Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202104215
Chemistry Europe
European Chemical Societies Publishing
www.chemeurj.org

Computational Insights into the Allosteric Effect and Dynamic Structural Features of the SARS-COV-2 Spike Protein

Qiao Xue,[a] Xian Liu,[a] Wenxiao Pan,[a] Aiqian Zhang,*[a, b, c] Jianjie Fu,*[a, b, c] and Guibin Jiang[a, b, c]

Abstract: COVID-19 caused by SARS-COV-2 is continuing to surge globally. The spike (S) protein is the key protein of SARS-COV-2 that recognizes and binds to the host target ACE2. In this study, molecular dynamics simulation was used to elucidate the allosteric effect of the S protein. Binding of ACE2 caused a centripetal movement of the receptor-binding domain of the S protein. The dihedral changes in Phe329 and Phe515 played a key role in this process. Two potential cleavage sites S1/S2 and S2′ were exposed on the surface after the binding of ACE2. The binding affinity of SARS-COV-2 S protein and ACE2 was higher than that of SARS-COV. This was mainly due to the mutation of Asp480 in SARS-COV to Ser494 in SARS-COV-2, which greatly weakened the electrostatic repulsion. The result provides a theoretical basis for the SARS-COV-2 infection and aids the development of biosensors and detection reagents.

Introduction

A new member of the coronavirus family, severe acute respiratory syndrome coronavirus 2 (SARS-COV-2), has spread worldwide.[1] Coronavirus disease 2019 (COVID-19) caused by SARS-COV-2 had led to over 332 million confirmed cases and over 5.5 million deaths as of January 2022, as reported by the World Health Organization (WHO) (https://covid19.who.int). To date, the numbers of confirmed COVID-19 cases and fatalities has far exceeded those of the previous outbreaks caused by other coronaviruses, such as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), and the transmission speed is much faster than that of these two.[2] COVID-19 has symptoms such as fever, cough, and shortness of breath. In more severe cases, the infection can cause pneumonia, acute respiratory distress syndrome, and even death.[3] There are also a considerable number of people who are asymptomatic after being infected with SARS-COV-2, which has greatly increased SARS-COV-2 spread and complicated detection.[4]

During coronavirus spread, the spike (S) protein, which is located on the raised portion of the SARS-COV-2 virus surface, plays a key role in recognizing host cell receptors and in the fusion process of viral and cellular membranes.[5] The S protein forms a complex of three monomers, and each monomer contains 1273 amino acids.[6] The S protein can be cleaved into an amino (N)-terminal S1 subunit and a carboxyl (C)-terminal S2 subunit.[7] The S1 subunit is the main host cell receptor-binding region, and the S2 subunit is the region that forms the fusion peptide.[8] The S1 subunit includes four domains: the N-terminal domain (NTD), receptor-binding domain (RBD), C-terminal domain 1 (CTD1), and C-terminal domain 2 (CTD2). A previous study showed that the membrane fusion process of coronaviruses, such as SARS-COV, begins with the binding of the S1 subunit to angiotensin-converting enzyme 2 (ACE2), after which a protease cleaves the S1/S2 or S2′ cleavage site.[9] A dramatic conformational change occurs in this process to form the post-fusion state of the S protein, which initiates membrane fusion.[10] However, for the S protein of SARS-COV-2, we still know very little about its structure and allosteric characteristic. Research on its interactions with host cell receptors and the relevant molecular mechanism is urgently needed.

The current progress on SARS-COV-2 structural information is still very limited. Wrapp and Walls et al. obtained the crystal structure of S protein in different states by the cryo-EM method, which is of great significance for understanding the structure of S protein.[6,11] Yan et al. also obtained the structure of the S protein RBD and ACE2 and neutral amino acid transporter B0AT1 by the cryo-EM method, which is helpful for understanding the recognition of SARS-COV-2 in host cell infection.[12] Moreover, Yao and Ke et al. revealed the architecture of the

[a] *Dr. Q. Xue, X. Liu, W. Pan, Prof. A. Zhang, Prof. J. Fu, Prof. G. Jiang*
State Key Laboratory of Environmental Chemistry and Ecotoxicology
Research Center for Eco-Environmental Sciences
Chinese Academy of Sciences
Beijing 100085 (P. R. China)
E-mail: aqzhang@rcees.ac.cn
jjfu@rcees.ac.cn

[b] *Prof. A. Zhang, Prof. J. Fu, Prof. G. Jiang*
College of Resources and Environment
University of the Chinese Academy of Sciences
Beijing 100049 (P. R. China)

[c] *Prof. A. Zhang, Prof. J. Fu, Prof. G. Jiang*
School of Environment
Hangzhou Institute for Advanced Study
University of the Chinese Academy of Sciences
Hangzhou, 310000 (P. R. China)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202104215

SARS-COV-2 surface, which exhibits the distribution of S protein.[13] Yuan et al. provided structural information on the antibody interaction with SARS-COV-2.[14] All these reports are beneficial for understanding the infection mechanism of SARS-COV-2. These structural discoveries still remain at several nodes in the SARS-COV-2 infection process, and it is not enough to analyze the recognition and binding mechanism of SARS-COV-2 and host receptors.

During the recognition process of ACE2 by the S protein, the S1 subunit of the S protein will first change from the "closed" conformation to the "up" conformation, which is also named the prefusion state.[9,11] Then, the S1 subunit in the "up" conformation will recognize and bind to the host target ACE2. To date, there are several SARS-COV S protein crystal structures in different conformations, including the "closed", "up", and ACE2-bound states. To carry out studies of SARS-COV-2 on the interaction with ACE2 and the subsequent allosteric effect of the S protein, we combined the prefusion state of the S proteins of SARS-COV and SARS-COV-2 as templates to build single and ACE2-bound S protein structures. Molecular dynamics (MD) simulations were performed on the structures to monitor the stability and change of S protein. The potential cleavage site was analyzed, and its structural character was analyzed. Then, the interaction mechanism between the S protein and ACE2 was studied, and the results were compared between SARS-COV and SARS-COV-2. Finally, the allosteric effect of S protein induced by the binding of ACE2 was analyzed. The study provides a structural basis and molecular mechanism of the S protein and provides a theoretical basis for comprehensively understanding the characteristics of SARS-COV-2 and for developing related vaccines and detection kits.

Results

Structural basis of the SARS-COV-2 S protein

The homotrimer structure of the S protein is similar in SARS-COV and SARS-COV-2. In the prefusion conformation of the S protein trimer, two monomers are in the "closed" state and are named monomers A and B. Another monomer in the "up" state is named monomer C. In this state, the RBD of monomer C is convex compared with the other monomers. The global and per-residue model quality was assessed and presented a good quality evaluation (Figures S1 and S2 in the Supporting Information). A 250-ns MD simulation was performed to evaluate the structural stability. The root mean square deviation (RMSD) curve presented a relatively high value (5 Å) but a narrow fluctuation range (Figure S3A). Root mean square fluctuation (RMSF) showed that the fluctuations among residues had clear differences (Figure S3B). Most of the residues had RMSF values less than 4 Å, whereas several surface residues had more than 5 Å RMSF values. The narrow fluctuation range of the RMSD and the low-value region of RMSF showed that the global structure is stable, while the increased RMSD value and the high value region of RMSF showed that local regions had structural changes. In addition, there are 42 disulfide bonds in the S protein structure, which are beneficial for maintaining the stability of the structure (Figure S4).

To identify structural changes, cluster analysis was performed (Figure S5). By comparing the initial conformation with the representative stable conformation, it could be seen that the main structural differences existed in the S1 subunit and that the S2 subunit remained stable during simulation (Figure S5.A). The main structural change in the S1 subunit was concentrated in the RBD. The structural change of monomer A and B was inconspicuous. Both the S1 and S2 subunits of monomer C had higher RMSD values than those of monomers A and B. As the RMSD of S1 of monomer C was high, we needed to confirm whether the up state has changed. We defined a long axis with the outer and inner sides of the RBD structure, and defined the central axis with the normal vector perpendicular to the trimer plane (Figure S5). The angle between the two axes was monitored during the simulation. The angles of monomers A and B were relatively consistent and remained at about 75°, while the angle of monomer C was stable at about 40°. This result showed that although the monomer C has a relatively large RMSD value, it still remained stable in the "up" state. Generally, the highly fluctuating region had a good induced-fit effect on the protein-protein interaction. We speculate that the variable and convex RBD region could help SARS-COV-2 recognize and bind ACE2. More importantly, the simulation results show that after SARS-COV-2 enters the human body, the S protein that can be stabilized in the "up" state for a long time has a higher recognition efficiency for ACE2 than that in the "closed" state, which is important for the high infection efficiency of SARS-COV-2.[15]

Dynamic structural features of cleavage sites

Fusion activation is usually initiated by proteolytic cleavage. Investigating the cleavage mechanism of S protein is important to understand the viral pathogenesis process. As little is known about the SARS-COV-2 S protein, we analyzed the dynamic structural features of cleavage sites obtained by the ProP 1.0 server.[16] The cleavage sites were also confirmed by previous work.[17] The first cleavage site (–RRAR–) was located between residues 682 and 685, which was not observed in the SARS-COV S protein. A previous study showed that the SARS-COV protein had a second cleavage site named S2', which played an important role in the SARS-COV S protein fusion process. Alignment of the sequences showed that SARS-COV-2 also had this site. This cleavage site was located on residues 814 and 815 (–KR–). Structural information showed that the positions of the S1/S2 sites in three monomers were different (Figure 1B). The S1/S2 site in monomer C was closer to the NTD domain than that in monomers A and B. The S1/S2 site in monomer C had clearly lower *B*-factor values than that in monomers A and B, which indicated that this site in the "up" conformation was more stable than that in the "closed" conformation. The electrostatic potential of the S1/S2 site also presented diversity in the two conformations. In the "closed" conformation, the S1/S2 site showed a spike-like positive potential surface. However,

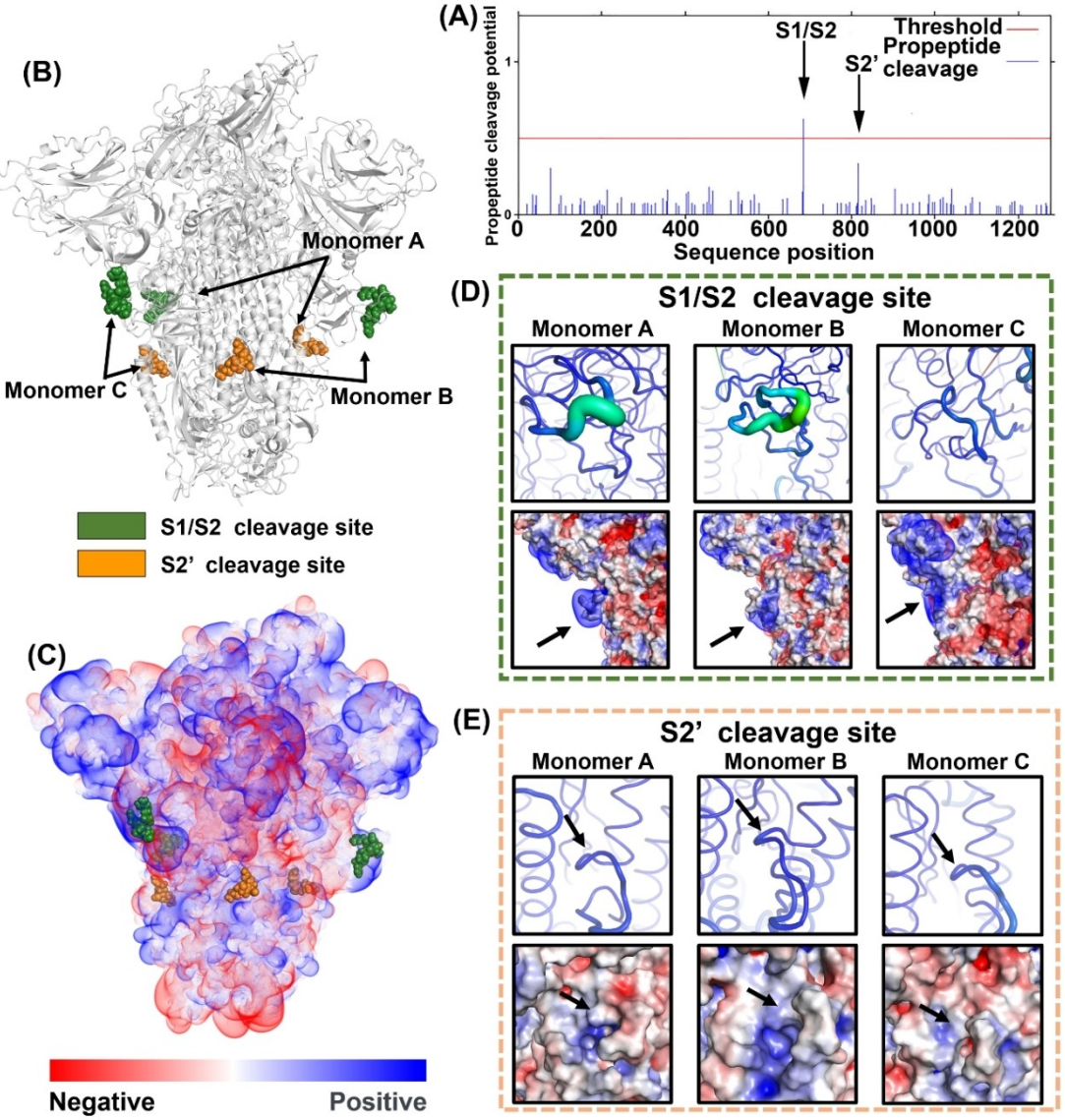

Figure 1. A) The cleavage sites S1/S2 and S2′ in the S protein sequence. B) The positions of the cleavage sites. C) The electrostatic potential surface of the S protein; the cleavage sites are highlighted. D) Top: the *B*-factors of the S1/S2 site in each monomer; a thicker line indicates a higher value. Bottom: the corresponding electrostatic potential surfaces. E) Top: the *B*-factors of the S2′ site in each monomer. Bottom: the corresponding electrostatic potential surfaces.

in the "up" conformation, the positive electrostatic potential surface of the S1/S2 site was smooth and connected to the positively charged region of the NTD (Figure 1C and D). The position of the S2′ site was similar in each monomer. The structure of the site had small fluctuations, and the electrostatic potential surface of the site was located in the sunk area (Figure 1C and E). The results indicate that the S1/S2 site was more likely to be the initial cleavage site. The S2′ site may not be cleaved from the beginning which is similar as SARS-COV. Compared with SARS-COV, the additional cleavage site of the S protein in SARS-COV-2 further enhance its infectivity in the human body, which is also presented in other work.[17b]

Binding mechanism comparison of the SARS-COV-2 and SARS-COV S proteins with ACE2

The binding free energy between the S protein and ACE2 was evaluated by MM-GBSA (Table 1). Clearly, the binding affinity between the S protein and ACE2 is higher for SARS-COV-2 (-26.35 kcal mol^{-1}) than that for SARS-COV (-18.91 kcal mol^{-1}). These results reflect that when the virus enters the human body, SARS-COV-2 is more likely to bind to host cell receptors than SARS-COV, which may be the reasons that SARS-COV-2 is more infectious than SARS-COV. By decomposing energy into different terms, it can be seen that electrostatic energy and polar solvation energy are the most significant differences between the two coronavirus S proteins. The electrostatic

Table 1. Binding free energy between the S protein and ACE2 obtained by MM-GBSA.

	Binding free energy [kcal/mol] SARS-COV-2	SARS-COV
E_{vdw}	−88.43 ± 5.31	−97.69 ± 6.14
E_{ele}	−932.18 ± 42.87	−588.20 ± 46.53
E_{GB}	1007.18 ± 42.15	680.55 ± 43.90
E_{surf}	−12.91 ± 0.76	−13.57 ± 0.70
E_{polar} [1]	75.00	92.35
$E_{nonpolar}$ [2]	−101.34	−111.26
ΔH	−26.35 ± 6.63	−18.91 ± 11.02

1. $E_{polar} = E_{ele} + E_{GB}$; 2. $E_{nonpolar} = E_{vdW} + E_{surf}$.

interaction between the S protein of SARS-COV-2 and ACE2 (−932.18 kcal mol^{-1}) is much stronger than that of the SARS-COV protein and ACE2 (−588.20 kcal mol^{-1}). Due to the large unfavorable polar solvation energy, the total polar binding free energy (E_{polar}) is unfavorable for binding. However, the total polar binding free energy in SARS-COV-2 (75.00 kcal mol^{-1}) was more favorable than that in SARS-COV (92.35 kcal mol^{-1}). In both coronaviruses, the dominant binding affinity comes from nonpolar binding free energy ($E_{nonpolar}$). The nonpolar energy in SARS-COV (−111.26 kcal mol^{-1}) is more favorable than that in SARS-COV-2 (−101.34 kcal mol^{-1}). Through structural analysis, the binding interface of S protein is a long loop in the RBD, and the interface of ACE2 is an α-helix. The flexibility of the loop makes the S protein have an induced-fit effect in the binding process (Figures 2 and S3). The surface properties of the S proteins of the two coronaviruses are quite different. The electrostatic potentials of the interface surfaces for ACE2, the S protein in SARS-COV and that in SARS-COV-2 were negative, negative and positive, respectively (Figure 2B). Notably, there is a favorable positive–negative electrostatic interaction between SARS-COV-2 and ACE2 and an unfavorable negative-negative electrostatic interaction between SARS-COV and ACE2.

The binding free energy was further decomposed into each residue to obtain the key residues of the binding process. The residues with an energy contribution greater than −1 kcal mol^{-1} were important (Tables S1 and S2). The S proteins of SARS-COV-2 and SARS-COV had 13 and 14 important residues, respectively. For SARS-COV-2, Gln493 and Tyr505 offered the most significant energy contributions. The two phenylalanine residues, Phe456 and Phe486, which do not exist in SARS-COV, also played important roles. The key residues of ACE2, including Asp355, Tyr41, Thr27, Phe28 and Gln24, were similar when binding to either SARS-COV-2 or SARS-COV. Lys417 was a newly appeared residue in SARS-COV-2. It formed a salt bridge with Asp30 of ACE2, which was verified by other report.[18] This favorable electrostatic interaction was reflected in the surface electrostatic potential of S protein and ACE2 (Figure 2B). The corresponding residue of SARS-COV was Val404, which caused the significant difference in the electrostatic potential surface of the S proteins of the two coronaviruses. We further analyzed the residues that are unfavorable for binding (Figure 2C and Table S3). The electrostatic interaction between Asp480 (10.02 kcal mol^{-1}) of SARS-COV and Asp38 (6.22 kcal mol^{-1}) and Glu35 (3.97 kcal mol^{-1}) of ACE repels the two molecules, which results in a very large unfavorable binding energy. However, in SARS-COV-2, Asp480 is replaced by Ser494, which has a short and noncharged side chain. In contrast to that of Asp480, the side chain of Ser494 does not point toward ACE2 and enlarges the distance to Glu35 of ACE2 from approximately 5 Å to approximately 9 Å (Figure S6). Thus, the corresponding binding energy decreased from 10.02 to 0.43 kcal mol^{-1} for Asp480 and Ser494, respectively. Correspondingly, the energies of Asp38 and Glu35 in ACE2 decreased from 6.22 and 3.97 kcal mol^{-1} to −0.39 and −0.56 kcal mol^{-1}, respectively. This result is consistent with the surface electrostatic potential (Figure 2B). Compared with that with SARS-COV, Ser19 of ACE2 also had a smaller unfavorable binding affinity when ACE2 bound to SARS-COV-2. In addition, Lys390 in SARS-COV presented an unfavorable energy contribution, and the corresponding Arg403 in SARS-COV-2 decreased the unfavorable binding affinity. In fact, nearly all of the unfavorable residues in SARS-COV decreased the binding energy in the SARS-COV-2-ACE2 complex (Table S3). Compared with SARS-COV, the mutation of several amino acids in SARS-COV-2 enhanced the binding affinity to host target ACE2.

The centripetal movement of ACE2 and the RBD to the S protein trimer

When ACE2 bound to the S protein in the "up" state, the ACE2 and RBD regions had obvious structural change. Through the overlap of the initial conformation with the conformation at 50 ns, it can be seen that the ACE2 and RBD regions had obvious centripetal movement (Figure 3). The RMSD values of ACE2 and the three S protein monomers did not change much, but the RMSD value of the overall complex changed significantly. This indicated that the movement of the RBD and ACE2 was an integral movement rather than the internal dissociation (Figure S7). The structural biology research has obtained the conformation of the SARS-COV S protein before and after binding to ACE2, which was named as "up" conformation and the "ACE2-bound 1" conformation, respectively.[10] We superimposed the initial conformation with the 50 ns conformation and superimposed the SARS-COV S protein in the "up" state with that in the "ACE2-bound 1" state. The allosteric effects between SARS-COV-2 and SARS-COV were similar (Figure 3A and B). After binding to ACE2, the RBD regions of SARS-COV and SARS-COV-2 had centripetal movement. The essential dynamics analysis (EDA) also showed that during 0–50 ns, the main movement of ACE2 and the RBD was from outside to inside (Figure 3C). To quantitatively characterize this movement, we monitored the distance between the RBD of monomers C and B on the opposite side. After 5 ns, the distance between Asn481(C) and Asn370(B) began to gradually decrease from 30 to 10 Å within 34 ns. Then, the distance fluctuated at about 10 Å (Figure 3F). All these results showed that after the binding of ACE2, the S protein of SARS-COV-2 quickly changed from the "up" state to the "ACE2-bound 1" state. Since we performed long time MD simulation for the S protein trimer without ACE2 bound, the RBD remained stable during the simulation time

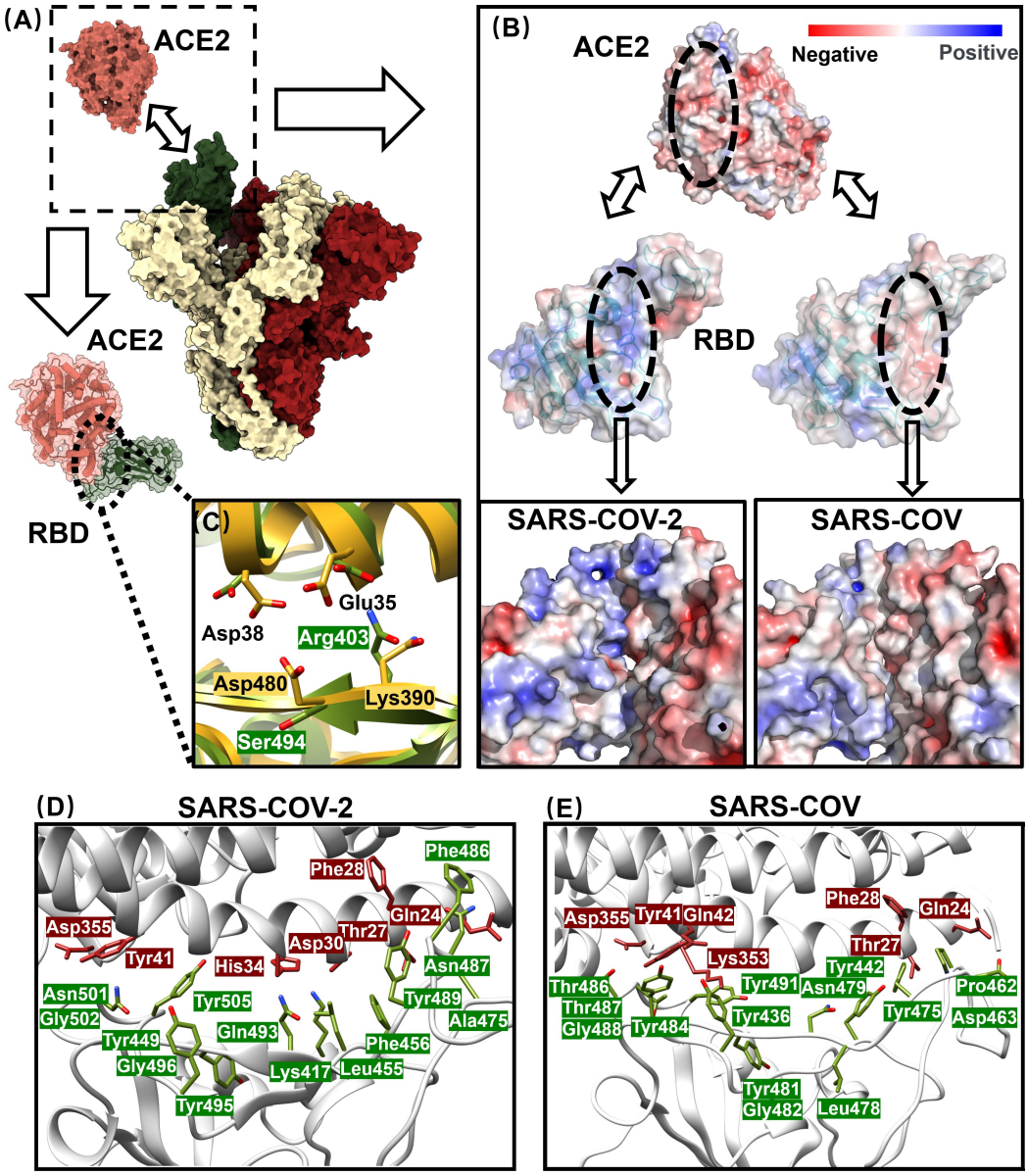

Figure 2. A) Binding diagram of ACE2 (pink) and the S protein; monomers A, B and C are red, yellow, and green, respectively. B) The electrostatic potential surface of the interface (labeled by black lines) between ACE2 and the S protein RBD, and the bottom shows the bound state: left: SARS-COV-2, right: SARS-COV. C) Residues with highly unfavorable energy contributions in ACE2-bound SARS-COV (yellow) and the corresponding residues in SARS-COV-2 (green). D) The key residues in ACE2-bound SARS-COV-2. E) The key residues in ACE2-bound SARS-COV-2. (protein: brown; key residues of ACE2: red; key residues of SARS-COV-2/SARS-COV: green).

(Figure S5). We confirmed that the centripetal movement of the RBD was triggered by the binding of ACE2.

To explore the cause of this allosteric effect, we analyzed the binding energy between ACE2 and the S protein (Figure 3G and H). The results showed that both electrostatic energy and van der Waals energy gradually became stronger after the binding of ACE2. The van der Waals energy changed from −60 to approximately −80 kcal mol^{-1}, and the electrostatic energy changed from −140 to approximately −190 kcal mol^{-1}. The surface electrostatic potential analysis showed that the positively charged area of the CTD1 tended to move to the negatively charged area of the opposite monomer B (Figure S8). The long-range impact of binding energy was also shown in other work.[17b] These results indicate that the centripetal movement caused by the binding of ACE2 is a thermodynamically favorable spontaneous movement and that its movement can help enhance the binding affinity of S protein to ACE2.

To further investigate the molecular mechanism of this allosteric effect, we used various methods to study the intermolecular interactions of the allosteric process. The anisotropic network model (ANM) results showed that the main movement areas were concentrated in the RBD and ACE2 (Figure 3D). The mobility analysis obtained by EDA showed that the RBD, ACE2 had high mobility, which is consistent with the

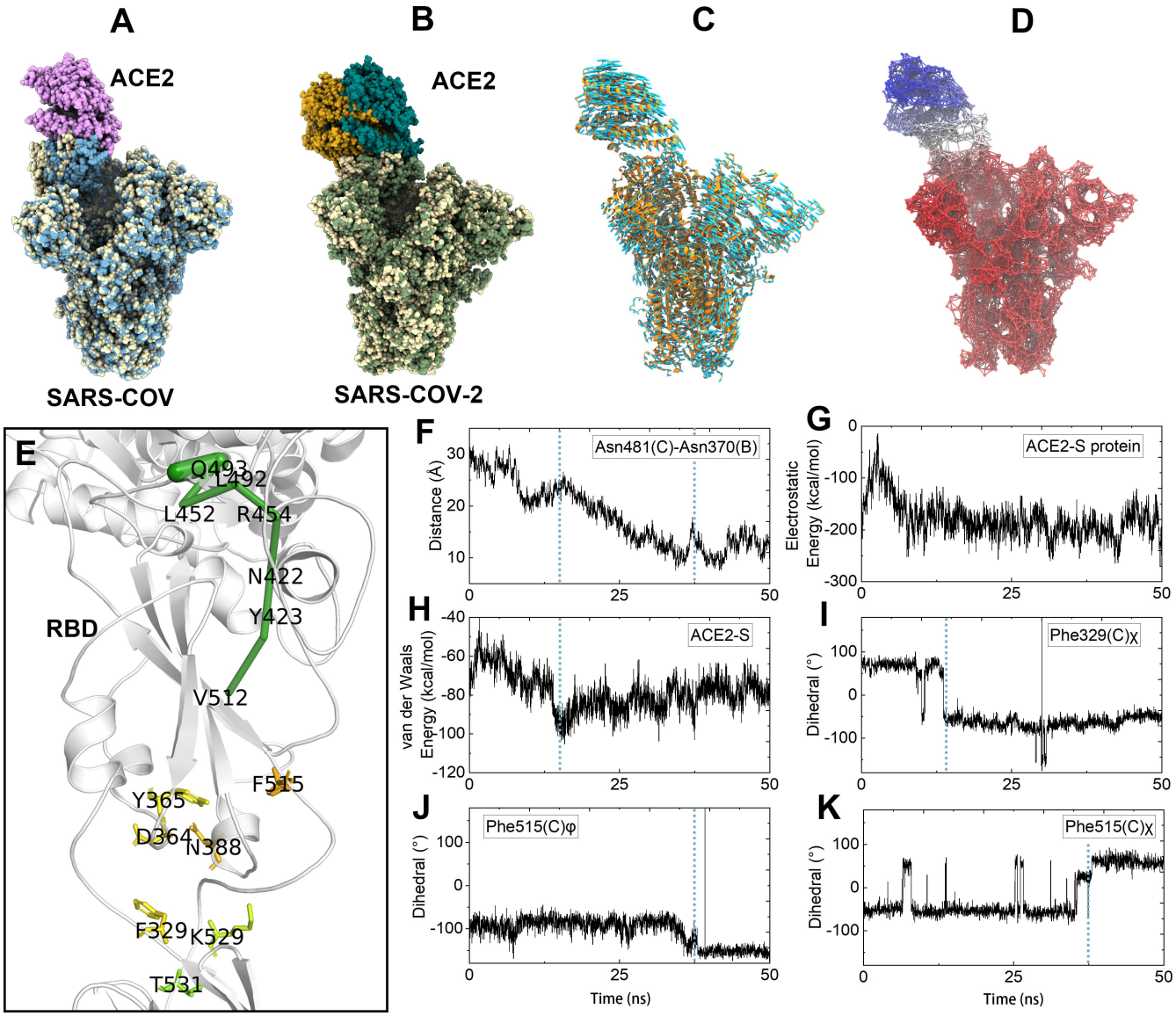

Figure 3. A) Comparison of the "up" state (yellow) and "ACE2-bound 1" (S protein: blue, ACE2: pink) state of SARS-COV. B) Comparison of the initial binding state (S protein: yellow, ACE2: orange) and "ACE2-bound 1" state (S protein: green, ACE2: cyan) of SARS-COV-2. C) EDA of ACE2-bound S protein (protein: orange, direction of movement: cyan arrows). D) ANM analysis. Large inter-residue fluctuations are blue and small fluctuations are red. E) The residues of RBD with obvious dihedral changes are shown in a stick model. The allosteric pathway of RBD is portrayed in green. F) Distance between Asn481 of monomer C and Asn370 of monomer B. G) Electrostatic binding energy between ACE2 and the S protein. H) Van der Waals binding energy between ACE2 and the S protein. I) The χ dihedral angle of Phe329. J) The ϕ dihedral of Phe515. K) The χ dihedral of Phe515. The important time points are labeled in blue lines.

ANM result (Figure S9). A new method, "Ohm", which is based on the perturbation propagation algorithm, is used to identify and characterize the allosteric communication of proteins.[19] The allosteric potential of amino acids is reflected by the allosteric coupling intensity (ACI) values. Among the RBD residues, 53 residues had ACI values larger than 0.7, which represents a high potential allosteric effect (Table S4). The dihedral, which was an important mark of structural change, was calculated for all 53 residues. The results showed that the dihedral of several residues, including Phe329, Asp364, Tyr365, Asn388, Phe515, Lys529 and Thr531, changed during the simulation (Figures 3I, J, K and S10). In these residues, Phe515 and Phe329, which are located at the two terminals of the RBD, play important roles in the movement of the RBD. It should be noted that these dihedral angle changes were concentrated at two points in time. The first time was at approximately 15–17 ns; the χ angle of Phe329 as well as the ϕ or ψ angles of Lys529 and Thr531 changed. The second time was at approximately 34 ns; the ϕ or ψ angles of Asn364, Tyr365 and Phe515 and the χ angle of Asn388 Phe515 changed. These two time points are in response to changes in binding energy and the distance between monomers C and B (Figure 3F, G and H). In addition, Phe329, Asp364, Tyr365 and Phe515 had high positive correlations. Phe329, Lys529 and Thr531 had positive correlations (Figure S11). Notably, all these residues are located near the terminus of the RBD with CTD1 and the NTD. Phe329,

Asp364, Tyr365, Phe515, Lys529 and Thr531 all located in the loop region of RBD. Their dihedral angle changes affected the structure of RBD, but did not influence the main secondary structures of the RBD. The change in dihedral angle was a key factor in inducing conformational transformation. After characterizing the key residues that affect the allosteric effect, another question is how ACE2 binding affects these residues. To clarify this question, we analyzed the allosteric pathway through Ohm (Figure 3E). On the one hand, the binding of ACE2 directly affected the nearby residues. On the other hand, the binding site affected internal amino acid allosteric regulation through the allosteric pathway of "Gln493-Leu452-Leu492-Arg454-Asn422-Tyr423-Val512". Val512 was located in the lower part of the secondary structure region of RBD and had very high ACI values (0.88), exhibiting high allosteric character. Val512 had high positive correlations with Phe329, Asp364, Tyr365 and Phe515 (Figure S11).

From the above analysis, we can infer that ACE2 binding affects the dihedral changes in residues in the RBD from the outside to the inside through the allosteric pathway and finally triggers the RBD movement. We not only observed the whole process of the S protein transitioning from the "up" state to the "ACE2 bound 1" state but also indicated the molecular mechanism of this process and the inherent thermodynamic factors.

The structural movement after "ACE2-bound 1" state

For SARS-COV, after reaching the "ACE2-bound 1" state, the RBD and ACE2 regions swing outward again. The final swing amplitude even exceeds the "up" state, reaching the "ACE2-bound 2" state.[10] This process should be complicated; otherwise, the RBD and ACE2 should move directly from the "up" state to the "ACE2-bound 2" state instead of going through the "ACE2-bound 1" state first. Considering the similarity of SARS-COV-2 and SARS-COV, SARS-COV-2 might also encounter unknown energy barriers during its movement from the "ACE2-bound 1" to the next state. Therefore, after reaching the "ACE2-bound 1" state, we adopted the GaMD method to enhance the sampling efficiency. The RMSD values of the classic MD to GaMD present clear difference for the two methods (Figure S12). EDA showed that the main movement was the lateral swing of the ACE2-RBD regions after the "ACE2-bound 1" state (Figure 4). In addition, the NTD regions, focusing on residues 17–43, and 57–267, had obvious allosteric movement, especially in monomers B and C. The movement further affected the conformations of the cleavage site. For monomer A, the S1/S2 cleavage site became relatively smooth, while it was convex in the "up" state. The S2′ cleavage site was still not exposed on the surface (Figures 1 and 4). For monomer B, the S1/S2 cleavage site was totally convex. The S2′ cleavage site was covered by the loop of the S2 subunit. For monomer C, the S1/S2 cleavage site became convex. The most significant difference was that the S2′ cleavage site was completely exposed on the surface. The electrostatic potential surface clearly reflected the above results (Figure 4).

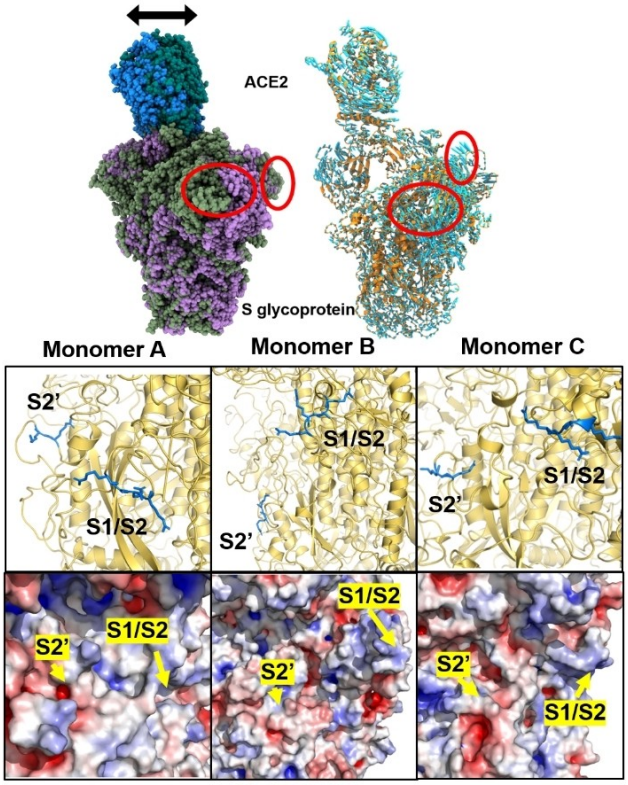

Figure 4. Top left: Comparison of the "ACE2-bound 1" conformation (S protein: green, ACE2: cyan) and the final state based on GaMD (S protein: purple, ACE2: blue). Right: the EDA (protein: orange, the movement direction: cyan arrow). Middle: The positions of the two cleavage sites S1/S2 and S2′ (blue) in the final state. Bottom: The electrostatic potential surfaces of S1/S2 and S2′ in the final state.

Discussion

COVID-19 is an ongoing pandemic caused by SARS-COV-2. Unlike other coronaviruses, SARS-COV-2 has shown greater spread in environment and pathogenicity to human health, which has brought difficulties to the prevention and control of the epidemic. In view of the importance of the S protein in the SARS-COV-2 infection process, we performed MD simulations on the S protein in the "up" state and the "ACE2-bound" complex. The current crystal structures show that the S protein has a "closed" state and an "up" state.[11] The S protein must be in the "up" state to be able to bind to the host receptor. The simulation results show that the S protein can be stabilized in the "up" state, which is obviously beneficial to its recognition and binding with host receptors ACE2. The binding of ACE2 initiate the centripetal movement of the S protein RBD in the "up" state and finally change to the "ACE2-bound 1" state, which is similar as SARS-COV.[10] The subsequent long-range GaMD simulation showed that the "ACE2-bound 1" failed to transform into the "ACE2-bound 2" conformation, which is not similar to SARS-COV. There might be a large energy barrier between the "ACE2-bound 1" state and the next state. In any case, this state is an important intermediate state in the SARS-COV-2 infection process, and our results indicate that SARS-

COV-2 can be stabilized in this "semi-fusion" state for a long time. Many news reports and studies have reported that a large number of people who are infected by SARS-COV-2 might be asymptomatic.[4,20] The long incubation period of COVID-19 and its asymptomatic infection suggest that SARS-COV-2 may not fuse quickly after entering the body. According to our results, it is possible that when SARS-COV-2 enters the body to bind ACE2, it quickly changes into the "ACE2-bound 1" state and stay in this state. Before entering the next stage from "ACE2-bound 1", the S protein-ACE2 complex might take a long time or require other factors to participate, which need further research. Recently, multiple new SARS-COV-2 variants has been found in many countries. In the United Kingdom, a variant, B1.1.17, emerged that might be associated with increased risk of death.[21] The mutation sites in this variant, including His69/Val70 and Tyr144, presented obvious allosteric movement. The variants B.1.351 and P.1, which were first found in South Africa and Brazil, respectively, have also been detected in many other countries. The mutation sites in these variants, including Leu18, Thr20, Phe26, Asp80, Asp138, Arg190, Asp215, Glu484, Lys417 and His655, all had obvious structural movement after ACE2 binding. The result of GaMD showed that these positions had high mobility after the "ACE2 bound 1" state, which indicated a correlation between the mutants and the ACE2 binding. The influence of the mutants on the spread and infection of SARS-COV-2 need further investigation.

Compared with S protein of SARS-COV, the ability of S protein evolves in SARS-COV-2. The newly emerging "-RRAR-" site at S1/S2 and the surface exposure of two sites in the "ACE2-bound 1" state can help SARS-COV-2 to be more easily recognized by proteases and initiate the fusion process than SARS-COV. By comparing the binding process of SARS-COV and SARS-COV-2 with ACE2, we found that the binding affinity of SARS-COV-2 to ACE2 was significantly improved. This difference was mainly due to the mutation of Asp480 of the S protein in SARS-COV to Ser494 in SARS-COV-2, which greatly weakened the electrostatic repulsion with Glu35 and Asp38 of ACE2. The analysis of binding free energy not only helps us to identify the important amino acids in the binding process, but also reminds us that there are still unfavorable binding amino acids in the S protein of SARS-COV-2, indicating that the virus may still continue to evolve. In addition, the analysis of amino acids can help engineer ACE2, which is useful to develop more effective biosensors or detection reagents. The modification of ACE2 can be performed by mutating amino acids with low binding affinity to high binding affinity or by mutating the unfavorable amino acids to favorable amino acids. Chan et al. reported that the mutation of T27Y and H34A can enhance the binding ability of ACE2 with the S protein, and these amino acids were identified as important residues in our analysis.[22] In addition, our results have found a variety of other amino acids that are unfavorable for binding. Whether these amino acid mutations can produce proteins with stronger binding affinity remains to be further studied. In conclusion, our results reveal the molecular mechanism of the allosteric effect of the S protein before and after ACE2 binding, which provides a theoretical basis for the investigation of the latent, infectious, and pathogenic mechanisms of SARS-COV-2. The results are also helpful for the development of biosensors and detection reagents.

Conclusion

In view of the importance of the S protein in the SARS-COV-2 infection process, we performed MD and GaMD simulations on the S protein in the "up" state and the "ACE2-bound" complex. The results showed that the S protein could be stabilized in the "up" state. In this process, the S1/S2 site was convex in the two "closed" monomers, whereas the site in the "up" monomer was smooth. S2′ was hidden inside the S protein. Comparing the binding process of SARS-COV and SARS-COV-2 with ACE2, we found that the binding affinity of SARS-COV-2 to ACE2 was significantly improved. This was mainly due to mutation of Asp480 of the S protein in SARS-COV to Ser494 in SARS-COV-2, which greatly weakened the electrostatic repulsion with Glu35 and Asp38 in ACE2. In addition, this study also identified key residues in the binding process of S protein and ACE2. Finally, this study found that the binding of ACE2 could promote the centripetal movement of the S protein RBD in the "up" state and finally change to the "ACE2-bound 1" state, which was similar to the changes in SARS-COV. The binding of ACE2 affected the internal residues of the RBD, especially Phe329 and Phe515, through a long-range allosteric pathway. The dihedral changes in these residues play an important role in the movement of the RBD. A subsequent long-range GaMD simulation showed that the "ACE2-bound 1" conformation failed to transform into another conformation; this is different with SARS-COV. The two potential cleavage sites, S1/S2 and S2′ on "up" monomer were completely exposed on the surface during the process. The NTD, which contained several pathogenic mutation sites, had high structural fluctuation in this process. In conclusion, the results revealed the molecular mechanism of the allosteric effect of the S protein before and after ACE2 binding, which provides a theoretical basis for the investigation of the latent, infectious, and pathogenic mechanisms of SARS-COV-2. The identification of key amino acids also aids the development of more effective biosensors and detection reagents.

Experimental Section

See the Supporting Information for computational details.

Acknowledgements

This study was supported by the National Key Research and Development Program of China (2018YFA0901101, 2020YFA0907500), the National Natural Science Foundation of China (22022611, 91743204, 92043302), the Strategic Priority Research Program of Chinese Academy of Sciences (XDPB2005)

and the Youth Innovation Promotion Association of Chinese Academy of Sciences (2018052).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: ACE2 · allosterism · molecular dynamics · SARS-COV-2 · spike proteins

[1] A. E. Gorbalenya, S. C. Baker, R. S. Baric, R. J. de Groot, C. Drosten, A. A. Gulyaeva, B. L. Haagmans, C. Lauber, A. M. Leontovich, B. W. Neuman, D. Penzar, S. Perlman, L. L. M. Poon, D. V. Samborskiy, I. A. Sidorov, I. Sola, J. Ziebuhr, *Nat. Microbiol.* **2020**, *5*, 536–544.
[2] D. Klingelhofer, M. Braun, D. Bruggmann, D. A. Groneberg, *J. Glob. Health* **2020**, *10*, 020508.
[3] Z. Y. Wu, J. M. McGoogan, *JAMA J. Am. Med. Assoc.* **2020**, *323*, 1239–1242.
[4] Q. X. Long, X. J. Tang, Q. L. Shi, Q. Li, H. J. Deng, J. Yuan, J. L. Hu, W. Xu, Y. Zhang, F. J. Lv, K. Su, F. Zhang, J. Gong, B. Wu, X. M. Liu, J. J. Li, J. F. Qiu, J. Chen, A. L. Huang, *Nat. Med.* **2020**, *26*, 1200–1204.
[5] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Kruger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N. H. Wu, A. Nitsche, M. A. Muller, C. Drosten, S. Pohlmann, *Cell* **2020**, *181*, 271–280.
[6] A. C. Walls, Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Veesler, *Cell* **2020**, *181*, 281–292.
[7] a) W. B. Tai, L. He, X. J. Zhang, J. Pu, D. Voronin, S. B. Jiang, Y. S. Zhou, L. Y. Du, *Cell. Mol. Immunol.* **2020**, *17*, 613–620; b) G. Lu, Y. Hu, Q. Wang, J. Qi, F. Gao, Y. Li, Y. Zhang, W. Zhang, Y. Yuan, J. Bao, B. Zhang, Y. Shi, J. Yan, G. F. Gao, *Nature* **2013**, *500*, 227–231; c) F. Li, W. H. Li, M. Farzan, S. C. Harrison, *Science* **2005**, *309*, 1864–1868.
[8] S. Belouzard, V. C. Chu, G. R. Whittaker, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5871–5876.
[9] A. C. Walls, M. A. Tortorici, J. Snijder, X. Xiong, B.-J. Bosch, F. A. Rey, D. Veesler, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11157–11162.
[10] W. Song, M. Gui, X. Wang, Y. Xiang, *PLoS Pathog.* **2018**, *14*, e1007236.
[11] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, *Science* **2020**, *367*, 1260–1263.
[12] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, *Science* **2020**, *367*, 1444–1448.
[13] a) H. Yao, Y. Song, Y. Chen, N. Wu, J. Xu, C. Sun, J. Zhang, T. Weng, Z. Zhang, Z. Wu, L. Cheng, D. Shi, X. Lu, J. Lei, M. Crispin, Y. Shi, L. Li, S. Li, *Cell* **2020**, *183*, 730–738; b) Z. Ke, J. Oton, K. Qu, M. Cortese, V. Zila, L. McKeane, T. Nakane, J. Zivanov, C. J. Neufeldt, J. M. Lu, J. Peukes, X. Xiong, H.-G. Kräusslich, S. H. W. Scheres, R. Bartenschlager, J. A. G. Briggs, *Nature* **2020**, *588*, 498–502.
[14] M. Yuan, N. C. Wu, X. Zhu, C.-C. D. Lee, R. T. Y. So, H. Lv, C. K. P. Mok, I. A. Wilson, *Science* **2020**, *368*, 630–633.
[15] Z. Lv, Y.-Q. Deng, Q. Ye, L. Cao, C.-Y. Sun, C. Fan, W. Huang, S. Sun, Y. Sun, L. Zhu, Q. Chen, N. Wang, J. Nie, Z. Cui, D. Zhu, N. Shaw, X.-F. Li, Q. Li, L. Xie, Y. Wang, Z. Rao, C.-F. Qin, X. Wang, *Science* **2020**, *369*, 1505–1509.
[16] P. Duckert, S. Brunak, N. Blom, *Protein Eng. Des. Sel.* **2004**, *17*, 107–112.
[17] a) A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Veesler, *Cell* **2020**, *181*, 281–292; b) B. Qiao, M. Olvera de la Cruz, *ACS Nano* **2020**, *14*, 10616–10623.
[18] S. Polydorides, G. Archontis, *Biophys. J.* **2021**, *120*, 2859–2871.
[19] J. Wang, A. Jain, L. R. Mcdonald, C. Gambogi, N. V. Dokholyan, *Nat. Commun.* **2020**, *11*, 3862.
[20] J. He, Y. Guo, R. Mao, J. Zhang, *J. Med. Virol.* **2021**, *93*, 820–830.
[21] B. Meng, S. A. Kemp, G. Papa, R. Datir, I. A. T. M. Ferreira, S. Marelli, W. T. Harvey, S. Lytras, A. Mohamed, G. Gallo, N. Thakur, D. A. Collier, P. Mlcochova, L. M. Duncan, A. M. Carabelli, J. C. Kenyon, A. M. Lever, A. De Marco, C. Saliba, K. Culap, E. Cameroni, N. J. Matheson, L. Piccoli, D. Corti, L. C. James, D. L. Robertson, D. Bailey, R. K. Gupta, *Cell Rep.* **2021**, *35*, 109292.
[22] K. K. Chan, D. Dorosky, P. Sharma, S. A. Abbasi, J. M. Dye, D. M. Kranz, A. S. Herbert, E. Procko, *Science* **2020**, *369*, 1261–1265.

Manuscript received: November 24, 2021
Accepted manuscript online: December 27, 2021
Version of record online: January 19, 2022