RESEARCH ARTICLE

PROTEINS WILEY

Why are ACE2 binding coronavirus strains SARS-CoV/SARS-CoV-2 wild and NL63 mild?

Puneet Rawat[1] | Sherlyn Jemimah[1] | P. K. Ponnuswamy[2] | M. Michael Gromiha[1]

[1]Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, India

[2]Department of Physics, Bharathidasan University, Tiruchirapalli, India

Correspondence
M. Michael Gromiha, Indian Institute of Technology Madras, Chennai, India.
Email: gromiha@iitm.ac.in

Abstract

Coronaviruses are responsible for several epidemics, including the 2002 SARS, 2012 MERS, and COVID-19. The emergence of recent COVID-19 pandemic due to SARS-CoV-2 virus in December 2019 has resulted in considerable research efforts to design antiviral drugs and other therapeutics against coronaviruses. In this context, it is crucial to understand the biophysical and structural features of the major proteins that are involved in virus-host interactions. In the current study, we have compared spike proteins from three strains of coronaviruses NL63, SARS-CoV, and SARS-CoV, known to bind human angiotensin-converting enzyme 2 (ACE2), in terms of sequence/structure conservation, hydrophobic cluster formation and importance of binding site residues. The study reveals that the severity of coronavirus strains correlates positively with the interaction area, surrounding hydrophobicity and interaction energy and inversely correlate with the flexibility of the binding interface. Also, we identify the conserved residues in the binding interface of spike proteins in all three strains. The systematic point mutations show that these conserved residues in the respective strains are evolutionarily favored at their respective positions. The similarities and differences in the spike proteins of the three viruses indicated in this study may help researchers to deeply understand the structural behavior, binding site properties and etiology of ACE2 binding, accelerating the screening of potential lead molecules and the development/repurposing of therapeutic drugs.

KEYWORDS

ACE2 binding, coronavirus spike protein, COVID-19, mutational analysis, NL63, SARS, surrounding hydrophobicity

1 | INTRODUCTION

Coronaviruses are giant enveloped viruses with positive-sense RNA, divided into four genera: α-, β-, γ-, and δ-types. The former two inflict only mammals and the latter two are primarily present in birds. Among them, four stains (229E, NL63, OC43, and HKU1) were identified to be present in humans, causing mild symptoms. Recently, three new strains (SARS, MERS, and SARS CoV-2) affected human population and specifically, SARS CoV-2 affected more than 60 million people globally and caused over 1.4 million deaths until November 2020. Earlier studies have shown that the α-coronavirus NL63 and the two β-coronaviruses, SARS-CoV and SARS-CoV-2 bind differentially to the common human host cell receptor, angiotensin-converting enzyme 2 (ACE2).[1-4] The spike protein (S protein) present on the membranes of these coronaviruses is mainly responsible for the receptor binding and fusion with the human cell membrane, making this protein a key target for potential therapies and diagnostics.

The NL63 virus is widely spread within human population and generally causes minor symptoms, such as fever, dry cough, runny

nose, and other respiratory illnesses.[5,6] The recently emerged pandemics have high case fatality rates (CFR), which were ∼11% for SARS-CoV and ∼3.61% for SARS-CoV-2.[7,8] The CFR for the SARS-CoV-2 pandemic was ∼5% by the end of June 2020 according to statistics shown by Our World in Data (https://ourworldindata.org/grapher/coronavirus-cfr).[9] However, the higher transmission rate of SARS-CoV-2 has made it more deadly than the SARS-CoV strain.[10] Wu et al[11] compared the sequences and structures of the receptor-binding domains of NL63 and SARS-CoV spike proteins with ACE2[12] and observed very little sequence conservation and structural homology between them, which indicated that these two viruses might have evolved independently to bind to ACE2 receptor. Recently, the 3D-structure was determined for the SARS-CoV-2 trimeric spike protein[13] and RBD of the spike protein in complex with ACE2 receptor.[4] Sequence and structural comparisons of the SARS-CoV and SARS-CoV-2 strains indicate high similarity between these strains. The domain B (S^B) of the spike protein of these two strains, which contains ACE2 receptor-binding site, shows ∼75% sequence identity.[14] However, the SARS-CoV spike protein binding antibodies S230, m396, and 80R do not bind to SARS-CoV-2 spike protein despite such high similarities.[13]

In this work, we performed a comparative analysis of spike protein structures from NL63, SARS-CoV, and SARS-CoV-2, focusing on the receptor-binding domains to aid drug design and provide explanations for the observed differences and similarities. The study concentrates on three main aspects: sequence/structure conservation, hydrophobic clustering behavior and affinity change upon mutations in binding domains to bring out reasons for varying severity, transmission rate and longevity among ACE2 binding coronavirus strains. It is observed that the NL63 spike protein does not show much sequence or structure conservation with SARS-CoV or SARS-CoV-2, although there are conserved residues in the binding interface of the spike proteins of all three strains. Also, the interaction energies indicate that any mutation on these residues could reduce the binding affinity of the ACE2 bound complex. Moreover, the conserved Gly residue in the binding interface remains essential for ACE2 binding and had low surrounding hydrophobicity in all three strains. The analysis of interface residues in the RBD regions shows that the surrounding hydrophobicity of the interface is very high for the severe SARS-CoV strain of coronavirus.

2 | MATERIALS AND METHODS

2.1 | Sequence and structures

A simple search query "coronavirus" in the PDB database[15] generates result page with more than 800 structures, which includes all sorts of structural information such as proteins/protein fragments, protein complexes with natural ligands, therapeutics or neutralizing agents related to coronaviruses. Currently, six human coronavirus strains (three mild and three severe strains) have 3D structures available in the PDB database for the spike protein receptor-binding domains (RBDs) (Table 1). We have selected all the three ACE-2 binding coronavirus strains, one mild (NL63) and two severe (SARS-CoV and SARS-CoV-2) in the current analysis. The binding site residues at the interface between RBD of spike protein and ACE2 receptor are identified using a distance cut-off of 4 Å suggested by Lan et al[4] The structures were visualized and analyzed using PyMOL (https://pymol.org/2/) and the area of the binding interface was calculated using "get_area" command in PyMOL.[16] The sequence alignment among the spike protein RBD sequences was performed using MAFFT server[17] with default values and visualized using JalView.[18] It is important to note that the crystal structures of RBD-ACE2 complex do not have a full-length RBD region in all three strains. Hence, most of the RBD region analyses are done using free spike proteins.

2.2 | Computation of surrounding hydrophobicity

We computed the surrounding hydrophobicity of residues in the free form and complex structures of proteins using the method of Manavalan and Ponnuswamy.[19] It is computed using the formula,

$$H_t(i) = \sum_{j=1}^{20} n_{ij} h_j \qquad (1)$$

where $H_t(i)$ is surrounding hydrophobicity of ith residue of the protein. n_{ij} is the total number of surrounding residues of type j around residue i within 8 Å. h_j is hydrophobicity value for the j residue type (in kcal/mol) obtained from thermodynamic transfer experiments.[20,21] In essence, the sounding hydrophobicity of a particular residue in a

TABLE 1 Coronavirus strains considered for the present study

CORONA virus	Coronavirus type	Condition	RBD region	Receptor	PDB ids	
					Spike protein	RBD Complex[a]
NL63	Alpha	Mild	476-616	ACE2	5SZS	3KBH
OC43	Beta	Mild	318-624	HLA class I	6NZK	–
HKU1	Beta	Mild	310-622	Unknown	5I08	–
MERS-CoV	Beta	Severe	382-503	DPP4	5X5F	4L72
SARS-CoV	Beta	Severe	306-527	ACE2	5WRG	2AJF
SARS-CoV-2	Beta	Severe	319-541	ACE2	6VXX	6M0J

[a]Receptor binding region (RBD) of the C-terminal domain (CTD) in complex with respective receptors.

protein indicates how tightly or loosely packed at its position. Accordingly, we obtain average hydrophobicity indices (HI) for the 20 residues taking into account of all the residues in every protein.

The surrounding hydrophobicity profiles were calculated for the spike protein, RBD region and interface residues of the respective strains. It is important to note that the surrounding hydrophobicity values for the disordered and flexible residues are not considered in the analysis due to the unavailability of atomic coordinates in PDB file. However, these residues are expected to have a dynamic yet lower surrounding hydrophobicity value than the average due to high solvent exposure to attain flexibility. The surrounding hydrophobicity calculations for the RBD-ACE2 complex does not include the residues from the ACE2 protein.

2.3 | Computation of interaction energy and stability

Mutated structures for the spike RBD-ACE2 complex were generated using FoldX software.[22] We removed heteroatom coordinates from the PDB files and retained one subunit of the spike protein RBD and ACE2. The PDB files were then rectified using the "RepairPDB" command in FoldX. The residues present at the interface of spike protein and ACE2 protein were mutated systematically using "BuildModel" command, and the interaction energy between the proteins was calculated using "AnalyseComplex" command. The change in interaction energy (ΔIE) was further calculated using following equation:

$$\Delta IE = IE_{mutant} - IE_{wild-type} \tag{2}$$

where IE_{mutant} is interaction energy of the complex with point mutation and $IE_{wild-type}$ is the interaction energy of wild-type complex. Similarly, we computed the change in stability upon mutation using FoldX[22] and CUPSAT[23] (http://cupsat.tu-bs.de/).

3 | RESULTS AND DISCUSSION

3.1 | Sequence comparison of spike proteins

SARS-CoV and SARS-CoV-2 spike RBDs show a high sequence identity of 73.1% using MAFFT.[17] However, the spike protein RBD region of *Alpha*-coronavirus NL63 showed a low sequence identity with

Beta-coronavirus strains SARS-CoV (23.7%) and SARS-CoV-2 (25%). The interface residues for SARS-CoV and SARS-CoV-2 are positioned more towards the C-terminal side, unlike NL63 interface residues that are towards the N-terminal (Figure 1). These observations are in agreement with the hypothesis that NL63 alpha-coronavirus and the SARS-CoV beta-coronavirus might have evolved independently to bind the ACE2 receptor.[11]

3.2 | Structural comparison of spike proteins

The spike proteins of SARS-CoV-2 (PDB 6VXX) and SARS-CoV (PDB 5WRG) strains have higher structure similarity (RMSD = 2.15 Å) compared to NL63 strain (RMSD = 36.2 Å) (Figure 2A). RMSD values of the RBD region, in the free spike protein and bound to ACE2 receptor, were significantly low (0.7-1.06 Å) for three coronavirus strains. However, the binding interface in the RBD region of the spike protein is usually buried in the core structure and exposes itself only to bind the ACE2 receptor.[14,24,25] The change in RBD conformation between closed (PDB 6VXX) and open (PDB 6VYB) state for the SARS-CoV-2 spike protein is shown in Figure 2B, which clearly shows that binding interface gets exposed due to rotational motion of the whole RBD region. The beta-viruses SARS-CoV and SARS-CoV-2 contain disordered/flexible regions in the RBD domain. However, SARS-CoV-2 has such residues in the binding interface as well. The presence of flexible region(s) in the interface significantly influences the binding by either favoring or disfavoring it.[26] The experimental studies have shown that the binding affinity of SARS-CoV-2 is higher than SARS-CoV strain with ACE2 receptor.[4]

We further calculated the interface surface area of the spike protein-ACE2 complexes. NL63 spike protein shared the least surface area (1190 Å²), as previously reported.[11] SARS-CoV and SARS-CoV-2 had higher surface areas of 1816 Å² and 1861 Å², respectively. The high contact surface area in these two proteins positively correlate with enhanced binding affinity.[27] Hence, this analysis shows that mild NL63 strain has less binding affinity with ACE2 compared to the other two strains.

3.3 | Conserved residues in spike protein

In protein-protein interactions, the hotspot residues at the interface are generally conserved and contribute towards the interaction energy.[28] We have identified the conserved and semi-conserved
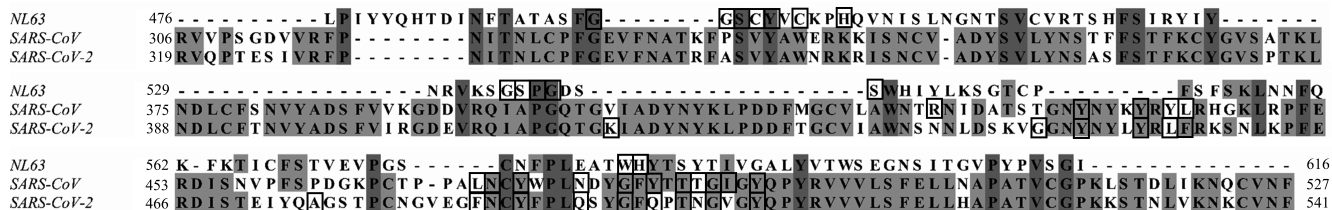

FIGURE 1 Sequence alignment for the receptor-binding domain of spike proteins from NL63, SARS-CoV and SARS-CoV-2 coronavirus strains. Interface residues are highlighted in the alignment
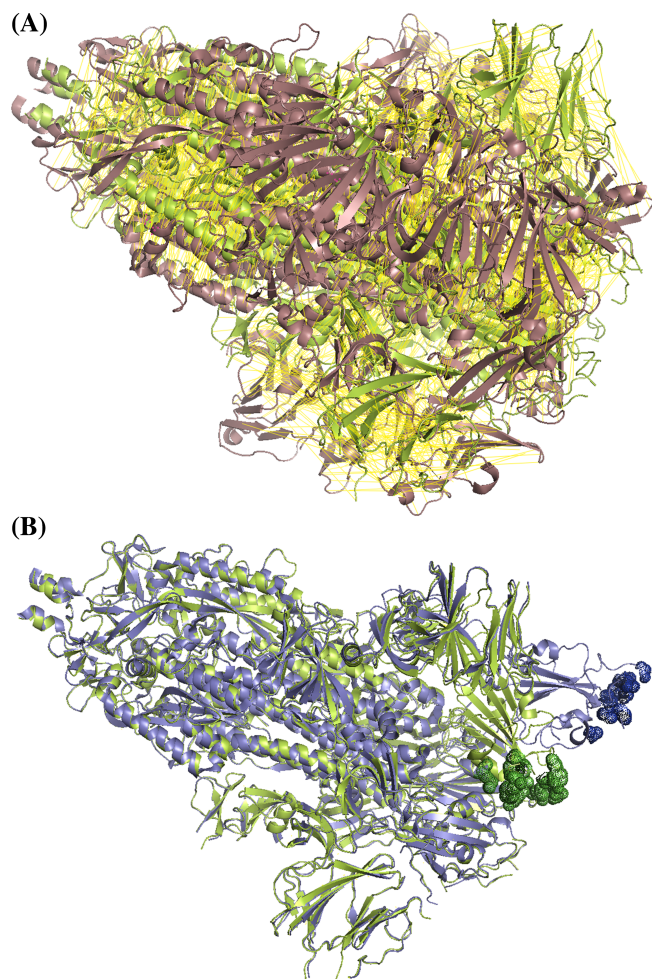
FIGURE 2 Structural alignment of corona virus spike proteins: A, NL63 (magenta) with closed state of SARS-CoV-2 spike protein (green), and, B, closed (green) and open (blue) states of SARS-CoV-2 spike proteins (interface residues are shown in dotted sphere) [Color figure can be viewed at wileyonlinelibrary.com]

residues based on (1) occupancy of the same residue in a particular position in all the three spike proteins in multiple sequence alignment and (2) the interacting residues (same amino acid residue and residues with similar physicochemical characteristics, respectively) between spike proteins and ACE2. The identified conserved and semi-conserved residues are highlighted in Table 2. The conserved Gly residue in all three strains (G537 in NL63, G488 in SARS-CoV and G502 in SARS-CoV-2) has the same contacts (Lys353 and Gly354) in ACE2. The semi-conserved residues have similar physicochemical characteristics. The semi-conserved residue, Tyr (Y498 in NL63, Y436 in SARS-CoV and Tyr449 in SARS-CoV-2) shows a common binding to a negatively charged residue in ACE2 (E37 or D38).

Further, the aromatic residues, Tyr (Tyr491 in SARS-CoV and Tyr505 in SARS-CoV-2 strains) or Trp585 (in NL63) present in the close proximity of conserved Gly residue, also form common contacts with Gly354 in ACE2. However, Trp585 residue in NL63 spike protein is a distant residue in sequence (48 residues father in NL63 compared to three residues farther in SARS strains, in the C-terminal side

relative to conserved Gly residue) (Figure S1). The conserved and semi-conserved residues in the spike proteins are highlighted in Table 2.

The conserved residues identified in the SARS-CoV and SARS-CoV-2 are compared to other β-coronaviruses (HKU1, OC43, and MERS-CoV) using multiple sequence alignment to identify the mutations. We noticed that MERS-CoV misses the ACE2 binding region (Figure S2), which might induce MERS-CoV spike protein to bind with a different receptor, DPP4.[29] The HKU1 and OC43 spike proteins have Tyr449 residue mutated to Trp and Tyr505 residue mutated to Thr when compared to the SARS-CoV-2 strain. These mutations are not observed in the ACE2 binding coronavirus strains. Although OC43 strain has conserved Gly residue present, its host receptor remains unknown. The HKU1 spike protein also lacks the conserved Gly residue, which explains its binding to a different receptor, HLA class I antigen (Figure S2).[30]

3.4 | Comparison of surrounding hydrophobicity of spike proteins

3.4.1 | Comparison of surrounding hydrophobicity indices (HI) of 20 amino acids

The surrounding hydrophobicity indices (HI) of 20 amino acid residues in the respective spike proteins of three viruses NL63, SARS-CoV, and SARS-CoV-2 showed a near-equal Pearson correlation (r) of 0.89 for NL63/SARS-CoV-2 and 0.85 for SARS-CoV/SARS-CoV-2 strains. These indices for the whole spike protein are surprisingly similar for coronavirus strains, especially for NL63 protein, which does not show sequence/structure similarity with other two coronavirus strains. However, the RBD regions of NL63 and SARS-CoV strains showed correlation (r) of 0.26 (P-value = .29) for NL63/SARS-CoV-2 and 0.51 (P-value = .03) for SARS-CoV/SARS-CoV-2 strains. The Met residues were less frequent in the RBD region (only one Met417 present in SARS-CoV), hence not considered in the correlation analysis to avoid overfitting. The hydrophobicity indices of 20 amino acids are shown in Figure S3. We further analyzed the residue-wise surrounding hydrophobicity of the RBD region and compared the free spike protein binding interface and ACE2 bound spike protein interface.

3.4.2 | Residue wise surrounding hydrophobicity analysis

A direct comparison of residue-wise surrounding hydrophobicity is only possible for the RBD regions of SARS-CoV and SARS-CoV-2 spike proteins. The RBD region of alpha-coronavirus NL63 spike protein is shorter (only 141 residue-long) and does not align properly to other two coronavirus strains. However, it has more clearly defined stretches of similar hydrophobicity environment (wider peaks and valleys) (Figure 3A-C). On the other hand, RBD regions of spike proteins

TABLE 2 Residues present in the binding interface of the spike protein (RBD)-ACE2 complex

NL63		SARS-CoV		SARS-CoV-2	
Spike	**ACE2**	**Spike**	**ACE2**	**Spike**	**ACE2**
G494	H34	R426	Q325, E329	K417	D30
G495	H34	**Y436**	**D38, Q42**	G446	Q42
S496	D30, N33, H34	Y440	H34	**Y449**	**D38, Q42**
C497	E37, R393	Y442	K31	Y453	H34
Y498	**E37, K353, G354**	L443	T27	L455	H34
C500	A387	L472	L79, M82	F456	T27, D30
H503	H34	N473	Q24, Y83	A475	Q24, T27
G534	Y41	Y475	T27, F28, K31, Y83	F486	L79, M82, Y83
S535	Y41, K353	N479	H34	N487	Q24, Y83
P536	Y41, Q325, G326, N330, D355	G482	K353	Y489	T27, F28, K31
G537	**K353, G354**	Y484	Y41, Q42, L45	Q493	K31, H34, E35
S540	T324, Q325	T486	Y41, L45, N330, D355, R357	G496	K353
W585	**G354, F356**	T487	Y41, K353	Q498	Y41, Q42
H586	P321, N322, T324	**G488**	**K353, G354, D355**	T500	Y41, N330, D355, R357
		I489	Q325	N501	Y41, K353
		Y491	**E37, K353, G354**	G502	K353, G354
				Y505	**E37, K353, G354, R393**

Note: (1) The conserved positions are highlighted in bold. (2) In the highlighted part, E37 and D38 (negatively charged residues in ACE2); W585 in NL63, Y491 in SARS-CoV and Y505 in SARS-CoV-2 (aromatic residue in RBD interface) are considered semi-conserved. (3) The flexible residues are underlined in SARS-CoV-2.

of SARS-CoV and SARS-CoV-2 strains are 222 and 223 residue long, respectively.

The average surrounding hydrophobicity was ~15 kcal/mol for the spike protein RBD region for all coronavirus strains, which was used as a cut-off to define the hydrophobic or hydrophilic environments. However, a small proportion of residues present in the RBD region of SARS-CoV (seven residues) and SARS-CoV-2 (30 residues) contain flexible/disordered residues[4,12] and lack surrounding hydrophobicity values. For the ACE2 bound complex, NL63 (11 residues) and SARS-CoV (six residues) RBD regions had some flexible/disordered residues. Although surrounding hydrophobicity values in SARS-CoV and SARS-CoV-2 were almost the same as the average (~15 kcal/mol), the surrounding hydrophobicity of NL63 decreased significantly to 13.4 kcal/mol. The reduction in average surrounding hydrophobicity indicates structural changes and a probable increase in solvent exposure for some residues within RBD region for NL63, upon binding to ACE2. A comparative analysis of residue-wise surrounding hydrophobicity of SARS-CoV and SARS-CoV-2 shows that aligned counterpart of flexible regions spanning 455-461 and 469-488 in SARS-CoV-2 are buried in SARS-CoV spike protein (Figure 3D).

Tsai et al[31] reported that hydrophobicity plays a dominant role at binding interfaces, but it is not as strong as the interior of protein. We obtained similar results, where average surrounding hydrophobicity of the binding interface is lower than the RBD region (~15 kcal/mol) in all three strains. At the binding interface, the average surrounding hydrophobicity is lower in the SARS-CoV-2 strain for free protein (10.73 kcal/

mol compared to 11.6 and 13.9 kcal/mol for NL63 and SARS-CoV, respectively) even after not considering seven flexible residues, which are likely to have less than average surrounding hydrophobicity. The flexible residues in the interface become ordered after binding to ACE2. However, NL63 strain has the least average surrounding hydrophobicity (10.8 kcal/mol compared to 13.7 and 12.3 kcal/mol for SARS-CoV and SARS-CoV-2, respectively) after binding to the ACE2 receptor (Figure 4A,B). SARS-CoV had the highest surrounding hydrophobicity for the interface in both free and bound form, which might help in protein-protein interaction. The increases of surrounding hydrophobicity in the SARS-CoV-2 spike protein is mainly due to the stabilization of flexible residues in the binding interface after binding to ACE2 (Table S1). However, NL63 and SARS-CoV showed a small decrease in surrounding hydrophobicity after binding to ACE2 receptor, probably due to more solvent exposure of the binding interface after rotation of the RBD region outwards for ACE2 binding.

The analysis of change in residue-wise surrounding hydrophobicity of the binding interface showed that the surrounding hydrophobicity of Tyr442 (21.9 kcal/mol) and Leu443 (19.8 kcal/mol) was relatively higher even when bound to ACE2 in SARS-CoV. However, the flexible residue counterparts Leu455 and Phe456 in SARS-CoV-2 showed the surrounding hydrophobicity value of 18.7 kcal/mol in the bound form. Similarly, Leu472 and Asn473 in SARS-CoV spike protein have almost equal surrounding hydrophobicity (~10 kcal/mol) in both free and complex forms. However, flexible residue counterparts, Phe486 and Asn487 in SARS-CoV-2 have lower surrounding

(A) Surrounding Hydrophobicity plot for NL63 strain

(B) Surrounding Hydrophobicity plot for SARS strain
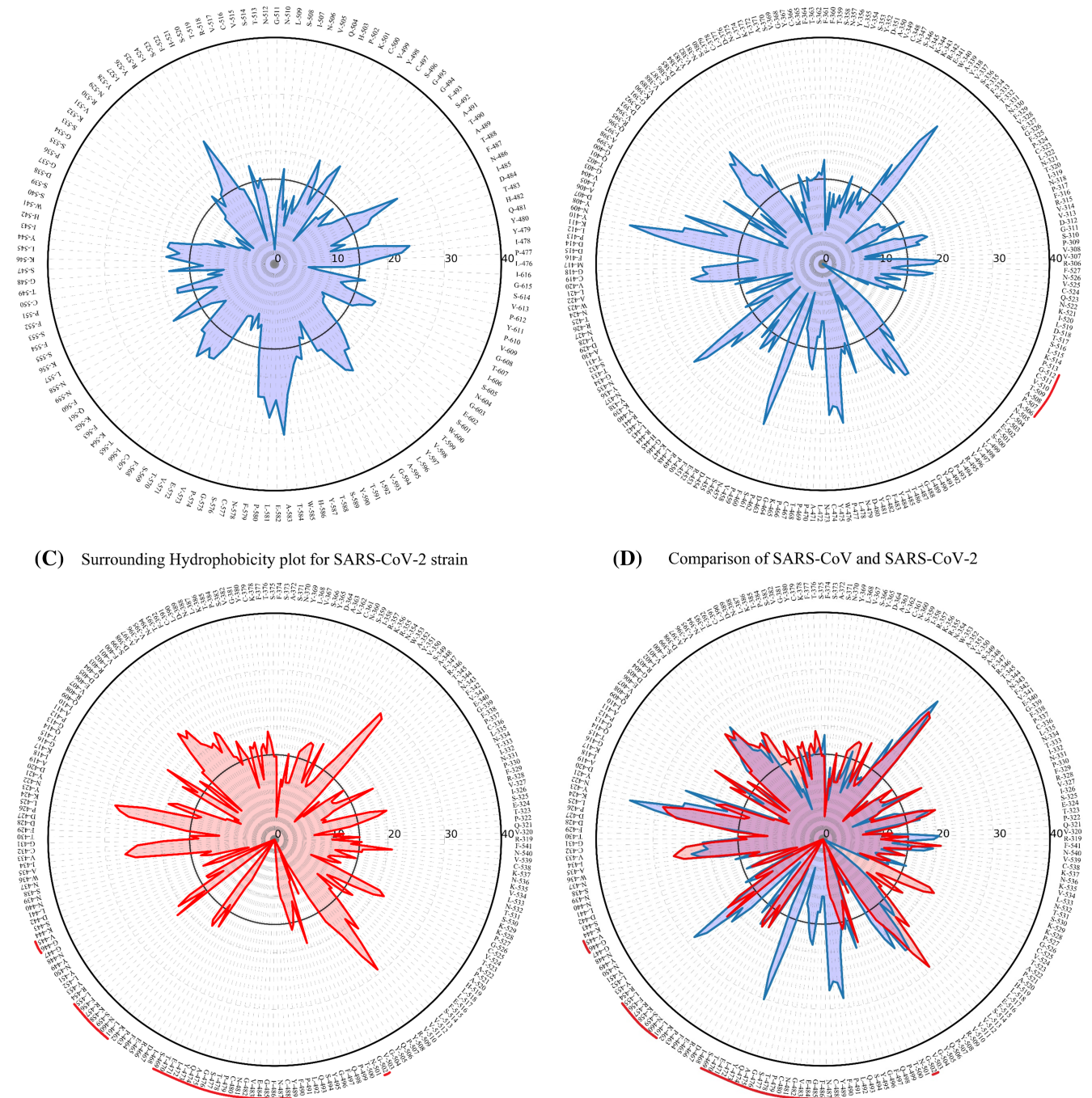
(C) Surrounding Hydrophobicity plot for SARS-CoV-2 strain

(D) Comparison of SARS-CoV and SARS-CoV-2

FIGURE 3 Surrounding hydrophobicity of the receptor-binding domain of the different strains of coronavirus (A) NL63, (B) SARS-CoV, and (C) SARS-CoV2. D shows the comparison of hydrophobicity profiles for SARS-CoV (blue) and SARS-CoV-2 (red). The flexible/disordered region is highlight with red color in the polar axis

hydrophobicity (~8 kcal/mol) in the bound form. A similar trend is observed with Gly488 in SARS-CoV (6.2 kcal/mol), whose counterpart Gly502 in SARS-CoV-2 has lower surrounding hydrophobicity (4.9 kcal/mol; Table S1). In conclusion, SARS-CoV-2 has lower surrounding hydrophobicity compared to SARS-CoV strain in the binding interface. The higher hydrophobic environment at the interface for SARS-CoV strains can improve its binding to ACE2.[31]

3.4.3 | Comparison of conserved/semi-conserved binding interface residues

The conserved residues Tyr498 and Trp585 in NL63 have relatively more surrounding hydrophobicity in both free protein and in complex with ACE2 (Table S2). However, for the beta-coronavirus, these residues have less surrounding hydrophobicity. The conserved Gly
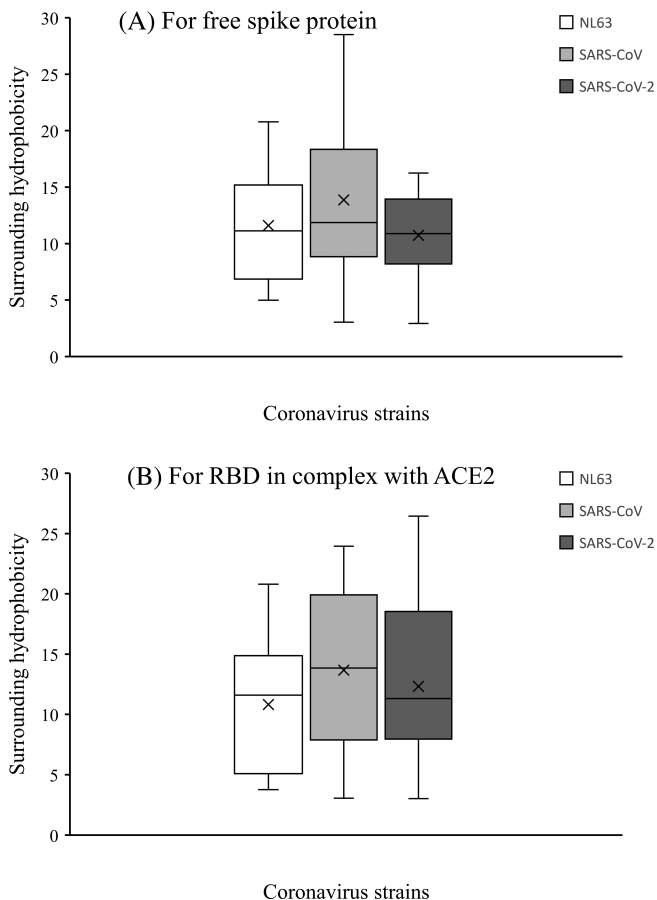
FIGURE 4 Average surrounding hydrophobicity of the residues in the interface of spike protein-ACE2 receptor in, A, free form and, B, complex with ACE2 receptor; x denotes the mean

residue in all three strains has a significantly lower value of surrounding hydrophobicity, which shows that this residue is positioned in a hydrophilic environment and significantly exposed in all strains. In conclusion, the surrounding hydrophobicity of the conserved residues in three strains follows the ascending order of SARS-CoV-2 < SARS-CoV < NL63.

3.5 | Mutational analysis

The mutational analyses are widely used to study the binding interface, especially for their contribution to interaction energy.[32] For example, a recent study has shown that a point mutation (D614G) in the spike protein of SARS-CoV-2 coronavirus can significantly affect the viral infectivity and neutralization sensitivity.[33] Hence, we have performed a mutational analysis for the residues at the RBD-ACE2 interface to identify the hotspot residues essential for the interaction. The interaction energies for the wild-type RBD-ACE2 complex calculated using FoldX software were −6.99 kcal/mol for NL63, −18.81 kcal/mol for SARS-CoV, and −16.89 kcal/mol for SARS-CoV-2. These predicted interaction energies are directly proportional to the severity of the coronavirus strains. Further, we have calculated

the change in interaction energy (ΔIE) for the interface residues in the RBD of all three strains (Table 3). The analysis of changes in interaction energy upon point mutation showed that most of the residues in the interface reduce the binding affinity the complex upon mutation, which indicates their importance for binding. However, change in interaction energy was relatively higher for Ser535, Gly537, and Ser540 residues in NL63; Tyr487, and Gly488 in SARS-CoV; Leu455, Ala475, Asn487, Gly496 and Gly502 residues in SARS-CoV-2 (Table 3). The conserved Gly residue significantly reduced the binding affinity in all three strains (maximum change in interaction energy of 19.6, 15.92 and 14.45 kcal/mol for NL63, SARS-CoV and SARS-CoV-2, respectively). The stability analysis using FoldX and CUPSAT also revealed that the mutation of conserved Gly destabilizes the protein (Tables S3 and S4), showing its importance for both stabilizing and interacting with ACE2.

4 | CONCLUSIONS

The severity of a disease can be attributed to the sequence or structure features of the involved host-pathogen proteins. Here, we have analyzed the relationship between the severity of coronavirus strains and the inherent structural and sequence features of the spike protein. The least interaction area, surrounding hydrophobicity and interaction energy for the interface residues are the main reasons for the mild severity of human coronavirus NL63. SARS-COV is more severe than SARS-COV-2 mainly due to similar size of the interface area, but absence of flexible residues (flexible residues can increase the entropic cost to binding), highest hydrophobic environment at the interface for better interaction with ACE2 and highest interaction energy. The wet lab studies have shown that SARS-CoV-2 has a better binding affinity than SARS-CoV.[4] However, an experimental study on 200 neutralizing antibodies obtained from convalescent patients has also concluded that binding affinity does not necessarily correlate with effective neutralization.[34] Our study has also identified conserved residues at the interface of the spike protein for all the three coronavirus strains, which might act as a recognition site for ACE2 receptor. These conserved interaction sites can help in effective targeting of the ACE2 binding site by therapeutics in SARS-CoV as well as SARS-CoV-2 strain.

With millions of people infected worldwide by SARS-CoV-2 over a short period, COVID-19 has attained the status of a global pandemic. A proper understanding of the biophysical properties of spike protein as indicated in this study and related molecular mechanisms will undoubtedly help in the development of therapeutics against the novel coronavirus.

TABLE 3 Change in binding affinity of the spike protein (RBD)-ACE2 complex upon mutation

Residues	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Count[a]	Max
NL63																						
G-494	0.07	0.41	1.1	1.63	0.93		3.15	0.38	0.13	2.12	-0.95	1.75	0.64	1.48	0.98	0.35	0.01	0.72	1.98	1.45	18	3.15
G-495	-0.42	-0.08	0.85	-0.18	-0.29		-0.13	-0.11	-0.44	-0.28	-0.71	0.07	-0.19	-0.01	-0.35	0.74	0.55	0.19	-0.45	-0.47	5	0.85
S-496	0.2	0.49	1.27	1.51	-0.34	1.01	1.21	-0.07	-0.01	0.39	-0.31	0.92	-0.53	0.68	0.52	-0.35	-0.06	0.34	-0.81	-0.22	11	1.51
C-497	0.34		0.17	-0.03	-0.71	-0.43	-0.19	-1.23	-1.18	-1.28	-1.75	-1.76	-0.39	-1.08	-2.43		-0.78	-1.39	-0.65	-0.28	2	0.34
Y-498	0.88	1.07	1.76	0.43	-0.29	1.44	1.36	1.6	0.68	0.22	0	1.48	0.59	1.28	0.19	1.71	0.08	0.88	0.73		18	1.76
C-500	-1.12		-0.56	-0.55	0.05	-0.73	-0.57	-0.89	-0.38	-0.72	-1.29	-0.59	0.25	-0.5	-0.66	-0.67	-0.55	-0.54	-0.66	0.04	3	0.25
H-503	0.01	-0.03	-0.05	0	0.03	0.01		-0.03	0.02	0.03	-0.06	0.02	0.03	0.08	0.09	0.01	0.09	0.04	-0.05	0.03	14	0.09
G-534	-0.32	-0.47	-0.3	-0.45	-1.17		-0.6	-0.79	-0.54	-0.94	-1.01	-0.42	-0.23	-0.46	-0.43	-0.21	-0.68	-0.79	-1.62	-1.08	0	-0.21
S-535	0.83	0.75	5.17	4.63	-0.58	1.25	0.69	3.13	8.11	4.83	2.37	6.35	0.64	3.56	7.85		2.11	5.78	8.55	1.95	18	8.55
P-536	0.72	0.55	2.21	1.2	0.13	1	1.29	0.14	0.73	-0.05	0.02	0.74		1.53	0.32	1.3	1.76	0.52	-0.67	-0.49	16	2.21
G-537	2.14	4.22	7.16	6.76	8.89		8.49	8.02	15.6	6.97	9.5	5.83	11	6.67	19.6	5.86	7.99	6.21	18.41	14.28	19	19.6
S-540	0.48	-0.29	0.74	-0.07	9.08	0.66	8.93	-0.69	1.08	0.91	-0.37	-0.68	1.16	0.97	0.49		-0.37	-0.25	18.14	13.96	12	18.14
W-585	1.8	1.66	1.93	1.97	1.47	1.76	1.85	1.6	2.02	1.72	1.68	1.66	1.77	1.93	1.95	1.71	1.79	1.72		0.97	19	2.02
H-586	0.67	0.46	2.07	1.51	0.37	0.73		0.65	0.31	-0.22	-0.77	1.08	1.03	0.92	0.82	0.72	0.38	0.29	-1.64	5.35	16	5.35
SARS-CoV																						
R-426	0.33	0.31	0.54	0.59	0.21	0.35	-0.04	0.35	0.22	0.32	0.31	0.3	0.36	0.33		0.35	0.31	0.31	0.33	0.33	18	0.59
Y-436	2.39	2.49	2.57	2.73	1.97	2.46	1.69	2.45	2.36	2.47	2.51	2.39	2.4	2.79	2.26	2.46	2.48	2.42	2.38		19	2.79
Y-440	0.44	0.1	0	-0.02	-0.19	0.44	0.14	0.08	0.16	0.05	-0.06	0.09	0	0.24	0.12	0.43	0.11	0.11	-0.29		15	0.44
Y-442	0.76	0.46	-0.87	0.29	-0.15	1.79	0.98	0.97	1.84	-0.18	-0.68	1.27	0.99	0.8	0.89	1.83	0.88	0.38	0.63		15	1.84
L-443	0.77	0.68	0.99	0.85	0.19	0.84	1.36	0.59	1.03		-0.16	0.7	0.88	1.04	0.6	0.79	0.55	0.62	2.1	-0.64	17	2.1
L-472	2.34	1.58	2.49	2.21	1.34	2.57	1.6	0.59	1.4		-0.36	2.37	1.85	1.38	1.59	2.55	1.75	1.14	1.65	2.12	18	2.57
N-473	1.01	1.1	-0.4	1.07	1.15	1.61	1.3	0.49	0.53	0.32	-0.32		1.28	1.36	1.38	1.15	1.03	1.12	1.12	0.64	17	1.61
Y-475	1.96	1.81	3.08	3.01	0.27	2.03	1.74	1.55	1.67	0.72	-0.02	2.24	2.01	2.32	1.12	2.28	1.79	1.9	0.42		18	3.08
N-479	0.51	0.04	0.2	0.53	-0.89	0.28	0.11	-0.59	-1.2	-0.89	-1.49		0.02	0.22	-0.86	0.47	-0.07	-0.46	-0.69	0.77	10	0.77
G-482	0.46	0.89	4.41	4.45	1.96		2.04	3.04	1.75	2.32	2.69	2.79	1.43	3.24	2.42	1.05	2.01	3.95	2.29	1.83	19	4.45
Y-484	1.48	1.16	2.43	2.38	-0.25	1.9	0.87	0.44	1.66	0.34	0.46	1.46	1.15	2.25	1.77	0.88	1.39	0.54	-0.33		17	2.43
T-486	0.14	0.02	1.43	0.65	-0.1	0.68	-0.5	1.93	-0.23	-0.56	-0.64	-0.38	0.2	-0.98	-0.1	-0.16		0.76	-0.37	0.34	9	1.93
T-487	0.52	0.67	2.2	1.03	1.68	2.06	12.44	0.6	2.77	3.08	1.78	2.32	1.04	2.53	7.25	0.78		-0.17	14.55	20.16	18	20.16
G-488	2.43	2.58	4.05	3.73	2.65		2.69	4.67	2.77	3.12	2.45	1.99	15.92	2.89	3.06	2.52	3.89	7.3	3.26	2.42	19	15.92
I-489	0.53	0.11	0.7	0.35	-0.52	0.4	0.26		0.01	-0.57	-0.22	0.13	-0.41	0.13	-0.12	0.21	0.15	-0.04	0.23	-0.01	12	0.7
Y-491	2.66	2.49	3.25	2.98	1.22	3.03	2.68	2.72	2.27	1.97	1.63	2.7	3.12	2.69	2.47	3.57	1.99	2.69	0.97		19	3.57
SARS-CoV-2																						
K-417	0.7	1.05	0.81	0.73	0.48	1	0.38	0.66		1.04	0.67	0.78	0.74	0.79	1.01	0.68	0.94	0.46	1.09	-0.73	18	1.09
G-446	0.05	0.05	0.02	0.53	0.57		0.05	0.02	0.08	0.02	-0.06	0.03	-0.02	-0.02	-0.72	0.04	0.04	0	0.06	0.03	15	0.57
Y-449	0.94	0.69	0.57	0.64	0.75	0.41	1.13	0.47	0.7	0.18	0.5	0.85	1.35	0.3	0.01	0.38	1.3	1.33	1.04		19	1.35
Y-453	0.14	0.15	0.12	-0.02	-0.38	0.14	0.02	-0.08	0.51	-0.08	-0.14	0.14	0.16	0.13	-1.09	0.14	0.15	0.16	0.04		13	0.51
L-455	1.59	1.1	2.08	2.26	4.88	1.16	2.13	0.91	0		0.27	1.92	1.36	2.31	1.6	1.26	1.17	1.08	3.04	6.58	19	6.58

TABLE 3 (Continued)

Residues	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Count[a]	Max
F-456	2.73	2.21	1.74	2.9		2.86	0.25	2.09	1.32	1.61	1.64	2.52	2.72	1.45	1.63	2.66	2.24	2.05	2.26	2.35	19	2.9
A-475		-0.08	0.12	1.14	4.42	0.53	2.7	0.84	4.72	3.13	0.53	-0.19	0.52	1.76	5.87	0.56	-0.08	3.55	10.36	10.58	16	10.58
F-486	2.96	2.31	2.82	2.05		3.07	2.3	1.58	2.08	0.94	0.52	2.67	2.57	2.29	1.96	3.12	2.57	2.03	1.17	-0.34	18	3.12
N-487	1.8	0.86	-0.3	1.08	6.52	2.05	1.88	1.49	1.61	0.82	1.26		2.1	1.19	1.77	1.66	1.41	1.35	2.53	9.94	18	9.94
Y-489	1.91	1.41	2.53	1.8	-0.43	2.13	1.42	0.64	1.1	0.79	0.53	2.14	1.76	1.65	2.41	1.18	1.96	0.93	-0.34		17	2.53
Q-493	0.19	-0.09	0.62	0.47	-2.11	-0.5	-0.02	-0.32	0.22	-0.9	-1.27	0.1	0.52		-0.52	0.36	0.06	-0.19	0.65	-2.25	9	0.65
G-496	-0.36	0.88	2.15	2.47	11.74		11.11	4.84	1.99	4.73	2.25	3.29		8.14	3.05	-0.68	2.71	1.19	4.34	5.83	16	11.74
Q-498	-1.11	-1.25	-1.15	-1.42	-2.59	-0.6	-1.66	-2.24	-1.05	-2.49	-2.41	-1.52	-0.92		-1.57	-1.81	-0.9	-1.84	-2.34	-3.36	0	-0.6
T-500	0.22	0.4	0.21	0.63	-1.38	-0.56	-0.83	-0.19	-1.74	-0.67	-0.86	-0.45	-0.06	-0.49	-0.28	0.06		-0.21	-0.7	-0.15	5	0.63
N-501	-1.16	-1.01	-1.3	-2.32	4.52	-0.87	-0.5	-1.44	-0.7	-3.16	-1.29		-0.69	-1.77	0.17	-0.99	-1.59	-2.02	1.76	3.04	4	4.52
G-502	2.26	2.19	3.56	3.28	2.1		2.31	3.35	2.46	1.97	2.37	2.63	14.45	2.75	2.38	3.17	4.53	3.38	2.59	2.2	19	14.45
Y-505	0.63	1.4	2.12	2.2	0.85	1.59	2.48	0.89	2.71	0.57	0.68	1.06	1.33	2.95	2.26	1.55	1.79	0.71	0.94		19	2.95

Note: Energy values are given in kcal/mol. The wild-type interaction energies are (1) NL63: −6.99 kcal/mol, (2) SARS-CoV: −18.81 kcal/mol, and (C) SARS-CoV-2: −16.89 kcal/mol.
[a]Number of mutations which increase the change in interaction energy (given in bold font).

ORCID

M. Michael Gromiha https://orcid.org/0000-0002-1776-4096

REFERENCES

1. Hofmann H, Pyrc K, Van Der Hoek L, Geier M, Berkhout B, Pöhlmann S. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc Natl Acad Sci*. 2005;102(22):7988-7993.
2. Li W, Moore MJ, Vasilieva N, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*. 2003;426(6965):450-454.
3. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273.
4. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020;581(7807):215-220.
5. Fouchier RA, Hartwig NG, Bestebroer TM, et al. A previously undescribed coronavirus associated with respiratory disease in humans. *Proc Natl Acad Sci*. 2004;101(16):6212-6216.
6. Van Der Hoek L, Pyrc K, Jebbink MF, et al. Identification of a new human coronavirus. *Nat Med*. 2004;10(4):368-373.
7. Chan-Yeung M, Xu RH. SARS: epidemiology. *Respirology*. 2003;8: S9-S14.
8. Khafaie MA, Rahim F. Cross-country comparison of case fatality rates of COVID-19/SARS-COV-2. *Osong Public Health Res Perspect*. 2020;11(2):74-80.
9. Roser M, Ritchie H, Ortiz-Ospina E, Hasell J. "Coronavirus pandemic (COVID-19)". https://ourworldindata.org/grapher/coronavirus-cfr. Accessed July 1, 2020.
10. Stadnytskyi V, Bax CE, Bax A, Anfinrud P. The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. *Proc Natl Acad Sci*. 2020;117(22):11875-11877.
11. Wu K, Li W, Peng G, Li F. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc Natl Acad Sci*. 2009;106(47):19970-19974.
12. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*. 2005;309(5742):1864-1868.
13. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020;367(6483):1260-1263.
14. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*. 2020;181(2):281-292.e6.
15. Burley SK, Berman HM, Bhikadiya C, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res*. 2019;47(D1):D464-D474.
16. The PyMOL Molecular Graphics System, Version 1.7.4, Schrödinger, LLC.
17. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780.
18. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189-1191.

19. Manavalan P, Ponnuswamy PK. Hydrophobic character of amino acid residues in globular proteins. *Nature*. 1978;275(5681):673-674.

20. Nozaki Y, Tanford C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions establishment of a hydrophobicity scale. *J Biol Chem*. 1971;246(7):2211-2217.

21. Jones DD. Amino acid properties and side-chain orientation in proteins: a cross correlation approach. *J Theor Biol*. 1975;50(1):167-183.

22. Delgado J, Radusky LG, Cianferoni D, Serrano L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics*. 2019;35(20):4168-4169.

23. Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res*. 2006;34 (suppl_2):W239-W242.

24. Chakraborti S, Prabakaran P, Xiao X, Dimitrov DS. The SARS coronavirus S glycoprotein receptor binding domain: fine mapping and functional characterization. *Virol J*. 2005;2(1):73.

25. Walls AC, Tortorici MA, Frenz B, et al. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat Struct Mol Biol*. 2016;23(10):899-905.

26. Grünberg R, Nilges M, Leckner J. Flexibility and conformational entropy in protein-protein binding. *Structure*. 2006;14(4):683-693.

27. Chen J, Sawyer N, Regan L. Protein–protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci*. 2013;22(4):510-515.

28. Ma B, Elkayam T, Wolfson H, Nussinov R. Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci*. 2003;100(10):5772-5777.

29. Raj VS, Mou H, Smits SL, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*. 2013;495(7440):251-254.

30. Chan CM, Lau SK, Woo PC, et al. Identification of major histocompatibility complex class IC molecule as an attachment factor that facilitates coronavirus HKU1 spike-mediated infection. *J Virol*. 2009;83(2):1026-1035.

31. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Sci*. 1997;6(1):53-64.

32. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. *Science*. 1995;267(5196):383-386.

33. Hu J, He CL, Gao Q, Zhang GJ, Cao XX, Long QX, Deng HJ, Huang LY, Chen J, Wang K, Tang N. The D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity and decreases neutralization sensitivity to individual convalescent sera. *bioRxiv* 2020. https://doi.org/10.1101/2020.06.20.161323.

34. Wec AZ, Wrapp D, Herbert AS, et al. Broad neutralization of SARS-related viruses by human monoclonal antibodies. *Science*. 2020;369 (6504):731–736. https://doi.org/10.1126/science.abc7424.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.