Drug binding dynamics of the dimeric SARS-CoV-2 main protease, determined by molecular dynamics simulation

Teruhisa S. Komatsu[†*1], Noriaki Okimoto[†1,2], Yohei M. Koyama[1],
Yoshinori Hirano[1,2], Gentaro Morimoto[1,2], Yousuke Ohno[1], and Makoto Taiji[*1,2]

[1]Laboratory for Computational Molecular Design
[2]Drug Discovery Molecular Simulation Platform Unit
RIKEN Center for Biosystems Dynamics Research (BDR)
6-2-3 Furuedai, Suita, Osaka 565-0874, Japan

†These authors contributed equally to the work.
*To whom correspondence should be addressed.
    Teruhisa S. Komatsu and Makoto Taiji
    Laboratory for Computational Molecular Design, RIKEN BDR
    6-2-3 Furuedai, Suita, Osaka 565-0874, Japan.
    Tel. +81-6-6872-4850, Fax. +81-6-6872-4818
    E-mail: teruhisa.komatsu@riken.jp, taiji@riken.jp

Abstract

We performed molecular dynamics simulation of the dimeric SARS-CoV-2 (severe acute respiratory syndrome corona virus 2) main protease (M^{pro}) to examine the binding dynamics of small molecular ligands. Seven HIV inhibitors, darunavir, indinavir, lopinavir, nelfinavir, ritonavir, saquinavir, and tipranavir, were used as the potential lead drugs to investigate access to the drug binding sites in M^{pro}. The frequently accessed sites on M^{pro} were classified based on contacts between the ligands and the protein, and the differences in site distributions of the encounter complex were observed among the ligands. All seven ligands showed binding to the active site at least twice in 28 simulations of 200 ns each. We further investigated the variations in the complex structure of the active site with the ligands, using microsecond order simulations. Results revealed a wide variation in the shapes of the binding sites and binding poses of the ligands. Additionally, the C-terminal region of the other chain often interacted with the ligands and the active site. Collectively, these findings indicate the importance of dynamic sampling of protein-ligand complexes and suggest the possibilities of further drug optimisations.

Introduction

The pandemic of the new corona virus disease, COVID-19, is an urgent global issue. Currently, many research groups are trying to find effective medicines by repurposing approved drugs, using clinical, experimental, and computational approaches[1,2]. However, till date, no therapeutic agent has been approved to be effective against COVID-19 (except remdesivir in Japan). Here, we report the drug binding process of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) main protease (M^{pro}, 3CL hydrolase), using large-scale molecular dynamics (MD) simulations. The M^{pro} protein is essential for processing the precursor polyprotein for replication of the virus. Owing to its crucial role, M^{pro} is one of the major targets for development of anti-SARS-CoV-2 drugs. The first X-ray crystal structure of M^{pro} was released on February 5, 2020[3]. Since then, the number of experimental structures has increased to over 100. For drug repurposing for SARS-CoV-2 M^{pro}, protease inhibitors of human immunodeficiency virus (HIV) are expected to be effective since HIV protease shows similar function as SARS-CoV-2. Many HIV protease inhibitors have been developed and clinical trials of the repurposed HIV protease inhibitors for COVID-19 are currently ongoing (e.g. ChiCTR2000029603). Among these inhibitors, China's National Health Commission has recommended the use of HIV-1 protease inhibitors, lopinavir and ritonavir, as an ad hoc treatment for pneumonia caused by SARS-CoV-2. However, the results from an urgent randomised clinical trial, evaluating the efficacy of lopinavir–ritonavir in patients with COVID-19 in Wuhan, China, showed that no benefit was observed with lopinavir–ritonavir treatment beyond standard care for hospitalised adult patients[4]. Another HIV protease inhibitor, nelfinavir, is also one of the drug candidates against COVID-19. Nelfinavir showed suppression of growth of SARS-CoV in a cell-based experiment[5], and it was shown to be effective for SARS-CoV-2[6,7] in vitro. Although the mechanisms that underlie the inhibitory action of nelfinavir on SARS-CoV remain to be identified, the high sequence similarity (about 96%) between the M^{pro} of SARS-CoV-2 and that of SARS-CoV[8] led us to hypothesise promising activity of nelfinavir against SARS-CoV-2 M^{pro}. In addition, other HIV protease inhibitors such as indinavir, darunavir, and saquinavir have been proposed as drug candidates against SARS-CoV-2, using computational studies[9–12]. These HIV protease inhibitors, some of which are already being tested in clinical trials, are repurposed drug candidates. However, their efficacies are yet to be confirmed.

In this study, we aimed to investigate the dynamics of binding process of various HIV protease inhibitors to SARS-CoV-2 M^{pro}. We performed all-atom MD simulations of

the systems with the dimeric M^{pro} and seven HIV protease inhibitors (darunavir, indinavir, lopinavir, nelfinavir, ritonavir, saquinavir, and tipranavir), solvated in saline solution. Large-scale simulations starting from ligand unbound states (28 simulations of 200 ns for each ligand) has been done using the massively parallel supercomputer. The results allowed us a systematic investigation of the ligand access on the protein surface of M^{pro}, and the frequently accessed sites on M^{pro} were classified, based on the contact between the ligand and the protein. These potential drug binding sites could be useful for further drug development/repositioning. Furthermore, we performed microsecond-scale simulations for the 20 protein-ligand complexes using the special-purpose computer MDGRAPE-4A[13], which is designed for long-term MD simulations, besides the conventional supercomputers. Results revealed that the active site has a high flexibility and allows various binding poses of these ligands.

Results

Identification of ligand binding sites

First, the identification of the potential sites for drug binding was performed on the X-ray crystal structure[3] of dimeric M^{pro}, using the site finder module of Molecular Operating Environment (MOE)[14]. We found three representative drug binding sites on M^{pro} and named them as "sites 1-3" (Fig. 1). Site 1 was the orthosteric active site that had the catalytic residues, His41 and Cys145. Site 2, which was the largest site, was located near the interface of two domain III, and Site 3 was located at domain I.

Next, we investigated the access of drugs to these binding sites, in addition to other sites, on the fluctuating surface of M^{pro} by observing the dynamical trajectories obtained by direct MD simulations of dimeric M^{pro} with seven HIV inhibitors, darunavir, indinavir, lopinavir, nelfinavir, ritonavir, saquinavir, and tipranavir (Supplementary Fig. A1). When a drug accesses the M^{pro} surface, it is likely to form an encounter complex that is not tightly bound to the protein. We hypothesised that the stability and frequency of formation of the encounter complex would reflect the likelihood of the binding process between the drug and the protein. Hence, we thoroughly investigated the formation of the encounter complex by 28 simulations of 200 ns for each ligand (Supplementary Fig. B1).

To analyse the formation ratio of the encounter complex, we calculated the contact map of each ligand to the protein, as shown in Fig. 2. These were calculated by using last 100 ns (500 time points at every 0.2 ns) of each 200 ns simulation, and the threshold of a contact pair was set to 0.35 nm. The contact events on both chains of the dimer were accumulated, and a contact frequency was calculated as the number of events
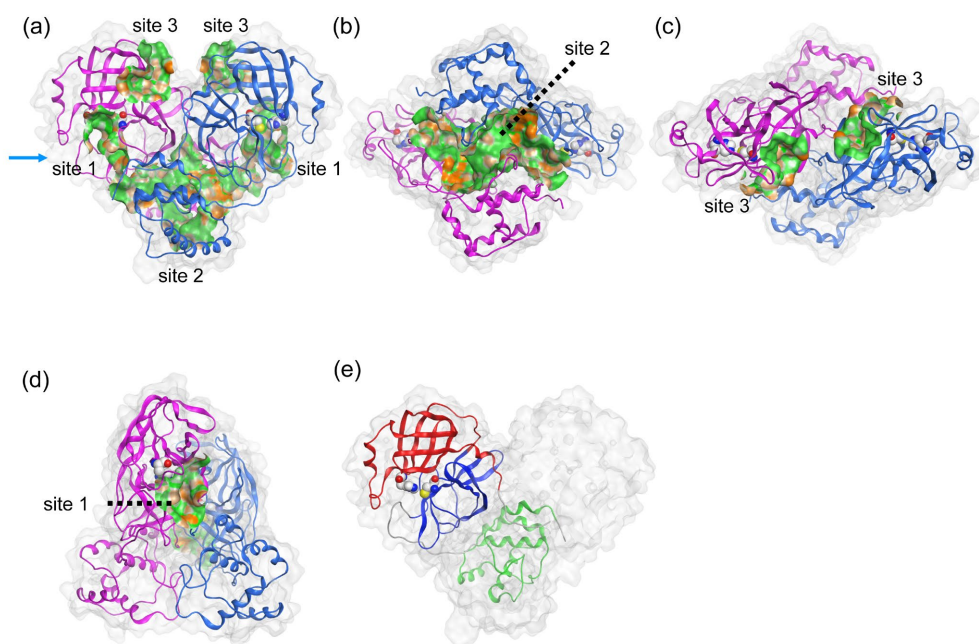
Figure 1. Structural information of dimeric M^pro. The respective protomers of dimeric M^pro are shown in magenta and blue cartoon representations. The catalytic dyad, His41 and Cys145, are represented by the space-filling model. The three possible ligand binding sites were predicted. (a) side view, (b) bottom view, (c) top view, (d) the view from the arrow of (a), and (e) the three domains of protomer (monomer) structure. Each protomer was composed of three domains: domains I, II, and III are residues 8-101 (red), 102-184 (blue), and 201-303 (green), respectively[3]. The lipophilic and hydrophilic regions in each site are depicted with green and orange.

divided by the total number of samples. The results showed that most of the contacts located at the predicted binding sites shown in Fig. 1. The frequent contacts with the active site (site 1) were observed for indinavir, nelfinavir, ritonavir, and tipranavir (Figs. 2a-b). Adjacent to the active site, a frequently visited site existed at the border of the chains, indicated as "site 4" in Fig. 2. It was a new site that was not considered as a major binding site in Fig. 1. Frequent visits to site 4 were observed for all ligands, except lopinavir. The contact with site 2 (Fig. 2c) were frequent for lopinavir, ritonavir and saquinavir, modest for darunavir, indinavir and nelfinavir, and weak for tipranavir. In domain III (Fig. 1e), another shallow site "site 5" was observed for lopinavir, nelfinavir, and tipranavir (Fig. 2b-c). The contact frequency to site 3 generally was low, except for lopinavir (Fig. 2d). We should be careful about the interpretation of these absolute values of contact strength since we could observe only a few unbinding events and they did not reflect the correctly quantified values in equilibrium.

To classify the frequently accessed sites, we performed hierarchical clustering analysis of the contact maps for whole trajectories of 200 ns (1,000 time points at every 0.2 ns) (Supplementary information C and D). Based on the obtained classification of the
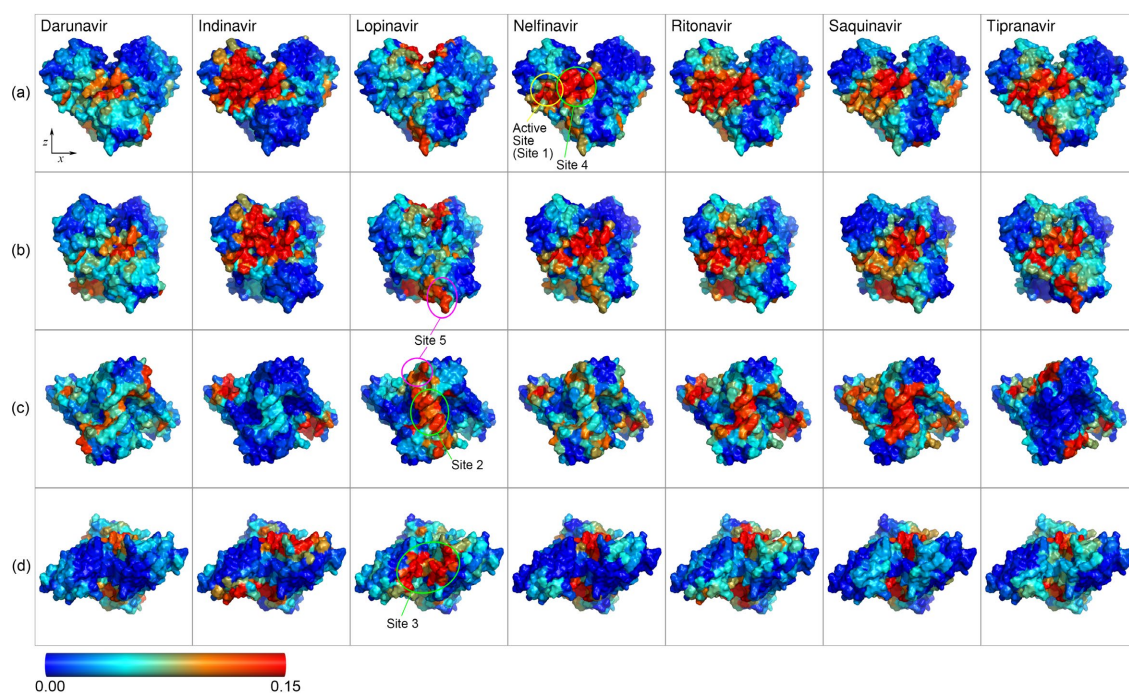
Figure 2. Heat maps of the contact frequencies for the seven ligands. Row (a): front view, (b): side view (rotate 30° around z axis), (c): bottom view (from -z axis), (d): top view (from +z axis). Frequent contacts at the active site were observed in indinavir, nelfinavir, ritonavir, and tipranavir. The major contacts at site 2 were lopinavir, ritonavir, and saquinavir, while site 3 was frequently visited only by lopinavir.

sites, the number of transition events in the trajectories and the binding free energies (estimated by the MM-GB/SA method[15,16]) are summarised in Table 1. The time course of transitions among the classified sites and the binding free energies for each trajectory of the seven ligands are shown in Supplementary Figs. E1-7. As also recognised in Fig. 2, site 3 was only frequently visited by lopinavir. Site 5 experienced a comparable number of on and off events, which was attributed to its relative shallowness compared to the other sites. The active site (site 1), site 2, and 4 were rather stable for most of the ligands. Among the seven ligands, indinavir, nelfinavir, ritonavir, and tipranavir had similar profiles, with rather high counts of contact events to the active site and site 4. Ritonavir also bound to site 2 frequently, and a few unbinding events from the active site were observed. Darunavir, lopinavir, ritonavir, and saquinavir had also visited the active site, but the number of events and the free energies were competitive with the other sites: site 2, 3 (lopinavir only), and 4. Since it is difficult to compare the binding energies of the different ligands by MM-GB/SA, the selection of the best candidate was not possible. However, by comparing ΔG of the active site with respect to the other sites for each ligand in Table 1, indinavir, nelfinavir, and tipranavir could be considered as possible candidates for further drug optimisation.

Table 1. The number of events and the binding free energies estimated by the MM-GB/SA method in 28 trajectories of 200 ns for each ligand.

		Active Site	Site 2	Site 3	Site 4	Site 5	Others
Darunavir	On/Off	3/1	6/1	1/0	4/1	7/6	8[10]/20
	ΔG Ave	-27.3 (6.4)	-20.2 (6.8)	-17.0 (3.6)	-22.8 (7.5)	-16.7 (5.4)	-15.6 (7.2)
	Min	-34.4	-29.5	-18.6	-36.7	-24.7	-31.7
Indinavir	On/Off	9/3	2/0	1/1	10/5	2/2	8[8]/21
	ΔG Ave	-29.5 (7.8)	-24.8 (8.2)	-17.3 (6.1)	-27.4 (6.9)	-17.7 (4.1)	-22.8 (9.8)
	Min	-49.6	-39.7	-25.7	-45	-25.9	-55.8
Lopinavir	On/Off	4/1	4/0	9/4	2/1	7/3	8[9]/25
	ΔG Ave	-22.3 (4.9)	-24.4 (7.6)	-23.4 (7.6)	-18.7 (5.1)	-19.0 (5.5)	-21.5 (11.2)
	Min	-34.8	-38	-34.2	-27.8	-29.5	-46.8
Nelfinavir	On/Off	5/0	4/1	0/0	8/1	4/1	3[9]/21
	ΔG Ave	-28.4 (8.7)	-22.0 (5.4)	-11.8* (3.8)	-24.5 (6.5)	-27.1 (6.5)	-21.3 (9.0)
	Min	-47.6	-31	-12.8	-41.2	-36.9	-42.6
Ritonavir	On/Off	8/3	5/0	2/1	7/1	7/6	10[7]/28
	ΔG Ave	-35.7 (8.6)	-32.6 (8.4)	-21.0 (5.1)	-36.6 (12.6)	-27.6 (9.0)	-26.9 (10.2)
	Min	-54	-54.6	-32.3	-64.5	-40.3	-52.8
Saquinavir	On/Off	5/1	7/1	2/2	8/2	4/2	3[8]/21
	ΔG Ave	-30.2 (6.7)	-25.8 (11.0)	-11.8* (5.2)	-28.7 (9.2)	-17.9 (5.5)	-17.7 (8.6)
	Min	-40.5	-45.3	-20.9	-46.8	-29.3	-37.9
Tipranavir	On/Off	9/2	0/0	3/2	6/1	7/3	7[9]/24
	ΔG Ave	-27.1 (8.0)	2.0 (*N*=1)	-21.8 (2.9)	-21.2 (5.8)	-18.6 (7.8)	-17.6 (8.3)
	Min	-44.8	-6.5	-32.4	-33.9	-36.6	-39.2

The classification was based on the hierarchical clustering analysis. The events were counted after smoothing, by taking majority for 71 data points, sampled every 0.2 ns. The number in square brackets in the others column indicates the number of trajectories that stayed on it during whole simulation period. The yellow sites had the largest fraction at final states (the largest number of "on" events minus "off" events), except for the others category, the blue sites had the smallest ratio of off/on events, and the green sites satisfied both the conditions. The rows "ΔG" indicates the average (with the standard deviation in parenthesis) and minimum binding free energy in a unit of kcal/mol. The lowest energies among sites 1-5 are marked as yellow. The average energies were calculated from the last 100 ns, except for values marked by asterisk, which were taken from the full trajectories. The minimum energies were taken from the averaged energies over 15 continuous points, every 0.2 ns of the whole 200 ns trajectory.
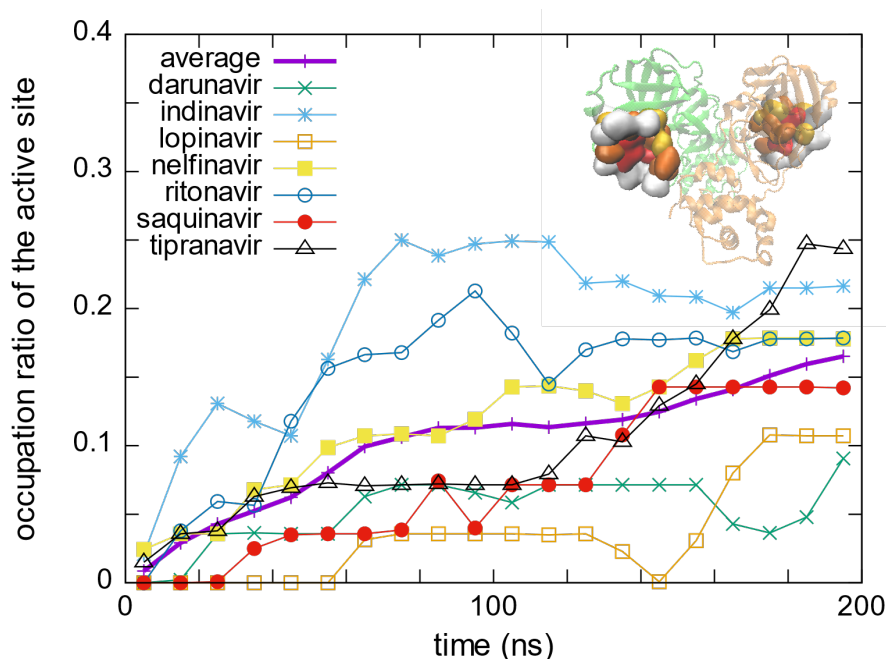
Figure 3. Time course of occupation ratio at the active sites. The figure shows mean occupation ratio averaged in a time span of every 10 ns over 28 trajectories for each ligand. The average over seven ligands is also shown with thicker line. Inset shows the active site residues (see Supplementary Fig. D1). Occupation ratios for the other sites are shown in Supplementary Fig. F1.

Figure 3 shows the time course of the occupation ratio in the active sites (see Supplementary Fig. F1 for the other sites). On an average, among the seven ligands, the ratio increased in time, which suggested that the active sites in fact could be potential drug binding sites for the ligands. It required a minimum simulation duration of 100 ns to observe these tendencies, e.g. tipranavir reached the active site after 100 ns. Although the current number of samples for each ligand does not allow a detailed comparison of the binding affinities to the active sites among the ligands, we expected that the tendency of binding (Fig. 3) reflects the likelihood to form encounter complexes at the active sites, i.e. tipranavir and indinavir were more likely to bind to the active site of M^{pro}, compared to lopinavir and darunavir.

While MD-based free energy calculation is necessary to provide precise selections of drug candidates, it requires proper complex structures. We extended the duration of simulations up to a microsecond (μs) for 20 trajectories, arbitrarily selected from the trajectories in which the ligands attached to the active site at 200 ns. For these microsecond trajectories, the binding free energies were estimated by MM-GB/SA (Supplementary information G). Results showed that the minimum and average binding free energies decreased for all ligands (except for the slight increases in the minimum energies of nelfinavir and the average one of saquinavir). This observation indicated that
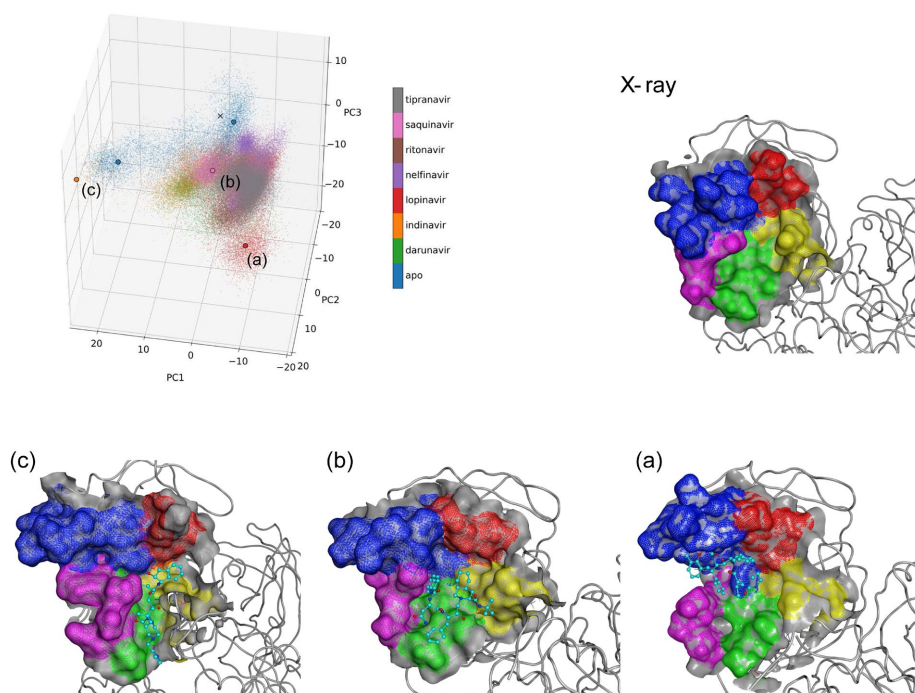
Figure 4. Conformational diversity of the M^{pro} active site. The projection of the first three principal components (PC1-3) is shown. The contribution ratios of first three principal components (PC1, PC2, and PC3) were 36.5%, 18.2%, and 8.5%, respectively. The conformations of the active site, corresponding to the cross mark and (a)-(c) in the projection, are shown: the cross mark and (a)-(c) are the crystal structure (PDB ID: 6lu7) and simulation structures of lopinavir-bound M^{pro}, saquinavir-bound M^{pro}, and indinavir-bound M^{pro} systems, respectively. The five regions of the active site, (1) residues 24-27, (2) 41-54, (3) 140-145, (4) 163-168 and 172, and (5) 187-192, are coloured in red, blue, yellow, green, and magenta, respectively. The representative conformations for the active site of apo-M^{pro} system are shown in Supplementary Fig. H3.

the current simulation time of 200 ns was insufficient for the equilibrium analysis. In the next section, we analysed the active site structures and ligand binding poses to understand the key factors involved in binding and their dynamical properties.

Conformational variations upon ligand binding

To explore the conformational variations in the M^{pro} active site, a principal component analysis of Cα atoms of 37 amino acid residues (residue 24-27, 41-54, 140-145, 163-168, 172, and 187-192) contained in the active site (see Supplementary information H) was performed for 20 ligand-bound MD trajectories of 1 μs together with another 1.8 μs MD trajectory of M^{pro} without ligand (apo-M^{pro}). The projection of the first three principal components (PC1-3) characterised the conformational diversity of the active site (Fig. 4). Interestingly, the projection on the first two PCs, PC1 and PC2, mostly discriminated the MD trajectories of ligand-bound systems from that of the apo-M^{pro} system. Especially,

PC2 largely contributed to the discrimination (see Supplementary Fig. H2). Also, since these two eigenvectors were closely related to the motion of the loop regions (residue 41-54 and 187-192; Supplementary Fig. H2), it was suggested that the interactions of the ligands with the active site residues constrained the motion of these loop regions, especially along PC1. Furthermore, the conformations of the active site, obtained from the MD simulations, were clearly different from that of the crystal structure, and various conformations emerged even in the ligand-bound systems. Although it was difficult to classify the active site conformations upon binding of the different ligands systematically, its characteristic conformations could be visualised (examples shown in Fig. 4a and c, lopinavir-bound M^{pro} and indinavir-bound M^{pro} systems). These results indicated that the active site of M^{pro} was highly flexible and allowed conformational changes to accommodate various ligands.

Next, we investigated how each ligand interacted with the active site residues in the MD simulations. To detect the important protein-ligand interactions, we analysed the interaction fingerprint which could enable the identification of the existence of the ionic and the hydrogen bonds between the protein and the ligand for each snapshot in MD trajectories. Observing the representative appearance rate of the interaction fingerprints (RAIF) in Supplementary Table J1, the seven key active site residues were found to have comparatively large contribution (residues 44, 143, 166, 187-190 with RAIF >20%) and were speculated to play an important role in the ligand binding. In addition, by performing clustering analysis of the fingerprint, we picked three representative binding poses for each ligand from top three classified clusters. Figure 5 shows the typical binding poses with the seven key residues, as highlighted (Supplementary Figs. J1-7 show full set of binding poses). Based on the observation of the binding poses, we found that the various binding poses were contained in MD simulations of each ligand-bound M^{pro} system, and the variety of binding poses resulted from not only the initial conformation of the encounter complex but also from the conformational refinement and/or equilibrium dynamics within each 1 μs MD simulation.

The two key residues, Glu166 and Gln189, were especially noted because they commonly formed the effective interactions (RAIF > 20%) with multiple ligands. Glu166 formed effective interactions with five ligands (darunavir, indinavir, lopinavir, nelfinavir, and tipranavir) (Supplementary Table J1). The interaction was influenced mainly by conformational change of Glu166 side chain. Gln189 also formed effective interactions with three ligands (darunavir, ritonavir, and saquinavir), and also comparable interactions with indinavir (RAIF = 19.1%) and nelfinavir (RAIF = 18.8%). Thus, Gln189 could be utilised as a key residue for a broad range of ligands. Gln189, which is known to be quite
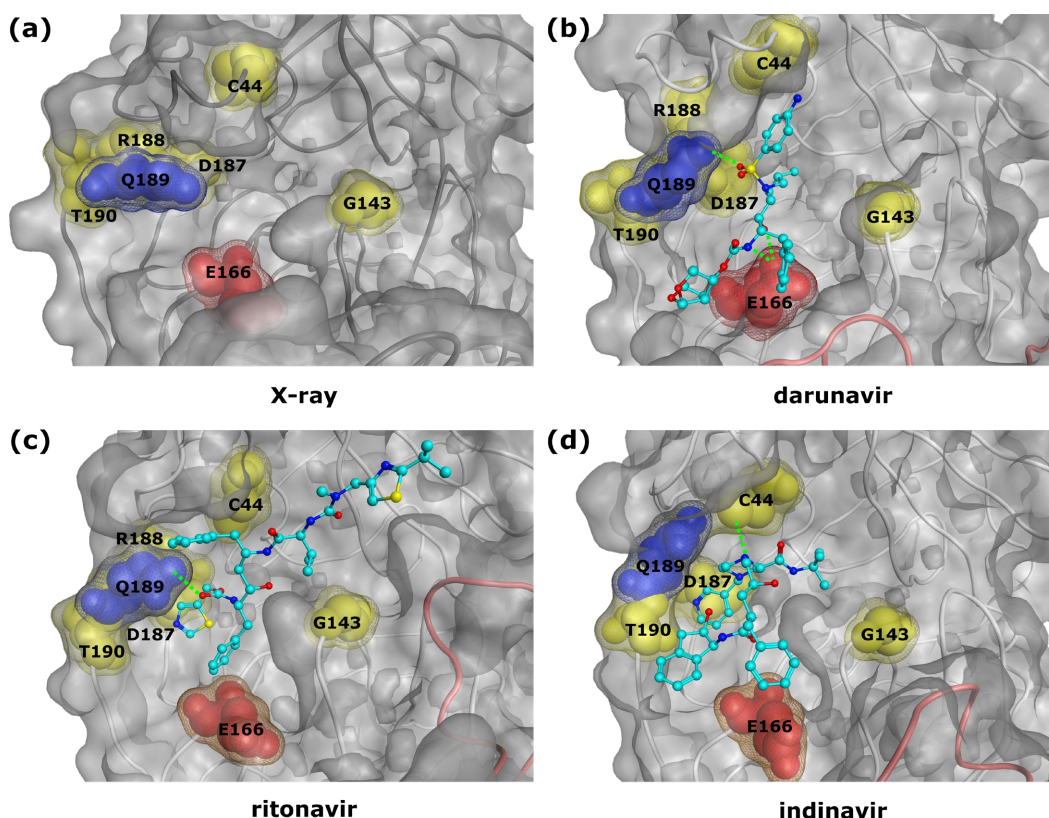
9

Figure 5. Key active site residues for ligand binding. (a) The seven key amino acids were determined with the representative appearance rate of the interaction fingerprints (RAIF) and are highlighted with red (Glu166), blue (Gln189), and yellow (Cys44, Gly143, Asp187, Arg188, and Thr190). First two residues were commonly utilised for multiple ligands, while others for single ligand. (see main text and Supplementary Table J1). This figure is depicted on the crystal structure. (b) The representative binding pose of darunavir-bound M^{pro} system is shown. The ligand interacted with the Glu166 and Gln189. (c) The representative binding pose of ritonavir bound M^{pro} system is shown. The ligand interacted with Gln189. (d) The representative binding pose of indinavir-bound M^{pro} system is shown. The ligand interacted with Cys44. In (b)-(d), the hydrogen bonds are shown with green dotted lines.

flexible in the M^{pro} of SARS-CoV[17], belongs to the loop region (residue 187-192) that is closely related to the PC1-3 eigenvectors (Supplementary Fig. H2). Gln189 could keep effective interactions with a variety of ligands by utilising the flexibilities in its side chain and the backbone of the loop. Each of the other five key residues formed an effective interaction with only one of the seven ligands. The three residues: Asp187, Arg188, and Thr190, in the same loop of Gln189 might have less chance to form an effective interaction because of their location within the loop and the preferred orientation of their side chains. The remaining two residues: Cys44 and Gly143, formed the effective interactions with darunavir (Supplementary Table J1), which suggested potential utilisation of these key residues for a specific class of drugs. Besides the seven key active site residues noted above, Met49 in the loop region (residue 41-54) formed the moderate

interactions (RAIF of 5.0-16.5%) commonly with the seven ligands. As Met49 is known to play an important role to accommodate a substrate peptide for the M^{pro} of SARS-CoV[18], it might largely contribute to the molecular recognition for drug development against M^{pro}. While there were commonly utilised interactions, we also found some of the interactions formed uniquely to a ligand in Supplementary Fig. J8, which would imply broad possibility of conformational variation in the active site specifically induced by the ligands. These analyses indicated that there were diverse interaction patterns of the respective ligands in the flexible active site of M^{pro}.

Discussion

We investigated the access to the drug binding sites in SARS-CoV-2 M^{pro}, using seven HIV inhibitors as potential lead drugs. The frequently accessed sites on the M^{pro} were classified based on the contact between the ligand and the protein. Although the limited length of the simulations may give statistics only for encounter complexes, it can provide a list of potential drug binding sites which can be employed for further drug development/repositioning. The microsecond-scale simulations of the active site complexes of M^{pro} and ligands revealed a wide variation in the shapes of the active site and also in the binding poses of the ligands. This suggested that the surface of the M^{pro} is rather flexible, and conformational change due to induced fit between the ligand and the protein was a dominant factor affecting the binding processes. Thus, MD simulation could be an effective tool to investigate the ligand binding on the current M^{pro} system.

The conventional molecular-docking approach, that is, rigid receptor-flexible ligand docking, uses a single protein structure (for example, the X-ray crystallographic structure). However, it was difficult to adopt it for a target like the M^{pro} active site with high flexibility. Additionally, the protein structural information obtained from our MD simulations could be utilised for structure-based drug design strategies, including the ensemble docking. For a detailed comparison of the binding affinity among ligands, an estimation of the precise protein-ligand complex structure is crucial and can be achieved by MD simulations with longer duration. Moreover, the results suggested that the non-specific binding to sites other than the active site should be taken into account while designing drugs.

To explore longer temporal behaviour of ligand binding at the active site of the M^{pro}, three pilot simulations for indinavir, nelfinavir, and tipranavir were performed by extending three trajectories to reach 6 µs (Fig. 6 and Supplementary Fig. K1). It was observed that the ligands stayed at the active site. The ligand binding poses were mostly
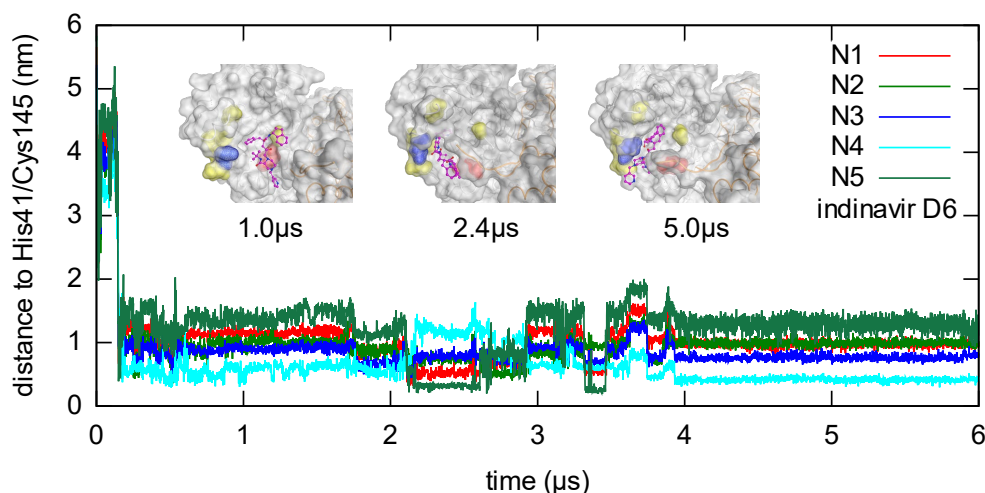
Figure 6. Ligand flipping during 6 μs MD simulations. Minimum distances between a nitrogen atom in the ligand (indinavir) and the catalytic residues His41/Cys145 are plotted. These distances reflect binding pose of the ligand, and thus flipping events are visualised as a sudden switching of these distances. Inset shows ligand poses at 1.0, 2.4, and 5.0 μs, where surfaces of the active site are coloured as in Fig.5.

stable for μs scale; however, it was observed that the ligands exhibited flipping of the binding pose a few times in 6 μs. These observations suggest that these binding states might be rather loose, and such loose binding states may be a typical or essential class of the small-molecule ligand binding state to the M^{pro} active site. There is a possibility to develop a more effective drug that can bind tightly to the M^{pro} active site with high enthalpy difference. However, loose binding drugs may still have a chance to work effectively to inhibit the M^{pro}. We hope systematic investigation of longer temporal behaviour in future researches will give us perspective view of the drug binding dynamics.

Another noteworthy finding was the interaction between the ligands and the C-terminal residues of the other chain of the dimer, observed for all ligands. Figure 7 shows the typical snapshot of such interactions in case of indinavir. In the extended simulations of 1 μs, the C-terminal residue, Gln306, stayed within a minimum distance of 0.35 nm to the ligand for 8% of the time duration (from 200 ns to 1,000 ns, averaged over 20 simulations). In addition, in several cases the region entered the active site (Supplementary information L). Similar observations were noted in the clustering analysis of protein-ligand contacts and interactions. Since M^{pro} is known to autoprocess the N- and C-terminals of the precursor protein of itself[19,20], it is reasonable to observe the interaction of the C-terminal region and the active site. This fact also suggested that the observed C-terminal interaction might stabilise the substrate or drug interactions. The observed dynamic interaction revealed by the MD simulations would be another important factor to be considered in drug design.
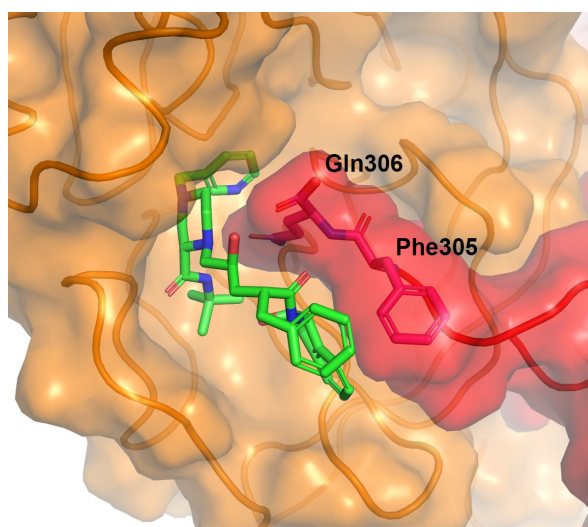
Figure 7. C-terminal region interacting with indinavir at the active site. Similar interactions were observed for the other ligands. The chain in orange corresponds to the chain that has the active site occupied by indinavir, and the one in red (C-terminal) corresponds to another chain of the dimer. This snapshot was taken at 2.4 μs from the trajectory shown in Fig. 6.

The analysis on ligand access to the surface of the M^{pro} provided several drug binding sites, in addition to the orthosteric active site. Among these sites, sites 2 or 3, located at the interface of the dimeric M^{pro}, which were relative frequently accessed by lopinavir, ritonavir, and saquinavir, may be worth exploring. While the ligand binding to these two sites may not directly inhibit the enzymatic reaction, there is a possibility to influence the dimerisation and structural stabilisation of the dimeric M^{pro}. It would be interesting to elucidate the roles of these sites by simulating the binding of the other drug molecules and the substrate peptides. Further, a study on the detailed mechanism to recognise the specific amino acid sequences in the active site is an interesting target of MD simulations in drug designing, including allosteric inhibitors. Our next goal is to propose an atomic level drug design strategy against the M^{pro} by integrating the dynamical information of various binding process.

Methods

To simulate the binding process of seven HIV protease inhibitors (darunavir, indinavir, lopinavir, nelfinavir, ritonavir, saquinavir, and tipranavir) to the SARS-CoV-2 M^{pro}, the initial structures were built based on the X-ray crystal structure of the holo SARS-CoV-2 M^{pro} (Protein Data Bank[21] entry: 6LU7[3]). Since the crystal structure was the M^{pro} with the covalently bound inhibitor, the structure of the inhibitor-unbound M^{pro} (apo-M^{pro}) was prepared by removing the covalently bound inhibitors. In addition, each HIV protease inhibitor was initially placed at least 1.5 nm apart from the active sites of SARS-CoV-2 M^{pro} (see Supplementary Fig. B1). The initial structure of the M^{pro} with an HIV inhibitor was then solvated in a cubic box of TIP3P water

molecules. In addition, chlorine and sodium ions (0.154 M) were added to neutralise the system. The system of apo-M^{pro}, without covalently bound ligand, was also prepared in a similar manner. Amber FF14SB[22] was used for the M^{pro} protein, and general amber force field (GAFF)[23] was used for seven HIV protease inhibitors. The partial charges for the ligands were calculated at the RHF/6-31G(d) level with Gaussian16[24] and the restrained electrostatic potential method[25,26]. ParmEd[27] was used to convert Amber topologies to GROMACS[28] formats. VMD[29] and PyMOL[30] were used for visualisation.

All MD simulations were performed on the massively parallel supercomputer HOKUSAI Big Waterfall (BW) system at RIKEN ICS, or the special-purpose, specialised for faster calculation of MD, computer MDGRAPE-4A[13] (the advanced version of MDGRAPE-4[31]) at RIKEN BDR. All hydrogen bonds and TIP3P waters were constrained with the methods summarised in Table 2. The periodic boundary conditions were applied to the system, and the long-range Coulomb interactions were treated with the method described in Table 2, with a direct space cutoff distance of 1.3 nm. After the energy minimisation of each solvated system, the system was heated to 300 K for 1ps with the integration time step of 0.5 fs, without constraints. Then, 100 ps MD simulations, with a time step of 2 fs, under NPT ensemble ($P = 1$ bar and $T = 310$ K), were performed to adjust the size of the simulation box. For each HIV inhibitor, the above relaxation protocol was applied for 4-fold variation in the initial location of the inhibitor (Supplementary information B), and 7-fold variation produced by randomising velocities with different random seeds. As a result, totally 28 initial conditions were prepared for the 200 ns production runs on HOKUSAI-BW (with a time step of 2.5 fs) under NVT ensemble (T=310K). The trajectories of each system were saved for every 100 ps (2000 conformations in each MD trajectory), and most of the analysis were done on snapshots every 200 ps. In addition, 20 trajectories were picked for extended simulation of 1 μs duration (11 on MDGRAPE-4A and 9 on HOKUSAI-BW), and 3 trajectories in these trajectories were extended further to 6 μs duration on MDGRAPE-4A. Pilot runs in the early stage of this study were also performed on MDGRAPE-4A.

Table 2. Summary of numerical methods.

Machine	Intel PC cluster (HOKUSAI Big Waterfall system)	Special-purpose computer MDGRAPE-4A
Software	GROMACS[28] 2020.1/2018.8	Inhouse code, partly derived from GROMACS[28]
Long range Coulomb	SPME[33]	TME (see Supplementary information Z)
Constraint	LINCS[34](h-bonds) SETTLE[35](TIP3P)	RATTLE[36](h-bonds) SETTLE[35](TIP3P)
Thermostat	the canonical sampling velocity rescaling (CSVR)[37]	Same as left

Data availability

Raw trajectory data analysed in this paper and movie examples are available at the zenodo repository[32].

References

[1]   Share Chemistry - American Chemical Society. https://www.acs.org/content/acs/en/covid-19/share-chemistry.html.

[2]   Amaro, R. E. & Mulholland, A. J. A Community Letter Regarding Sharing Bimolecular Simulation Data for COVID-19. *J. Chem. Inf. Model.* (2020) doi:10.1021/acs.jcim.0c00319.

[3]   Jin, Z. *et al.* Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* (2020) doi:10.1038/s41586-020-2223-y.

[4]   Cao, B. *et al.* A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2001282.

[5]   Yamamoto, N. *et al.* HIV protease inhibitor nelfinavir inhibits replication of SARS-associated coronavirus. *Biochem. Biophys. Res. Commun.* (2004) doi:10.1016/j.bbrc.2004.04.083.

[6]   Ohashi, H. *et al.* Multidrug treatment with nelfinavir and cepharanthine against COVID-19. *bioRxiv* (2020) doi:10.1101/2020.04.14.039925.

[7]   Yamamoto, N., Matsuyama, S., Hoshino, T. & Yamamoto, N. Nelfinavir inhibits replication of severe acute respiratory syndrome coronavirus 2 in vitro. *bioRxiv* (2020) doi:10.1101/2020.04.06.026476.

[8]   Chen, Y. W., Yiu, C. P. B. & Wong, K. Y. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CLpro) structure: Virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Research* (2020) doi:10.12688/f1000research.22457.1.

[9]   Li, Y. *et al.* Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening. *bioRxiv* (2020) doi:10.1101/2020.01.28.922922.

[10]   Xu, Z. *et al.* Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. *bioRxiv* (2020) doi:10.1101/2020.01.27.921627.

[11]   Contini, A. Virtual screening of an FDA approved drugs database on two COVID-19 coronavirus proteins. (2020) doi:10.26434/chemrxiv.11847381.

[12]   Farag, A., Wang, P., Ahmed, M. & Sadek, H. Identification of FDA Approved Drugs Targeting COVID-19 Virus by Structure- Based Drug Repositioning. *chemrxiv* (2020) doi:10.26434/chemrxiv.12003930.v3.

[13]   Komatsu, T. S. *et al.* Japanese published unexamined patent application No. 2020-64560. (JP 2020 - 64560A).

[14]   ULC, C. C. G. Molecular Operating Environment (MOE), 2013.08. *1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7* (2018) doi:10.1038/ja.2010.129.

[15]   Kollman, P. A. *et al.* Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* (2000) doi:10.1021/ar000033j.

[16]   Gohlke, H. & Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* (2004) doi:10.1002/jcc.10379.

[17]   Zhang, L. *et al.* α-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.* (2020) doi:10.1021/acs.jmedchem.9b01828.

[18] Muramatsu, T. *et al.* SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. U. S. A.* (2016) doi:10.1073/pnas.1601327113.

[19] Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. Coronavirus main proteinase (3CLpro) Structure: Basis for design of anti-SARS drugs. *Science (80-.).* (2003) doi:10.1126/science.1085658.

[20] Muramatsu, T. *et al.* Autoprocessing mechanism of severe acute respiratory syndrome coronavirus 3C-like protease (SARS-CoV 3CLpro) from its polyproteins. *FEBS J.* (2013) doi:10.1111/febs.12222.

[21] Berman, H. M. *et al.* The Protein Data Bank (www.rcsb.org). *Nucleic Acids Res.* (2000) doi:10.1093/nar/28.1.235.

[22] Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* (2015) doi:10.1021/acs.jctc.5b00255.

[23] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* (2004) doi:10.1002/jcc.20035.

[24] Frisch, M. J. *et al. Gaussian16 Revision C.01*. *Gaussian16 Revision C.01* (2016).

[25] Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* (1993) doi:10.1021/j100142a004.

[26] Cornell, W. D. *et al.* A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* (1995).

[27] ParmEd/ParmEd: Parameter/topology editor and molecular simulator. https://github.com/ParmEd/ParmEd.

[28] Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* (2015) doi:10.1016/j.softx.2015.06.001.

[29] VMD - Visual Molecular Dynamics. http://www.ks.uiuc.edu/Research/vmd/.

[30] schrodinger/pymol-open-source: Open-source foundation of the user-sponsored PyMOL molecular visualization system. https://github.com/schrodinger/pymol-open-source.

[31] Ohmura, I., Morimoto, G., Ohno, Y., Hasegawa, A. & Taiji, M. MDGRAPE-4: A special-purpose computer system formolecular dynamics simulations. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* (2014) doi:10.1098/rsta.2013.0387.

[32] Komatsu, T. S. *et al.* Molecular dynamics trajectories for SARS-CoV-2 Mpro with 7 HIV inhibitors. https://doi.org/10.5281/zenodo.3766084 (2020) doi:10.5281/zenodo.3766084.

[33] Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* (1995) doi:10.1063/1.470117.

[34] Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* (1997) doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[35] Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* (1992) doi:10.1002/jcc.540130805.

[36]   Andersen, H. C. Rattle: A 'velocity' version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* (1983) doi:10.1016/0021-9991(83)90014-1.

[37]   Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* (2007) doi:10.1063/1.2408420.

Acknowledgements

Author Contributions

TSK, NO, MT planed this research project. TSK, GM, YO contributed to numerical simulations. TSK, NO, YH contributed to the system preparation. TSK, NO, YMK, YH, MT contributed to analysis of the result. Whole authors contributed to discussion and manuscript preparation.

Competing Interest

The authors declare no competing interests.