JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

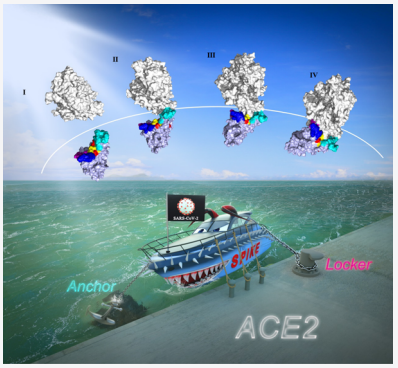
Anchor-Locker Binding Mechanism of the Coronavirus Spike Protein to Human ACE2: Insights from Computational Analysis

Yalong Cong,[⊥] Yinghui Feng,[⊥] Hui Ni, Fengdong Zhi, Yulu Miao, Bohuan Fang, Lujia Zhang,* and John Z. H. Zhang*

Cite This: https://doi.org/10.1021/acs.jcim.1c00241 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: COVID-19 has emerged as the most serious international pandemic in early 2020 and the lack of comprehensive knowledge in the recognition and transmission mechanisms of this virus hinders the development of suitable therapeutic strategies. The specific recognition during the binding of the spike glycoprotein (S protein) of coronavirus to the angiotensin-converting enzyme 2 (ACE2) in the host cell is widely considered the first step of infection. However, detailed insights on the underlying mechanism of dynamic recognition and binding of these two proteins remain unknown. In this work, molecular dynamics simulation and binding free energy calculation were carried out to systematically compare and analyze the receptor-binding domain (RBD) of six coronavirus' S proteins. We found that affinity and stability of the RBD from SARS-CoV-2 under the binding state with ACE2 are stronger than those of other coronaviruses. The solvent-accessible surface area (SASA) and binding free energy of different RBD subunits indicate an "anchor-locker" recognition mechanism involved in the binding of the S protein to ACE2. Loop 2 (Y473-F490) acts as an anchor for ACE2 recognition, and Loop 3 (G496-V503) locks ACE2 at the other nonanchoring end. Then, the charged or long-chain residues in the β-sheet 1 (N450-F456) region reinforce this binding. The proposed binding mechanism was supported by umbrella sampling simulation of the dissociation process. The current computational study provides important theoretical insights for the development of new vaccines against SARS-CoV-2.



1. INTRODUCTION

Since December 2019, an outbreak of pneumonia (COVID-19) related to a novel coronavirus named SARS-CoV-2 has caused worldwide infections, leading to significant mortality. To date, scientists throughout the world are focusing on investigating the infecting mechanisms and finding effective medicine and therapy against SARS-CoV-2. SARS-CoV-2 and the other coronaviruses that have caused large-scale pandemics[1,2] belong to the genus *Betacoronavirus* and were proposed to originate from bat and transmit to humans via intermediate hosts, like the camel, civet, and pangolin.[3−6] The genome of SARS-CoV-2 has been shown to share a high similarity with other coronaviruses, such as RatG13 (96%), bat-SL-CoVZXC21 (88%), and bat-SL-CoVZC45 (88%), and those of pangolin-originating coronaviruses (88−92% for pangolin 1 and pangolin 2S variants etc.) but showed a relatively lower similarity with SARS-CoV (80%).[6−8] Despite the high genome similarity of SARS-CoV-2 with other coronaviruses, the sequence similarity of spike glycoproteins (S proteins) in these coronaviruses varies. In particular, the S proteins of the above coronaviruses all attack epithelial (lung) cells and type II pneumocytes in the host, using the angiotensin-converting enzyme 2 (ACE2) as a receptor.[9,10] Shi and co-workers have suggested that both the recombination and random mutagenesis of the S proteins (like SARS-CoV) could directly endow some coronaviruses with the ability to infect humans.[11−13] On the other hand, Letko et al.[14] has demonstrated that it is the replacement of the entire SARS-CoV receptor-binding domain (RBD), not the introduction of the 14 contact residues (402, 426, 436, 440, 442, 472, 473, 475, 479, 484, 486, 487, 488, and 491) on SARS-CoV (PDB id: 2AJF), that imparts the infectious ability of the S protein from coronaviruses of SLZXC21 and Bat-BM48−31. All of these apparently contradictory insights demonstrate that more detailed and comprehensive investigations on the characteristics of these S proteins are required. In addition, such studies could offer insights into the infection mechanism employed by the virus for human hosts.

The S protein comprises the S1 and S2 subunits; S1 is involved in ACE2 recognition via the RBD and S2 plays a major role during the fusion of the viral-cell membrane.[15] The S2 subunit is more conserved in most S proteins, whereas the residues in the S1 subunit, especially in the receptor-binding motif (RBM) that directly involved in binding to human ACE2,

Received: March 10, 2021

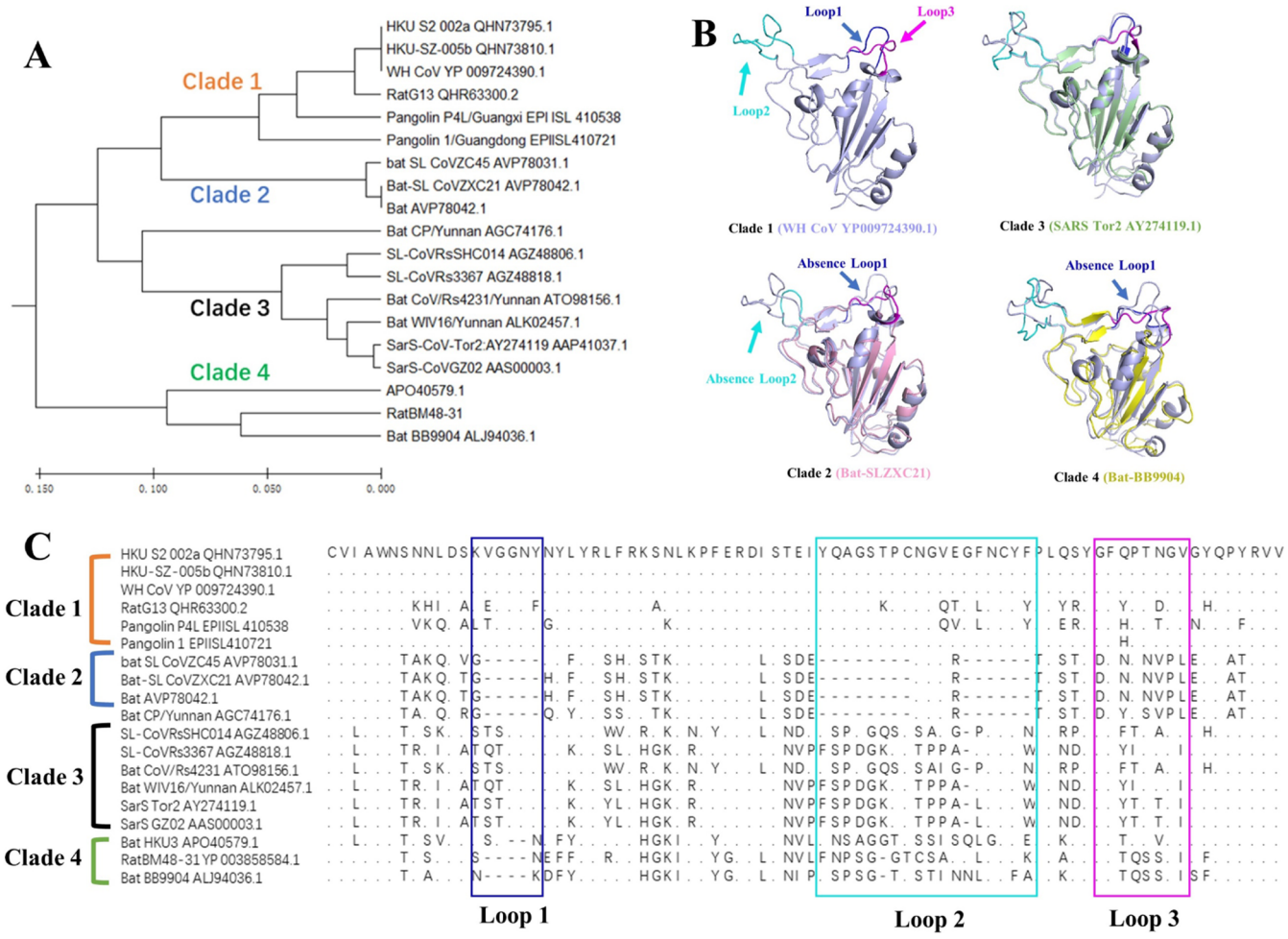


Figure 1. (A) Phylogenetic tree of S proteins from different coronaviruses, classified into four clades; clade 1: SARS-CoV-2, RatG13, and pangolin-P4L; clade 2: SL-CoVZC45 and SL-CoVZXC21; clade 3: SARS-CoV (Tor2 and GZ02) and SARS-like CoV (RsSHC014, Rs3367, and WIV6); and clade 4: Bat-HKU3, Bat-BM48-31, and Bat-BM9904. (B) Structure alignment of the S proteins from SARS-CoV-2 and other clades. The loops are colored in blue (Loop 1), cyan (Loop 2), and magenta (Loop 3). (C) Sequence alignment of the S proteins in the different clades.

are significantly different.[3,5,7,12,13] To gain a better understanding of the S protein-ACE2 binding, the structures of the full-length ACE2, S proteins, and that of RBD binding to the peptidase domain (PD) in ACE2 for both SARS-CoV-2 and SARS-CoV have been reported.[16−21] Both the S proteins of SARS-CoV-2 and SARS-CoV were trimeric and shared the same structure (1.2 Å root-mean-square deviation (RMSD) similarity for the Cα atoms) and binding mode with ACE2.[16,17] However, most of the antibodies against SARS-CoV were found to be ineffectual against SARS-CoV-2,[22] and the K_d value of ACE2 binding with the SARS-CoV-2 S protein was appreciably lower than those with SARS-CoV.[17] By analyzing the structures of ACE2 and RBDs, these differences were suggested to originate from the residue-level differences in the RBM and the strength of their hydrophilic interactions.[18,19] However, the holistic contributions of the RBM subunits like loops or β-sheets in the recognition of ACE2 remain unclear. In addition, there is no consensus on the mechanism of recognition and binding of the RBD to ACE2, which has severely limited further studies related to pharmaceutical and therapeutic applications.

Wrapp et al.[20] found experimentally that the SARS-CoV-2 S protein binds ACE2 with higher affinity than the SARS-CoV S protein. However, Xu et al.[23] calculated a lower binding affinity of the SARS-CoV-2 S protein toward ACE2, utilizing the structures obtained from the homology modeling method. The inconsistency in the experiment and calculation could be explained by the fact that the authors in the latter study[23] used a single frame of the homology modeled structure. At the same time, the error of the free energy calculation method is also the cause of the result. Therefore, the mechanistic basis of the strong affinity of ACE2 and the SARS-CoV-2 S protein needs to be further investigated. In this work, molecular dynamics (MD) simulation was used to explore the mechanism of recognition and binding of the S protein to ACE2. The binding free energy, including enthalpy and entropy, was used to estimate the binding affinity. Enthalpy calculation was performed with the widely used molecular mechanics/generalized Born surface area (MM/GBSA) method;[24−26] and entropy was calculated with the interaction entropy (IE) method[27] developed by our research group.[28−33] In addition, the alanine scanning method[34] was used to calculate the contribution of individual residues to the binding affinity.[35]

2. RESULTS AND DISCUSSION

2.1. Structural and Sequence-Level Differences in S proteins from Different Coronaviruses. The entry of the coronavirus into human cells has been proven to be mediated by the binding of the S protein and human ACE2. Here, the binding

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

mode of the S protein with ACE2 is outlined, based on the published structures of SARS-CoV-2 (Figure S1).[16−21] The S protein is trimeric and comprises two highly conserved subunits S1 and S2.[16,20] The RBD in one of the three monomers of the trimeric S protein can undergo a hinge-like movement to transition between "up" and "down" configurations. ACE2 can only bind to RBD with the "up" configuration (Figure S1a), and the "down" conformation is inaccessible to ACE2.[22] Moreira el al. found that the high-frequency contacts between the NTD (N-terminal domain) and RBD are responsible for the local conformational stability, and they play an important role in transition of the configurations.[36] The diversity of the residue composition in RBM (Figure S1c) has been proposed to lead to the significant differences in the affinity of S proteins from SARS-CoV and SARS-CoV-2 toward ACE2.[20]

To study the difference in residues of the RBD region, based on classification rules reported by Letko et al.,[14] the S proteins from different coronaviruses were further classified into four clades, as shown in Figure 1. The S proteins from SARS-CoV-2 and SARS-CoV share a 50% similarity in the RBM region, with a high frequency of random mutagenesis (Figure 1C) and they were clustered into clade 1 and 3 (Figure 1A), respectively. The sequences of RBMs in clades 2 and 4 were markedly different, with Loops 1 and 2 absent in clade 2, and Loop 1 absent in clade 4. Letko et al.[14] have earlier shown that the S protein from SL-CoVZC45, SL-CoVZXC21 (clade 2), and Bat-BM48−31 (clade 4) could be endowed with the ability to infect cells by introducing the whole RBM of SARS-CoV, not the corresponding 14 residues that were involved in contacting, demonstrating the importance of the conformation of the RBM in the infection. To systematically investigate the binding mode of different S proteins and ACE2 and the influence of conformational differences in the RBM to the ACE2 recognition, six RBDs from the four clades were further studied, as shown in Table 1.

Table 1. Features and Binding Free Energy (kcal/mol) of Six RBDs of the S Protein from Different Coronaviruses

source	clade	loop absence	ΔG_{bind}
SARS-CoV-2	1	No	−25.27
RatG13	1	No	−22.11
Pangolin-P4L	1	No	−16.04
Bat-SLZXC21	2	Loop 1/2	−4.53
SARS-CoV	3	No	−20.14
Bat-BB9904	4	Loop 1	−13.47

The absence of Loop 1 (6 residues) and Loop 2 (18 residues) in clades 2 and 4 made their corresponding RBMs shorter, with lesser flexibility at the binding interface (Figure 1B). In addition, the topology of the Loop 2 region varies in most S proteins, suggesting a contribution of flexibility during the ACE2 recognition and binding. Therefore, the topology or sequence of RBM subunits might determine the ACE2-S protein binding, thereby influencing the infection abilities of these viruses.

2.2. Binding Free Energy and Hot-Spot Residues. The binding free energy of ACE2 to the RBDs (from SARS-CoV-2, RatG13, Pangolin-P4L, Bat-SLZXC21, SARS-CoV, and Bat-BB9904) were calculated to explore the infection ability of different S proteins (Table 1). Among them, the RBD from SARS-CoV-2 has the highest binding affinity with ACE2, Bat-SLZXC21 without Loop 1/2 and Bat-BB9904 without Loop 1 have the lowest affinity. Especially for Bat-SLZXC21, the absence of the loop leads to a significant decrease in binding affinity. This indicates that Loop 2 can significantly affect the binding of ACE2 and the S protein. Wrapp et al.[20] have experimentally found that the SARS-CoV-2 S protein binds to ACE2 with higher affinity than the SARS-CoV S protein. Our calculation result is consistent with this experimental observation, which verifies the validity of the calculation method.

To provide more detailed and microlevel information about the binding mode, the binding free energy of residues within 5 Å of the binding interface is listed in Figure 2.[37] Hot-spot residues contributing more than −2 kcal/mol are filled in red.[38−41] The RBD of RatG13 has the most, seven hot-spot residues, while that of Pangolin-P4L, Bat-SLZXC21, and Bat-BB9904 have the least, three hot-spot residues. In addition, most of these hot-spot residues are long-chain amino acids that are charged or contain benzene rings. Therefore, the salt bridges and Pi−Pi interaction can be considered in the development of the corresponding antibodies and inhibitors to improve affinity. For SARS-CoV-2 seen in Figure 2A, there are five hot-spot residues (R403, L455, F486, Y489, and Y505) with binding free energy greater than −2 kcal/mol. For SARS-CoV seen in Figure 2E, it is the residues of K390, Y442, Y475, N479, Y484, and Y491 that contribute significantly. Although the number of hot-spot residues on SARS-CoV-2 is slightly less than that of SARS-CoV, the binding free energy of most hot-spot residues on SARS-CoV-2 is stronger than that of SARS-CoV, which is an important reason that caused higher affinity of SARS-CoV-2. Particularly, some studies[42,43] that ignore the entropy effect have found that N501 of SARS-CoV-2 is a hot-spot residue. In our calculation, without the entropy change, the enthalpy change of the N501 residue is −2.49 kcal/mol (Table S1), which can be regarded as a hot-spot residue. However, after considering the contribution of entropy (−0.67 kcal/mol), the binding free energy of N501 becomes −1.81 kcal/mol, which can only be regarded as a warm-spot residue. This explains to a certain extent the phenomenon that the well-known N501 mutation will lead to stronger affinity.[44]

Besides, the binding free energy and hot-spot residues of ACE2 are also listed in Figure 3. Similarly, almost all of the hot-spot residues of ACE2 are long-chain amino acids that are charged or contain benzene rings. While for the same ACE2 protein to binding with different RBDs, the hot-spot residues are different, which indicates that the diversity of RBD causes differences of hot-spot residues in ACE2. For the SARS-CoV-2 system shown in Figure 3A, there are five hot-spot residues (Y41, K353, H34, Y83, and Q84) with the binding free energy greater than −2 kcal/mol. Among them, the Y41 always plays an important role as the hot-spot residue in those six systems.

2.3. Structural Stability Analysis. Considering that there are four RBD structures from homology modeling, the RMSDs of the whole protein backbone and binding interface protein backbone of ACE2 and RBD are calculated separately (Figures S2 and S3) to verify the structural stability of simulation. After about 40 ns, RMSD basically remained stable, which indicates that the simulation has reached convergence. In particular, the RMSD fluctuation of the binding interface is significantly smaller compared to that of the whole protein, which indicates the stability of the binding.

The average native contact of heavy atoms in the RBD during the last 50 ns simulation was calculated as a function of time to measure the stability of the entire RBD (Figure 4A). A contact of residues was defined if the distance of heavy atoms is less than 5 Å. The stability of the contact in these six systems shows the convergence of the simulation. Interestingly, the native contact of the three systems (SARS-CoV-2, RatG13, and Pangolin-P4L)

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

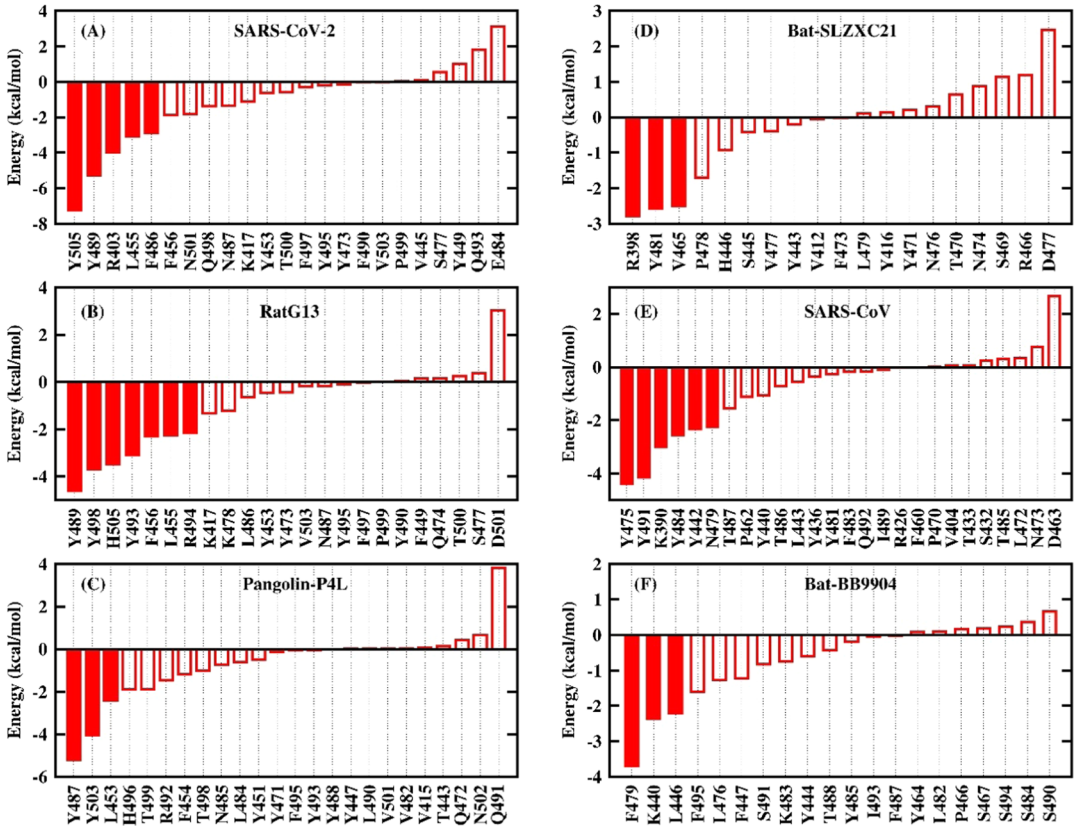


Figure 2. Binding free energy of residues within 5 Å of the binding interface in the different RBDs. (A) RBD from SARS-CoV-2; (B) RBD from RatG13; (C) RBD from Pangolin-P4L; (D) RBD from Bat-SLZXC21; (E) RBD from SARS-CoV; and (F) RBD from Bat-BB9904. Hot-spot residues contributing more than −2 kcal/mol are filled in red. The detailed results are shown in Tables S1−S6.

of clade 1 is significantly higher than those of other clades (Bat-SLZXC21, SARS-CoV, and Bat-BB9904). Among them, that of SARS-CoV-2 is the highest and those of Bat-SLZXC21 and Bat-BB9904 are the least due to the absence of Loop 1/2. In addition, the native contact map of each residue pair in RBD is plotted in Figure 5, and the overall native contact maps of the six RBDs are similar. There is always the native contact between β-sheet 1 and β-sheet 2 at the binding interface, which is conducive to the stability of the binding interface. However, the Loop 2 regions of the six RBDs are different. Then, the average native contact of heavy atoms in the Loop 2 is calculated in Figure 4B. Similarly, the native contact of clade 1 in the Loop 2 region is significantly higher than those of other clades, and that of RSARS-CoV-2 is also the highest, and that of Bat-SLZXC21 is almost zero due to the absence of Loop 2. Considering that the calculation of the native contact depends on the selection of the cutoff criterion, the different cutoffs (5, 6, and 7 Å) are used to verify the reliability of the calculation. The native contacts with cutoffs of 6 and 7 Å are shown in Figures S4 and S5, respectively. The results are consistent with the cutoff of 5 Å shown in Figure 4, and the native contact of SARS-CoV-2 is always more than those of other sources under different cutoff criteria. Overall, the more contact in SARS-CoV-2 means greater stability than those of other sources, among them, the diversity of Loop 2 is an important factor affecting stability.

To further compare and analyze the stability of different RBDs, the isotropic temperature factor (B-factor) of the protein backbone was calculated and shown in Figure 6. The experimental B-factors of SARS-CoV-2 and SARS-CoV are presented in Figure 6A,E, respectively, and the other systems are from homology modeling structures without experimental values. The calculated values and the experimental values fluctuate basically the same, which show the reliability of our simulation. Particularly, D376-N381 residues of SARS-CoV shown in Figure 6E are not resolved due to too much fluctuation and the calculated B-factor in this region is also very large. The hot-spot residues are marked by blue triangles and their B-factor are all relatively small, which indicates that these hot-spot residues can bind to ACE2 stably. Besides, the B-factor of RBD from SARS-CoV-2 is overall smaller than those of other RBDs. This indicates that the RBD from SARS-CoV-2 is more stable in structure than those from other sources, which is consistent with the contact analysis.

In particular, the B-factor between SARS-CoV and SARS-CoV-2 is further compared and analyzed in Figure 7A. For most residues, the B-factor was very similar, while that of the Loop 2 region showed a large difference, with a lower value for SARS-CoV-2 compared to SARS-CoV. This indicates that this region is more stable in SARS-CoV-2. Loop 2 is one of the regions in the RBD closest to ACE2 and plays an essential role in the binding of RBD and ACE2. The flexibility of the loop structure is relatively strong compared to those of other structures. In the evolution of S proteins, the rigidity of the Loop 2 region has increased, resulting in more stable structures of RBD, which, in turn, leads to the higher binding affinity of RBD and ACE2.

The sequence alignments of RBD between SARS-CoV and SARS-CoV-2 (Figure 7A) suggested that the residues' difference of Loop 2 is mainly reflected on N480-N487, which is located between a disulfide bond. To further validate the above conclusions, we exchanged the T467-N473 stretch of SARS-

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

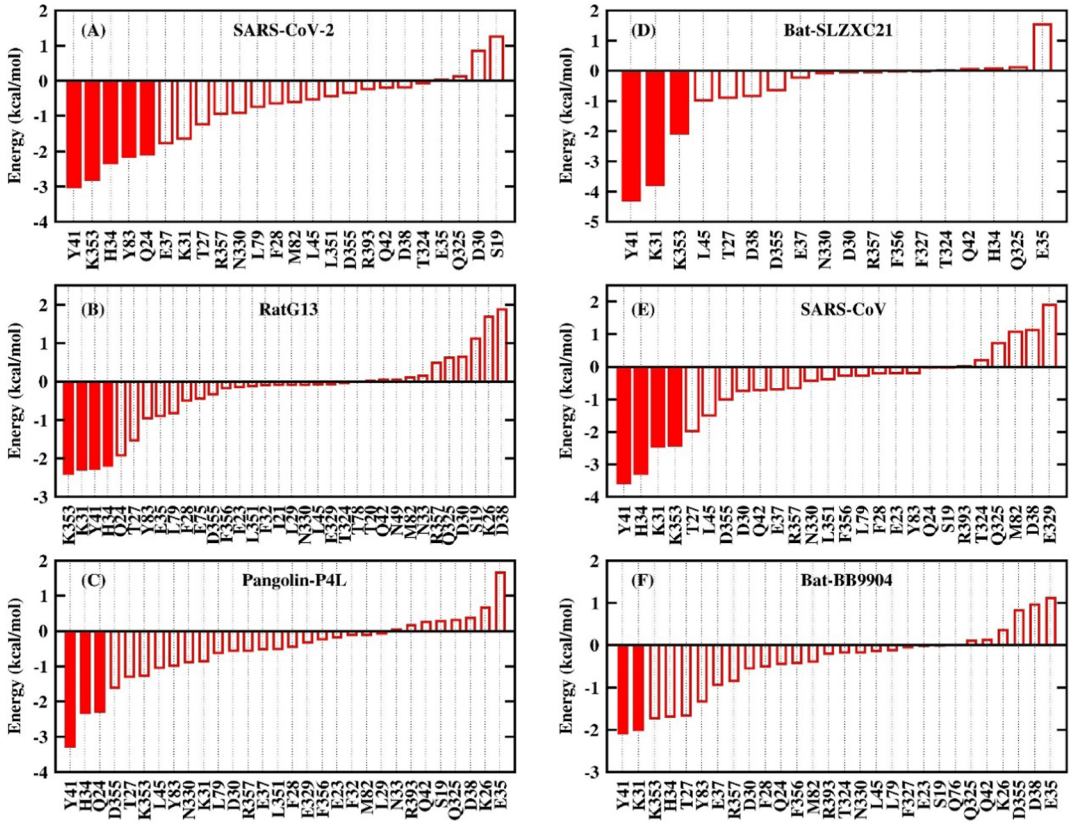


Figure 3. Binding free energy of residues within 5 Å of the binding interface in ACE2. (A) SARS-CoV-2 system; (B) RatG13 system; (C) Pangolin-P4L system; (D) Bat-SLZXC21 system; (E) SARS-CoV system; and (F) Bat-BB9904 system. Hot-spot residues contributing more than −2 kcal/mol are filled in red. The detailed results are shown in Tables S7−S12.

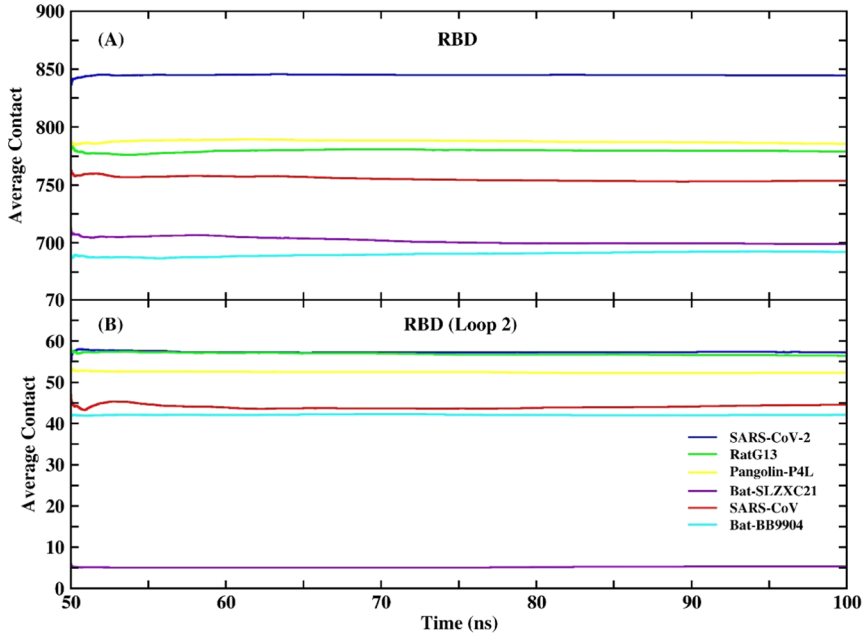


Figure 4. (A) Average native contact of heavy atoms in the RBD during the last 50 ns simulation and (B) average native contact of heavy atoms in Loop 2 during the last 50 ns simulation. A contact of residue was defined if the distance of heavy atoms is less than 5 Å.

CoV and the N480-N487 stretch of SARS-CoV-2 on the two RBDs. The two structures generated by this exchange were named SARS-CoV_exchange and SARS-CoV-2_exchange, respectively. Then, MD simulations based on the exchanged structure were rerun with the same parameters. After the exchange, the relative strength of the B-factor in the Loop 2 region (Figure 7B) has also been exchanged, showing that the Loop 2 region of SARS-CoV-2 indeed provides a stable structure. The difference in the Loop 2 region between SARS-CoV and SARS-CoV-2 is an important factor in determining the

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

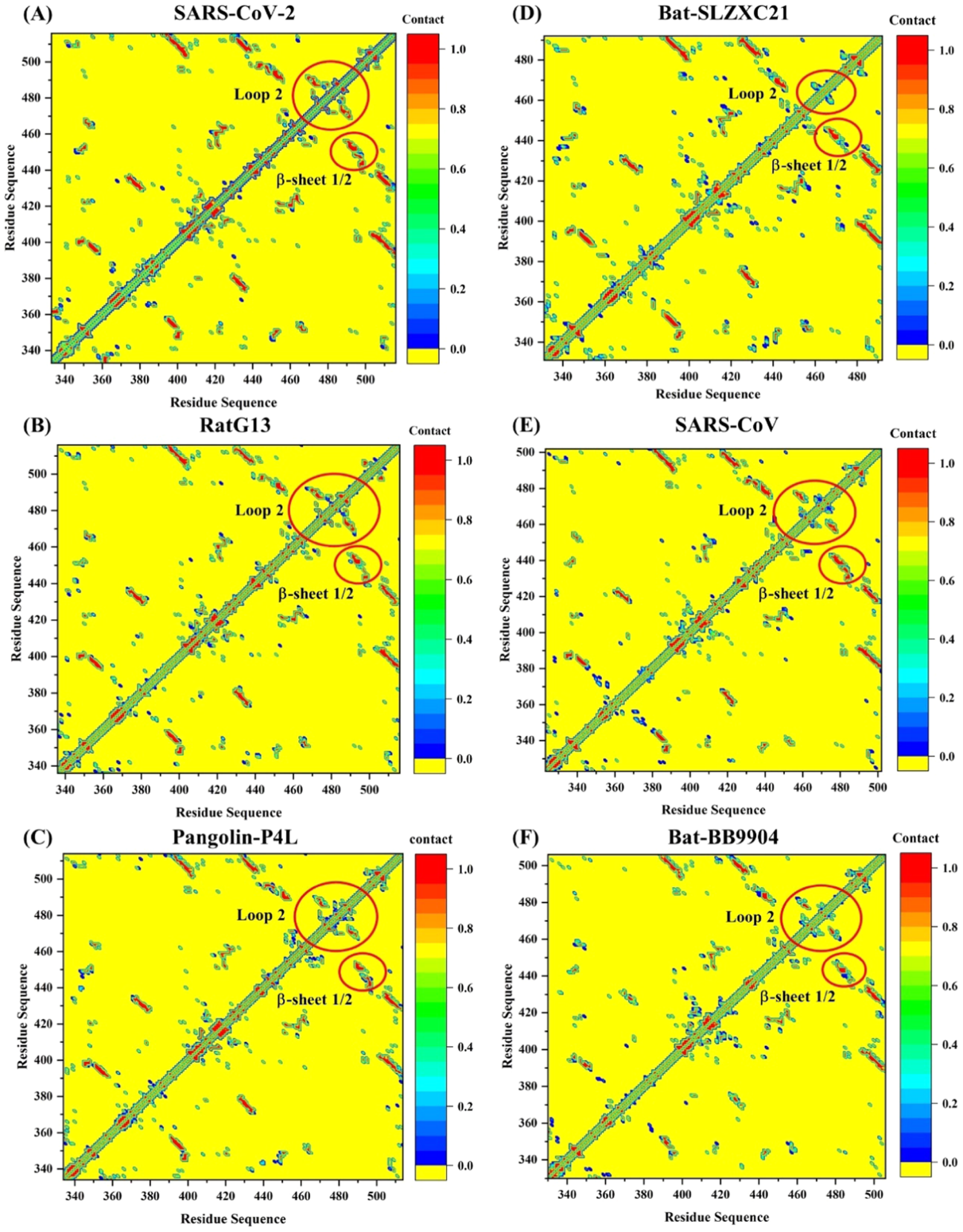


Figure 5. Native contact map of each residue pair in the different RBDs. (A) RBD from SARS-CoV-2; (B) RBD from RatG13; (C) RBD from Pangolin-P4L; (D) RBD from Bat-SLZXC21; (E) RBD from SARS-CoV; and (F) RBD from Bat-BB9904. A contact of residue was defined if the distance of heavy atoms is less than 5 Å.

strength of the binding affinity. Too high flexibility of the Loop 2 region would hinder the stable binding of RBD to ACE2, while too high rigidity would not be conducive to the specific recognition of ACE2 by RBD. Therefore, the evolution of S proteins appears to be seeking a balance of flexibility and rigidity of the Loop 2 region.

Barros el al.[45] found that ACE2's intrinsic flexibility could promote a large swinging motion of the ACE2-S1 complex, providing a mechanical force for the approximation of the two membranes and shedding of S1 toward fusion of the S2 domains into the receptor cell. Besides, stability of the binding interface guarantees the high affinity of ACE2 and RBD. Therefore, the B-

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

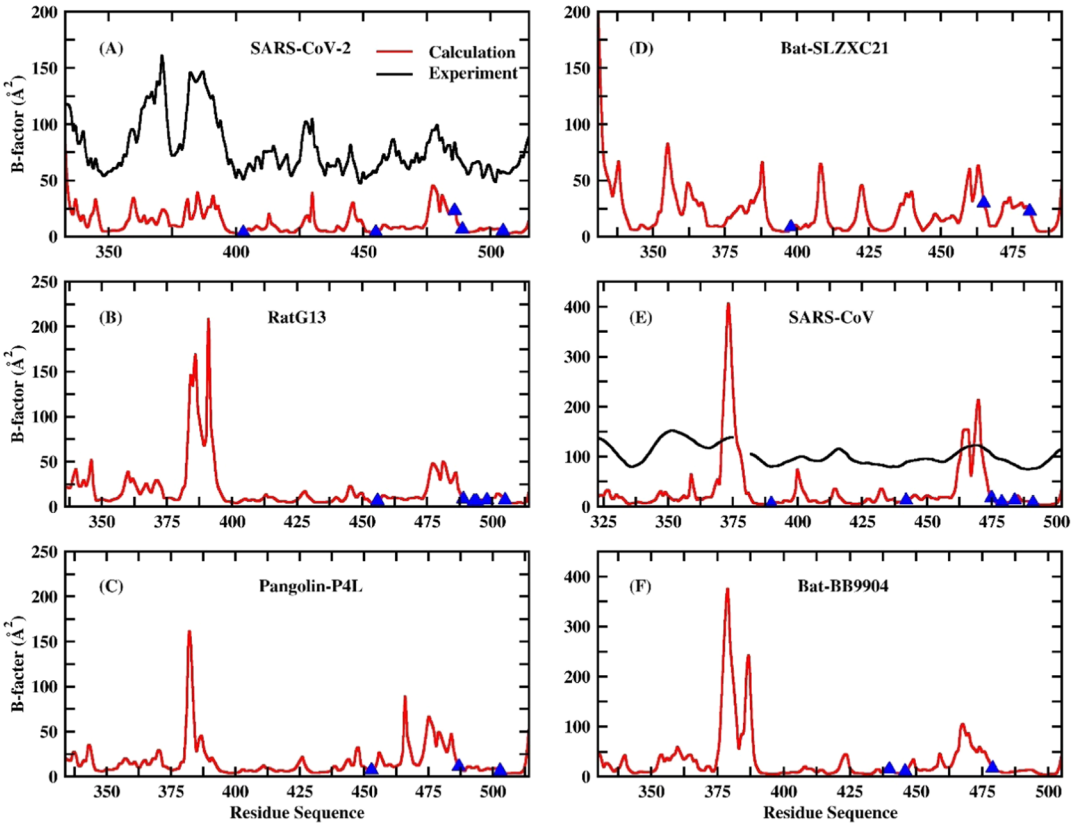


Figure 6. B-factors of the protein backbone in the different RBDs. (A) RBD from SARS-CoV-2; (B) RBD from RatG13; (C) RBD from Pangolin-P4L; (D) RBD from Bat-SLZXC21; (E) RBD from SARS-CoV; and (F) RBD from Bat-BB9904. The hot-spot residues are marked in blue triangles. The experimental B-factors of SARS-CoV-2 and SARS-CoV are presented in (A) and (E), respectively, and the other systems are from homology modeling structures without experimental values.

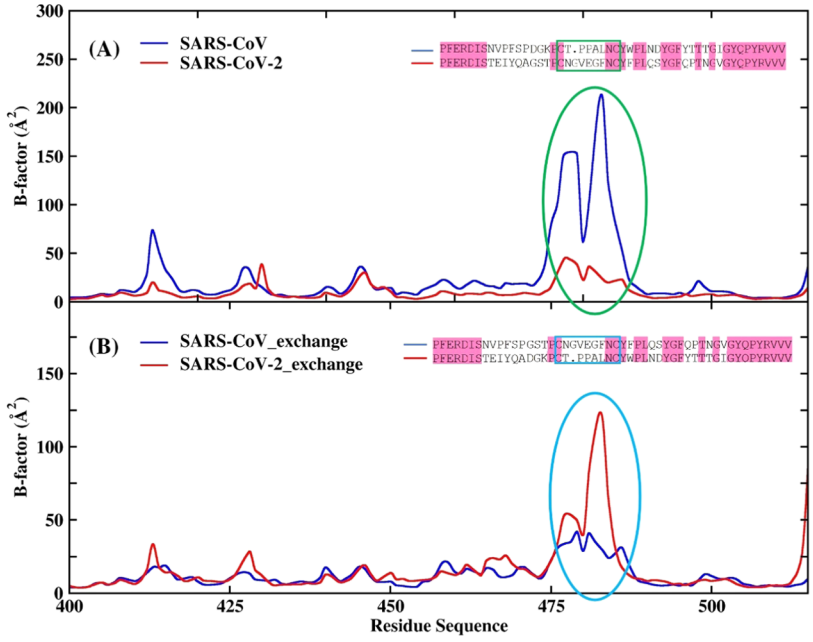


Figure 7. Comparative analysis of B-factors for the RBD protein backbone between SARS-CoV and SARS-CoV-2. The significant differences in the Loop 2 (Y473-F490) region are marked in (A) green and (B) blue. The abscissa represents the residue numbers corresponding to the RBD of SARS-CoV-2. The inset shows the sequence of the two RBDs.

factor of the ACE2 protein is further calculated (Figure 8) to explore the role of structural flexibility in binding RBD. The overall trends of the B-factor of these six systems are very similar. In particular, the binding interface is magnified, and the residues within 5 Å of the RBD are marked. The B-factor of the binding interface is indeed relatively small, which indicates the stability

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

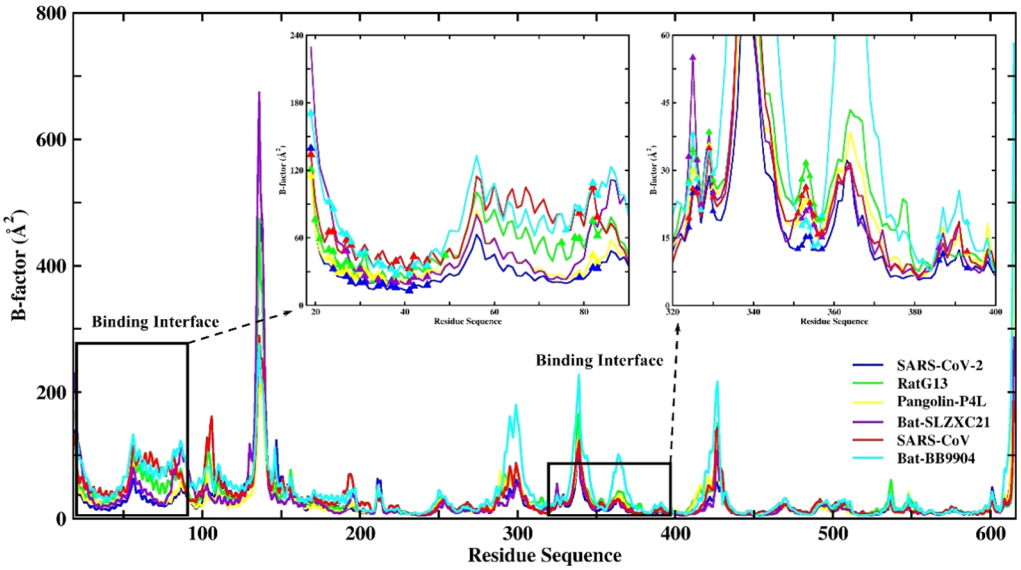


Figure 8. B-factors of the ACE2 protein backbone in the six systems. The binding interface is magnified, and the residues (less than 5 Å from the RBD) of the binding interface are marked.

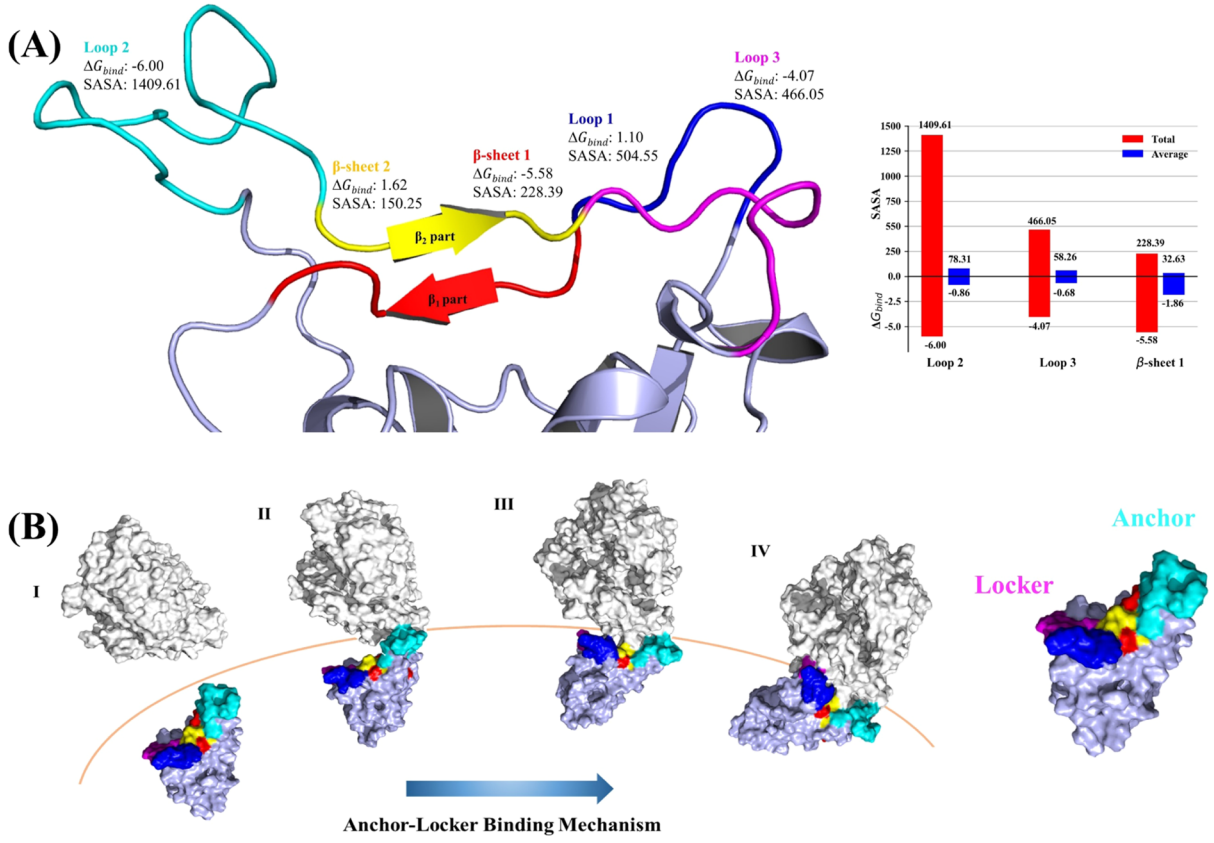


Figure 9. (A) Contributions toward binding free energy and the solvent-accessible surface area (SASA) of the five RBD subunits in SARS-CoV-2. These subunits consist of Loop 1 (K444-Y449), Loop 2 (Y473-F490), Loop 3 (G496-V503), β-sheet 1 (N450-F456), and β-sheet 2 (P491-Y495). The energy is in kcal/mol and SASA is in Å^2. (B) Schematic representation of the anchor-locker binding mode of the SARS-CoV-2 S protein RBD to the host ACE2. ACE2 and RBD are rendered in gray and slate, respectively. The anchor is cyan in color and locker is in magenta.

of the binding interface. Further, the B-factor of binding interface residues from SARS-CoV-2 is smaller than those of other coronaviruses. The high stability of the binding interface is conducive to providing higher affinity, which is consistent with the calculation of the native contact and binding free energy.

2.4. "Anchor-Locker" Binding Mechanism of the RBD to ACE2. To further investigate the functions of RBD subunits involved in establishing direct contacts with ACE2, the contributions of Loop 1, Loop 2, Loop 3, β-sheet 1, and β-sheet 2 regions toward binding free energy were calculated. For SARS-CoV-2, as shown in Figure 9A, Loop 2 (cyan, Y473-F490)

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. Contribution of the Five Subunits in the Six RBDs from Different Coronaviruses toward Binding Free Energy[a]

source	clade	loop absence	Loop 1	Loop 2	Loop 3	β-sheet 1	β-sheet 2
SARS-CoV-2	1	No	1.10	−6.00	−4.07	−5.58	1.62
RatG13	1	No	0.14	−6.51	−0.61	−4.98	−5.34
Pangolin-P4L	1	No	−0.14	−6.21	−4.85	−4.10	−2.24
Bat-SLZXC21	2	Loop 1/2	N/A	N/A	1.64	−1.55	1.98
SARS-CoV	3	No	−0.02	−1.67	−4.77	−3.93	−2.52
Bat-BB9904	4	Loop 1	N/A	−4.56	−0.67	−4.03	−0.47

[a]All values are in kcal/mol.

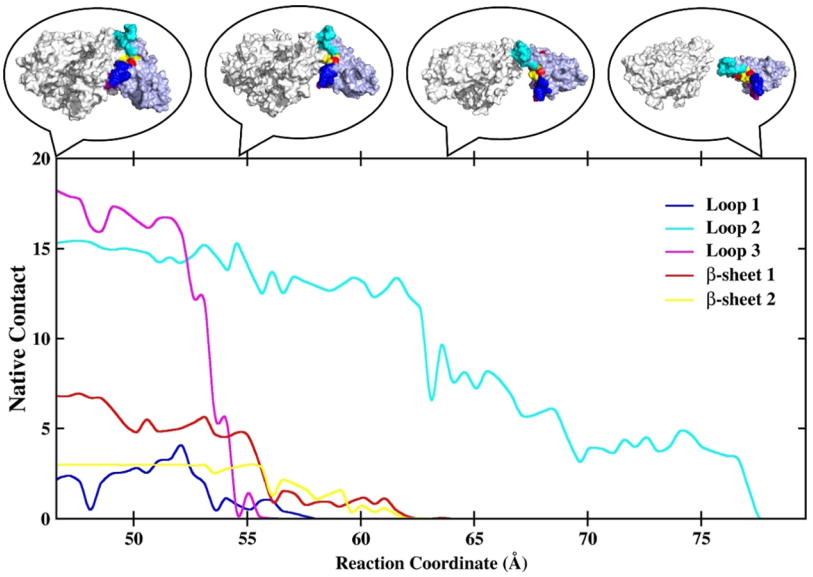


Figure 10. Average native contact between ACE2 and the five subunits (Loop 1: blue, Loop 2: cyan, Loop 3: magenta, β-sheet 1: red, and β-sheet 2: yellow) as the function of the reaction coordinate in each window of the umbrella sampling simulation. The upper structure comes from the conformation corresponding to the reaction coordinate during the umbrella sampling simulation.

and Loop 3 (magenta, G496-V503) exhibited considerable binding free energies (−6.00 and −4.07 kcal/mol, respectively), demonstrating their significant involvement in ACE2 binding. In contrast, Loop 1 (blue, K444-Y449, 1.10 kcal/mol) and β-sheet 2 (yellow, P491-Y495, 1.62 kcal/mol) regions do not seem to contribute to the binding. Interestingly, though the β-sheet 1 (red, N450-F456, −5.58 kcal/mol) was slightly away from ACE2, long-chain residues in this β-sheet 1 (NYLYRLF) exhibited extremely significant affinity. As a result, the β-sheet 1 region also plays an important role in stabilizing the middle portion of the RBD. The strong binding free energy of Loop 2, Loop 3, and β-sheet 1 demonstrated their important roles in the recognition and binding of ACE2.

To study the exact contributions of the five subunits, the solvent-accessible surface area (SASA) analysis was also conducted, as shown in Figure 9A. The SASA of the β-sheets 1 and 2 (150.25 and 228.39 Å^2, respectively) was significantly lower than those of the Loops 1, 2, and 3 (504.55, 1409.61, and 466.05 Å^2, respectively), suggesting that the middle portion of the RBD (β-sheet parts) is difficult to bind to ACE2 first. The β-sheet 1 region may, therefore, be responsible for the reinforcement of the binding, following the recognition of Loops 2 and 3 to ACE2. In addition, at the two ends of the RBD, Loop 2 exhibited higher binding affinity and SASA (−6.00 kcal/mol and 1409.61 Å^2, respectively) compared to Loop 3 (−4.07 kcal/mol and 466.05 Å^2, respectively), demonstrating a greater importance of Loop 2 in ACE2 recognition. Considering that the binding free energy and SASA depend on the number of residues, the average contribution of residues is also calculated on Figure 9A. The average binding free energy of residues in Loop 2, Loop 3, and β-sheet 1 are all significant, and the average SASA of the residue in β-sheet 1 is obviously smaller than that of Loop 2 and Loop 3. Simultaneously, the SASA of per-residue is shown in Figure S6, and the SASA of the β-sheet 1 regions is also obviously smaller than those of Loop 2 and Loop 3. Based on the structural flexibility, binding free energy, and SASA, Loop 2 was proposed to act as an "anchor" in recognition and binding of ACE2 and Loop 3 was proposed to be a "locker", stabilizing the other end of the RBD. The β-sheet 1 region is responsible for enhancing the binding after the recognition at Loops 2 and 3 on the two sides.

In short, the recognition and binding of ACE2 by the SARS-CoV-2 S protein was proposed to occur by an anchor-locker mechanism catalyzed by Loop 2, Loop 3, and the β-sheet 1 regions (Figure 9B). Loop 2 acts as an anchor in ACE2 recognition in the first step, providing the greatest SASA and binding free energy. Loop 3 works as the locker to lock the other end of the RBD, thus completing the primary stabilization of the S protein adjacent to ACE2. Once binding at the two ends of the RBD was stabilized, the charged or long-chain residues in the β-sheet 1 regions were pulled closer to the α-helices of ACE2 and reinforced the binding.

To verify the reproducibility of results, simulations were repeated twice again under the same parametric settings. The binding free energies of the five subunits and the whole RBD in the three simulations are shown in Table S13. The results of the

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

three simulations are very similar in the acceptable variations, and Loop 2, Loop 3, and β-sheet 1 always play an important contribution toward the binding of ACE2 and RBD, which demonstrates the reliability of the anchor-locker mechanism.

Besides, the contributions of the five subunits in the six different RBDs toward binding free energy are shown in Table 2. For Bat-BB9904, despite the absence of Loop 1, the contribution of these subunits toward binding free energy is basically consistent with that of SARS-CoV-2. For RatG13, Pangolin-P4L, and SARS-CoV, the difference from SARS-CoV-2 is that the β-sheet 2 subunit also provides a significant contribution toward binding free energy, which does not make the binding mode change essentially. The anchor-locker mechanism is still applicable to the above four coronaviruses, while it does not apply to Bat-SLZXC21 due to the absence of Loop 2 (anchor). This leads to the conclusion that the affinity of RBD in Bat-SLZXC21 and ACE2 is much lower than those of other coronaviruses.

2.5. Dissociation of the RBD of the SARS-CoV-2/ACE2 Complex. To further validate the anchor-locker binding mechanism proposed here, the unbinding process of RBD of SARS-CoV-2 and ACE2 was tracked through umbrella sampling simulations. As the reverse of the binding process, the unbinding process can largely reflect the dynamics of the binding process. The average native contact between ACE2 and the five subunits is shown as a function of the reaction coordinate (r), as shown in Figure 10. In the initial binding state, the native contact of Loops 2 and 3 was the strongest and that of Loop 1 and β-sheet 2 was the weakest, which is basically consistent with the order of binding free energies shown in Figure 9A. As the dissociation progresses, the native contact of the Loop 3 region was disrupted first, which means that the Loop 3 region on one end of the RBD unbinds from ACE2. Subsequently, the β-sheets 1 and 2 in the middle were unbound from ACE2. Finally, the Loop 2 region on the other end dissociated from ACE2, and the native contact of Loop 1 remained low throughout the dissociation process. To visualize the dissociation process clearly, snapshots from all sampling windows were combined to generate an animation, which is showed in Video S1. The unbinding process shown in the animation is in complete agreement with our analysis of native contacts, which strongly supports the anchor-locker binding mechanism proposed here.

3. CONCLUSIONS

In conclusion, MD simulation was employed to systematically compare and analyze the dynamics of the binding process between ACE2 and RBDs of six coronavirus S proteins, with the overall aim of unraveling the theoretical framework underlying the binding pattern of the two. The simulation found that there is always the native contact between β-sheet 1 and β-sheet 2 at the binding interface, which is conducive to the stability of the binding interface, and the native contact of SARS-CoV-2 is significantly higher than those of other coronaviruses. The more contact of SARS-CoV-2 means greater stability, which is verified by B-factor analysis, and in turn leads to the higher binding affinity of RBD and ACE2. The diversity of Loop 2 is an important factor affecting stability and binding affinity.

Besides, Loop 2, along with Loop 3, on opposite ends of the S protein RBD was found to play an important role in the recognition and binding of S proteins to human ACE2, analogous to an anchor and a locker, respectively. Based on these insights, we propose the anchor-locker binding mechanism, which is strongly supported by the unbinding process observed under umbrella sampling simulations. The importance of the RBD subunits in ACE2 recognition was in agreement with existing experimental observations, showing that missing of subunit or residue-level differences in the RBD could significantly impact the infection abilities of coronaviruses.[14,20] This work would provide important theoretical guidance for an in-depth research in coronavirus S proteins and would be crucial in the development of vaccines and prevention of pneumonia.

4. METHODS

4.1. Sequence and Structure Analysis. For a detailed analysis of the S proteins, especially the RBD domain, from different coronaviruses, the sequence alignment and phylogenetic analysis of several S proteins was conducted with CLUSTALW (https://www.genome.jp/tools-bin/clustalw). The analysis included three SARS-CoV-2 S proteins (WH-CoV/YP-009724390.1, HKU-S2-002a/QHN73795.1, and HKU-SZ-005b/QHN73810.1), two from SARS-CoV (SARS-CoVGZ02/AAS00003.1 and SARS-CoVGZ02-AAS00003.1), and those from other coronaviruses (Pangolin-P4L/EPI-ISL-410538, SL-CoV-RatG13/QHR63300.2, Bat-SL-CoVZXC21/AVP78042.1, Bat-SL-CoVZC45/AVP78031.1, SL-CoV-RsSHC014/AGZ48806.1, Bat-CoV/Rs4231/ATO98156.1, Bat-WIV16/ALK02457.1, SL-CoVRs3367/AGZ48818.1, RatBM48-31/YP-003858584.1, Bat-BB9904/ALJ94036.1, Bat-CP/AGC74176.1, Bat-CoV/AVP78042.1, and Bat-HKU3/APO40579.1). Structures of four RBDs from RatG13, Pangolin-P4L, Bat-SLZXC21, and Bat-BB9904 were homology modeled with the crystal structures of RBD from SARS-CoV or SARS-CoV-2 as the template. The homology models were created using Discovery Studio 2016. The obtained structures were optimized for subsequent alignments using the CHARMM force field.[46] The analysis and graphic processing were conducted with Pymol software.[47]

4.2. MD Simulations. The initial structures of the four complexes from different coronaviruses (RatG13, Pangolin-P4L, Bat-SLZXC21, and Bat-BB9904) were obtained by homology modeling, as described above. The two initial crystal structures of RBDs from SARS-CoV and SARS-CoV-2 bound with ACE2 were reported by Li et al.[18] (PDB id: 2AJF) and Lan et al.[17] (PDB id: 6M0J), respectively. There is one glycosylation site (Asn343) on RBD of SARS-CoV-2 and three glycosylation sites (Asn90, Asn322, and Asn53) on PD of ACE2. They are all far away from the binding surface and do not affect the binding of ACE2 and RBD.[48] Therefore, the glycosylation of the RBD and ACE2 were not considered.

Parallel MD simulations were performed to study the binding mechanism of ACE2 and S proteins. All water molecules and metal ions (away from the binding pocket) were removed from the crystal structure. The prime module[49] of Schrödinger 2015-1[50] and the leap module of AMBER18[51] were used to fill in the missing residues and atoms, respectively. The AMBER14SB force field was used to generate the protein parameters. The TIP3P water box with a 12 Å buffer was chosen as the solvent, then sodium ions were added to ensure charge neutralization in the system. The steepest descent and conjugate gradient methods were used to optimize the structure of solvents and solutes and for subsequent energy minimization. Then, the whole system was heated to 300 K, and the protein was subjected to a constraint of 10 kcal/(mol·Å^2). The SHAKE algorithm[52] was employed to impose constraints on chemical bonds involving hydrogen atoms. Finally, simulations were performed for 100 ns without any restrictions in the NPT

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

ensemble. The last 50 ns trajectory was used for the calculation of binding free energy. The MM/GBSA method extracted averagely 100 frames of snapshots for calculation of enthalpy, and the IE method extracted averagely 50 000 frames of snapshots for calculation of entropy to achieve energy convergence.

4.3. MM/GBSA Method for Enthalpy Calculation. The binding free energy was calculated by the following equation

$$\Delta G_{\text{bind}} = G(\text{complex}) - (G(\text{ACE2}) + G(\text{S})) \quad (1)$$

During the calculation, the ΔG_{bind} was divided into two parts corresponding to changes in enthalpy and entropy

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S \quad (2)$$

where ΔH and $-T\Delta S$ represent the contribution of enthalpy and entropy, respectively, to the binding process. These terms were calculated using the MM/GBSA and IE methods, respectively. In the MM/GBSA method, the ΔH term was calculated by the following equation

$$\Delta H = \Delta E_{\text{ele}} + \Delta E_{\text{vdW}} + \Delta G_{\text{GB}} + \Delta G_{\text{np}} \quad (3)$$

where ΔE_{ele}, ΔE_{vdW}, ΔG_{GB}, and ΔG_{np} represent the electrostatic interaction, van der Waals (vdW) interaction, electrostatic solvation free energy, and nonpolar solvation free energy, respectively. The ΔG_{GB} was estimated by the GB model[53,54] with "igb = 2", and ΔG_{np} was estimated using the following formula

$$\Delta G_{\text{np}} = \gamma \times \text{SASA} \quad (4)$$

where γ is constant with value of 0.005 kcal/(mol·Å^2). The SASA represents the solvent-accessible surface area, which was calculated using the MSMS program.[55]

4.4. IE Method for Entropy Calculation. The gas-phase free energy was calculated by the following equation

$$\Delta G_{\text{gas}} = \langle E_{\text{pp}}^{\text{int}} \rangle + KT \ln\langle e^{\beta \Delta E_{\text{pp}}^{\text{int}}} \rangle \quad (5)$$

where β represents $1/KT$ and $E_{\text{pp}}^{\text{int}}$ represents the interaction energy of protein−protein binding, which includes both electrostatic and vdW interactions. In addition, the gas phase free energy can be expressed as

$$\Delta G_{\text{gas}} = \langle E_{\text{pp}}^{\text{int}} \rangle - T\Delta S \quad (6)$$

therefore,

$$-T\Delta S = KT \ln\langle e^{\beta \Delta E_{\text{pp}}^{\text{int}}} \rangle \quad (7)$$

4.5. Estimating Contributions of Individual Residues. The alanine scanning method introduced mutations at a specific residue (x) by replacing it with alanine (a). Considering that the alanine side chain is composed of a simple alkyl group, which is assumed not to have a measurable contribution toward binding free energy values. Therefore, the contribution of the individual residue (x) could be calculated from the difference of the binding free energy before and after the alanine mutation

$$\Delta\Delta G_{\text{bind}}^{x \to a} = \Delta G_{\text{bind}}^{x} - \Delta G_{\text{bind}}^{a} \quad (8)$$

where, $\Delta G_{\text{bind}}^{a}$ and $\Delta G_{\text{bind}}^{x}$ represent the binding free energy of the alanine mutant and wild-type protein, respectively. The dielectric constants in the MM/GBSA method were set to 1, 3, and 10 for nonpolar, polar, and charged residues, respectively,[56] before introducing the mutation. Equation 8 is also applicable for the calculation of enthalpy change, entropy change, and each energy term. In our work, all residues on the S protein within a 5 Å distance cutoff of the binding interface were individually subjected to alanine scanning. Assuming that the residues farther than 5 Å interact weakly with the ACE2 protein, the total binding free energy was obtained by adding the contributions of all of the residues within 5 Å of binding interface

$$\Delta G_{\text{bind}} = \sum_{x} \Delta\Delta G_{\text{bind}}^{x \to a} \quad (9)$$

4.6. Umbrella Sampling Calculation. Umbrella sampling[57] is one of the most commonly used methods for calculating pathway free energy. This method keeps the system under study in a high potential energy state by adding an artificial constraint, to achieve the purposes of biased sampling. In this work, umbrella sampling simulations were performed through a series of windows along the reaction coordinates to explore the unbinding pathway of ACE2 and $\text{RBD}_{\text{SARS-CoV-2}}$. The reaction coordinate was defined as the distance between the centroids of the ACE2 protein and the $\text{RBD}_{\text{SARS-CoV-2}}$ backbone, and the values ranged from 46.6 to 79.6 Å with 0.5 Å intervals and 67 windows. The 500 ps equilibration and 1 ns sampling runs were performed with a weak force constant of 10 kcal/(mol·Å^2) for all windows. The last snapshot of each window was used as the initial structure for the next window.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00241.

Binding manners and structures of S proteins and ACE2 (Figure S1); RMSDs of the whole protein backbone and binding interface protein backbone of ACE2 and RBD (Figures S2−S3); average native contact of heavy atoms in the RBD with cutoffs of 6 and 7 Å (Figures S4−S5); SASA of per-residues for RBD of SARS-CoV-2 (Figure S6); binding free energy of computational alanine scanning of RBD from six different coronaviruses (Table S1−S6); binding free energy of computational alanine scanning of ACE2 in six different coronaviruses systems (Tables S7−S12); and binding free energies of five subunits and the whole RBD in the three simulations (Table S13) (PDF)

Animation of the dissociation process between RBD of SARS-CoV-2 and ACE2 (Video S1) (MP4)

■ AUTHOR INFORMATION

Corresponding Authors

Lujia Zhang − *Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, Shanghai Key Laboratory of Green Chemistry & Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China; NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China;* orcid.org/0000-0003-3566-917X; Email: ljzhang@chem.ecnu.edu.cn

John Z. H. Zhang − *Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, Shanghai Key Laboratory of Green Chemistry & Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China; NYU-ECNU Center for Computational Chemistry at NYU Shanghai,*

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Shanghai 200062, China; Department of Chemistry, New York University, New York 10003, United States; orcid.org/0000-0003-4612-1863; Email: John.zhang@nyu.edu

Authors

Yalong Cong − *Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, Shanghai Key Laboratory of Green Chemistry & Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China;* orcid.org/0000-0002-0910-811X

Yinghui Feng − *Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, Shanghai Key Laboratory of Green Chemistry & Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

Hui Ni − *College of Food and Biology Engineering, Jimei University, Xiamen, Fujian 361021, China;* orcid.org/0000-0001-5827-1717

Fengdong Zhi − *Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, Shanghai Key Laboratory of Green Chemistry & Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

Yulu Miao − *Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, Shanghai Key Laboratory of Green Chemistry & Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

Bohuan Fang − *Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, Shanghai Key Laboratory of Green Chemistry & Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c00241

Author Contributions

[⊥]Y.C. and Y.F. contributed equally to this work.

Notes

The authors declare no competing financial interest.

Data and Software Availability All data and software are available. Output data are available in the Supporting Information and software used are as follows: AMBER18: http://ambermd.org/. Pymol: https://pymol.org/2/. Schrödinger: https://www.schrodinger.com/. Xmgrace: https://plasma-gate.weizmann.ac.il/Grace/

■ ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (Grant No. 2019YFA0905200), the National Natural Science Foundation of China (Grant No. U1805235, 21933010, 91753103), and NYU-ECNU Center for Computational Chemistry at NYU Shanghai. The authors also thank the Supercomputer Center of East China Normal University (ECNU Public Platform for Innovation 001) for providing us computational time.

■ REFERENCES

(1) Drosten, C.; Gunther, S.; Preiser, W.; van der Werf, S.; Brodt, H. R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R. A. M.; Berger, A.; Burguiere, A. M.; Cinatl, J.; Eickmann, M.; Escriou, N.; Grywna, K.; Kramme, S.; Manuguerra, J. C.; Muller, S.; Rickerts, V.; Sturmer, M.; Vieth, S.; Klenk, H. D.; Osterhaus, A.; Schmitz, H.; Doerr, H. W. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **2003**, *348*, 1967−1976.

(2) Zaki, A. M.; van Boheemen, S.; Bestebroer, T. M.; Osterhaus, A. D. M. E.; Fouchier, R. A. M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* **2012**, *367*, 1814−1820.

(3) Cui, J.; Li, F.; Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17*, 181−192.

(4) Lam, T. T.-Y.; Jia, N.; Zhang, Y.-W.; Shum, M. H.-H.; Jiang, J.-F.; Zhu, H.-C.; Tong, Y.-G.; Shi, Y.-X.; Ni, X.-B.; Liao, Y.-S.; Li, W.-J.; Jiang, B.-G.; Wei, W.; Yuan, T.-T.; Zheng, K.; Cui, X.-M.; Li, J.; Pei, G.-Q.; Qiang, X.; Cheung, W. Y.-M.; Li, L.-F.; Sun, F.-F.; Qin, S.; Huang, J.-C.; Leung, G. M.; Holmes, E. C.; Hu, Y.-L.; Guan, Y.; Cao, W.-C. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **2020**, *583*, 282−285.

(5) Zhang, T.; Wu, Q.; Zhang, Z. Pangolin homology associated with 2019-nCoV. *bioRxiv* **2020**, DOI: 10.1101/2020.02.19.950253.

(6) Liu, P.; Jiang, J.-Z.; Wan, X.-F.; Hua, Y.; Li, L.; Zhou, J.; Wang, X.; Hou, F.; Chen, J.; Zou, J.; Chen, J. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog.* **2020**, *16*, No. e1008421.

(7) Wu, F.; Zhao, S.; Yu, B.; Chen, Y.; Xu, L.; Holmes, E. C.; Zhang, Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265−269.

(8) Xiao, K.; Zhai, J.; Feng, Y.; Zhou, N.; Zhang, X.; Zou, J.-J.; Li, N.; Guo, Y.; Li, X.; Shen, X.; Zhang, Z.; Shu, F.; Huang, W.; Li, Y.; Zhang, Z.; Chen, R.-A.; Wu, Y.-J.; Peng, S.-M.; Huang, M.; Xie, W.-J.; Cai, Q.-H.; Hou, F.-H.; Chen, W.; Xiao, L.; Shen, Y. Isolation and Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* **2020**, *583*, 286−289.

(9) Li, W.; Moore, M. J.; Vasilieva, N.; Sui, J.; Wong, S. K.; Berne, M. A.; Somasundaran, M.; Sullivan, J. L.; Luzuriaga, K.; Greenough, T. C.; Choe, H.; Farzan, M. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **2003**, *426*, 450−454.

(10) Qian, Z.; Travanty, E. A.; Oko, L.; Edeen, K.; Berglund, A.; Wang, J.; Ito, Y.; Holmes, K. V.; Mason, R. J. Innate immune response of human alveolar type II cells infected with severe acute respiratory syndrome-coronavirus. *Am. J. Respir. Cell Mol. Biol.* **2013**, *48*, 742−8.

(11) Menachery, V. D.; Yount, B. L.; Debbink, K.; Agnihothram, S.; Gralinski, L. E.; Plante, J. A.; Graham, R. L.; Scobey, T.; Ge, X.-Y.; Donaldson, E. F.; Randell, S. H.; Lanzavecchia, A.; Marasco, W. A.; Shi, Z.-L.; Baric, R. S. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **2015**, *21*, 1508−1513.

(12) Hu, B.; Zeng, L. P.; Yang, X. L.; Ge, X. Y.; Zhang, W.; Li, B.; Xie, J. Z.; Shen, X. R.; Zhang, Y. Z.; Wang, N.; Luo, D. S.; Zheng, X. S.; Wang, M. N.; Daszak, P.; Wang, L. F.; Cui, J.; Shi, Z. L. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* **2017**, *13*, No. e1006698.

(13) Ge, X.-Y.; Li, J.-L.; Yang, X.-L.; Chmura, A. A.; Zhu, G.; Epstein, J. H.; Mazet, J. K.; Hu, B.; Zhang, W.; Peng, C.; Zhang, Y.-J.; Luo, C.-M.; Tan, B.; Wang, N.; Zhu, Y.; Crameri, G.; Zhang, S.-Y.; Wang, L.-F.; Daszak, P.; Shi, Z.-L. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **2013**, *503*, 535−538.

(14) Letko, M.; Marzi, A.; Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **2020**, *5*, 562−569.

(15) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *183*, 1735.

(16) Gui, M.; Song, W.; Zhou, H.; Xu, J.; Chen, S.; Xiang, Y.; Wang, X. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* **2017**, *27*, 119−129.

(17) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 spike

receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, *581*, 215−220.
(18) Li, F.; Li, W.; Farzan, M.; Harrison, S. C. Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science* **2005**, *309*, 1864−1868.
(19) Sun, C.; Chen, L.; Yang, J.; Luo, C.; Zhang, Y.; Li, J.; Yang, J.; Zhang, J.; Xie, L. SARS-CoV-2 and SARS-CoV Spike-RBD Structure and Receptor Binding Comparison and Potential Implications on Neutralizing Antibody and Vaccine Development. *bioRxiv* **2020**, DOI: 10.1101/2020.02.16.951723.
(20) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260−1263.
(21) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367*, 1444−1448.
(22) Yuan, M.; Wu, N. C.; Zhu, X.; Lee, C.-C. D.; So, R. T. Y.; Lv, H.; Mok, C. K. P.; Wilson, I. A. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **2020**, *368*, 630−633.
(23) Xu, X.; Chen, P.; Wang, J.; Feng, J.; Zhou, H.; Li, X.; Zhong, W.; Hao, P. 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.* **2020**, *63*, 457−460.
(24) Constanciel, R.; Contreras, R. Self consistent field theory of solvent effects representation by continuum models: introduction of desolvation contribution. *Theor. Chim. Acta* **1984**, *65*, 1−11.
(25) Zou, X.; Yaxiong; Kuntz, I. D. Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model. *J. Am. Chem. Soc.* **1999**, *121*, 8033−8043.
(26) Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.
(27) Duan, L.; Liu, X.; Zhang, J. Z. H. Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein−Ligand Binding Free Energy. *J. Am. Chem. Soc.* **2016**, *138*, 5722−5728.
(28) Aldeghi, M.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Statistical Analysis on the Performance of Molecular Mechanics Poisson−Boltzmann Surface Area versus Absolute Binding Free Energy Calculations: Bromodomains as a Case Study. *J. Chem. Inf. Model.* **2017**, *57*, 2203−2221.
(29) Ben-Shalom, I. Y.; Pfeiffer-Marek, S.; Baringhaus, K.-H.; Gohlke, H. Efficient Approximation of Ligand Rotational and Translational Entropy Changes upon Binding for Use in MM-PBSA Calculations. *J. Chem. Inf. Model.* **2017**, *57*, 170−189.
(30) Zou, Y.; Qian, Z.; Sun, Y.; Wei, G.; Zhang, Q. Orcein-Related Small Molecule O4 Destabilizes hIAPP Protofibrils by Interacting Mostly with the Amyloidogenic Core Region. *J. Phys. Chem. B* **2017**, *121*, 9203−9212.
(31) Song, J.; Qiu, L.; Zhang, J. Z. H. An efficient method for computing excess free energy of liquid. *Sci. China: Chem.* **2018**, *61*, 135−140.
(32) Sun, H.; Duan, L.; Chen, F.; Liu, H.; Wang, Z.; Pan, P.; Zhu, F.; Zhang, J. Z. H.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. *Phys. Chem. Chem. Phys.* **2018**, *20*, 14450−14460.
(33) Cong, Y.; Huang, K.; Li, Y.; Zhong, S.; Zhang, J. Z. H.; Duan, L. Entropic effect and residue specific entropic contribution to the cooperativity in streptavidin-biotin binding. *Nanoscale* **2020**, *12*, 7134−7145.
(34) Massova, I.; Kollman, P. A. Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **1999**, *121*, 8133−8143.
(35) Liu, X.; Peng, L.; Zhou, Y.; Zhang, Y.; Zhang, J. Z. H. Computational Alanine Scanning with Interaction Entropy for Protein−Ligand Binding Free Energies. *J. Chem. Theory Comput.* **2018**, *14*, 1772−1780.
(36) Moreira, R. A.; Guzman, H. V.; Boopathi, S.; Baker, J. L.; Poma, A. B. Characterization of Structural and Energetic Differences between Conformations of the SARS-CoV-2 Spike Protein. *Materials* **2020**, *13*, No. 5362.
(37) Cong, Y.; Duan, L.; Huang, K.; Bao, J.; Zhang, J. Z. H. Alanine scanning combined with interaction entropy studying the differences of binding mechanism on HIV-1 and HIV-2 proteases with inhibitor. *J. Biomol. Struct. Dyn.* **2021**, *39*, 1588−1599.
(38) Yan, Y.; Yang, M.; Ji, C. G.; Zhang, J. Z. H. Interaction Entropy for Computational Alanine Scanning. *J. Chem. Inf. Model.* **2017**, *57*, 1112−1122.
(39) Wu, F.-X.; Yang, J.-F.; Mei, L.-C.; Wang, F.; Hao, G.-F.; Yang, G.-F. PIIMS Server: A Web Server for Mutation Hotspot Scanning at the Protein−Protein Interface. *J. Chem. Inf. Model.* **2021**, *61*, 14−20.
(40) Huang, D.; Qi, Y.; Song, J.; Zhang, J. Z. H. Calculation of hot spots for protein−protein interaction in p53/PMI-MDM2/MDMX complexes. *J. Comput. Chem.* **2019**, *40*, 1045−1056.
(41) Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280*, 1−9.
(42) Ghorbani, M.; Brooks, B. R.; Klauda, J. B. Critical Sequence Hotspots for Binding of Novel Coronavirus to Angiotensin Converter Enzyme as Evaluated by Molecular Simulations. *J. Phys. Chem. B* **2020**, *124*, 10034−10047.
(43) Lim, H.; Baek, A.; Kim, J.; Kim, M. S.; Liu, J.; Nam, K.-Y.; Yoon, J.; No, K. T. Hot spot profiles of SARS-CoV-2 and human ACE2 receptor protein protein interaction obtained by density functional tight binding fragment molecular orbital method. *Sci. Rep.* **2020**, *10*, No. 16862.
(44) Khan, A.; Zia, T.; Suleman, M.; Khan, T.; Ali, S. S.; Abbasi, A. A.; Mohammad, A.; Wei, D. Q. Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. *J. Cell. Physiol.* **2021**, 1−13.
(45) Barros, E. P.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Wang, Y.; Fallon, L.; Raguette, L.; Belfon, K.; Simmerling, C.; Amaro, R. E. The flexibility of ACE2 in the context of SARS-CoV-2 infection. *Biophys. J.* **2021**, *120*, 1072−1084.
(46) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.
(47) Schrödinger, L.. *The PyMOL Molecular Graphics System*, Version 1.8; New York, 2015.
(48) Delgado, J. M.; Duro, N.; Rogers, D. M.; Tkatchenko, A.; Pandit, S. A.; Varma, S. Molecular basis for higher affinity of SARS-CoV-2 spike RBD for human ACE2 receptor. *Proteins: Struct., Funct., Bioinf.* **2021**, 1−11.
(49) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 351−367.
(50) Schrödinger, L. *Maestro*, Release 2015-1; New York, 2015.
(51) Case, D. A.; Brozell, S. R.; Cerutti, D. S.; Cheatham, E.T., III, V. W. D. C.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2018*; University of California: San Francisco, 2018.
(52) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(53) Tsui, V.; Case, D. A. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275−291.

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(54) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383−394.
(55) Sanner, M. F.; Olson, A. J.; Spehner, J.-C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305−320.
(56) Li, Z.; Bao, J.; Qi, Y.; Zhang, J. Z. H. Computational approaches to studying methylated H4K20 recognition by DNA repair factor 53BP1. *Phys. Chem. Chem. Phys.* **2020**, *22*, 6136−6144.
(57) Kottalam, J.; Case, D. A. Dynamics of ligand escape from the heme pocket of myoglobin. *J. Am. Chem. Soc.* **1988**, *110*, 7690−7697.

https://doi.org/10.1021/acs.jcim.1c00241
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX