Functional Variants in Human ACE2 Can Decrease its Protein Stability and May Influence the Binding with SARS-CoV-2

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@njjust.edu.cn

Abstract. As for the transmission of human to human, the recent pandemic of COVID-19 is now spreading among the population, which has already led to 1.86 million deaths. As recently reported, it is the SARS-CoV-2 that causes COVID-19. Furthermore, human ACE2 is the receptor of the coronavirus. Nowadays, it has been reported that missense variants in ACE2 may affect the coronavirus susceptibility. In this work, we explored the functional impacts of missense variants in human ACE2. Briefly, we initially collected the variant in human ACE2, which have been labeled as important sites of ACE2 or the critical binding sites with the spike protein. Thereafter, applying the technology of protein structure homology modeling, we constructed the molecular spatial structure models of the variants. Next, variants molecular models of ACE2 were superimposed over the wild type of ACE2, to observe the structural changes. As experimental results demonstrated, the overall structures of ACE2 variants are similar. However, several variants (i.e., G173S, V184A, I233F, D355N, R357A, R357S, and G575V) in ACE2 are predicted to decrease the stability of human ACE2 protein and/or to be harmful to human health. Accordingly, the final findings could also provide a functional and structural basis for the potential pathogenicity of ACE2-driven viral infections.

1. Introduction

COVID-19 is now outbreaks worldwide, which is reported to be caused by the virus SARS-CoV-2 [1]. Compared to outbreaks caused by the earlier two strains (i.e., MERS-CoV and SARS-CoV) [2], COVID-19 is determined as a pandemic, which indicates the widespread inflection. As reported, until 30 November 2020, COVID-19 has transmitted around the globe with more than 86.10 million populations all over the world being infected, more than 1.86 million people have lost their lives [3].

Differing from the earlier strain of coronavirus, SARS-CoV-2 can infect humans [3]. Until now, scientific research reports have revealed that SARS-CoV-2 invades host human cells through an endocytic pathway, which begins with its spike protein interacting with human ACE2 protein [4]. As reported, it is believed that variants in virus spike protein can cause the transmission of cross-species [5]. Conversely, variants in ACE2 may result in susceptibility or resistance against the coronavirus infection [2]. That is, genetic variants in ACE2 may affect the broad spectrum severity coronavirus infection and leading the rapid spreading of COVID-19. As reported, SARS-CoV-2 has transmitted from human-to-human. Meanwhile, many research studies have been implemented to explore the impact of variants both in ACE2 and in the coronavirus [6-8]. Depending on such researches, it has been demonstrated that...
variants of human ACE2 and the coronavirus may be one kind of barrier for the inflecting of viruses across species. Moreover, variants in the coronavirus may lead to the widespread infection of COVID-19 and increasing its pathogenicity [2, 7, 8]. As known, in the biological evolutionary procedure, it is generally believed that a benign variant is a choice for species to adapt to the environment. Considering such theory, herein, one interesting hypothesis is proposed. Whether genetic variants in human ACE2 are deleterious to human health? And thereby affect its binding with the spike protein.

In this work, we mainly explored the aforementioned hypothesis and tried to explain its meaningfulness to human health. Briefly, by applying several protein sequences and structure analysis tools, we explored the functional impacts of ACE2 variants on human health, predicted variant influence of human ACE2 interacting with the spike protein. Eventually, the final findings could also provide a functional and structural basis for the potential pathogenicity of ACE2-driven viral infections.

2. Methodology

2.1. Missense Variants Collecting for Human ACE2

Figure 1(A) depicts the functional domains of the coronavirus spike protein with 1273 amino acids, which includes S1 and S2 subunits [6, 9]. As reported, it is the RBM (receptor-binding motif) in RBD (receptor-binding domain) that binds directly with human ACE2 [4, 8, 10, 11].

In this work, we collected the protein sequence of human ACE2 (i.e., Q9BYF1) from the UniPort database [12]. Q9BYF1 has 805 amino acids and mainly comprises two functional domains: peptidase M2 domain in N-terminal (residues 19-611) and collectrin domain in C-terminal (residues 612-740) [5]. As for the 3D structure corresponding to the human ACE2 protein sequence, we searched the RCSB PDB (Protein Data Bank) database [13] for its structural information. In RCSB PDB, the partial structure of human ACE2 protein has been resolved, such as PDBid: 1R42, 2AJF, 3D0G, 3SCK, 6MID, and others [14]. Among these structures, we selected 2AJF-A (chain A of 2AJF) as the wild type 3D structure of human ACE2, which contains residues 19~615 of ACE2. Notably, there is a peptidase M2 domain in 2AJF-A, as shown in Figure 1(B).

**Figure 1** Structural diagrams of the spike protein of SARS-CoV-2 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. Note: Figure 1(A) is adapted from Ref [6].)

In this work, we first collected missense variants in human ACE2 from the Uniport database [12], COSMIC database [15], and other recent research papers, such as Ref [5], Ref [4], and Ref [16]. Thereafter, appropriate selection criteria were employed to retain the missense variants located in the encoding regions. Noteworthy here, the aforementioned missense variants in human ACE2 that reported as binding sites and/or marked as important functional sites are mainly considered. Eventually, we obtained thirty genetic missense variants, which are listed in Table 1.

2.2. Impact Analysis of Missense Variants in Human ACE2

In this work, the stability and polymorphism of human ACE2 protein missense variants were predicted to explore the hypothesis aforementioned in Section 1. In terms of stability, we utilized the SDM (site-directed mutant) [17] and STRUM (structure mutation) [18] software to predict the stability of human ACE2 protein, after amino acid substitution. Briefly,
PolyPhen (i.e., R357A, R357S, M383I, I468V, A501T, and G575V) are predicted as possibly damaging by PredictSNP. Furthermore, among the aforementioned variants, four variants (i.e., M82N, M82T, Y83F) predicted with the positive ΔΔG value means that the new encoded ACE2 protein corresponding to each variant becomes destabilizing and stabilizing, respectively [17]. Precisely, corresponding to one missense variant, the positive ΔΔG value indicates a decrease in protein stability, whereas the negative ΔΔG value means an increase in the new encoded protein stability. As for the polymorphism of variants, we applied three prediction tools, i.e., polymorphism phenotyping-v2 (PolyPhen-2) [19], protein variation effect analyzer (PROVEAN) [20], and predict single nucleotide polymorphism (PredictSNP) [21], to predict the functional impact. The results of the stability and polymorphism of human ACE2 protein are listed in Table 1.

2.3. Protein Molecular Modeling

The thirty coding missense variants of human ACE2 were collected from the open database (such as Uniport database [12] and COSMIC database [13]) and the published literatures (i.e., Ref [5], Ref [4], and Ref [16]). In this work, in order to verify whether the thirty variants in human ACE2 can change protein structure and thereby affect its interacting with the spike protein, we implemented the following experiments. Briefly, we generated the structural models of thirty variants by applying Modeller 9.16 software [22]. Notably, in the generating procedure, we utilized the atomic coordinates of PDBid: 2AJF-A as the template for modeling. Thereafter, each model of variants was superimposed over the template model. In order to quantitatively measure the positional difference between the two models, we applied the SWISS-PdBSviewer 4.1.0 [23] software to estimate the root mean square deviation (RMSD) in the backbone, which is listed in the last column of Table 1.

3. Results and Discussions

3.1. Stability and Polymorphism Prediction of Human ACE2 Protein Missense Variants

As known, evolution is dynamic and continuous, that is, a benign variant in one protein is a choice for species to adapt to the environment in the biological evolutionary procedure, which manifests as a stronger resistance and immune response against the virus inflection [5]. As aforementioned in Section 1, the hypothesis we proposed: whether the genetic variants of ACE2 are deleterious to human health? Considering such a hypothesis, we estimated the potential pathological consequences caused by missense variants in human ACE2 protein from two aspects, i.e., protein stability and variant polymorphism.

As mentioned in Section 2.2, after the amino acid substitution, the stability and polymorphism of human ACE2 protein are predicted through five effective tools. Firstly, setting the temperature at 25°C and PH at 7, the SDM [17] and STRUM [18] software were applied to examine the stability of the new encoded human ACE2 protein. On the other hand, we utilized three polymorphism prediction tools (i.e., PolyPhen-2 [19], PROVEAN [20], and PredictSNP [21]) to predict the functional impact. The results of the stability and polymorphism of human ACE2 missense variants are documented in Table 1.

As shown in Table 1, in terms of protein stability, twelve variants (i.e., I21T, K31D, N33D, M82N, M82T, Y83F, P84S, P84A, P84Q, P84T, A501T, and G575V) in the SDM [17] returned file and nine variants (i.e., K26R, M82N, M82T, Y83F, D355A, R357A, M383I, I468V, and F592L) in the STRUM [18] returned file, are remarked to affect the new encoded human ACE2 protein stability. Furthermore, three missense variants (i.e., M82N, M82T, and Y83F) predicted with the positive ΔΔG value by both SDM [17] and STRUM [18]. Furthermore, among the aforementioned variants, four variants (i.e., D355A, R357A, M383I, and G575V) are predicted to be deleterious to human health by PredictSNP [21], PolyPhen-2 [19], and/or PROVEAN [20].

By observing the results from Table 1, in terms of the polymorphism after amino acid substitutions, thirteen variants (i.e., S19P, Y41A, G173S, V184A, I233F, D355A, D355N, R357A, R357S, M383I, I468V, A501T, and G575V) are predicted as possibly damaging by PolyPhen-2 [19], nine variants (i.e., Y41A, G173S, V184A, I233F, D355A, D355N, R357A,
R357S, and G575V) are predicted as deleterious by PROVEAN [20], and other nine variants (i.e., G173S, V184A I233F, D355A, D355N, R357A, R357S, M383I, and G575V) are predicted as deleterious by PredictSNP [21]. Notably, among the aforementioned variants, there are seven missense variants (i.e., G173S, V184A, I233F, D355N, R357A, R357S, and G575V) predicted as deleterious by three prediction tools at the same time.

Table 1. Prediction results of the stability and polymorphism of human ACE2 missense variants.

| dbsnp/ COSMICid | AAS       | Stability (∆ΔG) | Polymorphism       | RMSD |
|------------------|-----------|-----------------|--------------------|------|
| rs73635825;      | S19P      | 0.00 DS         | -1.14 IS          | PD   | -0.950 N | N | N | 0.149 |
| COSM63700085     | I21T      | 1.07 DS         | -0.49 IS          | B    | 0.936 N  | N | N | 0.119 |
| rs4646116       | K26R      | -0.43 IS        | 0.30 DS           | B    | -0.499 N | N | N | 0.249 |
| NA               | 0.02 DS   | -0.32 IS        | B                  | 1.257 N | N | N | 0.176 |
| NA               | 0.37 DS   | -1.44 IS        | B                  | -0.185 N | N | N | 0.124 |
| NA               | 0.02 DS   | -0.32 IS        | B                  | 1.257 N | N | N | 0.176 |
| NA               | 0.7 DS    | 0.07 DS         | B                  | 0.759 N  | N | N | 0.228 |
| rs76696587;      | M82I      | -0.11 IS        | 0.36 IS           | B    | 0.884 N  | N | N | 0.240 |
| COSM9066129      | M82T      | 1.19 DS         | 0.18 DS           | B    | 0.821 N  | N | N | 0.122 |
| NA               | 0.26 DS   | 0.47 DS         | B                  | 0.249 N  | N | N | 0.128 |
| NA               | 0.63 DS   | -0.10 IS        | B                  | -0.296 N | N | N | 0.137 |
| COSM5616273      | P84A      | 0.37 DS         | -0.44 IS          | B    | 1.700 N  | N | N | 0.167 |
| COSM6332640      | P84Q      | 0.80 DS         | -0.39 IS          | B    | 0.631 N  | N | N | 0.229 |
| rs759134032      | P84T      | 0.83 DS         | 0.39 DS           | B    | 0.642 N  | N | N | 0.188 |
| rs372352182;     | N149S     | -0.17 IS        | 0.57 IS           | B    | 0.103 N  | N | N | 0.180 |
| rs746034076;     | N159S     | -0.17 IS        | 0.67 IS           | B    | 0.905 N  | N | N | 0.225 |
| rs754511501;     | G173S     | -0.17 IS        | 0.27 IS           | PD   | 5.891 D  | D | D | 0.202 |
| rs758142853;     | V184A     | -0.55 IS        | 0.57 IS           | PD   | 3.705 D  | D | D | 0.234 |
| NA               | 0.37 IS   | 0.14 IS         | DS                 | 0.284 D  | D | D | 0.138 |
| NA               | 0.22 IS   | -0.04 IS        | IS                 | 1.941 N  | N | N | 0.133 |
| NA               | 0.14 DS   | 0.04 DS         | IS                 | 1.941 N  | N | N | 0.133 |
| NA               | 0.24 IS   | 0.12 IS         | IS                 | 4.463 D  | D | D | 0.206 |
| NA               | 0.22 IS   | -0.13 IS        | IS                 | 3.640 D  | D | D | 0.151 |
| NA               | 0.24 IS   | -0.12 IS        | IS                 | 4.463 D  | D | D | 0.206 |
| NA               | 0.22 IS   | -0.13 IS        | IS                 | 0.226 D  | D | D | 0.121 |
| NA               | 0.22 IS   | -0.12 IS        | IS                 | 0.226 D  | D | D | 0.121 |
| NA               | 0.06 DS   | -0.14 IS        | DS                 | 0.226 D  | D | D | 0.121 |
| NA               | 0.06 DS   | -0.14 IS        | DS                 | 0.226 D  | D | D | 0.121 |
| NA               | 0.96 IS   | -0.14 IS        | IS                 | 0.226 D  | D | D | 0.121 |
| NA               | 0.96 IS   | -0.14 IS        | IS                 | 0.226 D  | D | D | 0.121 |
| NA               | 3.00 IS   | -0.89 IS        | IS                 | 0.226 D  | D | D | 0.121 |
| NA               | 3.00 IS   | -0.89 IS        | IS                 | 0.226 D  | D | D | 0.121 |

Abbreviations: dbsnp, reference SNP ID; COSMICid, ID in COSMIC database; NA, not available; AAS, amino acid substitution; IS, increased stability; DS, decreased stability; PD, possibly damaging; B, benign; N, neutral; D, deleterious;

Up to now, there are some research reports of pathogenic consequences that are related to the aforementioned variants predicted as unstable and/or deleterious. The detail information is described as following. (1) As remarked in the Uniport database [12], missense variants Y41A and R357A are reported to strongly inhibit interaction with SARS-CoV spike glycoprotein [24]. Moreover, variants K31D, K353H, M82N, Y83F, and P84S are documented to abolish the interaction with SARS-CoV spike glycoprotein [24]. (2) As described in the "family & domains" part of the Uniport database [12], residue region 30-41 and 353-357 in human ACE2 are reported to have an interaction with SARS-CoV spike protein. Precisely, missense
variants Y41, K353, and R357 in ACE2 are quite important for binding with the spike protein in coronavirus [24]. Similar descriptions are also seen in the literature, for example, it has reported that residues near 31, 41, 82-84, 353, and 357 of ACE2 are critical for binding with the coronavirus spike protein [24, 25]. (3) In the Ref. [16], missense variants D38E, N149S, N159S, G173S, V184A, and I233F are reported occurring in different populations around the world, which have been identified from the “China Metabolic Analytics Project” and “1000 Genomes Project” database [26].

3.2. Protein Molecular Modeling for Human ACE2 Protein Missense Variants

We developed the molecular models for each variant for observing the structural differences of human ACE2 protein encoded by the missense variants. We then superimposed the developed models over the wild type human ACE2 protein model. Briefly, by applying the atomic coordinate of PDBid: 2AJF-A as a template model, we utilized the Modeller 9.16 software [22] to generate a 3D structural model for each variant. After modeling, each model of variants was superimposed over the template model, aiming to find out the differences in the structures. Meanwhile, we used the SWISS-PdBViewer 4.1.0 [23] to estimate the RMSD between the generated and template models in the backbone.

Figure 2. Structure of human ACE2 variants. Structural comparison of wild type (2AJF-A is shown in green) and 30 missense variants (in a different color) of human ACE2. Figure 2(A) and 2(B) are cartoon and ribbon representation. Figure 2(C) is the stick representation of thirty residues. Figure 2(D) is the enlarged region of the box in Figure 2(C).

As depicted in Figure 2(A) and 2(B), it is easily seen that no major and obvious structural differences exist between variants and the wild type of human ACE2. More specifically, compared with the wild type of ACE2, the overall structure of human ACE2 missense variants is largely conserved, with RMSD measure range from 0.11Å to 0.249Å. However, by observing the results from Figure 2(C) and 2(D), the micro-changes of the orientation of the mutant residue side chain (from the wild type residue to the mutant one) might alter intermolecular interaction with the residues in SARS-CoV-2. Thus, it is necessary to further explore the affinity changes of complex protein which comprises ACE2 and the spike protein.

4. Conclusions

It has been demonstrated that bioinformatics distinctive prediction tools, such as protein molecular modeling software and protein affinity prediction tools, can provide a significant insight of harmfulness variants in human proteins [27]. Built on the aforementioned insight, we utilized two protein-energy changes and three variants polymorphism analysis prediction software to assess the functional impact of thirty missense variants in human ACE2 protein.
By analyzing the experimental results, we found that as for those variants with "deleterious" or/and "instability" marks, most of them have no empirical research report that linked with human disease. Noteworthy here, as reported by scientific research, residues near 31, and 41, 82-84, 353, and 357 in ACE2 protein are critical for binding with the spike protein [24]. Considering such a report, we noticed missense variants S19P, K31D, Y41A, M82N, M82T, D355A, and D355N were predicted deleterious and decrease ACE2 stability.

Depending on the news reports, recent scientific research results, and the clinical manifestation, it is known that the recovery rate of COVID-19 is quite different [28-31]. For instance, as known, people over 60 years old are more susceptible to COVID-19, whereas young people are less susceptible [29-31]. As for the negative prognosis, it may be due to the existence of ACE2 variants [14], 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 protein stability, polymorphism, and observations of structure superimposing [14]. However, we believe that the final findings could also provide a functional and structural basis for the potential pathogenicity of ACE2-driven viral infections. In our future work, we will explore these thirty missense variants in human ACE2 protein that can affect the binding interaction with the spike protein of SARS-CoV-2.

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

References
[1] Weisblum Y et al 2020 bioRxiv, 07.21.214759.
[2] Cui J, Li F and Shi Z-L 2019 Nature Reviews Microbiology, 17(3) 181-192
[3] COVID-19 Coronavirus Pandemic [https://www.worldometers.info/coronavirus/]
[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] Liu Z et al 2020 Journal of Medical Virology, 92(6) 595-601
[7] Wan Y, Shang J, Graham R, Baric R S and Li F 2020 Journal of Virology, 94(7) e00127-20
[8] Lu G, Wang Q and Gao G F 2015 Trends in Microbiology, 23(8) 468-478
[9] Wrapp D, Wang N, Corbett K S, Goldsmith J A, Hsieh C L, Abiona O, Graham B S and McLellan J S 2020 Science, 367(6483) 1260-1263
[10] Letko M and Munster V 2020 BioRxiv, 01.22.915660.
[11] Li F 2015 Journal of Virology, 89(4) 1954-1964
[12] Bairoch A et al 2005 Nucleic Acids Research, 33(suppl_1) D154-D159
[13] Burley S K et al 2019 Nucleic Acids Research, 47(D1) D464-D474.
[14] MacGowan S A and Barton G J 2020 BioRxiv, 05.03.074781
[15] Tate J G et al 2019 Nucleic Acids Research, 47(D1) D941-D947
[16] Cao Y et al 2020 Cell discovery, 6(1) 1-4
[17] Worth C L, Preissner R and Blundell T L 2011 Nucleic Acids Research, 39(suppl_2) W215-W222
[18] Quan L, Lv Q and Zhang Y 2016 Bioinformatics, 32(19) 2936-2946
[19] Adzhubei I A, Schmidt S, Peshkin L, Ramensky V E, Gerasimova A, Bork P, Kondrashov A S and Sunyaev S R 2010 Nature Methods, 7(4) 248-249
[20] Choi Y and Chan A P 2015 Bioinformatics, 31(16) 2745-2747
[21] Bendel J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J and Damborsky J 2014 PLoS Computational Biology, 10(1) e1003440
[22] Eswar N, Webb B, Marti-Renom M A, Madhusudhan M, Eramian D, Shen My, Pieper U and Sali A 2006 Current protocols in bioinformatics, 15(1) 5.6.1-5.6.30
[23] Johansson M U, Zoete V, Michielin O and Guex N 2012 BMC Bioinformatics, 13(1) 173
[24] Li W et al 2005 The EMBO journal, 24(8) 1634-1643
[25] Hadi-Alijanvand H and Rouhani M 2020 *Journal of Proteome Research*, **19**(11) 4609-4623
[26] Consortium GP 2015 *Nature*, **526**(7571) 68-74
[27] Dong S, Sun J, Mao Z, Wang L, Lu Y L and Li J 2020 *Journal of Medical Virology*, **92**(9) 1542-1548
[28] Zhou F *et al* 2020 *The lancet*, **395**(10229) 1054-1062
[29] Rothan H A and Byrareddy S N 2020 *Journal of Autoimmunity*, **109** 102433
[30] Liu K, Chen Y, Lin R and Han K 2020 *Journal of Infection*, **80**(6) e14-e18
[31] Zheng Y Y, Ma Y T, Zhang J Y and Xie X 2020 *Nature Reviews Cardiology*, **17**(5) 259-260