JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

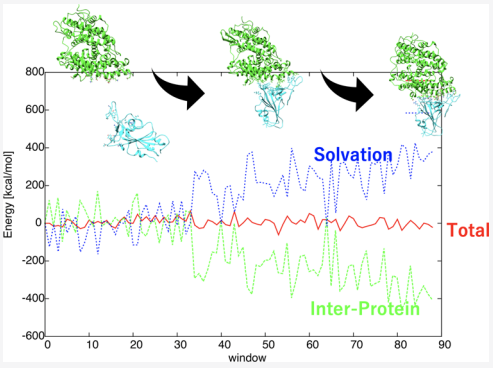
Computational Analysis of the SARS-CoV-2 RBD−ACE2-Binding Process Based on MD and the 3D-RISM Theory

Norio Yoshida,* Yutaka Maruyama, Ayori Mitsutake, Akiyoshi Kuroda, Ryo Fujiki, Kodai Kanemaru, Daisuke Okamoto, Alexander E. Kobryn, Sergey Gusarov, and Haruyuki Nakano

Cite This: J. Chem. Inf. Model. 2022, 62, 2889−2898 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The binding process of angiotensin-converting enzyme 2 (ACE2) to the receptor-binding domain (RBD) of the severe acute respiratory syndrome-like coronavirus 2 spike protein was investigated using molecular dynamics simulation and the three-dimensional reference interaction-site model theory. The results suggested that the protein-binding process consists of a protein−protein approaching step, followed by a local structural rearrangement step. In the approaching step, the interprotein interaction energy decreased as the proteins approached each other, whereas the solvation free energy increased. As the proteins approached, the glycan of ACE2 first established a hydrogen bond with the RBD. Thereafter, the number of interprotein hydrogen bonds increased rapidly. The solvation free energy increased because of the desolvation of the protein as it approached its partner. The spatial distribution function of the solvent revealed the presence of hydrogen bonds bridged by water molecules on the RBD−ACE2 interface. Finally, principal component analysis revealed that ACE2 showed a pronounced conformational change, whereas there was no significant change in RBD.



1. INTRODUCTION

The global spread of the novel severe acute respiratory syndrome-like coronavirus 2 (SARS-CoV-2) has continued to cause serious damage to human life, health, and economy. Many experimental and computational studies have been devoted to the investigation of the mechanism underlying SARS-CoV-2 infection and spread. Coronaviruses initiate cell entry via the binding of a viral spike protein to a receptor protein, angiotensin-converting enzyme 2 (ACE2), on the surface of cells.[1,2] The spike protein is located on the surface of the viral particle and is formed by a trimer of proteins consisting of two subunits, S1 and S2. The domain that binds to ACE2, which is called the receptor-binding domain (RBD), is located at the C-terminus of the S1 subunit. The RBD of the spike protein selectively recognizes ACE2. In turn, the binding of ACE2 to the RBD triggers various processes of invasion and proliferation. Therefore, the binding of spike proteins to ACE2 is an essential initial process of SARS-CoV-2 infection.

Because of its importance, the binding process of ACE2 to RBD has been studied with special attention. Molecular dynamics (MD) simulation is a promising tool to investigate the process at the atomic level, and many studies on the ACE2 and RBD interactions have been reported.[3−14] A comprehensive study of the ACE2−RBD-binding process was conducted by Barros et al.[10] They performed an all-atom MD simulation of the full-length glycosylated ACE2 homodimer including membrane and explicit solvent and revealed the importance of the flexible motion in the head-transmembrane domain linker region and helix mobility in the membrane of ACE2. In addition, it was suggested that specific glycans contribute to the inhibition of RBD binding. MD simulation studies of the full-length glycosylated spike protein were also conducted.[11,12,14] These studies revealed the role of the glycans on open−close or up−down conformational change of the spike protein. Differences in the interaction and structure between SARS-CoV-2 and SARS-CoV have also been intensively studied.[4,5,7,9] Those studies revealed the importance of the mutation of specific residues of RBD. Interactions with other coronavirus spike proteins have also been studied, and specific interactions of SARS-CoV-2 have been identified.[15] These interaction analyses have pointed out the importance of hydrogen bonding between amino acids on the RBD−ACE2-binding interface, as well as the importance of the hydration structure changes and hydrogen-bond bridging afforded by water molecules.[5,16] The design of RBD−ACE2-binding inhibitors is also being actively studied.[17−20] Chemical modification of amino acids in the spike protein RBD and

Received: February 15, 2022
Published: May 18, 2022

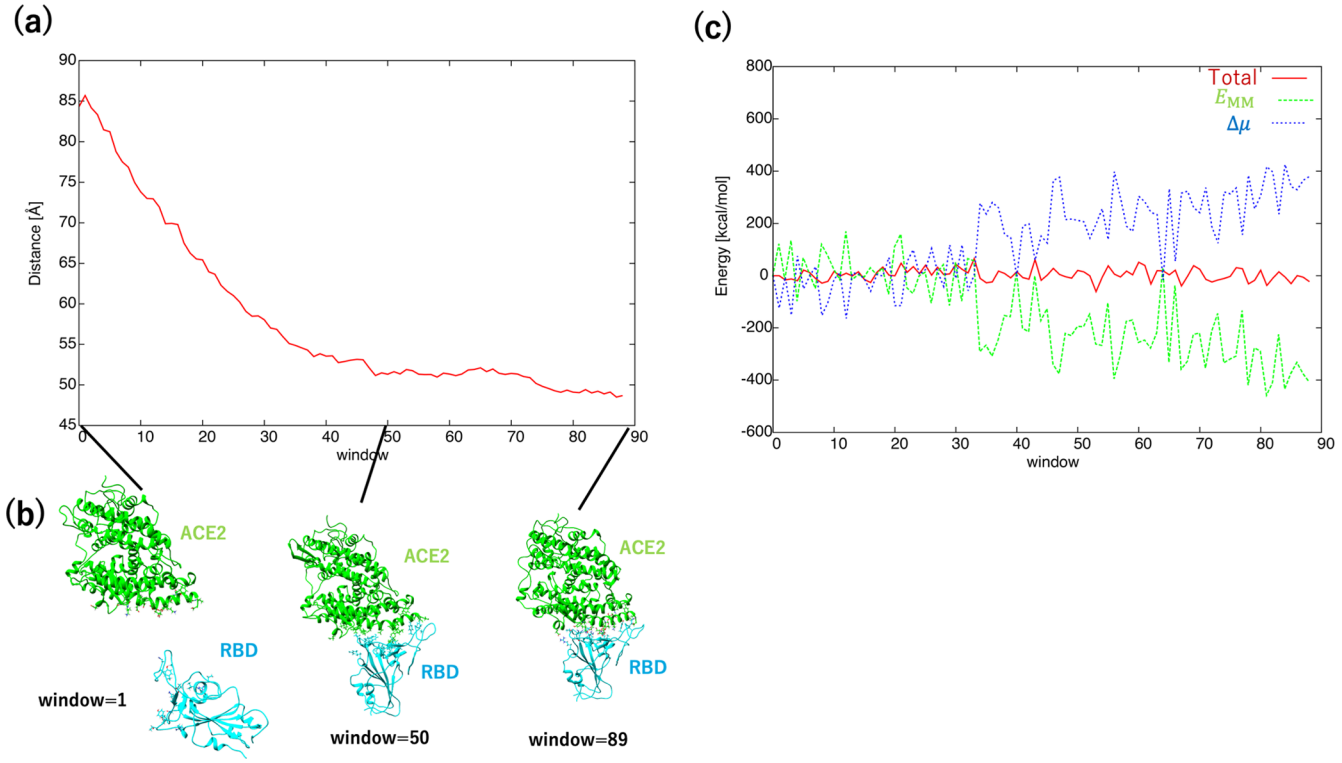


Figure 1. Interprotein distance and free energy changes along the binding pathway. (a) Distance between the center of mass of ACE2 and RBD. (b) Representative snapshots of protein structures taken from the umbrella sampling trajectories at windows = 1, 50, and 89. (c) Total free energy, protein structure energy, and SFE plotted in red-, green-, and blue-colored lines, respectively.

the use of nanobodies to inhibit binding have also been examined.[21,22] MD simulations are useful for analyzing the binding affinity of mutants, and new analyses are being rapidly performed with each successive appearance of SARS-CoV-2 variants, including the most recent Omicron variant.[4,23−25] As mentioned above, RBD−ACE2 binding has been investigated in various ways,[5,16] and a large amount of knowledge has been accumulating regarding the protein−protein interaction and the hydration structure after binding; however, studies focusing on the solvent effect on the entire binding process are scarce.

In the present study, the binding process of the SARS-CoV-2 spike protein RBD to ACE2 was investigated using MD and the three-dimensional reference interaction-site model (3D-RISM) theory. The 3D-RISM theory is a statistical mechanics theory that is applicable to the solvation of biomolecules.[26−29] This theory has been applied to various biological and chemical processes of biomolecules in solution, including the RBD−ACE2-binding process.[5,16,30,31] In a previous study, we applied this theory to investigate the role of solvation in the binding process and reported the solvation structure and the thermodynamics of the RBD−ACE2 structure, which was optimized based on the experimentally observed structure. In the present study, we employed MD simulation to examine the structural fluctuation of the proteins during the binding process. In general, the determination of protein-binding pathways is computationally exceedingly expensive and, thus, requires state-of-the-art sampling methods and huge computational resources. For the RBD−ACE2-binding process, D. E. Shaw Research provided several representative pathways of binding that were evaluated using long-time-scale MD simulations and accelerated weighted ensemble techniques performed on the Anton 2 supercomputer.[32] In the present study, we employed one of the trajectories provided by D. E. Shaw Research as a reference for the computational efficiency. According to the original trajectories, we assumed that the RBD takes already "up" conformation and only monomer RBD is considered. Although previous studies have suggested that both interprotomer and intraprotomer interactions affect the conformational changes of spike proteins,[11,12,14] which in turn affect the RBD−ACE2 binding, such interactions are ignored in this study. Therefore, the focus here is on the interaction between the RBD monomer and ACE2. Umbrella sampling was performed with an explicit solvent based on the reference trajectory, and fluctuating protein structures along the pathway were obtained. For the obtained protein structures, the 3D-RISM theory was applied to evaluate the solvation structure and thermodynamics. Based on these results, we clarified the correlation between protein structural changes and solvation thermodynamics during the RBD−ACE2-binding process.

2. METHOD

In the present study, an energetical analysis of the RBD−ACE2-binding process was conducted using the following procedure. First, the structural sampling of RBD−ACE2 was performed by MD simulation using an umbrella sampling algorithm. The structures of RBD−ACE2 were extracted from the obtained trajectory, and the solvation properties, such as solvation free energy (SFE), partial molar volume (PMV), and solvation structure, were determined using the 3D-RISM theory. We also analyzed the structural changes of RBD−ACE2 via a principal component analysis (PCA).[33−35] The details of the computational conditions are described below.

2.1. Structure Sampling. We started the sampling simulation from the structures proposed by D. E. Shaw

https://doi.org/10.1021/acs.jcim.2c00192
J. Chem. Inf. Model. 2022, 62, 2889−2898

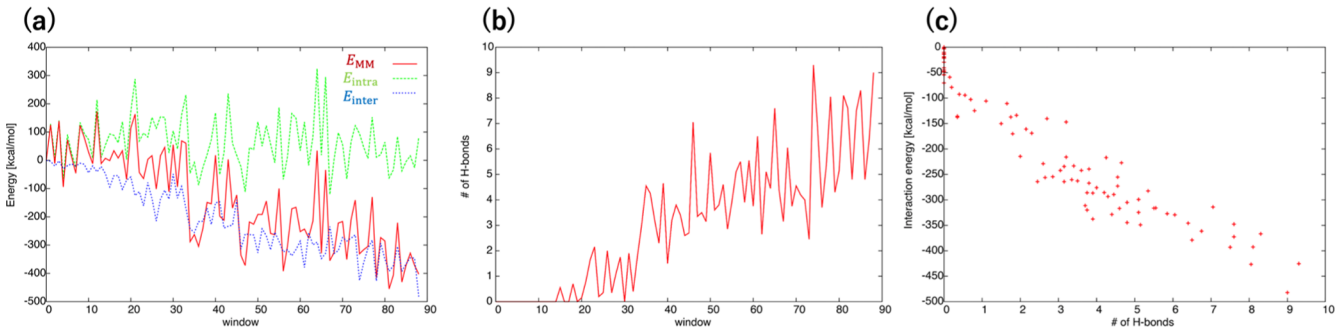


Figure 2. Changes of the structural energy and its components, the number of interprotein hydrogen bonds, and their correlations. The protein structure energy, intraprotein energy, and interprotein interaction energy are plotted in red-, green-, and blue-colored lines in (a). The number of interprotein hydrogen bonds and the correlation between the number of hydrogen bonds and the interaction energy are plotted in (b,c), respectively. The plotted values are the average of each window.

Research. The trajectory data set DESRES-ANTON-[10857295,10895671], which is available on the D. E. Shaw research website, was obtained via the accelerated weighted ensemble MD simulation of the RBD−ACE2 complex formation process.[32] The simulation employed the structure of PDB entry 6VW1 as a target system.[36] According to the trajectory data, ASN53, 90, 103, 322, and 546 of ACE2 and ASN343 of RBD were glycosylated. In the present study, only the glycans included in the PDB data were introduced. Four possible binding paths from the free energy surface were proposed, one of which (trajectory number 000713) was referred to as the initial structure for the present calculation. Based on the 89 structures that were contained in the binding path, we performed umbrella sampling with constraints on the distance between the centers of mass of RBD and ACE2. Sampling was performed at 10 ns for each window. Prior to sampling, a 1 ns equilibration simulation was performed. The simulations were conducted at 300 K and 1 bar in the NPT ensemble under the MC barostat and Langevin thermostat. Time step was set to 0.002 ps, and the SHAKE method was applied for hydrogen constrain. Amber ff14SB, GLYCAM 06j, and TIP3P force fields were employed for the proteins, glycans, and water, respectively.[37−39] A cubic water box with the periodic boundary condition was used, with box lengths of about 180 Å. Na^+ and Cl^- ions were added to a final concentration of 0.2 M.

2.2. 3D-RISM Analysis. We employed the 3D-RISM theory coupled with the Kovalenko-Hirata closure to evaluate the correlation functions and the SFE.[26−28] The SFE was calculated for 20 snapshots that were extracted every 500 ps from each trajectory. Here, all of the explicit water molecules and ions, with the exception of Zn^{2+}, were stripped from the snapshots. The same potential parameters as those used in the MD simulation were employed for all species. The temperature was 300 K, and a 0.2 M NaCl aqueous solution under ambient condition was assumed. The number of grid points in the 3D-RISM-KH calculations was 512,[3] with a spacing of 0.5 Å. All calculations were performed using the in-house 3D-RISM codes.[40,41]

3. RESULTS AND DISCUSSION

3.1. Binding Process. In Figure 1a, the distance between the center of mass of ACE2 and RBD was plotted against the umbrella sampling windows. Here, the averaged distances of 20 snapshots for each window were plotted. At the unbind state, window = 1, the interprotein distance was about 85 Å, and the distance decreased as the binding process progressed. At the binding state, window = 89, the interprotein distance reached a value of about 48 Å. At around window = 50, the interprotein distance seemed to have reached a plateau, with a distance around 50 Å. This behavior continued up to about window = 75, and the distance decreased to about 48 Å. Based on the distance change, the binding process can be considered to consist of two steps. The approaching step, in which the distance between the proteins decreases drastically, that is, windows = 1−49, is followed by a moderately changing step, that is, windows = 50−89. In Figure 1b, the snapshots of the protein structure taken from windows = 1, 50, and 89 are shown. At window = 50, the distance between the proteins was sufficiently close that the structure was similar to the bound state at window = 89. Therefore, the process after window = 50 would correspond to the arrangement of the local protein structure. In Figure 1c, the free energy of the system is plotted, which was defined by

$$G_{total} = E_{MM} + \Delta\mu \quad (1)$$

where E_{MM} and $\Delta\mu$ denote the protein structure energy and the SFE, respectively. The plotted values were relative to the average value of window = 1. The change in the total free energy was rather small compared with its components. The total free energy change at window = 89 from window = 1 was about −10 kcal/mol. Because of computational cost limitations, only 20 snapshots were calculated for each window; however, to evaluate the average value of the free energy with high accuracy, the initial and final states were calculated for 100 structures, and the free energy change was still about −10 kcal/mol. The components of the free energy, namely, the protein structure energy and the SFE, exhibited dramatic changes through the binding process. As the proteins approached each other, the conformational energy decreased, whereas the SFE increased. This was attributed to the penalty of free energy caused by desolvation that occurred as the distance between the proteins decreased. Because the first noticeable change in this process was observed at window = 34, it is likely that partial protein−protein contact occurred at this distance, which stabilized a large protein−protein interaction and caused a desolvation penalty. Subsequently, both the structure energy and the SFE changed gradually.

To examine the changes in interprotein interactions that occur during the binding process, the structural energy and number of hydrogen bonds were plotted (Figure 2). The structure energy of the protein can be split into the

https://doi.org/10.1021/acs.jcim.2c00192
J. Chem. Inf. Model. 2022, 62, 2889−2898

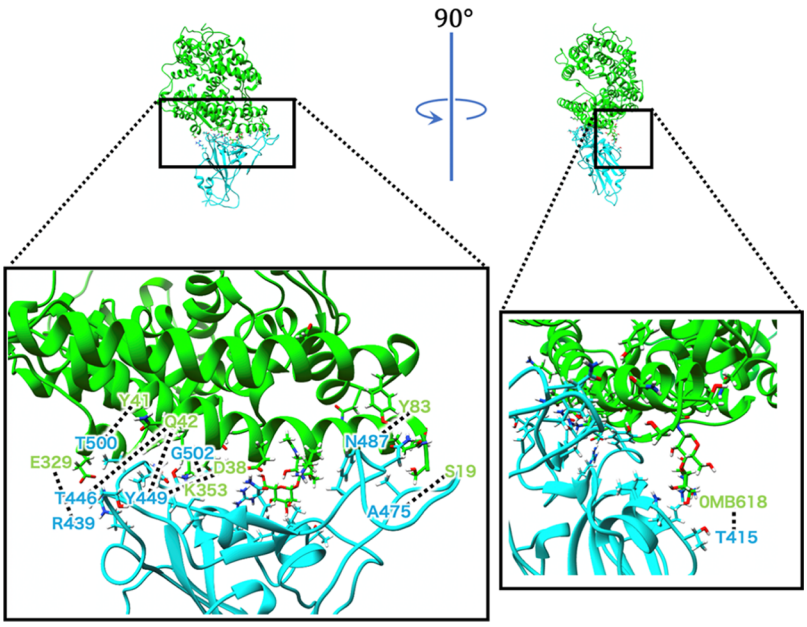


Figure 3. Illustration of the interprotein hydrogen bond between ACE2 and RBD for the snapshot taken from window = 89. The panel on the right shows the structure in the left panel rotated 90° around the vertical axis. 0MB618 denotes the β-L-mannose at the terminal of glycan connected to ACE2/ASN90.

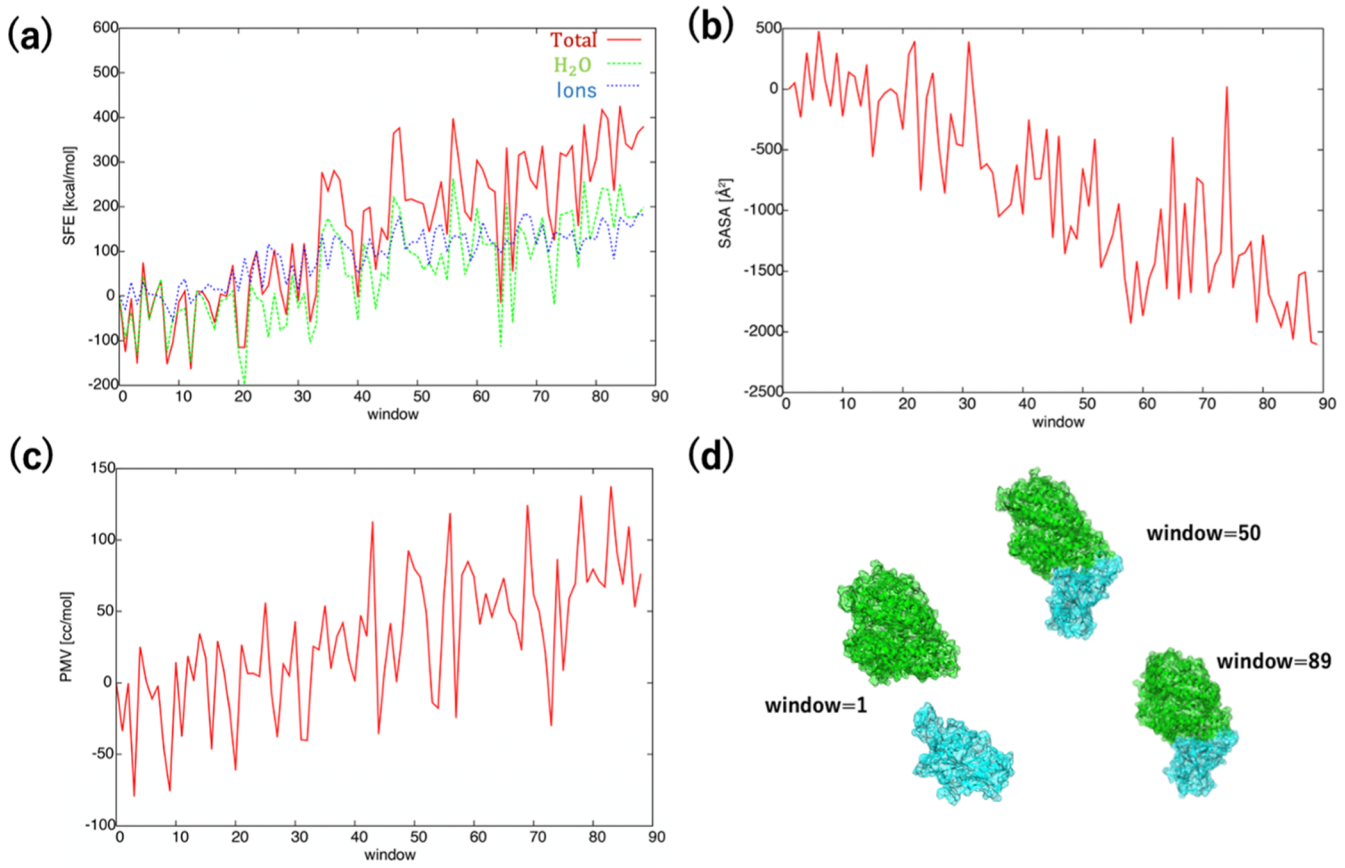


Figure 4. Properties related to the solvation structure. The SFE, the SASA, and the PMV are plotted in (a−c), respectively. All values plotted here are the relative value vs. those in window = 1. The van der Waals surface representations for the snapshots taken from windows = 1, 50, and 89 are depicted in (d).

intraprotein structure energy and interprotein interaction energy. As observed in Figure 2a, although the intraprotein structure energy fluctuated strongly, its central value appeared to remain almost unchanged during the protein-binding process. Conversely, the interprotein interaction energy provided a dominant contribution to the stabilization. In

https://doi.org/10.1021/acs.jcim.2c00192
J. Chem. Inf. Model. 2022, 62, 2889−2898

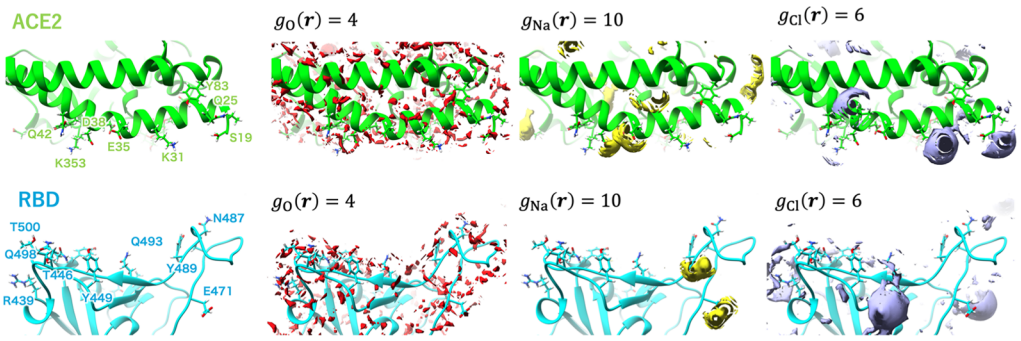


Figure 5. Isovalue surfaces of the 3D-DFs of the water oxygen, sodium ion, and chloride ion at window = 89 depicted in red-, yellow-, and purple-colored surfaces, respectively. The isovalues of each 3D-DF are indicated in each panel. The labels of the major amino acid residues involved in interprotein hydrogen bonding are shown in the leftmost panel.

particular, significant stabilization was observed around window = 34, which was attributed to the formation of interprotein hydrogen bonds, as shown in Figure 2b. The number of protein−protein hydrogen bonds increased rapidly around window = 34, followed by a further increase after window = 70. As observed in Figure 1a, after window = 50, the distance between the centers of mass changed only slightly, whereas the stabilization energy changed significantly. This may be attributed to the contribution of local structural changes to hydrogen-bond formation. Figure 2c shows a clear correlation between hydrogen-bond formation and interprotein interaction energy. However, it is known that the stabilization energy of one hydrogen bond is about 10 kcal/mol, and the stabilization of the interprotein interaction (about −400 kcal/mol) was too large for the number of hydrogen bonds (up to about 9) in this case. Therefore, the interprotein interaction might be stabilized by the electrostatic interaction and van der Waals (vdW) interaction, in addition to hydrogen bonds.

To address the local structure of the hydrogen bonds between ACE2 and RBD, lists of the hydrogen-bond pairs with high probability of formation are listed in Tables 1 and S1. At window = 50, the glycan of ACE2 formed hydrogen bonds with RBD residues. In addition, the frequency of hydrogen-bond formation between RBD/ASN501 and ACE2/LYS353 and between RBD/ASN487 and ACE2/SER19 or ACE2/GLN24 was also high. However, the hydrogen-bond fraction detected at window = 50 was less than 20%, even for the largest fraction one, suggesting that a strong interprotein interaction is not yet established at this stage. At window = 89, the hydrogen bond fraction increased, indicating a strong interprotein interaction. The residues forming the interprotein hydrogen bond between ACE2 and RBD at window = 89 in higher fractions are illustrated in Figure 3. ACE2/LYS353 and RBD/GLY502 were located almost at the center of the protein−protein interface, which has the highest fraction of hydrogen bonds. The hydrogen bonds of ACE2/TYR83-RBD/ASN487 and ACE2/SER19-RBD/ALA475 are located right at the edge of the protein−protein interface. These residues correspond to those in "hotspots" suggested by previous studies.[3] In this snapshot, ACE2/GLN42, RBD/THR446, and RBD/TYR449 appeared to be too far away to form hydrogen bonds, in contrast with their high hydrogen-bond-formation fraction. This suggests that the loop structure including RBD/THR446 and RBD/TYR449 is highly fluctuating. Most of the residues were located at the N-terminal helix of ACE2, and the loop consisted of ASN487 to TYR505 of RBD; however, interestingly, the glycan connected to ACE2/ASN90 and THR415 of RBD formed a hydrogen bond with high probability. This hydrogen bond showed highest probability at window = 50 and even earlier windows. In fact, the hydrogen bond formed at window = 34, as shown in Figure 2b, is between this glycan connected to ACE2/ASN90 and THR415 of RBD. This suggests that the hydrogen-bond formation between the flexible glycans of ACE2 and RBD amino acids plays an important role in the early process of RBD−ACE2 binding. Although a previous study observed ASN53 contacts in addition to ASN90, ASN53 contacts were weak in the present study.[10] This is considered to be due to the short glycans used in the present model compared with that study. For further investigation of the role of glycans, a more realistic model for glycans needs to be used.

3.2. Solvation. In this subsection, let us consider the solvation structure and its role in protein binding. In Figure 4a, the SFE and its components are plotted against the window steps. The SFE can be written as in the 3D-RISM framework. The summation on the right-hand side of the equation is taken for all the solvent species; thus, the SFE can be split into components corresponding to water and ions in the present case. The total SFE increased with the progression of the binding process, and the components of SFE also increased. At about window = 34, the SFE showed a drastic change, which was in accordance with the structural change of the protein, as mentioned above. The SFE of water showed a larger fluctuation compared with that of ions. It is indicated that the ions may coordinate to the relatively rigid structure of the protein. The increase in SFE was caused by the desolvation associated with protein binding. To investigate desolvation in terms of protein conformational changes, the solvent-accessible surface area (SASA) and the PMV are plotted in Figure 4b,c, respectively. The SASA was calculated using the "surf" protocol in cpptraj of the amber suite.[42] The PMV was calculated from the distribution function evaluated by the 3D-RISM theory.[43,44] The SASA decreased with the progression of the protein-binding process. Conversely, the PMV increased with protein binding. This was probably caused by the decrease in the solvent distribution in the first solvation shell because of the decrease in the SASA. The decrease in the interprotein distance resulted in an increase in interprotein interactions, rather than in protein−solvent interactions. Figure 4d shows the snapshots and vdW surfaces for windows = 1, 50, and 89. No contact between the vdW surfaces of the two proteins was shown in window = 1, whereas windows = 50 and 89 showed

https://doi.org/10.1021/acs.jcim.2c00192
J. Chem. Inf. Model. 2022, 62, 2889−2898

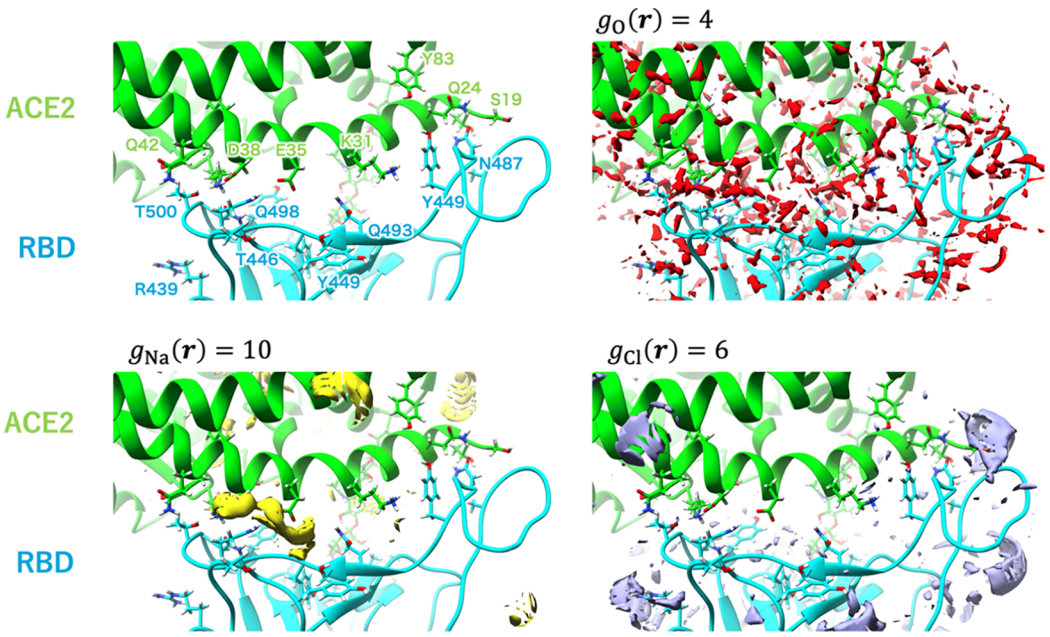


Figure 6. Isovalue surfaces of the 3D-DFs of the water oxygen, sodium ion, and chloride ion at window = 50 depicted in red-, yellow-, and purple-colored surfaces, respectively. The isovalues of each 3D-DF are indicated in each panel. The labels of the major amino acid residues involved in interprotein hydrogen bonding are shown in the left upper panel.

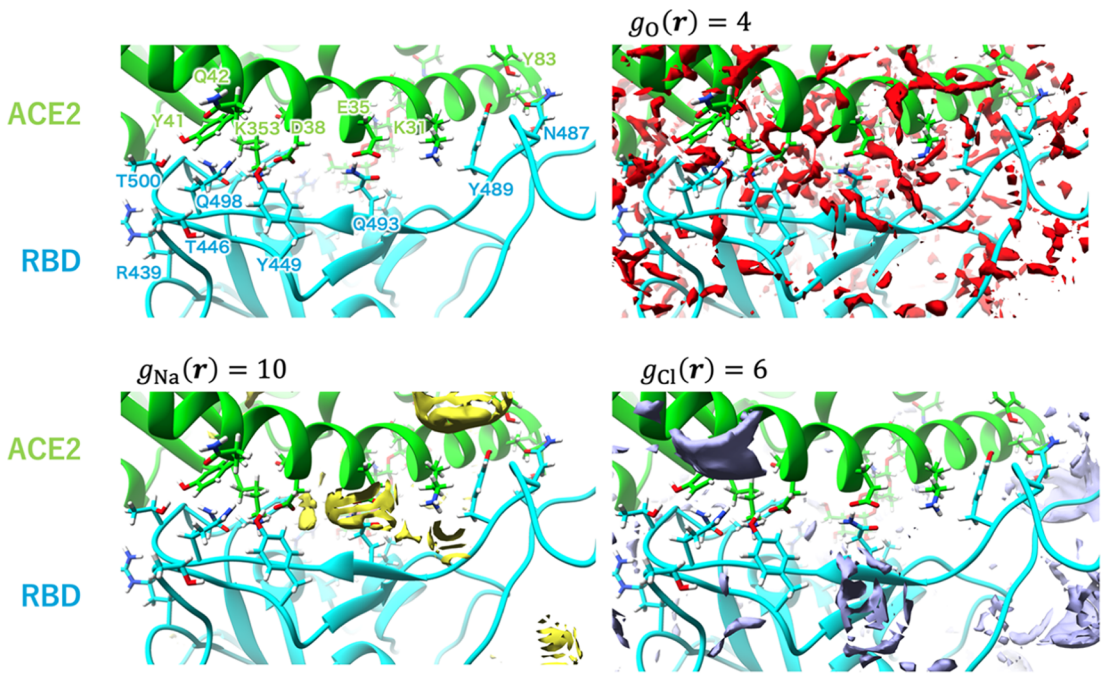


Figure 7. Isovalue surfaces of the 3D-DFs of the water oxygen, sodium ion, and chloride ion at window = 89 depicted in red-, yellow-, and purple-colored surfaces, respectively. The isovalues of each 3D-DF are indicated in each panel. The labels of the major amino acid residues involved in interprotein hydrogen bonding are shown in the left upper panel.

clear contact. These images support the discussion provided above.

$$\Delta\mu = \sum_{\nu}^{\text{solvent}} \Delta\mu_{\nu} \tag{2}$$

To visualize the interprotein interaction and the solvation structure of the protein, the snapshots and solvent distribution for windows = 1, 50, and 89 are shown in Figures 5−7. Figure 5 shows the structure near the RBD−ACE2 interface in the initial state (window = 1), displaying the atom model of the hydrogen-bond donor−acceptor residues listed in Table 1. At window = 1, the distance between ACE2 and RBD was sufficiently large for them to be fully solvated independently. The 3D distribution functions (3D-DFs) of solvent species were also considered. (See also Figure S1.) The water molecules were widely distributed on the protein surface of both ACE2 and RBD. The sodium ion peaks on the RBD side correspond to the coordination to RBD/GLU484 and RBD/GLU471. For chloride ions, there was a prominent peak at

https://doi.org/10.1021/acs.jcim.2c00192
J. Chem. Inf. Model. 2022, 62, 2889−2898

Table 1. List of the Five Most Frequently Formed Hydrogen Bonds at Windows = 50 and 89.[a]

acceptor residue	atom	donor residue	atom	fraction
		window = 50		
RBD		ACE2		
THR415	O	0MB618	H4O	0.18
ASN501	OD1	LYS353	HZ2	0.18
ASN501	OD1	LYS353	HZ3	0.18
ASN501	OD1	LYS353	HZ1	0.17
ASN487	OD1	GLN24	HE22	0.12
ACE2		RBD		
0MB618	O6	ARG408	HH11	0.12
0MB618	O3	GLN409	HE21	0.1
0MB618	O3	VAL417	H	0.1
GLN24	OE1	TYR489	HH	0.07
0MB618	O4	ARG408	HH11	0.07
		window = 89		
RBD		ACE2		
ASN487	OD1	TYR83	HH	0.58
ALA475	O	SER19	HG	0.54
THR415	O	0MB618	H3O	0.32
THR446	O	GLN42	HE21	0.29
TYR449	OH	GLN42	HE22	0.25
ACE2		RBD		
LYS353	O	GLY502	H	0.81
TYR41	OH	THR500	HG1	0.69
ASP38	OD2	TYR449	HH	0.51
GLU329	OE2	ARG439	HH11	0.43
GLU329	OE2	ARG439	HH21	0.36

[a]The cases in which RBD is the hydrogen-bond receptor and ACE2 is the donor and vice versa are listed. The evaluation was performed on snapshots of 100 frames, and the fractions in which hydrogen bonding was observed are shown. The atomic labels are according to the convention of the AMBER force field. 0MB618 denotes the β-L-mannose at the terminal of the glycan connected to ACE2/ASN90.

ACE2/LYS31 near the center of the interface. The left peak corresponded to the coordination to ACE2/LYS68, and the RBD showed coordination to RBD/LYS452 and 458, which is not involved in hydrogen bonding.

At window = 50 (Figures 6 and S2), the RBD−ACE2 interface was quite close, and hydrogen-bond formation was evident. Interestingly, there was sufficient distribution of water at the interface, and ion coordination to ACE2/GLU35 and ACE2/ASP38 was also maintained. Conversely, there was no distribution of chloride ions to ACE2/LYS31 at this threshold of the isovalue surface plot, indicating that the ion coordination was weakened.

At window = 89 (Figures 7 and S3), the RBD−ACE2 interface was well contacted and interprotein hydrogen bonds had formed. Some water distributions were maintained at the interface, although they were weaker compared with the former window. In particular, there were strong water peaks at ACE2/LYS31 and RBD/GLN493, 498. A broad sodium ion peak was observed at ACE2/GLU35, whereas the peak at ACE2/ASP38 was reduced compared with window = 50. Chloride ion peaks were rarely detected on the interface at this stage.

In Figure 8, the Placement algorithm was applied to investigate the position of explicit water molecules at the RBD−ACE2 interface.[45] Water molecules coordinated to ACE2/LYS31, GLU35, and LYS353 and RBD/TYR489 and GLN498 were observed. The water molecule on the left side was coordinated to three residues at the same time, namely, ACE2/ASP38, LYS353, and RBD/GLN498. Sodium ions were also examined in the same way. The sodium ion corresponding to the peak near ACE2/ASP38 in Figure 7 was observed, and coordination to RBD/TYR449 was also observed.

3.3. Structural Fluctuation. To examine the global conformational fluctuation of proteins caused by the binding, a PCA of protein structural change was performed.[33−35] For PCA, we used the cartesian coordinates of the heavy atoms of amino acids of a protein. We applied PCA to the coordinates of all snapshots for monomer ACE2 and RBD after removing the translational and rotational motions from the trajectories. The top few PCs were calculated to investigate conformational fluctuations.

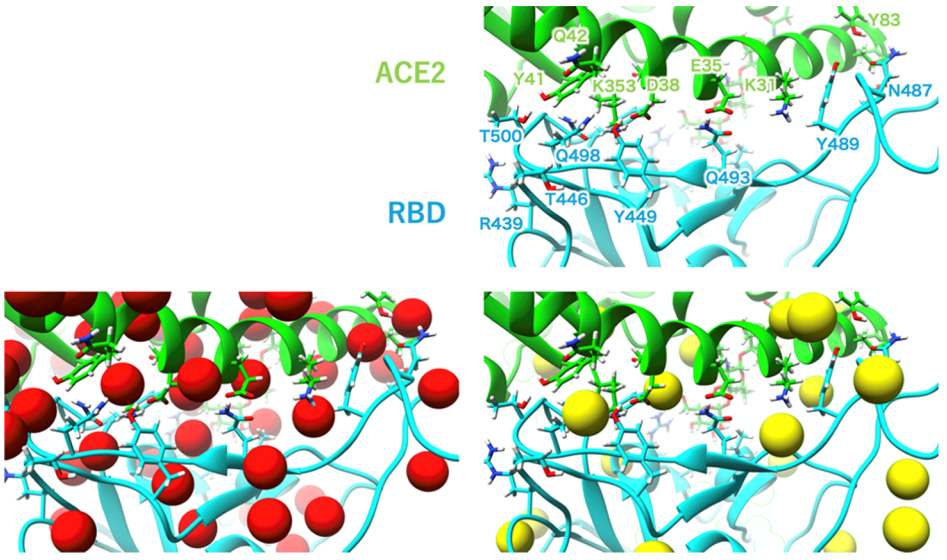


Figure 8. Explicit positions of the oxygen atoms of water and sodium ions determined using the Placement algorithm for the 3D-DFs at window = 89 depicted in red- and yellow-colored spheres, respectively. The labels of the major amino acid residues involved in interprotein hydrogen bonding are shown in the right upper panel.

https://doi.org/10.1021/acs.jcim.2c00192
J. Chem. Inf. Model. 2022, 62, 2889−2898

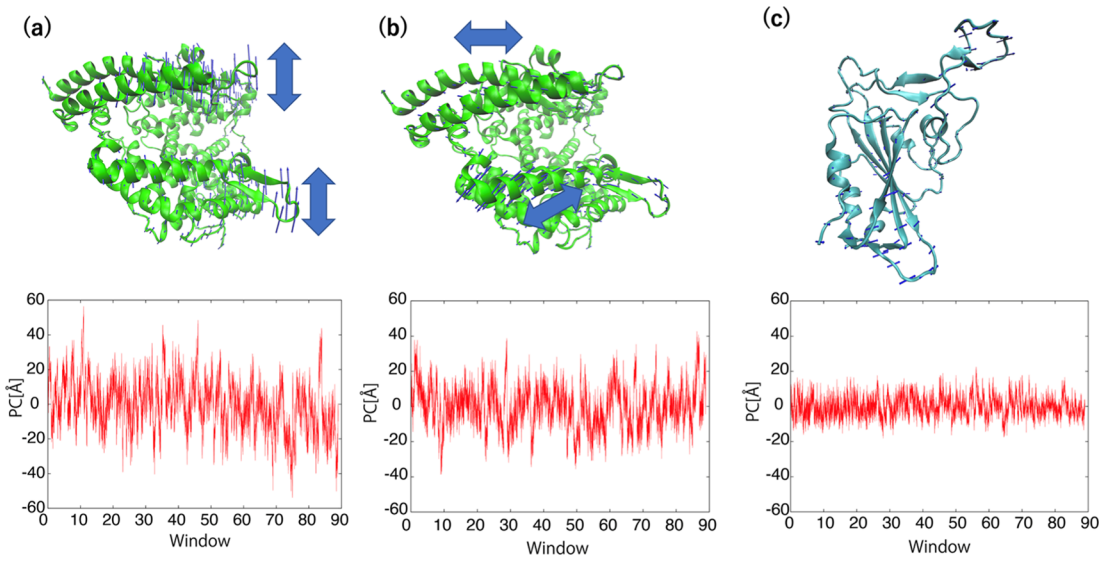


Figure 9. PCs of protein structural changes. (a,b) First and second PCs of ACE2, respectively and (c) first PC of RBD. The eigen vectors of PC are depicted by the thin arrows in the upper panels. The large bold arrows schematically indicate the direction of the conformational change. For ease of understanding, the size of the thin arrows representing PC in the upper panels has been enlarged 10 times that of the actual structural fluctuation.

The results of the first PC (PC1) and the second PC (PC2) for ACE2 and PC1 for RBD are shown in Figure 9a−c, respectively. ACE2 underwent a characteristic conformational change upon protein binding, whereas the RBD showed no specific change. PC1 of ACE2 corresponded to the open−close structure fluctuation of the hydrophobic pocket. The eigen vectors on the first and second helices of the C terminus were pointing downward, whereas those of the third and fourth helices were pointing upward, as shown in Figure 9a. As seen in the lower panel of Figure 9a, the PC seemed to change at around window = 50. After window = 50, PC took a negative value, indicating that the hydrophobic pocket opened slightly. The change corresponded to the interprotein hydrogen-bond formation discussed above, suggesting that there is a correlation between the open−close structural flexibility of the hydrophobic pocket of ACE2 and the binding process. The PC2 of ACE2 depicted in Figure 9b shows the sliding motion, namely, the eigen vectors of first and second helices of the C terminus were pointing to the right, whereas those of the third and fourth helices were pointing to the left. Conversely, the PC1 of RBD was rather small. The PC1 of RBD corresponded to the local structural fluctuation in the loop structures. The magnitude of the RBD structural fluctuation remained almost unchanged after complex formation.

4. SUMMARY

The binding process of ACE2 to the RBD of the spike protein of SARS-CoV2 was investigated using MD and the 3D-RISM theory.

The binding process can be regarded as consisting of two distinct steps, namely, a protein−protein approaching step and a local structural rearrangement step. In the approaching step, the interprotein interaction energy decreased as the proteins approached each other, whereas the SFE increased. As the distance between the proteins decreased to about 55 Å, interprotein hydrogen bonds began to form. At that time, it was revealed that the flexible glycan chain of ACE2 established hydrogen bonds with RBD. The importance of glycans has been highlighted in previous studies.[10−12,14] In the present analysis, only the glycans included in the PDB were considered, but more sophisticated models need to be used to examine the function of glycans in greater detail.[13] Thereafter, the number of interprotein hydrogen bonds increased rapidly, and eventually about 6−10 hydrogen bonds were established. The SFE increased because of the desolvation of the protein as it approached its partner. The SASA and PMV analyses suggest that the dehydration of hydrophilic residues on the protein surface contributed significantly to this increase in SFE. The spatial distribution function of the solvent revealed the presence of hydrogen bonds bridged by water molecules on the RBD−ACE2-binding interface. Furthermore, a PCA of the protein structures showed structural changes associated with complex formation. ACE2 exhibited a pronounced conformational change, such as the closing of the hydrophobic pocket, upon binding to RBD, whereas there was no significant change in RBD.

It is expected that the findings reported in this study will contribute to a deeper understanding of viral infection and replication, which is required to overcome the current human-scale crisis caused by coronaviruses.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00192.

Additional hydrogen bonding list and details of solvent distribution (PDF).

■ AUTHOR INFORMATION

Corresponding Author

Norio Yoshida − *Department of Chemistry, Graduate School of Science, Kyushu University, Fukuoka 819-0395, Japan; Department of Complex Systems Science, Graduate School of Informatics, Nagoya 464-8601, Japan;* orcid.org/0000-0002-2023-7254; Email: noriwo@nagoya-u.jp

https://doi.org/10.1021/acs.jcim.2c00192
J. Chem. Inf. Model. 2022, 62, 2889−2898

Authors

Yutaka Maruyama − *Department of Physics, School of Science and Technology, Meiji University, Kawasaki 214-8571, Japan*

Ayori Mitsutake − *Department of Physics, School of Science and Technology, Meiji University, Kawasaki 214-8571, Japan;* orcid.org/0000-0002-4194-7255

Akiyoshi Kuroda − *RIKEN Center for Computational Science, Kobe 650-0047, Japan*

Ryo Fujiki − *Department of Chemistry, Graduate School of Science, Kyushu University, Fukuoka 819-0395, Japan*

Kodai Kanemaru − *Department of Chemistry, Graduate School of Science, Kyushu University, Fukuoka 819-0395, Japan*

Daisuke Okamoto − *Department of Chemistry, Graduate School of Science, Kyushu University, Fukuoka 819-0395, Japan*

Alexander E. Kobryn − *Nanotechnology Research Centre, National Research Council Canada, Edmonton AB T6G 2M9, Canada*

Sergey Gusarov − *Nanotechnology Research Centre, National Research Council Canada, Edmonton AB T6G 2M9, Canada;* orcid.org/0000-0003-2033-705X

Haruyuki Nakano − *Department of Chemistry, Graduate School of Science, Kyushu University, Fukuoka 819-0395, Japan;* orcid.org/0000-0002-7008-0312

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c00192

Author Contributions

N.Y., Y.M., and A.M. performed the M.D. and 3D-RISM computations. R.F., K.K., and D.O. set up the M.D. and 3D-RISM computation, including parameter assignment. N.Y., Y.M., and A.K. implemented and tuned the 3D-RISM code on the supercomputer Fugaku system. N.Y., Y.M., A.M., A.K., and S.G. designed this study. All analyses were performed after discussion with all authors. The manuscript was written based on contributions from all authors. All authors have given approval of the final version of the manuscript.

Funding

This work was supported by JSPS KAKENHI (grant no 19K03768 and 20H03230) and the JST-SICORP program (grant no 21445860).

Notes

The authors declare no competing financial interest.

Requirements for Data and Software Availability: MD trajectory, 3D-RISM input/output, and parameter files obtained from the present calculations can be downloaded from the following repository: https://github.com/nyoriwo/SI_ci-2022-00192m.git. The 3D-RISM code for the multiple GPGPU system is available from the following repository: https://github.com/drmaruyama/3D-RISM-CUDA-MPI.git.

■ ACKNOWLEDGMENTS

Numerical calculations were conducted, in part, at the Research Center for Computational Science, Institute for Molecular Science, National Institutes of Natural Sciences, the Center for Computational Science, University of Tsukuba, and the supercomputer Fugaku, RIKEN, Japan. The RBD−ACE2-binding MD trajectory was provided by D. E. Shaw Research Technical Data, 2020.[32] Molecular graphics were depicted with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.[46]

■ ABBREVIATIONS

3D-RISM, three-dimensional reference interaction-site model; MD, molecular dynamics; ACE2, angiotensin-converting enzyme-2; RBD, receptor-binding domain

■ REFERENCES

(1) V'kovski, P.; Kratzel, A.; Steiner, S.; Stalder, H.; Thiel, V. Coronavirus Biology and Replication: Implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **2021**, *19*, 155−170.

(2) Korber, B.; Fischer, W. M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B.; Hastie, K. M.; Parker, M. D.; Partridge, D. G.; Evans, C. M.; Freeman, T. M.; de Silva, T. I.; McDanal, C.; Perez, L. G.; Tang, H.; Moon-Walker, A.; Whelan, S. P.; LaBranche, C. C.; Saphire, E. O.; Montefiori, D. C.; Angyal, A.; Brown, R. L.; Carrilero, L.; Green, L. R.; Groves, D. C.; Johnson, K. J.; Keeley, A. J.; Lindsey, B. B.; Parsons, P. J.; Raza, M.; Rowland-Jones, S.; Smith, N.; Tucker, R. M.; Wang, D.; Wyles, M. D. Tracking Changes in SARS-CoV-2 Spike: Evidence That D614g Increases Infectivity of the Covid-19 Virus. *Cell* **2020**, *182*, 812−827.

(3) Jawad, B.; Adhikari, P.; Podgornik, R.; Ching, W.-Y. Key Interacting Residues between RBD of SARS-CoV-2 and ACE2 Receptor: Combination of Molecular Dynamics Simulation and Density Functional Calculation. *J. Chem. Inf. Model.* **2021**, *61*, 4425−4441.

(4) Li, Z.; Zhang, J. Z. H. Quantitative Analysis of ACE2 Binding to Coronavirus Spike Proteins: SARS-CoV-2 Vs. SARS-CoV and Ratg13. *Phys. Chem. Chem. Phys.* **2021**, *23*, 13926−13933.

(5) Kumawat, N.; Tucs, A.; Bera, S.; Chuev, G. N.; Valiev, M.; Fedotova, M. V.; Kruchinin, S. E.; Tsuda, K.; Sljoka, A.; Chakraborty, A. Site Density Functional Theory and Structural Bioinformatics Analysis of the SARS-CoV Spike Protein and Hace2 Complex. *Molecules* **2022**, *27*, 799.

(6) Freitas, F. C.; Ferreira, P. H. B.; Favaro, D. C.; Oliveira, R. J. d. Shedding Light on the Inhibitory Mechanisms of SARS-CoV-1/Cov-2 Spike Proteins by ACE2-Designed Peptides. *J. Chem. Inf. Model.* **2021**, *61*, 1226−1243.

(7) Zou, J.; Yin, J.; Fang, L.; Yang, M.; Wang, T.; Wu, W.; Bellucci, M. A.; Zhang, P. Computational Prediction of Mutational Effects on SARS-CoV-2 Binding by Relative Free Energy Calculations. *J. Chem. Inf. Model.* **2020**, *60*, 5794−5802.

(8) Barton, M. I.; MacGowan, S. A.; Kutuzov, M. A.; Dushek, O.; Barton, G. J.; van der Merwe, P. A. Effects of Common Mutations in the SARS-CoV-2 Spike RBD and Its Ligand, the Human ACE2 Receptor on Binding Affinity and Kinetics. *eLife* **2021**, *10*, 10.

(9) Ali, A.; Vijayan, R. Dynamics of the ACE2-SARS-CoV-2/SARS-CoV Spike Protein Interface Reveal Unique Mechanisms. *Sci. Rep. Uk* **2020**, *10*, 14214.

(10) 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.

(11) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6*, 1722−1734.

(12) Harbison, A. M.; Fogarty, C. A.; Phung, T. K.; Satheesan, A.; Schulz, B. L.; Fadda, E. Fine-Tuning the Spike: Role of the Nature and Topology of the Glycan Shield in the Structure and Dynamics of the SARS-CoV-2 S. *Chem. Sci.* **2022**, *13*, 386−395.

(13) Gong, Y.; Qin, S.; Dai, L.; Tian, Z. The Glycosylation in SARS-CoV-2 and Its Receptor ACE2. *Signal Transduction Targeted Ther.* **2021**, *6*, 396.

(14) Sztain, T.; Ahn, S.-H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S.; Mashayekhi, G.; McCammon, J. A.; Ourmazd, A.; Frank, J.; McLellan, J. S.; Chong, L. T.; Amaro, R. E. A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. *Nat. Chem.* **2021**, *13*, 963−968.
(15) Cong, Y.; Feng, Y.; Ni, H.; Zhi, F.; Miao, Y.; Fang, B.; Zhang, L.; Zhang, J. Z. H. Anchor-Locker Binding Mechanism of the Coronavirus Spike Protein to Human ACE2: Insights from Computational Analysis. *J. Chem. Inf. Model.* **2021**, *61*, 3529−3542.
(16) Kobryn, A. E.; Maruyama, Y.; Velázquez-Martínez, C. A.; Yoshida, N.; Gusarov, S. Modeling the Interaction of SARS-CoV-2 Binding to the ACE2 Receptor Via Molecular Theory of Solvation. *New J. Chem.* **2021**, *45*, 15448−15457.
(17) Rajpoot, S.; Ohishi, T.; Kumar, A.; Pan, Q.; Banerjee, S.; Zhang, K. Y. J.; Baig, M. S. A Novel Therapeutic Peptide Blocks SARS-CoV-2 Spike Protein Binding with Host Cell ACE2 Receptor. *Drugs R&D* **2021**, *21*, 273−283.
(18) Sharma, G.; Song, L. F.; Merz, K. M. Effect of an Inhibitor on the ACE2-Receptor-Binding Domain of SARS-CoV-2. *J. Chem. Inf. Model.* **2022**, DOI: 10.1021/acs.jcim.1c01283.
(19) Wang, X.; Bie, L.; Gao, J. Structural Insights into the Cofactor Role of Heparin/Heparan Sulfate in Binding between the SARS-CoV-2 Spike Protein and Host Angiotensin-Converting Enzyme Ii. *J. Chem. Inf. Model.* **2022**, *62*, 656−667.
(20) Verkhivker, G. M.; Agajanian, S.; Oztas, D. Y.; Gupta, G. Allosteric Control of Structural Mimicry and Mutational Escape in the SARS-CoV-2 Spike Protein Complexes with the ACE2 Decoys and Miniprotein Inhibitors: A Network-Based Approach for Mutational Profiling of Binding and Signaling. *J. Chem. Inf. Model.* **2021**, *61*, 5172−5191.
(21) Ghasemitarei, M.; Privat-Maldonado, A.; Yusupov, M.; Rahnama, S.; Bogaerts, A.; Ejtehadi, M. R. Effect of Cysteine Oxidation in SARS-CoV-2 Receptor-Binding Domain on Its Interaction with Two Cell Receptors: Insights from Atomistic Simulations. *J. Chem. Inf. Model.* **2021**, 129−141.
(22) Golcuk, M.; Hacisuleyman, A.; Erman, B.; Yildiz, A.; Gur, M. Binding Mechanism of Neutralizing Nanobodies Targeting SARS-CoV-2 Spike Glycoprotein. *J. Chem. Inf. Model.* **2021**, *61*, 5152−5160.
(23) Wu, L.; Zhou, L.; Mo, M.; Liu, T.; Wu, C.; Gong, C.; Lu, K.; Gong, L.; Zhu, W.; Xu, Z. SARS-CoV-2 Omicron RBD Shows Weaker Binding Affinity Than the Currently Dominant Delta Variant to Human ACE2. *Signal Transduction Targeted Ther.* **2022**, *7*, 8. DOI: 10.1038/s41392-021-00863-2
(24) Cui, Z.; Liu, P.; Wang, N.; Wang, L.; Fan, K.; Zhu, Q.; Wang, K.; Chen, R.; Feng, R.; Jia, Z.; Yang, M.; Xu, G.; Zhu, B.; Fu, W.; Chu, T.; Feng, L.; Wang, Y.; Pei, X.; Yang, P.; Xie, X. S.; Cao, L.; Cao, Y.; Wang, X. Structural and Functional Characterizations of Infectivity and Immune Evasion of SARS-CoV-2 Omicron. *Cell* **2022**, 860−871.e13.
(25) Verma, J.; Subbarao, N. Insilico Study on the Effect of SARS-CoV-2 RBD Hotspot Mutants' Interaction with ACE2 to Understand the Binding Affinity and Stability. *Virology* **2021**, *561*, 107−116.
(26) Beglov, D.; Roux, B. An Integral Equation to Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997**, *101*, 7821−7826.
(27) Kovalenko, A.; Hirata, F. Three-Dimensional Density Profiles of Water in Contact with a Solute of Arbitrary Shape: A RISM Approach. *Chem. Phys. Lett.* **1998**, *290*, 237−244.
(28) Kovalenko, A.; Hirata, F. Self-Consistent Description of a Metal-Water Interface by the Kohn-Sham Density Functional Theory and the Three-Dimensional Reference Interaction Site Model. *J. Chem. Phys.* **1999**, *110*, 10095−10112.
(29) Hirata, F. *Molecular Theory of Solvation*; Kluwer: Dordrecht, 2003.
(30) Yoshida, N. Role of Solvation in Drug Design as Revealed by the Statistical Mechanics Integral Equation Theory of Liquids. *J. Chem. Inf. Model.* **2017**, *57*, 2646−2656.
(31) Yoshida, N.; Imai, T.; Phongphanphanee, S.; Kovalenko, A.; Hirata, F. Molecular Recognition in Biomolecules Studied by Statistical-Mechanical Integral-Equation Theory of Liquids. *J. Phys. Chem. B* **2009**, *113*, 873−886.
(32) Shaw, D. E. Molecular Dynamics Simulations Related to SARS-CoV-2. http://www.deshawresearch.com/resources_sarscov2.html/.
(33) Kitao, A.; Hirata, F.; Gō, N. The Effects of Solvent on the Conformation and the Collective Motions of Protein - Normal Mode Analysis and Molecular-Dynamics Simulations of Melittin in Water and in Vacuum. *Chem. Phys.* **1991**, *158*, 447−472.
(34) Ichiye, T.; Karplus, M. Collective Motions in Proteins - a Covariance Analysis of Atomic Fluctuations in Molecular-Dynamics and Normal Mode Simulations. *Proteins* **1991**, *11*, 205−217.
(35) Abagyan, R.; Argos, P. Optimal Protocol and Trajectory Visualization for Conformational Searches of Peptides and Proteins. *J. Mol. Biochem.* **1992**, *225*, 519−532.
(36) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural Basis of Receptor Recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221−224.
(37) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Xu, H.; Sherman, W.; York, D. M. Alchemical Binding Free Energy Calculations in Amber20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595−5623.
(38) Tessier, M. B.; DeMarco, M. L.; Yongye, A. B.; Woods, R. J. Extension of the Glycam06 Biomolecular Force Field to Lipids, Lipid Bilayers and Glycolipids. *Mol. Simulat.* **2008**, *34*, 349−364.
(39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(40) Maruyama, Y.; Hirata, F. Modified Anderson Method for Accelerating 3D-RISM Calculation Using Graphis Processor Unit. *J. Chem. Theory Comput.* **2012**, *8*, 3015−3021.
(41) Maruyama, Y.; Yoshida, N.; Tadano, H.; Takahashi, D.; Sato, M.; Hirata, F. Massively Parallel Implementation of 3D-RISM Calculation with Volumetric 3d-Fft. *J. Comput. Chem.* **2014**, *35*, 1347−1355.
(42) Weiser, J. r.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (Lcpo). *J. Comput. Chem.* **1999**, *20*, 217−230.
(43) Imai, T.; Kinoshita, M.; Hirata, F. Theoretical Study for Partial Molar Volume of Amino Acids in Aqueous Solution: Implication of Ideal Fluctuation Volume. *J. Chem. Phys.* **2000**, *112*, 9469−9478.
(44) Imai, T.; Nomura, H.; Kinoshita, M.; Hirata, F. Partial Molar Volume and Compressibility of Alkali−Halide Ions in Aqueous Solution: Hydration Shell Analysis with an Integral Equation Theory of Molecular Liquids. *J. Phys. Chem. B* **2002**, *106*, 7308−7314.
(45) Sindhikara, D. J.; Yoshida, N.; Hirata, F. Placevent: An Algorithm for Prediction of Explicit Solvent Atom Distribution-Application to Hiv-1 Protease and F-Atp Synthase. *J. Comput. Chem.* **2012**, *33*, 1536−1543.
(46) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.

https://doi.org/10.1021/acs.jcim.2c00192
J. Chem. Inf. Model. 2022, 62, 2889−2898