Biochemistry and Biophysics Reports 24 (2020) 100798

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

ACE2 coding variants in different populations and their potential impact on SARS-CoV-2 binding affinity

Fedaa Ali [a,1], Menattallah Elserafy [b,1], Mohamed H. Alkordi [c,*], Muhamed Amin [a,d,e,**]

[a] Department of Sciences, University College Groningen, University of Groningen, Hoendiepskade 23/24, 9718, BG Groningen, Netherlands
[b] Center for Genomics, Helmy Institute for Medical Sciences, Zewail City of Science and Technology, October Gardens, 6th of October, Giza, 12578, Egypt
[c] Center for Materials Science, Zewail City of Science and Technology, October Gardens, 6th of October, Giza, 12578, Egypt
[d] Rijksuniversiteit Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, Netherlands
[e] Nanotechnology Research Centre (NTRC), The British University in Egypt (BUE), El Sherouk City, Suez Desert Road, Cairo, 1183, Egypt

ARTICLE INFO

Keywords:
SARS-CoV-2 binding
ACE2 variants
Populations
Single nucleotide variants

ABSTRACT

The susceptibility of different populations to SARS-CoV-2 infection is not yet understood. Here, we combined ACE2 coding variants' analysis in different populations and computational chemistry calculations to probe the effects on SARS-CoV-2/ACE2 interaction. ACE2-K26R; which is most frequent in Ashkenazi Jewish population decreased the SARS-CoV-2/ACE2 electrostatic attraction. On the contrary, ACE2-I468V, R219C, K341R, D206G, G211R increased the electrostatic attraction; ordered by binding strength from weakest to strongest. The aforementioned variants are most frequent in East Asian, South Asian, African and African American, European, European and South Asian populations, respectively.

1. Introduction

Despite numerous reports on the clinical manifestation of the recent novel coronavirus disease 2019 (COVID-19); including pathogenicity, diagnosis, and recommended treatment regimes, understanding of the mechanisms of infection, the virus-host interactions, and transmission is still in its early stage [1]. The risk of certain populations for the infection is also still not well understood and remains under investigation. After the identification of the angiotensin-converting enzyme 2 (ACE2) as the SARS-CoV-2 receptor [2], a large number of reports have emerged trying to identify the pathology of the disease, and thus provide guidance for development of treatments. Some recent studies attempted to probe the comorbidity among infected patients, specifically patients with hypertension, diabetes mellitus, coronary heart diseases, and cerebrovascular disease that showed increased risk. Whether ACE2 polymorphisms linked to the aforementioned diseases or administration of ACE2 modulating drugs have larger impact on the severity of infection is still a question to be investigated [3].

Communications between the pathogen and the host is mainly governed by the different protein-protein electrostatic interactions [4]. These proteins need to bind in a specific way for protein-protein complexes formation and for the signals to be effectively transmitted. The existence of residues that are involved in energetically favorable interactions at the protein-protein interface leadsto enhanced interactions leading to viral-host cells fusion, marking the onset of the infection [2,5]. Therefore, we opted herein to investigate the nature of the interactions at the SARS-CoV-2/ACE2 interface, and how the subtle variations in the structure of ACE2, induced by certain mutations, can alter the binding of SARS-CoV-2 in different populations. This was attempted here by analyzing different ACE2 missense variants that code for ACE2-K26R, ACE2-I468V, ACE2-R219C, ACE2-K341R, ACE2-D206G, ACE2-G211R and the strength of their interaction with the spike protein (*S*-protein) of SARS-COV-2 using Molecular Dynamics and Monte Carlo Sampling, which are dominated by electrostatics. The electrostatic interactions are the dominant factor in protein-protein interactions [6], where classical treatment of these interactions in solution are described by the Poisson-Boltzmann equation (PBE). Our data points to a possible increase in the strength of the binding of the variants; in the order mentioned, with K26R showing the least favorable binding and G211R showing the most favorable binding. We analyzed the frequencies of the

* Corresponding author.
** Corresponding author. Department of Sciences, University College Groningen, University of Groningen, Hoendiepskade 23/24, 9718, BG Groningen, Netherlands.
E-mail addresses: malkordi@zewailcity.edu.eg (M.H. Alkordi), m.a.a.amin@rug.nl (M. Amin).
[1] These authors contributed equally to the manuscript.

https://doi.org/10.1016/j.bbrep.2020.100798
Received 20 July 2020; Received in revised form 11 August 2020; Accepted 12 August 2020
Available online 20 August 2020

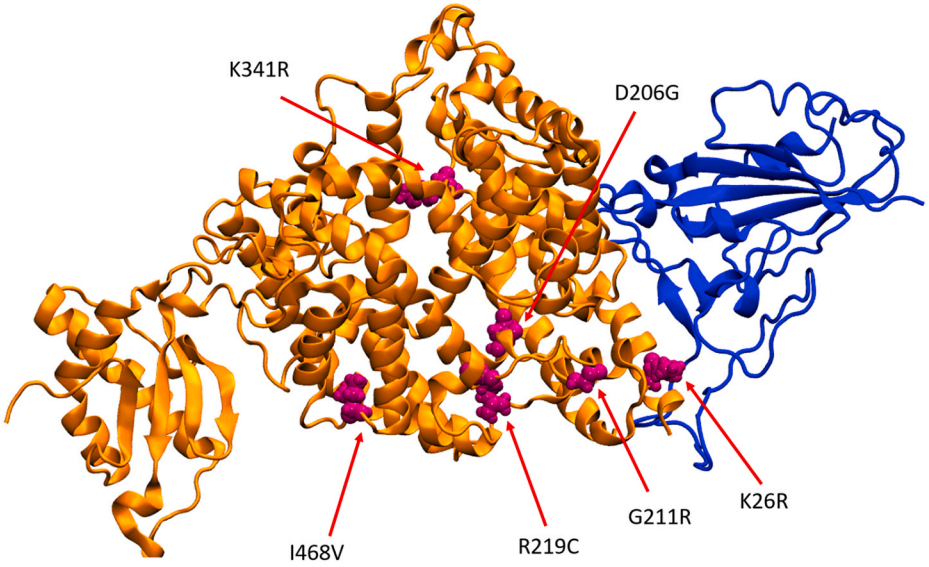

Fig. 1. The secondary structure of the ACE2 protein (shown in orange) and the RBD of the SARS-CoV-2 (shown in blue). The mutated amino acids are shown in magenta spheres (6M17)[5].

alleles coding for the variants in different populations, to identify the population in which each variant is more frequent. Our data helps in identifying variations of appreciable frequencies and the implications for each in affecting the strength of the virus-receptor interactions in individuals of different populations.

2. Materials and methods

Classical electrostatic calculations based on solving Poisson-Boltzmann (PB) equation are used to investigate the interaction energies between SARS-CoV-2 and ACE2 for different mutated ACE2 structures. All continuum electrostatics calculations were performed by the free available software "Multi-Conformer Continuum electrostatics (MCCE)" [7], where interaction energies are modeled by treating each residue as separate fragments with integer charges, which are interacting with each other by means of electrostatic and Lennard-Jones potentials [7]. The Poisson-Boltzmann (PB) equation is numerically solved by DELPHI software, which is integrated within the MCCE code, to compute self-energies of each conformer and pair-wise interaction energies between conformers and between conformers and backbone. Upon energy calculations, the energy look-up table is built, and theoretical titration is modeled by obtaining Boltzmann distribution of microstates, using the Monte-Carlo approach. The microstates of the structures are defined depending on rotamers and protonation states of the different amino acids. The protein interior is considered as highly polar media with low dielectric constant ($\varepsilon_{prot} = 4$), whereas the solvent (water) is treated as a continuum medium with high dielectric constant ($\varepsilon_{wat} = 80$).

Initial coordinates are defined according to the X-ray crystal structure of the wild type (WT) human ACE2 with the receptor-binding domain (RBD) of the S protein of SARS-CoV-2 (PDB ID: 6M17) from the protein data bank [2]. The structure of the Neutral amino acid transporter B^0AT1 was removed from the PDB file, as it is far from the SARS-CoV-2/ACE2 binding site. Furthermore, the structure of ACE2 was trimmed by removing residues from P733 to the end of the chain. The WT residues of ACE2 at positions 26, 206, 211, 219, 341 and 468 were mutated to Arg (R), Gly (G), Arg (R), Cys (C), Arg (R) and Val (V), respectively. The sidechains in the WT structure were replaced with the sidechains of the mutants using MCCE (Multi Conformer Continuum Electrostatics) (Fig. 1). A set of different conformers of the sidechains were generated for the WT and mutated structures by rotating the single bonds in the sidechains by 60° degrees. In addition, conformers were created for charged amino acids according to their possible protonation states. Then, these conformers were subjected to Monte Carlo sampling to obtain the Boltzmann distribution based on the electrostatic interactions, which are calculated using DELPHI. The most occupied conformers in the generated Boltzmann distributions were selected and subjected to molecular dynamics optimization using openMM. Finally, the interaction energies were computed for the optimized structures using MCCE. The electrostatics and van der Waals interactions between the amino acids in the RBD of SARS-CoV-2 and ACE2 were analyzed using python-based code.

3. Results and discussion

In an attempt to better understand the susceptibility of different populations to infection by SARS-CoV-2, we gathered data on ACE2 missense variants from different projects and databases that aggregate allele frequencies (AF). These projects and databases include Single Nucleotide Polymorphism Database (dbSNP) [8], 1000 genomes project phase 3 (1KGP3) [9], Allele Frequency Aggregator (ALFA project)[a], Exome Aggregation Consortium (ExAC) [10], genome Aggregation Database (gnomAD) [11], GO exome sequencing project (ESP)[b], Trans-Omics for Precision Medicine (TopMed) [12], in addition to a recent study reporting data from the China Metabolic Analytics Project (ChinaMAP) and other populations [13,14].

SARS-CoV-2 was reported to bind to human ACE2 via different ACE2 residues; Q24, D30, H34, Y41, Q42, M82, K353 and R357 [2]. We analyzed all ACE2 coding variants reported and selected the variants with the most allele frequencies (AF) in the African, African American, American, Ashkenazi Jewish, Asian (All, ChinaMAP, East Asian, Han Chinese South, Han Chinese Beijing and South Asian), European (All, European American, Finnish and non-Finnish) and Latino. The residues selected include K26R, D206G, G211R, R219C, K341R and I468V (see AF in populations in Table S1). Despite the fact that the chosen residues are non-contact residues that do not fall few angstroms away from the RBD residues, we prioritized allele frequency rather than proximity alone to assess the variants' potential impact on specific populations.

3.1. Variations in ACE2 affect its interaction with SARS-CoV-2

The electrostatic and the van der Waals contribution to the interaction energies of SARS-CoV-2/ACE2 were compared between single mutated and WT protein at pH = 7 (Fig. 2A). In contrary to K26R, most

A

The interaction energies between SARS-CoV-2 and ACE2				
	Electrostatics (Kcal/mol)	vdW (Kcal/mol)	Total (Kcal/mol)	Population
G211R	−12.4	−35.4	-47.8	European South Asian
D206G	−8.8	−36.1	-44.9	European
K341R	−14.2	−29.3	-43.4	African/African American
R219C	−4.6	−38.3	-42.9	South Asian
I468V	−4.7	−37.4	-42.1	East Asian
Native	−9.6	−30.1	−40.2	
K26R	−3.7	−34.4	-38.1	Ashkenazi Jewish
The entries are sorted from lowest to the highest interaction energies				

B

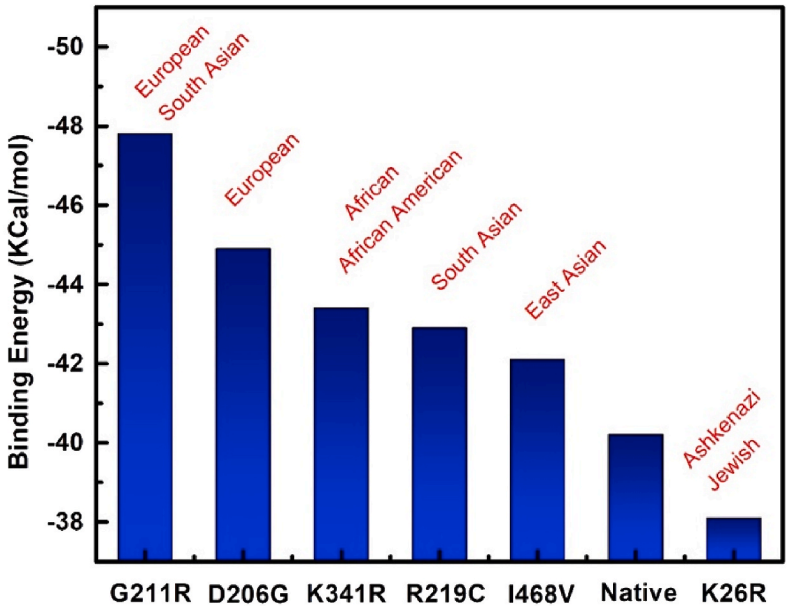

Fig. 2. The impact of ACE2 coding variants on SARS-CoV-2 binding A) Representation of Binding energies calculated and the corresponding population with the highest allele frequency. B) Graphical representation of (A).

of the mutated structures exhibit a negative shift in the total interaction energy compared to the WT structure (Native), i.e. more electrostatic attraction to the spike protein of SARS-CoV-2. Based on our calculations, G211R mutant is shown to induce the largest increase in the binding energy between SARS-CoV-2 and ACE2, where the binding is more favorable by ~7.6 kcal/mol than the WT. However, the K26R mutant causes a decrease in the binding energies by ~2.1 kcal/mol (Fig. 2A and B).

The substitution of Lys (K) to Arg (R) at positions 211 and 341 was shown to increase the electrostatic attraction between SARS-CoV-2 and ACE2 by 1.29-fold and 1.47-fold in comparison to the WT, respectively. Oppositely, substitution of R at position 26 was shown to reduce the electrostatic attraction between SARS-CoV-2 and ACE2 to 2.59 kcal/mol lower than that in WT. In addition, the D206G mutant showed similar electrostatic interactions to the WT. However, the R219C, I468V mutants showed significant decrease in the electrostatic attraction to by ~5 kcal/mol compared to the WT (Fig. 2A and B). SARS-CoV-2/ACE2 interface in the WT protein and corresponding mutants is shown to be dominated by van der Waals interactions, which account for more than 60% of the interaction energy. All the mutations except K341R induced an increase in the van der Waals attractions between the ACE2 and the SARS-CoV-2. The largest increase of ~8 kcal/mol is observed for the R219C variant (Fig. 2A).

Our results, shown in Fig. 2, demonstrate large variations in calculated binding energy in different ACE2 variants towards SARS-CoV-2 S protein. Although the calculated variations in binding energies are appreciable, as compared to the control, we hypothesize that even within the same population mutations in ACE2 in certain individuals can similarly dramatically influence the binding site configuration and ACE2/SARS-CoV-2 binding affinity.

3.2. How ACE2 variants can impact the susceptibility of different population to SARS-CoV-2 infection?

The K26R variant which decreases the ACE2-virus binding was found to be most frequent in the Ashkenazi Jewish population (1.2%). On the contrary, the Asian population had the lowest allele frequency for the single nucleotide variant (SNV) coding for K26R. In addition, the frequency of I468V which enhance ACE2 interaction with SARS-CoV-2 was observed in more than 1% of the East Asian population. Interestingly, I468V was not detected in the African, African American, American, Ashkenazi Jewish and Latino populations and was very rare in the European population. R219C was most frequent in the south Asian population (0.1%), but was not detected in other Asian subpopulations and was much less frequent or lacking in other populations. The K341R, which enhance the virus-receptor binding was most frequent in the African and African American populations; the average of the AF reported for the populations was ~0.4% and ~0.5%, respectively. Finally, the

D206G and G211R, which showed the strongest interaction with ACE2 were most abundant in the European population, mounting to 0.06% and ~0.2%, respectively. The G211R was also abundant in the South Asian population ~0.2%.

Our findings give indications regarding the populations that could be more prone to SARS-CoV-2 infection due to enhanced binding affinity. Nevertheless, there is a need for proper infection, death and recovery statistics to reach a definite conclusion. K26R and I468V are considered of intermediate frequencies (>1%) in the Ashkenazi Jewish and East Asian population, respectively. Therefore, we would refer to them as Single Nucleotide Polymorphism (SNPs) in the respective populations. However, the other variants are considered of rare frequencies (<1%) [15]. Nevertheless, in the era of Individualized Genomics, the individual differences in the same population should also be taken into account, as rare variants might exist affecting the virus binding to ACE2 in many individuals.

The possibility of having individuals baring more than one of the mutations that enhance the binding is also still an open question, that will be answered via sequencing of ACE2 in more individuals of the aforementioned populations. The presence of variants for the serine protease TMPRSS2; required for S protein priming, in different populations could also account for differences in viral entry and pathogenesis [5].

4. Conclusion

Our data suggest that ACE2 variants in different populations could cause significant variations in the binding affinity between SARS-CoV-2 and ACE2. Yet, experimental binding assays are still needed to confirm our claim.

To better understand the susceptibility of individuals from different populations to SARS-CoV-2 and their risk of infection, large scale sequencing projects should be performed. Integrating population genetics in SARS-CoV-2 studies will provide a new insight into the mystery of susceptibility, infection, pathogenicity and mortality in different regions. Sequencing of ACE2 in patients with the most severe conditions in every population will also provide a more reliable conclusion. The sequencing projects will not only help us test the association between ACE2 variants and risk/severity of infection in certain populations, it will also allow more accurate analysis of variants for other host genes involved in the viral entry and pathogenicity. Finally, our findings can potentially guide future attempts to devise small molecules inhibitors, designed specifically to disrupt the interaction between the strongest ACE2 binding variants and the S protein of SARS-CoV-2.

Author contribution

Fedaa Ali: Investigation, Validation, Formal analysis, Writing - review & editing, Menattallah Elserafy: Investigation, Validation, Formal analysis, Writing - review & editing, Mohamed Alkordi: Conceptualization, Methodology, Writing - original draft, Supervision, Muhamed Amin: Conceptualization, Methodology, Writing - original draft, Supervision.

Declaration of competing interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2020.100798.

References

[1] Y.R. Guo, Q.D. Cao, Z.S. Hong, Y.Y. Tan, S.D. Chen, H.J. Jin, K. Sen Tan, D. Y. Wang, Y. Yan, The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak- A n update on the status, Mil. Med. Res. 7 (2020) 1–10, https://doi.org/10.1186/s40779-020-00240-0.

[2] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, Science 367 (2020) 1444–1448, https://doi.org/10.1126/science.abb2762, 80-.

[3] L. Fang, G. Karakiulakis, M. Roth, Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? Lancet Respir. Med. 8 (2020) e21, https://doi.org/10.1016/S2213-2600(20)30116-8.

[4] F.B. Sheinerman, R. Norel, B. Honig, Electrostatic aspects of protein-protein interactions, Curr. Opin. Struct. Biol. (2000), https://doi.org/10.1016/S0959-440X(00)00065-8.

[5] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.H. Wu, A. Nitsche, M.A. Müller, C. Drosten, S. Pöhlmann, SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor, Cell 181 (2020) 271–280, https://doi.org/10.1016/j.cell.2020.02.052, e8.

[6] B. Honig, A. Nicholls, Classical electrostatics in biology and chemistry, Science (1995), https://doi.org/10.1126/science.7761829.

[7] Y. Song, J. Mao, M.R. Gunner, MCCE2: improving protein pKa calculations with extensive side chain rotamer sampling, J. Comput. Chem. (2009), https://doi.org/10.1002/jcc.21222.

[8] E.M. Smigielski, dbSNP: a database of single nucleotide polymorphisms, Nucleic Acids Res. (2000), https://doi.org/10.1093/nar/28.1.352.

[9] A. Auton, G.R. Abecasis, D.M. Altshuler, R.M. Durbin, D.R. Bentley, A. Chakravarti, J.A. Schloss, et al., A global reference for human genetic variation, Nature (2015), https://doi.org/10.1038/nature15393.

[10] K.J. Karczewski, B. Weisburd, B. Thomas, M. Solomonson, D.M. Ruderfer, D. Kavanagh, T. Hamamsy, M. Lek, K.E. Samocha, B.B. Cummings, D. Birnbaum, M.J. Daly, D.G. MacArthur, The ExAC browser: displaying reference data information from over 60 000 exomes, Nucleic Acids Res. (2017), https://doi.org/10.1093/nar/gkw971.

[11] K.J. Karczewski, L.C. Francioli, G. Tiao, B.B. Cummings, J. Alföldi, Q. Wang, D. G. MacArthur, et al., The mutational constraint spectrum quantified from variation in 141,456 humans, Nature (2020), https://doi.org/10.1038/s41586-020-2308-7.

[12] D. Taliun, D.N. Harris, M.D. Kessler, J. Carlson, Z.A. Szpiech, R. Torres, G. R. Abecasis, et al., Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program, BioRxiv, https://doi.org/10.1101/563866, 2019.

[13] Y. Cao, L. Li, Z. Feng, S. Wan, P. Huang, X. Sun, F. Wen, X. Huang, G. Ning, W. Wang, Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations, Cell Discov (2020), https://doi.org/10.1038/s41421-020-0147-1.

[14] Y. Cao, L. Li, M. Xu, Z. Feng, X. Sun, J. Lu, M. Wang, et al., The ChinaMAP analytics of deep whole genome sequences in 10,588 individuals, Cell Res. (2020), https://doi.org/10.1038/s41422-020-0322-9;

[15] L. Bomba, K. Walter, N. Soranzo, The impact of rare and low-frequency genetic variants in common disease, Genome Biol. 18 (2017) 1–17, https://doi.org/10.1186/s13059-017-1212-4;

a L. Phan, Y. Jin, H. Zhang, W. Qiang, E. Shekhtman, D. Shao, D. Revoe, R. Villamarin, E. Ivanchenko, M. Kimura, Z.Y. Wang, L. Hao, N. Sharopova, M. Bihan, A. Sturcke, M. Lee, N. Popova, W. Wu, C. Bastiani, M. Ward, J.B. Holmes, V. Lyoshin, K. Kaur, E. Moyer, M. Feolo, B.L. Kattman, ALFA: allele frequency aggregator National Center for Biotechnology Information, U.S. National Library of Medicine. www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/. (Accessed 10 March 2020);

b Exome Variant Server, NHLBI GO exome sequencing project (ESP), Seattle, WA, http://evs.gs.washington.edu/EVS/ [date (month, yr) accessed].