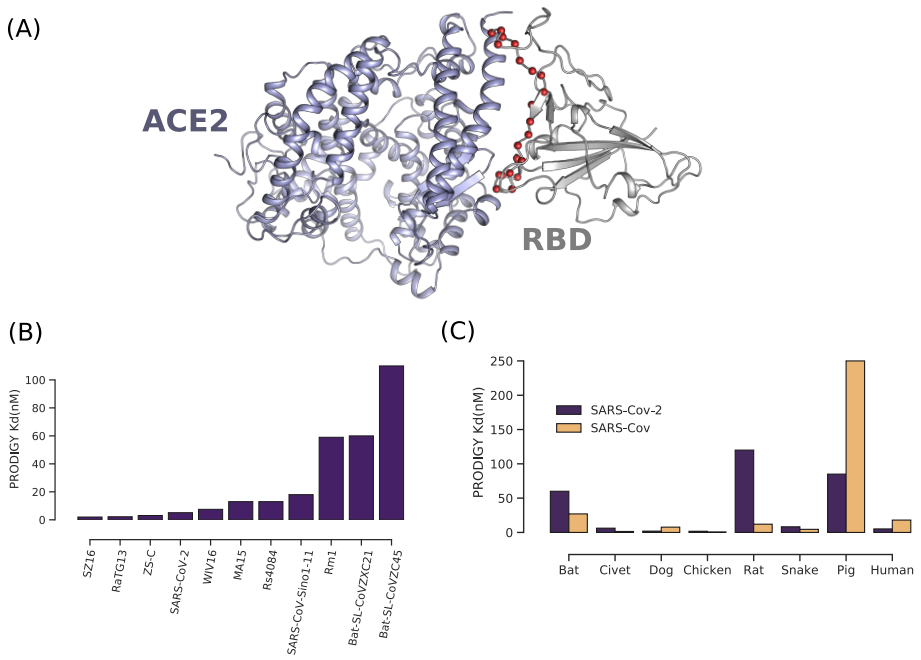

Figure 2. Homology based protein-protein docking of RBD/ACE2 and binding energy analysis of spike RBD with ACE2 receptor. (A) Homology based protein-protein docking complex of SARS-CoV-2 RBD with hACE2. The red spheres are the interface residues of the RBD. (B) Binding energies were converted to Kd values calculated for spike RBD from different coronavirus isolates interacting with hACE2. (C) Dissociation constant estimation calculated between SARS-CoV-2 RBD and homologous ACE2 from different model animals.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

0.1 Energy analyses of human SARS-CoV-2 and SARS-CoV with different animal ACE2 receptors

We made this analysis to investigate the tendency of SARS-CoV-2 and SARS-CoV to interact with different orthologous forms of ACE2 which is dictated by the divergence in their interacting surfaces. Homology based protein-protein docking was conducted to generate the interaction model of SARS-CoV-2 RBD with ACE2 receptor from different animal species (Figure 2C). We noticed that for the civet, dog, chicken and snake forms, the interaction energy is very low and very similar either for SARS-CoV-2 or SARS-CoV. Although the Kd are relatively low for the rat and bat forms interacting with SARS-CoV RBD, those of SARS-CoV-2 are high and go beyond 50 nM. On the other hand, it seems that the interaction with pig ACE2 is more favorable for SARS-CoV-2 isolate since the estimated Kd is 3 folds lesser compared to SARS-CoV-2.

Decomposition of the interaction energy

The MM-GBSA calculation allowed us to assign the contribution in the binding energy of each amino acid in the interface with hACE2. We carried this analysis using both sequences of the SARS-CoV-2 Wuhan-Hu-1 (Figure 3A) and the Sino1-11 SARS-CoV (Figure 3B) isolates. Residues F486, Y489, Q493, G496, T500 and N501 of SARS-CoV-2 RBD form the hotspots of the interface with hACE2 protein were investigated (we only consider values > 1 or < 1 kcal/mol to ignore the effect due to the thermal fluctuation). All these amino acids form three patches of interaction spread along the linear interface segment (Figure 3C): two from the N and C termini and one central. T500 establishes two hydrogen bonds using its side and main chains with Y41 and N330 of hACE2. N501 forms another hydrogen bond with ACE2 residue K353 buried within the interface. On the other hand, SARS-CoV RBD interface contains five residues (Figure 3D), L473, Y476, Y485, T487 and T488 corresponding to the equivalent hotspot residues of RBD from SARS-CoV-2 F487, Y490, G497, T501 and N502. Therefore, Q493 as a hotspot amino acid is specific to SARS-CoV-2 interface. The equivalent residue N480 in SARS-CoV only shows a non-significant contribution of 0.18 kcal/mol.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Flexibility analysis

Sequence analysis and the visual inspection of RBD/hACE2 complex might reflect the substitution of P499 in SARS-CoV-2 RBD as a form of adaptation toward a better affinity with the receptor. In order to further investigate its role, we performed a flexibility analysis using a reference structure (SARS-CoV-2 RBD containing P499) and an *in silico* mutated form P499T, a residue found in SARS-CoV and most of the clade 2. Our results show that the mutation caused a significant decrease in stability for nine residues of the interface corresponding to segment 482-491 (Figure 3E). Indeed, the RMSF variability per amino acid for this sequence increases compared to the reference structure.

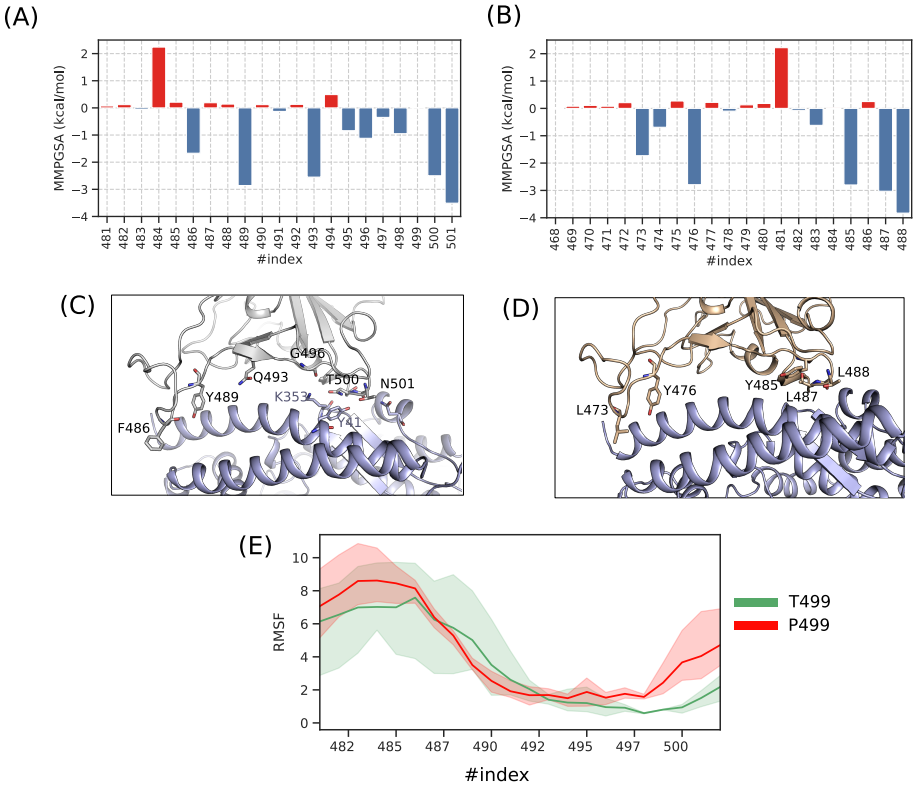

Figure 3. Analysis of the interaction between RBD and hACE2. Decomposition of the MMGBSA energy for each amino acid of the binding surface from SARS-CoV-2 (A) and SARS-CoV Sino1-11 isolate (B). Position of the hotspot residues of the complexes RBD-SARS-CoV-2/hACE2 (C) and RBD-SARS-CoV/hACE2 (D). (E) Flexibility of RBD interface residue expressed as Root Mean Square Fluctuation (RMSF) for two forms of RBD-SARS-CoV-2, T499 and P499.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Analysis of ACE2 variability and affinity with the virus

A total of eight variants of hACE2 that map to the interaction surface are described in the gnomAD database. All these variants are rare (Table 1) and mostly found in European non-Finnish and African populations. Considering both the enthalpy (ddG) and the vibrational entropy in our calculation (ddS), we found no significant change in the interaction energy (> 1 or < 1 kcal/mol) (Figure 4).

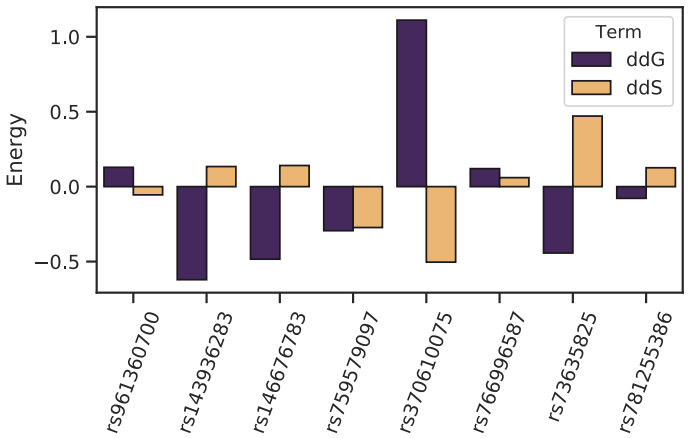

Figure 4. Estimation of the energy changing upon mutation for hACE2 variants. Both Enthalpy and entropy terms are computed.

Discussion

Since the Covid-2019 outbreak, several milestone papers have been published to examine the particularity of SARS-CoV-2 spike protein and its putative interaction with ACE2 as a receptor [25, 26]. In the current study, we focused our analysis on the interface segments of SARS-CoV-2 spike RBD interacting with ACE2 from different species by estimating interaction energy profiles.

We have studied the effect of eight variants of ACE2 in order to detect polymorphisms that may increase or decrease the virulence in the host. We concluded that if ACE2 is the only route for the infection in humans, variants interacting physically with RBD are not likely to disrupt the formation of the complex and would have a marginal effect on the affinity. Therefore, it

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

is unlikely that any form of resistance to the virus, related to the ACE2 gene, exists. However, this analysis merits to be investigated in depth in different ethnic groups for a better assessment of the contribution of genetic variability in host-pathogen interaction. The binding of SARS-CoV-2 RBD to different forms of ACE2 shows similar values compared to SARS-CoV which is in agreement with the ability of the virus to cross the species barrier. However, although we estimated low Kd value between SARS-CoV RBD and Gallus gallus ACE2, there have been no reported cases of SARS-CoV isolated from chicken. On the other hand, while Kd is high for SARS-CoV RBD interacting with porcine ACE2, there are only few reported cases of such type [20]. Indeed, host-pathogen interaction is a complex process unlikely to be controlled only by the binding of the spike protein with ACE2 [5]. For model animals where Kd values are very low, ACE2 analysis may play a key role in targeting the main animal reservoir of SARS-CoV-2. For high Kd values, other factors might regulate the infection including the implication of different receptors or the response of the immune system but that does not mean that the infection is unlikely to occur. On the other hand, whole-genome phylogenetic analysis of the different isolates included in this study is consistent with previous works that place the Wuhan-Hu-1 isolate close to Bat-SL-CoVZC45 and Bat-SL-CoVZXC21 isolates [14,15,22] within the Betacoronavirus genus. The use of RBD sequences, however, places the virus in a clade that comprises SARS-CoV related homologs including isolates from Bat and Civet. The clade swapping as seen in figure 1A, seems also to occur for RaTG13 and Rm1 isolated from bat. This is expected as the use of different phylogenetic markers may considerably affect the topology of the tree. However, given the functional implication of the spike RBD in host-pathogen infection, we have raised the question about where the virus obtained its RBD binding interface. The binding of the spike glycoprotein to ACE2 receptor requires a certain level of affinity. In the case where the RBD evolves from an ancestral form closer to that of Bat-SL-CoVZC45 and Bat-SL-CoVZXC21, we expected a decrease of the binding energy through the evolution process following incremental changes in the RBD. In such a scenario, we presume that there are other intermediary forms of coronavirus that describe such variation of the binding energy to reach a level where the pathogen can cross the species barriers and infect

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

humans with high affinity toward hACE2. On the other hand, our results show that the binding energy and the interface sequence of SARS-CoV-2 RBD are closer to SARS-CoV related isolates (either from Human or other species). Therefore a recombination event involving the spike protein that might have occurred between SARS-CoV and an ancestral form of the current SARS-CoV-2 virus might be also possible. This will allow for the virus to acquire a minimum set of residues for the interaction with hACE2. The recombination in the spike protein gene has been previously suggested by Wei et al in their phylogenetic analysis [6]. Thereafter, incremental changes in the binding interface segment will occur in order to reach a better affinity toward the receptor. One of these changes may involve P499 residue which substitution to threonine seems to drastically destabilize the interface segment and has a distant effect. Moreover, the decomposition of the interaction energy showed that 5 out of 6 hotspot amino acids in SARS-CoV-2 have their equivalent in SARS-CoV including N501. Contrary to what Wan et al [22] have stated, the single mutation N501T does not seem to enhance the affinity. Rather, the residue Q493 might be responsible for such higher affinity due to a better satisfaction of the Van der Waals by the longer polar side chain of asparagine. Indeed, when we made the same analysis while mutating Q493 to N493, the favorable contribution decreases from -2.55 kcal/mol to a non significant value of -0.01 kcal/mol, thus supporting our claim.

References

1. P. Aloy, H. Ceulemans, A. Stark, and R. B. Russell. The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.*, 332(5):989–998, Oct 2003.

2. J. F. Chan, S. Yuan, K. H. Kok, K. K. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C. Yip, R. W. Poon, H. W. Tsoi, S. K. Lo, K. H. Chan, V. K. Poon, W. M. Chan, J. D. Ip, J. P. Cai, V. C. Cheng, H. Chen, C. K. Hui, and K. Y. Yuen. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*, 395(10223):514–523, Feb 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

3. J. Delgado, L. G. Radusky, D. Cianferoni, and L. Serrano. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics*, 35(20):4168–4169, Oct 2019.

4. J. He, H. Tao, Y. Yan, S.-Y. Huang, and Y. Xiao. Molecular mechanism of evolution and human infection with the novel coronavirus (2019-ncov). *bioRxiv*, 2020.

5. D. Janies, F. Habib, B. Alexandrov, A. Hill, and D. Pol. Evolution of genomes, host shifts and the geographic spread of sars-cov and related coronaviruses. *Cladistics*, 24(2):111–130, 2008.

6. W. Ji, W. Wang, X. Zhao, J. Zai, and X. Li. Cross-species transmission of the newly identified coronavirus 2019-nCoV. *J. Med. Virol.*, 92(4):433–440, Apr 2020.

7. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski, , B. M. Neale, M. J. Daly, and D. G. MacArthur. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 2019.

8. K. Katoh, J. Rozewicki, and K. D. Yamada. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinformatics*, 20(4):1160–1166, 07 2019.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

9. P. Kaytes, L. Wood, N. Theriault, M. Kurkinen, and G. Vogeli. Head-to-head arrangement of murine type IV collagen genes. *J. Biol. Chem.*, 263(36):19274–19277, Dec 1988.

10. P. J. Kundrotas, Z. Zhu, J. Janin, and I. A. Vakser. Templates are available to model nearly all complexes of structurally characterized proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 109(24):9438–9441, Jun 2012.

11. M. Kurcinski, T. Oleniecki, M. P. Ciemny, A. Kuriata, A. Kolinski, and S. Kmiecik. CABS-flex standalone: a simulation environment for fast modeling of protein flexibility. *Bioinformatics*, 35(4):694–695, 02 2019.

12. J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, and X. Wang. Crystal structure of the 2019-ncov spike receptor-binding domain bound with the ace2 receptor. *bioRxiv*, 2020.

13. I. Letunic and P. Bork. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.*, 47(W1):W256–W259, Jul 2019.

14. R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi, and W. Tan. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*, 395(10224):565–574, Feb 2020.

15. D. Paraskevis, E. G. Kostaki, G. Magiorkinis, G. Panayiotakopoulos, G. Sourvinos, and S. Tsiodras. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infect. Genet. Evol.*, 79:104212, Apr 2020.

16. P. Rice, I. Longden, and A. Bleasby. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, 16(6):276–277, Jun 2000.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

17. C. H. Rodrigues, D. E. Pires, and D. B. Ascher. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.*, 46(W1):W350–W355, 07 2018.

18. A. Sali and T. L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, 234(3):779–815, Dec 1993.

19. M. Y. Shen and A. Sali. Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, 15(11):2507–2524, Nov 2006.

20. Z. Shi and Z. Hu. A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res.*, 133(1):74–87, Apr 2008.

21. K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.*, 30(12):2725–2729, Dec 2013.

22. Y. Wan, J. Shang, R. Graham, R. S. Baric, and F. Li. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J. Virol.*, Jan 2020.

23. G. Weng, E. Wang, Z. Wang, H. Liu, F. Zhu, D. Li, and T. Hou. HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res.*, 47(W1):W322–W330, Jul 2019.

24. S. Whelan and N. Goldman. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution*, 18(5):691–699, 05 2001.

25. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, Feb 2020.

26. X. Xu, P. Chen, J. Wang, J. Feng, H. Zhou, X. Li, W. Zhong, and P. Hao. Evolution of the novel coronavirus from the ongoing

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci*, 63(3):457–460, Mar 2020.

27. L. C. Xue, J. P. Rodrigues, P. L. Kastritis, A. M. Bonvin, and A. Vangone. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics*, 32(23):3676–3678, 12 2016.

28. R. Yan, Y. Zhang, Y. Guo, L. Xia, and Q. Zhou. Structural basis for the recognition of the 2019-ncov by human ace2. *bioRxiv*, 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.04.976027; this version posted March 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Population frequencies of hACE2 missense variants located on the interaction surface with SARS-CoV-2 RBD ($\times 10^{-5}$)

rs ID	European (non-Finnish)	African	Latino	Ashkenazi Jewish	East Asian	South Asian	Finnish	Other	Global
rs961360700	2.59	0	0	0	0	0	0		1.17
rs143936283	6.51	0	0	0	0	0	0	19.05	3.443
rs146676783	0	0.105	0	0	0	0	32.22	0	3.897
rs759579097	0	0.1056	0	0	0	0	0	0	0.9847
rs370610075	1.274	0	0	0	0	0	0	0	0.5755
rs766996587	0	26.23	0	0	0	0	0	0	2.442
rs73635825	0	332.3	0	0	0	0	0	18.82	31.29
rs781255386	0	0	7.303	0	0	0	0	0	1.091