In Silico Prediction of New Mutations That Can Improve the Binding Abilities Between 2019-nCoV Coronavirus and Human ACE2

Senbiao Fang, Ruoqian Zheng, Chuqi Lei, Jianxin Wang, Renyi Zhou, and Min Li

Abstract—The Coronavirus Disease 2019 (COVID-19) has become an international public health emergency, posing a serious threat to human health and safety around the world. The 2019-nCoV coronavirus spike protein was confirmed to be highly susceptible to various mutations, which can trigger apparent changes of virus transmission capacity and the pathogenic mechanism. In this article, the binding interface was obtained by analyzing the interaction modes between 2019-nCoV coronavirus and the human ACE2. Based on the "SIFT server" and the "bubble" identification mechanism, 9 amino acid sites were selected as potential mutation-sites from the 2019-nCoV-S1-ACE2 binding interface. Subsequently, a total number of 171 mutant systems for 9 mutation-sites were optimized for binding-pattern comparsion analysis, and 14 mutations that may improve the binding capacity of 2019-nCoV-S1 to ACE2 were selected. The Molecular Dynamic Simulations were conducted to calculate the binding free energies of all the 14 mutant systems. Finally, we found that most of the 14 mutations on the 2019-nCoV-S1 protein could enhance the binding ability between 2019-nCoV coronavirus and human ACE2. Among which, the binding capacities for G446R, Y449R and F486Y mutations could be increased by 20 percent, and that for S494R mutant increased even by 38.98 percent. We hope this research could provide significant help for the future epidemic detection, drug and vaccine development.

Index Terms—General, control structures and microprogramming, simulation, optimization

1 Introduction

Since the first case of Coronavirus Disease 2019 (COVID-19) was officially reported by the Chinese government on December 8, 2019, update to November 24, 2020, confirmed and dead cases have reached about 59.5 million and 1398 thousand, respectively. Now, COVID-19 has become a public health emergency triggering international concern. In reported cases, patients exhibited severe symptoms such as pneumonia, respiratory failure, acute respiratory distress syndrome (ARDS) and heart damage, which were all fatal complications of COVID-19, posing a vital threat to global health [1].

Various researches have been conducted for clarifying the pathogenesis of the COVID-19. The structures of coronavirus main protein (3CLpro), papain like protein (PLpro) and RNA dependent polymerase (RdRp) which were necessary for the translation and replication of RNA virus have been resolved [2], [3]. Based on these data, comprehensively in silico virtual screening and enzymologicla testing methods have been conducted, and the results indicated that drugs including Indinavir and lopinavir might be effective for COVID-19 [4], [5]. In addition, Remdesivir, an RNA-dependent polymerase inhibitor, which can inhibit the synthesis of viral nucleic acids and fight against virus showed certain potential in the treatment of the first novel coronavirus patient in the United States [6], [7]. In addition, on March 21, 2020, the American Food and Drug administration (FDA) speeded up the approval process of Remdesivir and Hydroxychloroquine Sulfate [8], [9], [10] as new treatment drugs for COVID-19. However, few effective vaccines have been developped for COVID-19 as current treatment methods including monitoring and maintaining patient function and providing effective oxygen therapy were just in a basic manner, focusing on treating symptoms, treating underlying diseases, preventing complications and secondary infections [11], [12], [13], [14].

Mutations were characterized with randomness, low frequency and reversibility in both eukaryotes and prokaryotes [15]. For many proteins, pathogenic mutations tend to occur in important structural and functional regions, and most polymorphisms are located in these functional areas. After 2019-nCoV enters into the human body, the viral spike glycoproteins (Spike proteins) on the host cell surface bind to the angiotensin-converting enzyme 2 (ACE2), and then the viral RNA invades into healthy human cells [16], [17], as shown in Fig. 1. Thus, the Spike proteins on cornoavirus play vital roles in virus replications [18]. In this article, the structural region of 2019-nCoV-S1 on Spike proteins which directly participate in the interactions with ACE2 were picked out for further analysis.

Viruses mutate very quickly, and mutations accumulate during the process of transmission from one individual to

- *The authors are with the Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410083, China. E-mail: {fangsb, zhengruoqian, jxwang, limin}@mail.csu.edu.cn, chuqilei227@gmail.com, zhourenyi@csu.edu.cn.*

Manuscript received 19 Aug. 2020; revised 27 Jan. 2021; accepted 6 Feb. 2021. Date of publication 9 Feb. 2021; date of current version 3 June 2022. (Corresponding author: Min Li.)
Digital Object Identifier no. 10.1109/TCBB.2021.3058265

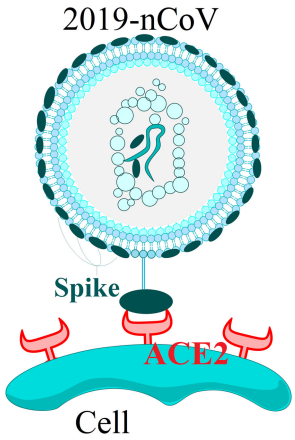


Fig. 1. Schematic diagram of 2019-nCoV virus binding to the human body.

another. On March 3, 2020, it has been reported that 149 mutations were found in the new 2019-novel coronavirus (2019-nCoV), among which the S-type mutation site is the most infectious [19]. Until March 21, 2020, one patient from Iceland was detected with two co-existing mutated types of coronaviruses, one of them was the novel mutated 2019-nCoV [20]. One study suggested that 2019-nCoV could increase its infectivity through the receptor binding domain recombination and a cleavage site insertion [21]. Intrapersonal variations and 6 different mutations in the spike glycoprotein (S protein) were observed from 11 patient-derived viral isolates which showed significant variation in cytopathic effects and viral load, up to 270-fold differences, when infecting Vero-E6 cells [22]. Therefore, the protein 2019-nCoV-S1 possesses mutation characteristics which can significantly affect the biological functions of the coronavirus. Thus, a comprehensive analysis of the coronavirus mutation characteristics will help for both understanding the binding and pathogenic mechanism and providing insights for subsequent vaccine research and development.

In this research, we first obtained the binding modes between the 2019-nCoV-S1 region and angiotensin-converting enzyme 2(ACE2) based on the 2019-nCoV-ACE2 crystal structure [23]. The binding interface between the coronavirus RBD-S1 region and the human ACE2 was obtained by analyzing existing crystal structure. Subsequently, tolerable mutations of all amino acid sites forming the binding interface between 2019-nCoV-S1 and human ACE2 were predicted, and 9 amino acids were found to be more susceptible to mutations. Finally, 171 single-point mutations were obtained by site-directed mutagenesis in the 9 amino acid sites. From qualitative analysis of the physical and chemical properties of the regions around the mutation sites, 14 mutations were found out for subsequent molecular dynamics simulations. Through comparative analysis of the binding free energies between each mutant and the wild-type complex, 4 mutations on 2019-nCoV-S1 region that may significantly improve the binding capacity of coronavirus to human ACE2 and finally trigger the stronger pathogenic.

2 Results

2.1 2019-nCoV-S1/ACE2 Complex Structure and Binding Pattern

Here, we used the crystal structure with high resolution (2.45 Å) from the RCSB PDB crystal database (PDB ID: 6M17) https://www3.rcsb.org/structure/6M17. When ACE2 binds with 2019-nCoV-S1, as shown in Figs. 2A and 2B, a total of 5 regions on ACE2 are involved and directly forming the interface: S19-T52, S77-Y83, M323-S331, G352-I358 and A386-R393.

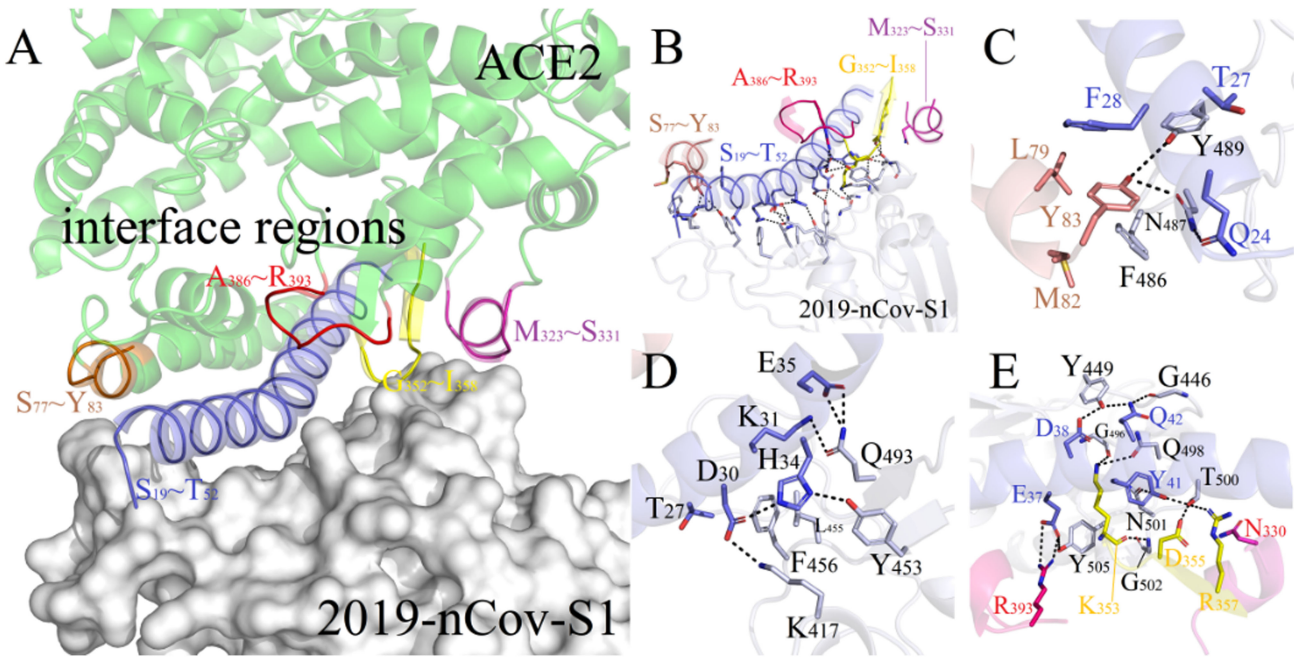


Fig. 2. Binding modes analysis between ACE2 and 2019-nCoV-S1. (A) Binding interface. The 2019-nCoV-S1 protein was shown as white surface. The ACE2 protein was colored by green, and binding interface regions were in different colors (S19∼T52: blue; S77∼Y83: brown; M323∼S331: maroon; G352∼I358: yellow and A386∼R393: red); (B) Key regions of the binding interface; (C) Binding modes of residues on S77∼Y83 region with ACE2; (D) Binding modes of residues on S19∼T52 region with ACE2; (E) Binding modes of residues on regions S77∼Y83, M323-S331, G352-I358, and A386-R393 with ACE2.

TABLE 1
RMSD Value Difference of Heavy Atoms on Resiudes Within 3.5Å From Mutated-Site Compared with Crystal Structure

	Positive			Negative		Pollar uncharged				Hydrophobic short sidechain						Hydrophobic long sidechain				
Mutations	R	H	K	D	E	S	T	N	Q	C	G	P	A	V	I	L	M	F	Y	W
G0446	0.201	0.147	0.197	0.141	0.171	0.170	0.134	0.198	0.206	0.146	-	0.148	0.136	0.151	0.169	0.134	0.152	0.216	0.157	0.165
Y449	0.532	0.518	0.578	0.561	0.566	0.567	0.569	0.561	0.557	0.569	0.571	0.573	0.568	0.566	0.570	0.561	0.576	0.588	-	0.529
F486	0.703	0.703	0.716	0.695	0.709	0.697	0.674	0.688	0.789	0.733	0.680	0.700	0.744	0718	0.657	0.680	0.662	-	0.653	0.635
N487	0.638	0.613	0.631	0.601	0.635	0.577	0.511	-	0.559	0.599	0.612	0.592	0.610	0.631	0.603	0.626	0.660	0.637	0597	0.627
Y489	0.534	0.528	0.535	0.519	0.534	0.512	0.519	0.529	0.519	0.525	0.522	0.527	0.514	0.509	0.510	0.531	0.520	0.533	-	0.506
Q493	0.655	0.637	0.651	0.635	0.637	0.632	0.646	0.631	-	0.639	0.614	0.633	0.633	0.639	0.634	0.639	0.637	0.619	0.633	0.628
S494	0.420	0.497	0.444	0.388	0.332	-	0.381	0.336	0.321	0.394	0.320	0.332	0.365	0.387	0.335	0.325	0.358	0.378	0.300	0.380
Y495	0.527	0.536	0.551	0.543	0.553	0.513	0.517	0.516	0.519	0.512	0.519	0.519	0.526	0.520	0.509	0.537	0.540	0.538	-	0.522
G496	0.666	0.629	0.660	0.666	0.662	0.619	0.618	0.636	0.644	0.633	-	0.611	0.612	0.619	0.636	0.629	0.617	0.627	0.625	0.625

In this complex system, there are a total of 16 amino acids in the 2019-nCoV-S1 region that directly act on ACE2: K417, G446, Y449, Y453, L455, F456, F486, N487, Y489, Q493, G496, Q498, T500, N501, G502 and Y505. However, a total of 19 amino acids are involved in the direct interactions with ACE2: Q24, T27, F28, D30, K31, H34, E35, E37, D38, Y41, Q42, L79, M82, Y83, N330, K353, D355, R357 and R393.

A dense interaction system is formed between the 2019-nCoV-S1 region and ACE2, as shown in Fig. 2B. Out of all the interactions, polar amino acids contribute the most of the energy through salt bridging and hydrogen bonding. Among the amino acids, K417 and adjacent D30 form a strong salt bridge. H34 forms a salt bridge with the nearest amino acid, D30 (Fig. 2D). A strong hydrogen bond network is formed between the 2019-nCoV-S1 region and ACE2, among which the typical amino acid is Q493, which could simultaneously form hydrogen bonds with K31 and E35 (as shown in Fig. 2D). N487 forms hydrogen bonds with surrounding amino acids Y83 and Q24 (Fig. 2C). For T450, although the side chain is short, the hydroxy-OH group at the end of the side chain provides a potential polar interaction with the adjacent amino acids R357, D355 and Y41, enabling T450 to form hydrogen bonds with all three amino acids. Consequently, the local binding structure is relatively tight (as shown in Fig. 2E).

In addition to the above polar effects, the proteins are equipped with good hydrophobic interactions. Among them, the typical representative was F486, which forms an excellent π-π aromatic stacking effect between the benzene ring of the amino acid side chain and the benzene ring of Y83 (as shown in Fig. 2C), stabilizing the microenvironment in this region to achieve a stable structural configuration. In addition, D30 (side chain carbon and backbone hydrophobic fragment), H34 (side chain flat ring) and K31 (side chain and backbone hydrophobic) on ACE2 form a local hydrophobic binding pocket which made hydrophobic effects with F456 on 2019-nCoV-S1 (Fig. 2D).

2.2 Prediction of Mutations in 2019-nCoV-S1

Site-directed mutations of specific amino acids in proteins are always accompanied with changes in biological functions. First, the mutation for 2019-nCoV-S1 protein was predicted by utilizing the online server Sorting Intolerant From Tolerant [24]. According to the prediction results (as shown in Table 1), 87 amino acids in the region of S438∼G527 on the 2019-nCoV-S1 with potential mutation possibilities were picked for further analysis. Among these mutations, 21 amino acids, distributing within V445∼Y449, V473∼C480 and V483∼F490 regions, with SIFT score less than 0.2 indicating a relatively higher mutation possibility. Coincidentally, 5 of these 21 amino acids (G446, Y449, F486, N487 and Y489) are key amino sites that directly interact with human ACE2, as shown in Fig. 3.

In addition, the good binding modes were observed at the interface when 2019-nCoV-S1 bound to human ACE2, as shown in Fig. 3B, which may directly explain the strong binding affinity of the coronavirus to the human body. In addition, one "bubble" area was found within the binding interface, indicating an extra geometric space for resiudes filling for binding affinity enhancement. If one mutation on 2019-nCoV-S1 protein occured to fill the gap in the "bubble" region, significant improvement of binding affinity might

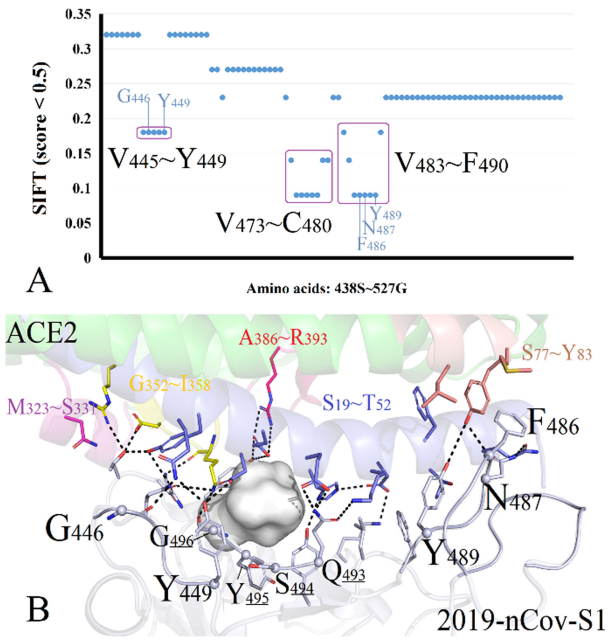


Fig. 3. Amino acdis predicted by SIFT and cavity visualization methods. (A) Amino acids substitution predicted by webserver SIFT (score $<$ 0.5) that affects the protein function. Residues (G446, Y449, F486, N487 and Y489) co-participated in forming binding interface were marked. (B) 3D diagram of 2019-nCoV-S1-ACE2 complex. Residues forming binding interface were shown as sticks. Cavity between 2019-nCoV-S1 and ACE2 providing opportunity for other amino acids (Q493, S494, Y495 and G496) insertion were shown as white surface.

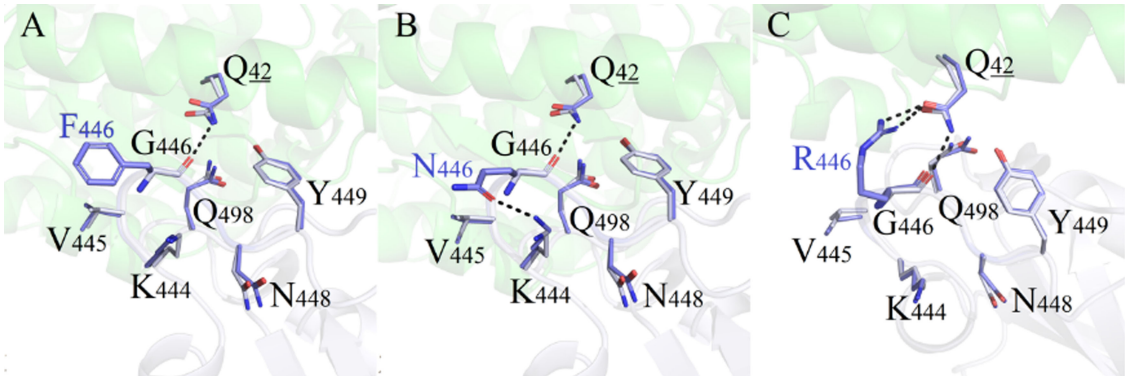


Fig. 4. Binding patterns comparison of G446 series mutant systems. 2019-nCoV-S1: white, sticks; ACE2 protein: green, sticks, underlined. The Q42 amino acid was located on the ACE2 protein, and other residues were located on the 2019-nCoV-S1 protein.

be detected. After careful comparasion, 4 crucial amino acids Q493, S494, Y495 and G496 adjacent to the "bubbles" in the 2019-nCoV-S1 were picked out.

As mentioned above, 5 amino acids (G446, Y449, F486, N487 and Y489) were predicted by the Sorting Intolerant From Tolerant server and 4 residues (Q493, S494, Y495 and G496) that compose the "bubble" within 2019-nCoV-S1-ACE2 binding interface were selected for subsequent mutation analysis.

2.3 Construction of Mutant Systems

In this study, we performed site-directed mutation of the 9 amino acid sites (G446, Y449, F486, N487, Y489, Q493, S494, Y495 and G496) based on the crystal structure by applying Pymol software [25] and obtained a total of 171 mutants.

2.4 MD Simulation Refinement of Mutant Systems

Subsequently, we optimized the structure of each mutant system with the software Amber16 [26]. In the process of molecular dynamics, we applied a 1500kcal/mol position limiting force to all of the heavy atoms of the protein, and the optimization time of each system was at least 1ns. Based on the 1ns molecular dynamics optimization results, the average complex structure on the 1ns trajectory was taken, and the structural changes before and after the mutation were analyzed.

When comparing the structural changes, the changes in the root mean square deviation (RMSD) [27] for the whole structure of the protein, especially the RMSD of all heavy atoms of amino acids within 5 Å of the mutation site, were primarily analyzed. The formula of RMSD was shown as in following formula (1):

$$RMSD(v,w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|v_i - w_i\|^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((v_{ix}-w_{ix})^2) + (v_{iy}-w_{iy})^2 + (v_{iz}-w_{iz})^2)}. \quad (1)$$

A large value of RMSD indicated that it will be put into further screening and analysis. For the amino acid sites conforming to the "mutation principle", molecular dynamics optimization simulation was carried out with an optimization time at least 6ns, and the average structure was used for in-depth analysis.

It can be seen from Table 1 that among the 171 mutant systems of selected 9 mutable sites on the 2019-nCoV-S1 protein, 54 percent of the mutant systems could cause regional structural changes near the mutation sites. Next, we analyze the 171 mutation sites one by one to find novel mutation system that may cause greater binding capacity of 2019-nCoV-S1 to the human body.

2.5 Comparative Analysis of 171 Mutants

2.5.1 *Comparative Analysis of G446 Mutants*

As shown in Fig. 4, the mutation site G446 is located in the peripheral region of the 2019-nCoV-S1/ACE2 binding interface, thus tiny conformational differences were observed for all complexes. Among 19 mutant systems, the RMSD values of amino acids within 5Å from the mutation site were less than 0.216. 3 systems (G446F, G446Q and G446R) presenting greater difference (>0.200) compared with the wild-type protein were selected for detailed analysis.

For mutant G446F, the sidechain on residue F446 forms hydrophobic interactions with surrounding residues including V445 (Fig. 4A). One new hydrogen bond is formed between K444 and N446 in mutant system G446Q (Fig. 4B). However, both the conformations present slight perturbations as the binding patterns between 2019-nCoV-S1 and ACE2 is not affected.

Regional structure is disturbed after mutation, as the mutated amino acid R446 form a new type of hydrogen bond with Q42, which is stronger than that of the wild-type weak hydrogen bond interaction between G446 and Q42 (Fig. 4C).

2.5.2 *Comparative Analysis of the Y449 Mutants*

This mutation site is located in the core region of the binding interface, and corresponding mutations are more likely to experience structural perturbations in terms of protein structure. Amino acid Y449 of 2019-nCoV-S1 forms good hydrogen bonds with neighboring Q42 and D38 on ACE2 (Fig. 5A). Thus the structural stability of mutants would be significantly affected by residues Q42 and D38.

The hydrophobic residues CGPAVILMF (Fig. 5B/C), noncharged polar amino acids STNQ and negatively charged residues DE (Fig. 5D) all could not compensate for the polarities of the hydrogen bonds between Y449 and Q42/D38. As a result, with the disappearance of wild-type polar interactions,

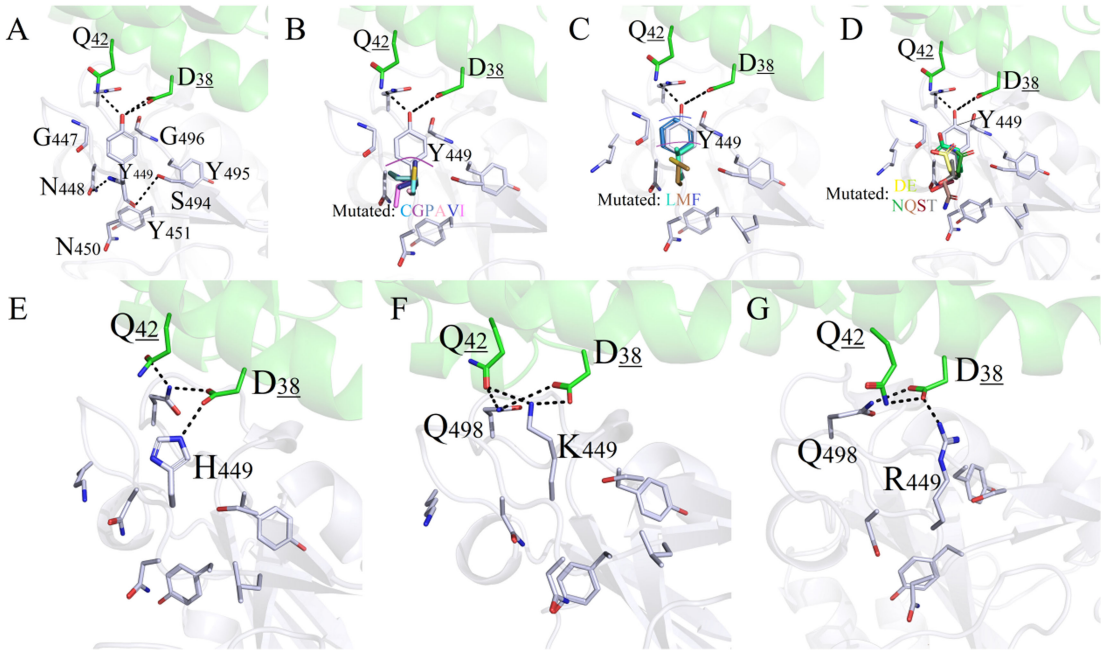


Fig. 5. Binding pattern comparisons of Y449 series mutants. (A) Interaction modes between Y449 and surrounding amino acids; (B) Binding modes comparsions between Y449 mutants (Y449C/G/P/A/V/I) and wild-type protein; (C) Binding interactions of Y449 mutant series (Y449L/M/F) and wild-type complex; (D) Binding interactions of the wild-type and Y449 mutants (Y449D/E/N/Q/S/T); (E) Interaction pattern of the wild-type and mutant Y449H; (F) Interaction pattern of the wild-type and mutant Y449K; (G) Binding modes of the wild-type and mutant Y449R.

the conformation of each mutant and corresponding RMSD value had experienced obvious variation.

The positive charged amino acids HKR are unable to form hydrogen bonds between Y449-Q42/D38, but could form new salt-bridge interactions with residue D38 on ACE2. In particular, mutated residues K and R could induce the positional deflection of Q42, and then the new polar contacts network (H-bonds and salt-bridges) are formed (Figs. 5E, 5F, 5G). These 3 mutants (Y449R/K/H) were selected for subsequent analysis.

2.5.3 Comparative Analysis of the F486 Mutants

For 19 mutants, the RMSD values fluctuated significantly from 0.635 to 0.789 as the mutation sites located at the core region of protein-protein binding interface. The physicochemical properties surrounding the mutated site F486 were analyzed as follows (Fig. 6A):

1) The upper part surrounded by amino acids L79, M82, Y83, P84, I88, I21, A25, etc. formed one hydrophobic binding region.
2) The bottom region contains polar amino acids N487 and Q24. However, one reasonable non-charged plane was formed between N487/Q24 themselves and Y83/Y489 residues, and no extra electric charge could be provided outwards to generate polarity with other residues.

Considering the binding characteristics of this regional pocket: structure of amino acids with opposite charge characteristics (positively charged RHK, negatively charged DE and polar non-charged STNQ) had experienced significantly swings as they could not match the original physical and chemical characteristics in wild-type complex(Fig. 6B).

In addition, due to the shorter sidechains of hydrophobic amino acids (CGPAVILM) compared with that of F486, the original geometric configuration of the wild-type system could not be complemented, which lead to the collapse of this binding region. All these elements results in the significant structural changes and weakening of the binding affinity (Fig. 6C).

Hydrophobic amino acids (W486 and Y486) with longer sidechains not only can partially fill the cavity excavated by the wild-type amino acid F486, but also induce newly physical and chemical properties with surrounding residues (Fig. 6D). Therefore, mutants F486W and F486Y were picked out for subsequent analysis.

2.5.4 Comparative Analysis of the Q493 Mutants

The amino acid Q493 of 2019-nCoV-S1 could form directly hydrogen bonds with residues E35 and K31 on ACE2 protein (Fig. 7A). Therefore, mutations could trigger local environmental changes and significant fluctuations of RMSD values which were greater than 0.6.

The side chains of the amino acids ACGILNPV after mutations are shorter; therefore, they could not make up for the polarity effects existing in wild system (Fig. 7A).

The hydrophobic amino acids FMWY, with long side chains, were conflicted with the hydrophilic residues E35 and K31, and the original interaction modes disappeared (Fig. 7B).

Mutations of Q493T and Q493S, both the hydroxyl group -OH on the two amino acids could form hydrogen bonds with E35 and K31 (Fig. 7C).

The amino acids D493/E493 could pull residue K31 from ACE2 closer to form new salt-bridging interactions (Fig. 7D), and the same situations were also found in mutants Q493K, Q493H and Q493R (Figs. 7E, and 7F). However, in terms of strong salt-bridging effects, the mutants Q493R, Q493H, Q493K, Q493E and Q493D were considered for subsequent binding energy analysis.

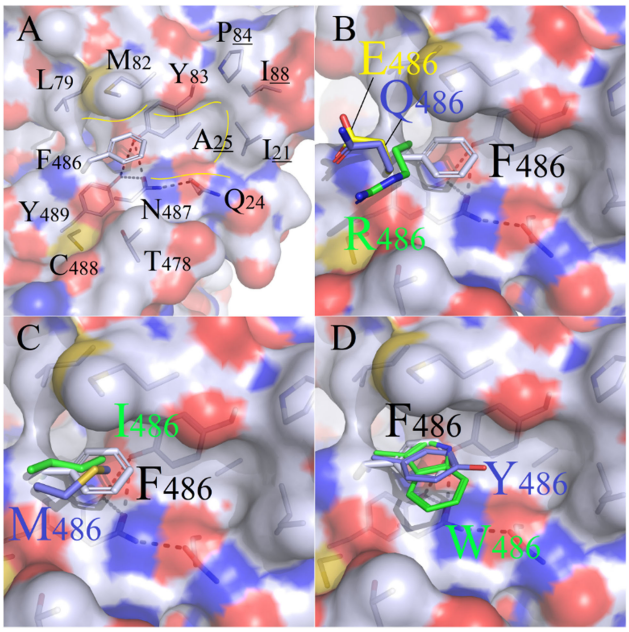


Fig. 6. Binding modes comparison of wild type and F486 series mutants. (A) Surface map of hydrophilic and hydrophobic properties around the F486 site (white: hydrophobic); (B) Selected residues: positive R486, negative E486 and uncharged polar Q486, compared with F486; (C) Hydrophobic short sidechain residues M486 and I486, compared with F486; (D) Hydrophobic long sidechain residues W486 and Y486, compared with F486.

In addition to the above results, comparative analysis of other 5 mutant systems (N487, Y489, S494, Y495 and G496) were presented in Supporting Information, and more information can be found on our website: http://bioinformatics.csu.edu.cn/mutation/.

2.6 Comparative Analysis of 171 Mutants

For the 9 amino acid sites, the corresponding 171 mutants of protein 2019-nCoV-S1 with potential mutagenic properties were conducted for detailed binding pattern variations analysis. Out of the 171 mutants, 14 (G446R, Y449R, Y449H, Y449K, F486Y, F486W, Q493R, Q493H, Q493K, Q493D, Q493E, S494R, S494K and Y495E) that may improve the binding capacity of 2019-nCoV-S1 to ACE2 were selected for subsequent "protein-protein" binding free energy analysis. The

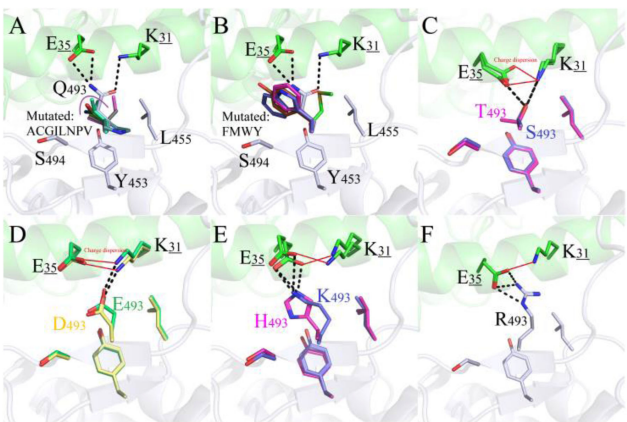


Fig. 7. Comparison analysis of the complex structures between wild-type and mutant systems.

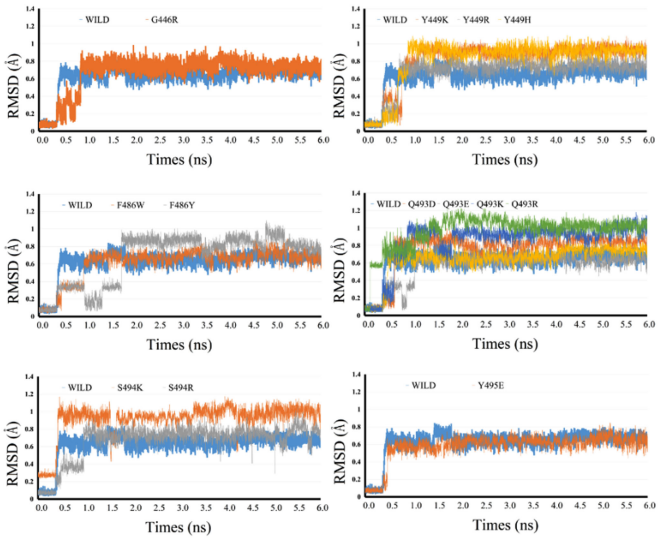


Fig. 8. Comparisons of RMSD values for mutated and the wild-type complex systems. The MD simulations reached equilibrium state within 1.5ns for all complex systems, and the overall structure fluctuation was less than 1.2 Å, indicating the excellent binding stability of 2019-nCoV-S1/ACE2 complex, which can be extremely harmful to the human body.

binding affinity between each 2019-nCoV-S1 mutant and human ACE2 protein was compared with that of the wild-type complex.

Molecular dynamic simulations of 14 novel mutants and wild-type systems were executed by the software Amber16 [26]. In order to maintain the global structural stability of each system, the complexes were restrained with a harmonic potential of the form $k\,(\Delta x)^2$ in which the force constant was equal to 1500 kcal/mol^{-1}Å^{-2}. Based on simulation trajectories, RMSD values for each complex system were extracted to detect the binding stability. Binding free energies of all 2019-nCoV-S1/ACE2 complexes were calculated by using the MMPBSA module in software Amber16 and AmberTools17.

The RMSD values for all complex systems were extracted, as shown in Fig. 8. It can be seen that the MD simulations had reached equilibrium state within 1.5ns for all complex systems, as well as the overall structure fluctuation was less than 1.2 Å. These results indicated good combination stability of coronavirus 2019-nCoV-S1 to the target protein ACE2, which can facilitate the infecting process. Based on the MDs simulations, 4000 snapshots were extracted from the last 4ns trajectory for the final average structure of complex.

The binding free energy of wild-type coronavirus 2019-nCoV-S1/ACE2 complex system is -55.0790 kcal/mol (Table II). This binding free energy is not significantly different from the first reported docking result -50.6 kcal/mol on January 21, 2020 [28], which demonstrated the reliability of our calculating results.

In addition, "protein-protein" binding affinities in 11 mutants (G446R, Y449K, Y449R, Y449H, F486W, F486Y, Q493R, Q493K, Q493E, S494R and Y495E) have improved comparing with that of the wild-type complex system. Among them, the binding free energies for mutant systems G446R, Y449R, F486Y and S494R had improved significantly. It is note worthing that the binding affinity for S494R is -76.5487 kcal/mol, which is 38.98 percent higher than that of the wild-type system. All these results indicate that

TABLE 2
Binding Free Energies Between 2019-nCoV-S1 and Human ACE2 in Mutated and Wild Complexes

Protein System	Complex (Kcal/mol)	Receptor (Kcal/mol)	Ligand (Keal/mol)	Total energy (Kcal/mol)	Imprive ratio
Wild	-73798.6202	-57139.9827	-16603.5585	**-55.0790**	-
G446R	-73758.5674	-57131.1416	-16560.9479	**-66.4779**	**20.69%**
Y449H	-73893.7173	-57175.3231	-16659.3992	-58.9950	7.11%
Y449K	-73834.5795	-57111.6168	-16661.8861	-61.0766	10.89%
Y449R	-73713.9782	-57124.3552	-16521.1992	**-68.4238**	**24.23%**
F486W	-73835.5647	-57115.2384	-16658.3607	-61.9657	12.50%
F486Y	-73801-4948	–57112.2393	-16621.9603	**-67.2952**	**22.18%**
Q493D	-73828.9582	-57054.2444	-16737.6009	-37,1129	-32.62%
Q493E	-73989.4812	-57184.4241	-16747.0621	-57.9949	5.29%
Q493H	-73875.4598	-57132.7830	-16691.7430	-50.9337	-7.52%
Q493K	-73906.2710	-57120.0036	-16729.5708	-56.6966	2.94%
Q493R	-73688.8159	-57110.2987	-16522.6746	-55.8427	1.39%
S494K	-73863.0269	-57117.1872	-16693.2454	-52.5943	-4.51%
S494R	-73789.9642	-57117.5904	-16595-8251	**-76.5487**	**38.98%**
Y495E	-73820.0100	-16666.5830	-16666.5830	-56,3820	2.37%

mutations on the 2019-nCoV-S1 can significantly enhance the binding strength between the 2019-nCoV and the human ACE2, which may have a more profound impact in this epidemic situation.

Among 19 mutants, 4 mutants including G446R, Y449R, F486Y and S494R with apparent higher binding affiniry comparing with corresponding wild-type were selected for futher analysis.

G446R. Regional conformation fluctutation for mutation G446R is detected, as shown in Fig. 9A. After mutation, the sidechain of residue R446 formed new polar interactions with Q42 on ACE2, resulting in the enhancement of protein-protein binding between 2019-nCoV-S1 and ACE2.

Y449R. As shown in Fig. 9B, local structure close to Y449R is significantly influenced. One new salt-bridge contact is formed between amino acids R449 on 2019-nCoV-S1 and E35 on ACE2, resulting in the increased binding capacity of the mutant system.

F486Y. In wild-type system, amino acids N487/Y489 from protein 2019-nCoV-S1 form hydrogen bonds with Y83/Q24 on ACE2, which weakens the hydrophobicity interactions between F486 and Y83/M82. After mutation, the chemical group -OH from the sidechain of residue Y486 form one weak polar interaction with the backbone -C = O on M82, which narrows the distance between F486 and Y83/M82 and enhances the binding ability of mutants (Fig. 9C).

S494R. Detailed analysis indicated that one new salt-bridge polar interactions is formed between residues R494 (2019-nCoV-S1) and E35 (human ACE2) in the mutant system, and this resulted in a sharp increase of the binding affinity by almost 40 percent between the mutated 2019-nCoV-S1 and ACE2 (Fig. 9D).

3 Methods

3.1 Mutation Sites Selection Principle

We adopted the following mutation principles when 2019-nCoV-S1 was mutated:

1) The basic conformation of the existing 2019-nCoV-S1 protein and ACE2 protein stayed unchanged, and the disturbance of structure near the mutation site were relatively distinct.
2) The current interaction pattern between amino acids in the "interaction interface" of the 2019-nCoV-S1-ACE2 protein complex was kept unchanged, and any mutation interfering with the existing interaction network system was not taken into considerations.
3) The binding affinity between 2019-nCoV-S1 and ACE2 could be enhanced.

3.2 Molecular Dynamics Simulation

The molecular dynamics simulation was carried out by Amber16 [26] using AMBER ff99sb force field for the complex. Hydrogen atoms were added to the initial 2019-nCoV-S1/ACE2 complex model using the leap module, setting ionizable residues as their default protonation states at a

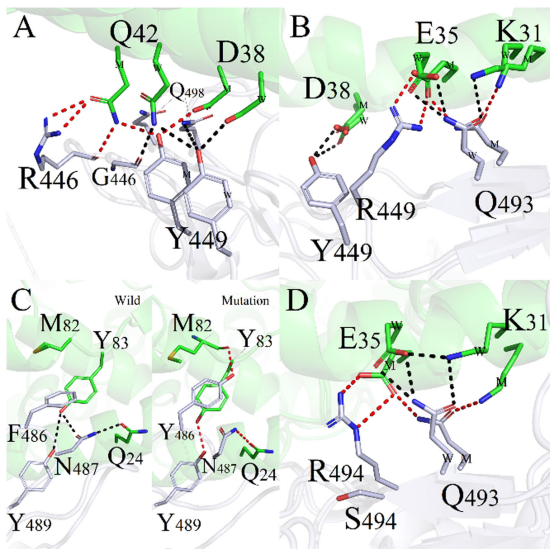


Fig. 9. Comparative analysis of protein structure binding patterns near mutant sites in mutant and wild proteins. Green: ACE2 protein; White: 2019-nCoV-S1 protein; Amino acid: sticks mode display; Wild amino acids: labeled as W, polar interactions between amino acids are marked with black and dashed lines; Mutated amino acids: labeled as M, polar interactions between amino acids are marked with red and dashed lines.

neutral PH value. The complex was solvated in a 10Å cubic periodic box of explicit TIP3P water model. The topology and coordinate structures were saved, and then simulation was performed. The simulation was proceeded as follows:

Energy minimization: first, the complex structure was restricted and the solvent water molecules were minimized. After that, protein structures were released to minimize the binding free energy of the entire system. The water molecules and counter ions were optimized using the steepest descent method of 2500 steps, followed by the conjugate gradient method for 2500 steps.

System balance: The entire system of 2019-nCoV-S1/ACE2 complex was optimized without any position constraint, followed by annealing simulation with a weak restraint ($k = 100kcal/mol^{-1}Å^{-2}$) for the complex. The entire 2019-nCoV-S1/ACE2 complex was then heated gradually in the NVT ensemble from 0 to 298K over 500ps.

Dynamic simulation: The $>$ 6ns MD simulation was performed under constant temperature at 298K with the NPT ensemble. Constant temperature was maintained using the Langevin thermostat with a collision frequency of $2ps^{-1}$. The cutoff distance between the Van der Waals energy and short-range electrostatic energy was 10Å. The PME method was used to calculate long-range electrostatic energy.

4 Conclusion

The binding interface between 2019-nCoV coronavirus and the human ACE2, plays a remarkable role in the mechanism of virus infection. The 2019-nCoV coronavirus protein was confirmed to be highly susceptible to various mutations, especially the mutations forming the 2019-nCoV-S1/ACE2 interface, would trigger apparent changes of virus transmission capacity and even the pathogenic mechanism. Predictions of mutations on 2019-nCoV-S1/ACE2 interface and how they might influence the binding modes remains unclear.

In this study, we analyzed and obtained the direct interaction interface between the coronavirus 2019-nCoV-RBD-S1 region and the human ACE2. Subsequently, all the amino acids involved in the "protein-protein" binding in the S1 region of the 2019-nCoV [29] were predicted with tolerable mutations, and ultimately, 9 amino acids with potential mutation sites and 171 single-point mutants were obtained. 14 mutants were selected for subsequent molecular dynamics simulation including the optimization of the biological computation structure and the differentiation of the physicochemical properties of the mutation sites. Through the comparative analysis of binding free energies between the 2019-nCoV-ACE2 mutant and the 2019-nCoV-ACE2 wild-type protein system, it was predicted that there are 8 amino acid mutation sites on the 2019-nCoV that may significantly improve the binding capacity of coronavirus to human ACE2, resulting in a stronger pathogenic characteristics.

Up to now, NCBI [30] has reported all the found mutations. There are 1538 mutations in total occurred on the 2019-nCoV-Spike protein and 13 mutations (K417N, G446S, G446V, Y449F, Y453F, L455F, F456L, F456Y, F486I, Y489H, Q493K, Q493L) located on the binding interface between 2019-nCoV-Spike and ACE2 proteins. Among the 13 mutations, the mutation effects on binding affinity of 8 mutations (K417N, G446V, Y449F, Y453F, L455F, F456L, F456Y, Y489H) are moderate, which indicate little difference of the toxicity between wild-type and mutants. Only the Q493K has high effects, which can increase of binding affinity with ACE2 [31], which is consistent with our prediction result.

In the global race for produceing novel coronavirus vaccine, these results are helpful to the study of the pathogenesis and infection mechanism of COVID-19, construction of animal models, new mutation detection kits and the development of new vaccines.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61832019, the Project (G20190018001), Hunan Provincial Science and Technology Program 2019CB1007, and the Fundamental Research Funds for the Central Universities, CSU (2282019SYLB004). Senbiao Fang and Ruoqian Zheng contributed equally to this work.

References

[1] Z. Xu *et al.*, "Pathological findings of COVID-19 associated with acute respiratory distress syndrome," *Lancet Respiratory Medicine*, vol. 8, pp. 420–422, 2020.

[2] J.-Y. Park *et al.*, "Chalcones isolated from angelica keiskei inhibit cysteine proteases of SARS-CoV," *J. Enzyme Inhibition Med. Chem.*, vol. 31, pp. 23–30, 2016.

[3] F. Riccio, S. K. Talapatra, S. Oxenford, R. Angell, M. Mazzon, and F. Kozielski, "Development and validation of RdRp screen, a crystallization screen for viral RNA-dependent RNA polymerases," *Biol Open*, vol. 8, 2019, Art. no. bio037663.

[4] Y. W. Chen, C. B. Yiu, and K. Y. Wong, "Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL pro) structure: Virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates," *F1000Res.*, vol. 9, pp. 129, 2020.

[5] H. Zhang *et al.*, "Deep learning based drug screening for novel coronavirus 2019-nCoV," *Interdisciplinary Sci.*, vol. 1, pp. 1–9, 2020.

[6] M. Wang *et al.*, "Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro," *Cell Res.*, vol. 30, pp. 269–271, 2020.

[7] G. Li and E. De Clercq, "Therapeutic options for the 2019 novel coronavirus (2019-nCoV)," *Nature Rev. Drug Discov.*, vol. 19, pp. 149–150, 2020.

[8] J. Yazdany and A. H. J. Kim, "Use of hydroxychloroquine and chloroquine during the COVID-19 pandemic: What every clinician should know," *Ann. Internal Medicine*, vol. 31, pp. 20–134, 2020.

[9] A. Cortegiani, G. Ingoglia, M. Ippolito, A. Giarratano, and S. Einav, "A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19," *J. Crit Care*, vol. 20, pp. 30390–30397, 2020.

[10] H. Lu, "Drug treatment options for the 2019-new coronavirus (2019-nCoV)," *Biosci. Trends*, vol. 14, pp. 69–71, 2020.

[11] J. Pang *et al.*, "Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): A systematic review," *J. Clin. Medicine*, vol. 9, no. 3, 2020, Art. no. 623.

[12] B. Shanmugaraj, A. Malla, and W. Phoolcharoen, "Emergence of novel coronavirus 2019-nCoV: Need for rapid vaccine and biologics development," *Pathogens*, vol. 9, no. 2, 2020, Art. no. 148.

[13] W. Zhang, Y. Zhao, andF. Zhang *et al.*, "The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): The Perspectives of clinical immunologists from China," *Clin. Immunology*, 2020, 214: 108393.

[14] K. K. Singh and A. Singh, "Diagnosis of COVID-19 from chest X-ray images using wavelets-based depthwise convolution network," *Big Data Mining Analytics*, vol. 4, no. 2, pp. 84–93, 2021.

[15] R. Lu *et al.*, "Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding," *Lancet*, vol. 395, pp. 565–574, 2020.

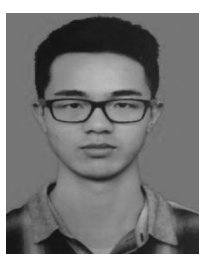
[16] Y. S. Wan, J. Shang, G. Rachel, S. B. Ralph, and F Li, "Receptor recognition by the novel coronavirus from wuhan: An analysis based on decade-long structural studies of SARS coronavirus," *J. Virol.*, vol. 94, pp. e00127–e00120, 2020.
[17] X. Peng, X. Xu, Y. Q. Li, L. Cheng, X. D. Zhou, and B. Ren, "Transmission routes of 2019-nCoV and controls in dental practice," *Int. J. Oral Sci.*, vol. 12, 2020, Art. no. 9.
[18] F. Wu *et al.*, "A new coronavirus associated with human respiratory disease in China," *Nature*, vol. 579, pp. 265–269, 2020.
[19] X. L. Tang *et al.*, "On the origin and continuing evolution of SARS-CoV-2," *Nature Sci. Rev.*, vol. 10, 2020, Art. no. 1093.
[20] Amanda Woods, Iceland scientists found 40 mutations of the coronavirus. NewYork Post, 2020. [Online]. Available: https://nypost.com/2020/03/24/iceland-scientists-found-40-mutations-of-the-coronavirus-report-says/
[21] H. Yi, "novel coronavirus is undergoing active recombination," *Clin. Infectious Diseases*, vol. 71, no. 15, pp. 884–887, 2020.
[22] H. P. Yao, X. Y. Lu, Q. Chen *et al.*, "Patient-derived mutations impact pathogenicity of SARS-CoV-2," *CELL-D-20-01124*, 2020. [Online]. Available: http://dx.doi.org/10.2139/ssrn.3578153
[23] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, and Q. Zhou, "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2," *Science*, vol. 367, pp. 1444–1448, 2020.
[24] N. L. Sim, P. Kumar, J. Hu, S. Henikoff, G. Schneider, and P. C. Ng, "SIFT web server: Predicting effects of amino acid substitutions on proteins," *Nucl. Acids Res.*, vol. 40, pp. 452–457, 2012.
[25] G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, "Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments," *J. Comput. Aided Mol. Des.*, vol. 27 no. 3, pp. 221–234, 2013.
[26] A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C Walker, "Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born," *J. Chem. Theory Comput.*, vol. 8, pp. 1542–1555, 2012.
[27] E. A. Coutsias, C. Seok, and K. A. Dill, "Using quaternions to calculate RMSD," *J. Comput. Chem.*, vol. 25, pp. 1849–1857, 2004.
[28] X. Xu *et al.*, "Evolution of the novel coronavirus from the ongoing wuhan outbreak and modeling of its spike protein for risk of human transmission," *Sci. China Life Sci.*, vol. 63, pp. 457–460, 2020.
[29] S. B. Fang, R. Q. Zheng, C. Q. Lei, J. X. Wang, R. Q. Zheng, and M. Li, "Key residues inf luencing binding affinities of 2019-nCoV with ACE2 in different species," *Brief. Bioinf.*, 2020, [Online]. Available: https://doi.org/10.1093/bib/bbaa329.
[30] D. L. Wheeler *et al.*, "Database resources of the national center for biotechnology information," *Nucl. acids Res.*, vol. 36, no. suppl_1, pp. D13–D21, 2007.
[31] W. M. Zhao *et al.*, "The 2019 novel coronavirus resource," *Yi Chuan*, vol. 42, no. 2, pp. 212–221, Feb. 20, 2020.

Senbiao Fang received the BS degree in chemistry from Xiangtan University, China, in 2011, and the MS degree in computational biology from Lanzhou University, China, in 2014. He is currently working towards the PhD degree in the School of Computer Science and Engineering, Central South University, China. His research interests include bioinformatics and system biology.

Ruoqian Zheng received the BS degree in network engineering from Xiangtan University, Xiangtan, China, in 2019. She is currently working towards the MS degree in computer science in Central South University, Changsha, China. Her current research interest is structure biology.

Chuqi Lei received the BS degree in computer science from Central South University, China, in 2020. She is currently working toward the PhD degree in computer science in Central South University, Changsha, China. Her current research interests include bioinformatics and system biology.

Jianxin Wang (Senior Member, IEEE) received the BEng and MEng degrees in computer engineering from Central South University, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University, China, in 2001. He is the dean and a professor with School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics and computer network. He has published more than 150 papers in various International journals and refereed conferences.



Renyi Zhou is currently working toward the graduate degree in the School of Computer Science and Engineering, Central South University, China. His current research interests include bioinformatics and representation learning.

Min Li received the PhD degree in computer science from Central South University, China, in 2008. She is currently a professor and vice dean of the School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. Her main research interests include bioinformatics and systems biology.