Missense Variants in Human ACE2 May Influence its Binding Interaction with SARS-CoV-2 and Infectivity of COVID-19

Fang Ge1* and Arif Muhammad2

1 School of Computer Engineering, Bengbu University, Caoshanlu 1866, Bengbu 233030, P.R. China
2 School of Computer Science and Engineering, Nanjing University of Science and Technology, Xiaolingwei 200, Nanjing 210094, P.R. China

*Email: gfang0616@njust.edu.cn

Abstract. The recent pandemic of COVID-19 is reported as a pandemic and spreads globally. As known, COVID-19 is caused by SARS-CoV-2 and human ACE2 has been reported as the receptor of SARS-CoV-2. Nowadays, as reported, variants in the coronavirus can lead to the transmission across species, in turn, variants of ACE2 may affect the susceptibility of SARS-CoV-2. In this work, we collected and selected critical missense variants in ACE2. Thereafter, we predicted the changes of protein-protein binding affinity, corresponding to each missense variant. According to the results, thirteen variants in human ACE2 exhibit obvious differences. More specifically, six variants (D38E, M82I, Y83F, K353H, R357A, and R357S) in ACE2 are predicted to enhance its interaction with the coronavirus spike protein. In comparison, seven variants (S19P, K31D, Y41A, M82N, M82T, D355A, and D355N) are predicted to inhibit such kind interaction. Accordingly, the final findings of our work may provide evidence that the potential relationship between COVID-19 susceptibility and human ACE2 genetic variants.

1. Introduction
COVID-19 has been declared as a pandemic and spreads worldwide. As known, until 30 November 2020, COVID-19 has infected more than 86.10 million people and caused more than 1.86 million deaths [1]. As scientific research reported, COVID-19 is caused by SARS-CoV-2 [2]. Furthermore, the coronavirus invades human cells through an endocytic pathway. This pathway starts from the coronavirus spike protein binding with ACE2 [3, 4]. As described in the literature, variants in the spike protein can give the ability of transmission, with the characteristic of high speed and cross-species [5]. Conversely, variants in ACE2 may result in susceptibility against virus infection [6].

According to recent scientific researches, variants in SARS-CoV-2 and ACE2 have been explored and lots of important conclusions have been turned out [7-9]. For example, variants in ACE2 may be the underlying reason for the widespread infection of COVID-19 [6, 8, 9]. Despite lots of scientific research have implemented to explore the clinical manifestation of COVID-19, including, the coronavirus and human host interactions, the understanding of virus infection mechanism, variants pathogenicity, to some extent, the investigations are still in its early stage [10-14]. In this work, one interesting hypothesis is proposed: does the missense variant in ACE2 affect the interaction with SARA-CoV-2? Herein, we mainly applied the SSIPe [15] software to calculate the affinity changes after the amino acid substitution at the protein-protein interaction. On the other hand, we utilized the discovery...
studio 3.5 visualizer [16] to monitor the hydrogen bond interaction of wild-type, as well as the mutant complex protein. The final findings of our work may provide evidence that the potential relationship between the COVID-19 susceptibility and ACE2 genetic variants.

2. Methodology

2.1. Missense Variants Collecting for Human ACE2

As reported, SARS-CoV-2 can encode more than twenty-seven proteins, spike protein (also named as spike glycoprotein) [17] is one kind of them. The spike protein is located in the outer membrane [7], whose functional domains are depicted in Figure 1(A).

![Figure 1](image-url) Structure of coronavirus RBD and human ACE2. (Abbreviations: NTD, N-terminal domain; FP, fusion peptide; SP, signal peptide; HR, heptad repeat 1, and heptad repeat 2; TM, transmembrane region [18]. Note: Figure 1(A) is adapted from Ref [7] and Figure 1(C) is adapted from Ref [19].)

As shown in Figure 1(A), the spike protein has 1273 residues, which comprises two parts, i.e., N-terminal S1 subunit and C-terminal S2 subunit [20]. In addition, there is a region named receptor-binding domain (RBD), with 200 amino acids. Further, the RBD region contains the core and external subdomains [21, 22]. As scientific research reported, the mainly binding region motif with human ACE2 is in the external subdomain of the RBD [4, 9, 23, 24]. More specifically, it is the receptor-binding motif (RBM) of RDB domain that interacts directly with human ACE2 [20]. Herein, we collected and selected variants in ACE2 from Uniport [25], COSMIC [26], and Ref [5], Ref [4], and Ref [27]. After the selecting procedure, we obtained thirty variants that mainly marked as the critical functional sites and/or binding sites with ACE2. The thirty variants are S19P, I21T, K26R, K31D, N33D, D38E, Y41A, M82N, M82I, M82T, Y83F, P84S, P84A, P84Q, P84T, N149S, N159S, G173S, V184A, I233F, K353H, D355A, D355N, R357A, R357S, M383I, I468V, A501T, G575V, and G592L.

2.2. Structure Analysis of the Complex Protein

Through X-ray diffraction or cryoelectron technologies, 3D structures of the coronavirus interacting with ACE2 have been resolved, for example, PDBid: 7C8D, 6LZG, 6VV1, 6M17, and others. Among these complex structures, we noticed that 6LZG is detected with a non-chimeric nature, as well as with a quite low wavelength (i.e., 2.50Å). Thus, we chose 6LZG as the complex structure. Noteworthy here, in 6LZG, there are two chains: partly of human ACE2 protein (i.e., chain A) and the RBD domain of coronavirus (i.e., chain B).

In this work, we utilized the SSIPe [15] software to calculate the affinity changes after the amino acid substitution at the protein-protein interaction. Furthermore, in order to explore the hydrogen bond changes of missense variants in human ACE2, we adopted the discovery studio 3.5 visualizer [16] to monitor the hydrogen bond interaction of wild-type, as well as the mutant complex protein. Meanwhile, we also recorded the interaction hydrogen changes...
between the wild type and the mutant ones.

3. Results and Discussions

As reported, protein usually interacts with other proteins for conducting their biological functions in cells [15]. That is, if one amino acid (AA) mutated, especially that locating at the protein interaction interface, it might alter the proteins binding stability and further impact their biological functions, which might thereby lead to various diseases [15]. As aforementioned in Section 1, the hypothesis we proposed is: does the missense variants in human ACE2 affect the binding interaction with the coronavirus spike protein? Considering such a hypothesis, by using the SSIPe [15] tool, we calculated the protein binding affinity changes after one amino acid substitution. Before calculating, two critical formulas should be introduced, shown as below.

\[
V G_{\text{bind}} = G_{\text{complex}} - G_{\text{monomers}} \tag{1}
\]

\[
V V G_{\text{bind,wt\rightarrow mut}} = G_{\text{bind,mut}} - G_{\text{bind,wt}} \tag{2}
\]

Equation (1) is the binding affinity of two proteins, which is measured by Gibbs free energy changes. Herein, \(G_{\text{monomers}}\) and \(G_{\text{complex}}\) are the Gibbs energy of the monomer protein and complex protein, respectively. In general, the greater the negative \(\Delta G_{\text{bind}}\) is, the stronger the monomer protein binding is.

As for missense variants of human ACE2, we calculated the binding affinity changes. That is, the Gibbs energy of variant complex protein (i.e., \(G_{\text{bind,mut}}\)) minus that of the wild type complex protein (i.e., \(G_{\text{bind,wt}}\)), as shown in Equation (2). Herein, in order to facilitate the writing, \(V V G_{\text{bind,wt\rightarrow mut}}\) in Equation (2) is abbreviated as \(\Delta \Delta G_{\text{bind}}\). According to the description in Ref. [15], after one amino acid substitution, if the Gibbs energy change of the mutant and wild type proteins binding affinity is negative (\(\Delta \Delta G_{\text{bind}} < 0 \text{kcal/mol}\)), it means variants in human ACE2 protein can enhance its interaction with the coronavirus. In contrast, if the change of Gibbs energy is positive (\(\Delta \Delta G_{\text{bind}} > 0 \text{kcal/mol}\)), it indicates that variants in human ACE2 protein may inhibit the interaction with the coronavirus.

Figure 2. The predicted \(\Delta \Delta G_{\text{bind}}\) scores for the mutated complex protein.

Figure 3. Structure of human ACE2 (in green) and the RBD domain of coronavirus (in cyan).

As aforementioned in Section 2.2, we selected PDBid: 6LZG as the wild type complex protein (also named as a dimer) of ACE2 and the RBD domain of coronavirus. In order to get protein-protein binding affinity changes, we put the three-dimensional coordinates of the dimer (i.e., 6LZG) and the missense variants information of ACE2, into the SSIPe web server [15]. After calculating, we depicted the returned \(\Delta \Delta G_{\text{bind}}\) score in Figure 2. As depicted in Figure 2, seven missense variants in human ACE2 protein have positive \(\Delta \Delta G_{\text{bind}}\) scores, while six variants are predicted with negative \(\Delta \Delta G_{\text{bind}}\) scores. That is, each of the seven variants (i.e., S19P, K31D, Y41A, M82N, M82T, D355A, and D355N) in ACE2 can decrease the binding affinity with the RBD. Conversely, each of the six variants (i.e., D38E, M82I,
Y83F, K353H, R357A, and R357S) in human ACE2 protein can increase its binding affinity with the coronavirus.

As for the missense variants of human ACE2 with positive bindG scores, it is believed that such variants can enhance the resistance against COVID-19. For example, as reported, variant S190P in human ACE2 protein is reported to offer resistance against evolving pathogenic viruses [5]. In this work, we mainly focused on the other six missense variants with negative predicted bindG scores, which are displayed in Figure 3, shown in sphere format and different colors. By observing the position of variants depicted in Figure 3, we found six aforementioned missense variants are all at the surface of human ACE2 protein. As known, variants occurring in proteins surface may more likely affect the affinity changes of binding with the coronavirus. Furthermore, according to the research results in the Ref [28], residues near 31, and 41, 82-84, 353, and 357 in ACE2 protein, are quite critical for the binding with spike protein of the coronavirus [28]. Notably, residue methionine 82, tyrosine 83, lysine 353, and arginine 357 shown in Figure 3 are all such kinds of residues. Accordingly, in the following part, we would explore the impact of the above six variants in human ACE2 protein from a more subtle perspective.

The returned files of the SSIPe [15] web-server comprise the predicted bindG scores, as well as the complex protein 3D structures which are corresponding to each missense variant. We displayed the six complex proteins predicted with negative bindG scores in Figure 4. Notably, six variants are colored and shown in sphere format. By observing the details from Figure 4, we can find that the overall structure of the complex proteins, corresponding to each variant, has no obvious difference. For further closer observation, we zoomed the mutated residue partially and displayed the residues in the 5Å range around the mutant site, along with the hydrogen bonds between atoms, depicted in the enlarged boxes.

From Figure 4, we can see clearly the residues in 5Å around the mutant site and the hydrogen bonding interaction among atoms. As shown in Figure 4(A), 5(D), 5(E), and 5(F), hydrogen bond interaction exists between two proteins. Noteworthy, if one amino acid substitution occurring in the 5Å range, the hydrogen bond interaction within such range may also change, which may further alter the binding affinity of two proteins. Especially when a strong interaction exists between the mutant residue in human ACE2 protein and residues in the coronavirus spike protein. In the next part, we would explore the hydrogen bond differences between the wild type ACE2 and the mutant ACE2, in the complex protein.
Figure 4. Cartoon displayed models of the complex structures, corresponding to six missense variants in human ACE2. The partially enlarged view in the dashed boxes is the residues around 5Å of the mutant site, along with the hydrogen bonds among atoms. Note: the human ACE2 protein is in green and the RBD is in cyan.

Compared with the wild type complex protein, whether the hydrogen interaction of the complex protein alters, after one amino acid substitution occurring in the human ACE2 protein? In order to verify such a question, the following experiments were implemented. Briefly, we initially collected the predicted complex structures provided by SSIPe [15] web-server and sought out fourteen interacted amino acid pairs between ACE2 and the spike protein reported in Ref. [5]. Thereafter, we utilized the discovery studio 3.5 visualizer [16] to monitor the hydrogen bond interaction of wild-type, as well as the mutant complex protein. To facilitate observation, herein, we labeled “√” if two residues have an intermolecular H-bond. Conversely, marked “×” if no H-bond existing. Eventually, the results are listed in Table 1. By observing the results in Table 1, we could get the following conclusions.

(1) In terms of intermolecular hydrogen bond interaction, we can find several conserved interaction pairs. For example, residues D30, Q42, S19, K353, and Y41 of human ACE2 protein, interacts with residues Y417, Y449, A475, Q498, G502, T500 (with Y41of ACE2), and N501 (also with Y41of ACE2) of the spike protein.

(2) As for residue D38 in wild-type ACE2, it has intermolecular H-bond interactions with
the residue Q449 of the spike protein. However, this interaction is absent in all six variants. Furthermore, after residue D38 mutants to E, it adds interaction with the residue Q498 of spike protein, whereas the wild-type, as well as other five variants lack such interaction. Notably, the binding affinity value of D38E predicted by SSIPe [15] is -0.6 kcal/mol\(^1\), which is the smallest among all the thirty variants. By summarizing the above discussions, it is believed that variant D38E makes human ACE2 protein easier bind with the coronavirus.

(3) The residue K353 of wild-type ACE2 established an intermolecular hydrogen bond interaction with the residue G502 of the spike protein. Meanwhile, as for the other six variants, such interaction also exists. Moreover, five of six variants (except variant K353H) add an interaction between K353 (residue of ACE2) and G496 (residue of the coronavirus). That is, the added interaction enhances the ACE2 binding interaction with the spike protein [21].

According to the previous related research, residue K353 of human ACE2 protein interacts with both spike proteins of SARS-CoV, as well as that of the coronavirus [5]. Noteworthy here, the binding affinity value of K353H predicted by SSIPe [15] is -0.46 kcal/mol\(^1\), which is the second smallest value among all the thirty variants. By summarizing, we believed that variant K353H in human ACE2 protein could enhance its binding with the spike protein, thereby increasing the COVID-19 susceptibility.

**Table 1.** Intermolecular hydrogen bonds in six missense variants and wild type of the complex proteins.

| Residue in SARA-CoV-2 spike protein | Residue in human ACE2 | WT(6LZG) | D38E | M82I | Y83F | K353H | R357H | R357S |
|------------------------------------|-----------------------|----------|------|------|------|-------|-------|-------|
| K417                               | D30                   | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| Y449                               | D38                   | ✓        | x    | x    | x    | x     | x     | x     |
| Q498                               | D38                   | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| Y449                               | Q42                   | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| A475                               | S19                   | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| N487                               | Y83                   | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| N487                               | Q24                   | ✓        | x    | x    | x    | ✓     | ✓     | ✓     |
| Q493                               | E35                   | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| G502                               | K353                  | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| G496                               | K353                  | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| Q498                               | Q42                   | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| T500                               | Y41                   | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| N501                               | Y41                   | ✓        | ✓    | ✓    | ✓    | ✓     | ✓     | ✓     |
| N487                               | Y83                   | ✓        | x    | x    | x    | x     | x     | x     |

(4) The intermolecular hydrogen between residue N487 in spike protein and residue Q24 in ACE2 exists only in the wild type and variant R357H, which may indicate that the interaction between K353 and Q24 is not critical for the binding of the two proteins.

**4. Conclusions**

In order to reveal whether variants in human ACE2 protein can lead to the COVID-19 susceptibility, or inhibit the coronavirus spike protein binding, the changes of binding affinity of ACE2 and spike protein are explored in this work. Briefly, by observing the predicted protein affinity results by SSIPe [15], we noticed that the aforementioned six missense variants (D38E, M82I, Y83F, K353H, R357A, and R357S) were predicted with negative binding affinity score. That is, after the amino acid substitution, these six missense variants were predicted to increase the binding of ACE2 with the coronavirus and the human ACE2 receptor, especially, such performance of missense variants D38E and K353H are more obvious.

According to the recent news, scientific reports, and the clinical manifestation, it is known that: the COVID-19 recovery rate is different between ages, nationalities, medical history, and others [29-32]. For example, elder people are more susceptible to COVID-19, whereas the younger people are less susceptible [30-32]. As for the negative prognosis, it may be due to the existence of ACE2 missense variants in elder people [3], such as D38E, M82I, Y83F,
K353H, R357A, and R357S. Due to the fact that in the biological evolutionary procedure, protein-protein interaction is dynamic. Thus, it is unfair to draw the conclusion for human ACE2 variants deleterious explanation, from simple observations of the changes in the binding affinity of two proteins [3]. However, we believe that our experimental results can provide evidence that the potential relationship between COVID-19 susceptibility and ACE2 genetic variants.

Acknowledgments
This work was supported by the Natural Science Foundation of Anhui Province of China (KJ2018A0572).

References
[1] COVID-19 Coronavirus Pandemic Available: https://www.worldometers.info/coronavirus/
[2] Weisblum Y et al 2020 bioRxiv, 07.21.214759.
[3] MacGowan S A and Barton GJ 2020 BioRxiv, 05.03.074781.
[4] Yan R, Zhang Y, Li Y, Xia L, Guo Y and Zhou Q 2020 Science, 367(6485) 1444-1448.
[5] Hussain M, Jabeen N, Raza F, Shabbir S, Baig A A, Amanullah A and Aziz B 2020 Journal of Medical Virology, 92(9) 1580-1586
[6] Cui J, Li F and Shi Z-L 2019 Nature Reviews Microbiology, 17(3) 181-192
[7] Liu Z et al 2020 Journal of Medical Virology, 92(6) 595-601
[8] Wu A et al 2020 Cell Host & Microbe, 27(3) 325-328
[9] Li H, Zhou Y, Zhang M, Wang H, Zhao Q and Liu J 2020 Military Medical Research, 7(1) 11
[10] Al-Tawfiq JA, Al-Homoud AH and Memish Z A 2020 Travel Medicine and Infectious Disease, 34 101615
[11] Chen G et al 2020 medRxiv, 130(5) 2620-2629
[12] Huang X, Zheng W, Pearce R and Zhang Y 2020 Bioinformatics, 36(8) 2429-2437
[13] Visualizer DS 2012 Accelrys Inc, San Diego, CA, USA
[14] Wu A et al 2020 Cell Host & Microbe, 27(3) 325-328
[15] Lu X et al 2017 Journal of Medical Virology, 89(3) 542-545
[16] Lan J et al 2020 Nature, 581(7807) 215-220
[17] Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C L, Abiona O, Graham B S and McLellan J S 2020 Science, 367(6483) 1260-1263
[18] Li F, Li W, Farzan M and Harrison SC 2005 Science, 309(5742) 1864-1868
[19] Lu G et al 2013 Nature, 500(7461) 227-231
[20] Letko M and Munster V 2020 BioRxiv, 01.22.915660.
[21] Li F 2015 Journal of Virology, 89(4) 1954-1964
[22] Bairoch A et al 2005 Nucleic Acids Research, 33(suppl_1) D154-D159
[23] Tate JG et al 2019 Nucleic Acids Research, 47(D1) D941-D947
[24] Cao Y et al 2020 Cell discovery, 6(1) 1-4
[25] Li W et al 2005 The EMBO journal, 24(8) 1634-1643
[26] Zhou F et al 2020 The lancet, 395(10229) 1054-1062
[27] Rothan HA and Byrareddy SN 2020 Journal of Autoimmunity, 109 102433.
[28] Liu K, Chen Y, Lin R and Han K 2020 Journal of Infection, 80(6) e14-e18.
[29] Zheng Y Y, Ma Y T, Zhang J Y and Xie X 2020 Nature Reviews Cardiology, 17(5) 259-260