Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Interaction of the spike protein RBD from SARS-CoV-2 with ACE2: Similarity with SARS-CoV, hot-spot analysis and effect of the receptor polymorphism

Houcemeddine Othman, Zied Bouslama, Jean-Tristan Brandenbourg, Jorge da Rocha, Yosr Hamdi, Kais Ghedira, Najet Srairi-Abide, Scott Hazelhurst

1. Introduction

The coronavirus SARS-CoV-2 (previously known as nCoV-19) has been associated with the recent epidemic of acute respiratory distress syndrome [2]. Recent studies have suggested that the virus binds to the ACE2 receptor on the surface of the host cell using spike proteins, and explored the binary interaction of these two partners [8,23]. In this work, we focused our analysis on the interface residues to get insight into four main subjects: (1) The architecture of the spike protein interface 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 multi-scale modelling approaches in combination with sequence and phylogenetic analysis.
2. Materials and methods

2.1. Sequences and data retrieval

Full genome sequences of 10 Coronavirus isolates were retrieved from NCBI Genbank corresponding to the following accession numbers: SARS coronavirus (Accession Number: K04763), SARS-like coronavirus (Accession Number: AY504128), SARS-like coronavirus (Accession Number: JX095129), SARS-like coronavirus (Accession Number: AY580288), SARS-like coronavirus (Accession Number: AY294780), SARS-like coronavirus (Accession Number: JX833183), SARS-like coronavirus (Accession Number: AY791350), SARS-like coronavirus (Accession Number: AY294780), and SARS-like coronavirus (Accession Number: AY791350). The sequences of the surface glycoprotein were extracted from the GenBank database.

2.2. Sequence analysis and phylogenetic tree calculation

MAFFT 7.650 was used to align the whole genome sequences and the protein sequences of viral RBDS [5] (Supplementary Materials 1). Prediction of the N-Glycosylation sites was made for all sequences using the NetNGlyc server (https://www.cbs.dtu.dk/services/NetNGlyc/). For the genome comparison, we used the Clustal Omega algorithm implemented in MEGA 6 software [16]. The General Time Reversible (GTR) model with 5 rate categories has been selected. Phylogenetic analysis of the different RBD sequences revealed two well supported clades. Clade 1 includes isolate N481 to N501 in SARS-CoV-2, RatG13, SARS-CoV-Sino1-11 isolates (Fig. 1A). Clade 2 includes SARS-CoV-2, RatG13, SARS-CoV-Sino1-11 isolates (Fig. 1A). SARS-CoV-2 and RatG13 sequences are the closest to the common ancestor of this clade. 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.

3. Results

3.1. Sequence and phylogenetic analysis

Phylogenetic analysis of the different RBD sequences revealed two well supported clades. Clade 1 includes isolate N481 to N501 in SARS-CoV-2, RatG13, SARS-CoV-Sino1-11 isolates (Fig. 1A). Clade 2 includes SARS-CoV-2, RatG13, SARS-CoV-Sino1-11 isolates (Fig. 1A). SARS-CoV-2 and RatG13 sequences are the closest to the common ancestor of this clade. 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.

3.2. Prediction of the RBD/hACE2 complex structure

To investigate whether the interface of the spike protein isolate 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 hACE2 receptor (PDB files for the complexes are listed in Supplementary Materials 1). 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 (Fig. 2A). The lowest sequence identity of the modeled spike proteins as well as those of any of the orthologous ACE2 sequences (Human, pig, rat, chicken and snake) do not fall below 63% toward their respective templates. At such values of sequence identities it is expected that the template and the target complexes share the same binding mode [6].
3.3. Analysis of the interaction energy scores of hACE2 with other virus isolates

We calculated the binding energy scores of the RBD from different virus isolates interacting with hACE2 (Fig. 2b). All three methods used for the calculation are in agreement that RBDs from bat-SL-CoVZC45, bat-SL-CoVZXC21 and Rm1 show the worst energy scores. While the binding energy score falls in the boundary limit of the incertitude margin for PRODIGY calculation (section 2, Supplementary material 2), the differences in the scores calculated by FoldX and MM-GBSA are not. Therefore we consider that such differences in energies compared to SARS-CoV-2 are consistent between the three methods. Except for FoldX, the affinity is predicted to be more favorable for RBD from SARS-CoV-2 compared to SARS-CoV. However, MM-GBSA only marginally discriminates between the two values.

3.4. Interaction of RBD from SARS-CoV-2 and SARS-CoV with different ACE2 orthologues

To investigate the tendency of SARS-CoV-2 and SARS-CoV to interact with different orthologous forms of ACE2, we analysed the divergence in their respective interacting surfaces. We have also mapped the putative glycosylation sites that overlap with the interface with RBD. Overall, the binding energy scores are similar between SARS-CoV-2 and SARS-CoV considering the estimation of error for each method. Variances are more important for the calculations made by FoldX and although of different formalism, MM-GBSA and PRODIGY scores are relatively in agreement. Compared to hACE2, only the Canidae form shows better energy scores both in PRODIGY and MM-GBSA for SARS-CoV-2. Moreover, We found that putative glycosylation sites overlap significantly with RBD interaction in Snake, Rat and Bat forms (section 3, Supplementary data 2). The docking also shows that key residues of RBD SARS-CoV-2 tend to interact with conserved residues on ACE2 (Fig. 3), which can explain the similar values of energy scores.

3.5. Decomposition of the interaction energy

MM-GBSA allowed us to assign the contribution of each amino acid in the interface with hACE2, in the binding energy score. We conducted this analysis using both sequences of the SARS-CoV-2 Wuhan-Hu-1 (Fig. 3A) and the Sin01-11 SARS-CoV (Fig. 3B) isolates. Residues F486, Y489, Q493, G496, T500 and N501 of SARS-CoV-2 RBD forming 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 (Fig. 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 (Fig. 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.

The similarity matrix analysis was conducted to assess the divergence of the interaction interface of RBDs qualitatively, i.e. the specific set of residues implicated in the interaction with ACE2, and quantity, i.e. the contribution of each residue in the binding energy score. The similarity matrix was calculated from free energy decomposition of interface residues of RBDs from SARS-CoV-2 and SARS-CoV in complex with ACE2 orthologous and reported as a network representation (Figs. 3E, 1 and 2 in Supplementary Materials 2). We noticed the existence of densely interconnected edges involving all the protein-protein complexes for SARS-CoV-2 and
SARS-CoV except those involving ACE2 from Sus scrofa and Rattus norvegicus. Complexes involving the RBD of SARS-CoV-2 show less intrinsic similarity compared to RBD of SARS-CoV. However, similarity scores tend to be uniform in the group involving ACE2 from human, civet, dog, bat, snake, and chicken. The complex including hACE2 does not seem to diverge from the rest of the members of the SARS-CoV-2 group such as the case of Sus scrofa and Rattus norvegicus.

3.6. 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 (Fig. 3F). Indeed, the RMSF variability per amino acid for this sequence increases compared to the reference structure.

3.7. 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 (Fig. 4A). 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 changes (> 1 or < 1 kcal/mol) in neither the folding energy of the complex (Fig. 4B) nor the interaction energy of the protein-protein partners (Fig. 4C).

4. 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 [21]. 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 virulence in the host. Our results showed 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 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 similar values of binding energy scores with different ACE2 orthologues suggest that the ability of binding to different ACE2 orthologous is preserved in many species either for SARS-CoV-2 or SARS-CoV. Therefore, the transition to the zoonotic form is trivial if that depends only on ACE2 as the primary route to the infection in both the intermediate and the final host. However, we know that such a process is very complex since it requires many protein-protein interactions to acquire the specific capacity of infecting and replicating in the host cells [18]. Consequently, it makes sense to assume that many other types of receptors or co-receptors may be critical to determine the capacity of crossing the species barrier. This has been already suggested for SARS-CoV [1] and similarly, SARS-CoV-2 may show the same feature. Moreover, our results show that the significant overlap of glycosylation sites with the protein-protein interface implies a likely interaction of SARS-CoV-2 progenitors with
receptors other than ACE2. Finally, recent transcriptomic profiling has suggested the possibility of multiple route infections via the interaction of many human receptors for SARS-CoV-2 [11].

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 [10,17] 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 Fig. 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, the significant divergence in the interfaces segments as a key molecular element contributing to the determination of the tree topology has driven our work toward studying their impact on the interaction with hACE2. 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 scores 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 score to reach a level where the pathogen can infect humans with high affinity toward hACE2. On the other hand, our results show that the binding energy score 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

---

**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 |
|-------------------|------------------------|---------|--------|------------------|------------|-------------|---------|-------|--------|
| rs961390700       | 2.59                   | 0       | 0      | 0                | 0          | 0           | 0       | 1.17  |        |
| rs143936283       | 6.51                   | 0       | 0      | 0                | 0          | 0           | 0       | 3.44  |        |
| rs146676783       | 0                      | 0.105   | 0      | 0                | 0          | 0           | 12.22   | 3.897 |        |
| rs759579097       | 0.1056                 | 0       | 0      | 0                | 0          | 0           | 0       | 0.9842|        |
| rs370610075       | 1.274                  | 0       | 0.1056| 0                | 0          | 0           | 0       | 0     | 0.5752|
| rs768996587       | 0                      | 0.2623  | 0      | 0                | 0          | 0           | 0       | 0     | 2.442  |
| rs73635825        | 0                      | 0       | 33.23  | 0                | 0          | 0           | 0       | 0     | 31.29  |
| rs781255386       | 0                      | 0       | 7.303  | 0                | 0          | 0           | 0       | 0     | 1.091  |

---

Fig. 4. Analyzing the interaction of SARS-CoV-2 RBD with different variants of hACE2. (A) Localization of the variants, labeled by the amino acid change and the dbSNP ID, on the interaction surface of hACE2 and RBD from SARS-CoV-2. Estimation of the changing upon mutation for hACE2 variants calculated for enthalpy (ddG) and entropy (ddS) terms of the folding energy calculated with DynaMut (B) and the interaction energy calculated with PRODIGY (D).
recombination in the spike protein gene has been previously suggested by Wei et al. in their phylogenetic analysis [4]. 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. [17] 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 $-2.55 \text{ kcal/mol}$ to a non significant value of $-0.01 \text{ kcal/mol}$, thus supporting our claim.

No major divergence of the interaction interface of SARS-CoV-2 RBD with hACE2 was noticed from the similarity matrix analysis. This suggests that the molecular elements required for the binding with the receptor might also be involved in the interaction with other orthologous forms of ACE2 and that these elements are not optimized specifically for the human form. Therefore, it is unlikely that the interface of RBD from SARS-CoV-2 is a result of human intervention via genetic engineering aiming to increase the affinity toward ACE2. For example, residue E484 contributes unfavorably to the binding energy with $2.24 \text{ kcal/mol}$ due to an electrostatic repulsion with E75 from hACE2. This residue is an apparent choice toward ACE2. One of these changes may involve Q493 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. [17] 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 $-2.55 \text{ kcal/mol}$ to a non significant value of $-0.01 \text{ kcal/mol}$, thus supporting our claim.

Declaration of competing interest

No conflicts of interest.

Acknowledgement

This work was supported by the South African National Research Foundation (NRF) (ID number SFH170626244782) and the Tunisian Ministry of Higher Education and Scientific Research. The author H. Othman would like to thank Shahine Othman for being understanding about his absence and for not being able to bring him the marshmallow candy because of the COVID-19 outbreak.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrc.2020.05.028.

References

[1] M. Bolles, E. Donaldson, R. Baric, SARS-CoV and emergent coronaviruses: viral determinants of interspecies transmission, Curr Opin Virol 1 (6) (Dec 2011) 624–634.
[2] J.F. Chan, S. Yuan, K.H. Kok, K.K. To, H. Chu, J. Yang, F. Xing, J. Lui, 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, 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) (Feb 2020) 514–523.
[3] J. Delgado, L.G. Radusky, D. Canferoni, L. Serrano, FoldX 5.0: working with RNA, small molecules and a new graphical interface, Bioinformatics 35 (20) (Oct 2019) 4168–4169.
[4] W. Ji, W. Wang, X. Zhao, J. Zai, X. Li, Cross-species transmission of the newly identified coronavirus 2019-nCoV, J. Med. Virol. 92 (4) (Apr 2020) 433–440.
[5] K. Katoh, J. Rozewicki, K.D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, Briefings Bioinf. 20 (4) (07 2019) 1160–1166.
[6] H. Othman et al. / Biochemical and Biophysical Research Communications 527 (2020) 702–708