



# Asymmetric dynamics of dimeric SARS-CoV-2 and SARS-CoV main proteases in an apo form: Molecular dynamics study on fluctuations of active site, catalytic dyad, and hydration water

Shinji Iida<sup>a,\*</sup>, Yoshifumi Fukunishi<sup>b</sup>

<sup>a</sup> Technology Research Association for Next-Generation Natural Products Chemistry, 2-3-26, Aomi, Koto-ku, Tokyo, Japan 135-0064

<sup>b</sup> Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 2-3-26 Aomi, Koto-ku, Tokyo, Japan 135-0064

## ARTICLE INFO

### Keywords:

SARS-CoV-2  
Cysteine protease  
Atomistic molecular dynamics simulation  
Catalysis  
Water molecule  
Dimerization

## ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been widely spread around the world. It is necessary to examine the viral proteins that play a notorious role in the invasion of our body. The main protease (3CL<sup>PRO</sup>) facilitates the maturation of the coronavirus. It is thought that the dimerization of 3CL<sup>PRO</sup> leads to its catalytic activity; the detailed mechanism has, however, not been suggested. Furthermore, the structural differences between the predecessor SARS-CoV 3CL<sup>PRO</sup> and SARS-CoV-2 3CL<sup>PRO</sup> have not been fully understood. Here, we show the structural and dynamical differences between the two main proteases, and demonstrate the relationship between the dimerization and the activity via atomistic molecular dynamics simulations. Simulating monomeric and dimeric 3CL<sup>PRO</sup> systems for each protease, we show that (i) global dynamics between the two different proteases are not conserved, (ii) the dimerization stabilizes the catalytic dyad and hydration water molecules behind the dyad, and (iii) the substrate-binding site (active site) and hydration water molecules in each protomer fluctuate asymmetrically. We then speculate the roles of hydration water molecules in their catalytic activity.

## 1. Introduction

Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) was originally identified in 2003 [1]. In 2019, the novel SARS coronavirus (SARS-CoV-2) emerged and has been spreading around the world. Both viruses cause severe pneumonia. The viruses encapsulate a single-stranded RNA in its viral envelope, and protrusions (i.e., spike proteins) project from the surface so that the virus can recognize certain receptors on cells.

The main protease (3-chymotrypsin like protease, 3CL<sup>PRO</sup>) plays an important role in the maturation of the SARS-CoV and SARS-CoV-2 coronaviruses. The protease is a cysteine protease and has a catalytic dyad formed by histidine (H41) and cysteine (C145), which are highly conserved among different coronaviruses (Fig. S1). The protease catalyzes the polyprotein translated from the viral RNA, breaking the polyprotein into functional pieces, which include the protease itself. In the present study, we called the protease “3CL<sup>PRO</sup>” and used the term “active site” as the one consisting of the physiological-ligand binding pocket and

the hydrolysis-reaction catalytic site.

Because the SARS-CoV and SARS-CoV-2 3CL<sup>PRO</sup> are a protease, designing binders to the active site is a direct way to deactivate the catalytic activity. On the basis of this manner, many investigations have been attempted to identify novel binders to the active site [2]–[7]. For instance, a Michael acceptor inhibitor (N3) was originally designed to inhibit SARS-CoV 3CL<sup>PRO</sup> [2]. N3 also binds to the active site for SARS-CoV-2 3CL<sup>PRO</sup> (PDBID: 6LU7 and 7BQY) [3].

Another way to impair the catalytic activity can be to inhibit the dimerization of the protease [8]. For dimeric SARS-CoV 3CL<sup>PRO</sup>, an X-ray experiment suggested that the substrate pocket S1 in the active site of a protomer was collapsed and the other pocket wasn't, and hence one protomer is active, but the other is inactive [9]. The analysis of MD trajectories for dyad conformation suggested that one protomer in the dimeric SARS-CoV 3CL<sup>PRO</sup> was inactive and a monomeric SARS-CoV 3CL<sup>PRO</sup> was inactive [10]. However, for dimeric SARS-CoV-2 3CL<sup>PRO</sup>, room-temperature X-ray crystallography demonstrated that the collapse of S1 pocket in the dimer did not occur; the authors pointed out that

\* Corresponding author.

E-mail address: [iida@protein.osaka-u.ac.jp](mailto:iida@protein.osaka-u.ac.jp) (S. Iida).

<https://doi.org/10.1016/j.bbadv.2021.100016>

previous crystal structures might have an artefact [11]. The dimer dissociation constant for both SARS-CoV and SARS-CoV-2 3CL<sup>PRO</sup> was estimated; also the catalytic efficiency of SARS-CoV-2 3CL<sup>PRO</sup> was slightly higher than SARS-CoV one [12]. Subsequently, the dissociation constant was revised by native mass spectrometry, the value of which was about  $0.14 \pm 0.03 \mu\text{M}$  [13]. The dimerization and the enzymatic activity was weakened by both the C- and N-terminal truncations of SARS-CoV 3CL<sup>PRO</sup> [14]. On the other hand, the C-terminal truncation, which ranged from 300 to 306, did not cause noticeable effects on the quaternary structure [15]. The stability of several known compounds on the active site was assessed for finding potent binding sites [16]. Their blind docking result implied that the dimer interface would be a potent target.

Although the relationship between the dimerization and the catalytic activity has been investigated so far [8], few studies demonstrate how they relate and why the dimerization matters for the catalytic activity. Furthermore, it is unclear whether the results of SARS-CoV 3CL<sup>PRO</sup> are reusable even for SARS-CoV-2 3CL<sup>PRO</sup>. It is thus necessary to investigate the relationship and the difference between the two proteases. It should be confirmed whether SARS-CoV-2 3CL<sup>PRO</sup> behaves in the same way as SARS-CoV 3CL<sup>PRO</sup>; and whether it is inactive in a monomeric form and is active in a dimeric form. No study revealed, from a statistical and structural viewpoint, the reason why the monomer is inactive, but the dimer is active. The present study aims to unveil the relationship between the protease's structure/dynamics and its catalytic activity.

In the present study, we demonstrate that the dimerization of the proteases stabilizes their catalytic dyad and hydration water behind the dyad. We apply atomistic molecular dynamics simulations to calculate the global dynamics, dyad stability, and hydration water dynamics. We find that the global dynamics of SARS-CoV-2 3CL<sup>PRO</sup> differ from those of SARS-CoV 3CL<sup>PRO</sup>. For both dimeric proteases, each active site complies with different structural probability distributions. The catalytic dyad is retained in a dimeric state but not in a monomeric state, and hydration water molecules behind the dyad stay longer in a dimeric state than in a monomeric state.

**Table 1**

Summary of simulation.  $N_{\text{atoms}}$  is the number of atoms in a simulation system,  $t_{\text{each}}$  is the total simulation time of each run, and  $N_{\text{runs}}$  is the number of runs. All simulations were performed at 310 K.

Target	State	No	Ref	PDB ID	$N_{\text{atoms}}$	$N_{\text{runs}}$	$t_{\text{each}}$ , $\mu\text{s}$
SASRS-CoV 3CL <sup>PRO</sup>	Dimeric	1	[9]	1UK3 <sup>a</sup>	87,554	5	1.0
		2	[9]	1UJ1 <sup>b</sup>	79,123	5	1.0
	Monomeric	3	[9]	1UK3A <sup>c</sup>	68,521	5	1.0
		4	[9]	1UK3B <sup>d</sup>	70,073	5	1.0
SASRS-CoV-2 3CL <sup>PRO</sup>	Dimeric	5	N/A	HM <sup>e</sup>	78,938	5	1.0
		6	[18]	6LU7 <sup>f</sup>	79,466	10	1.0
	Monomeric	7	[18]	6LU7 <sup>f</sup>	62,290	5	1.0
		8	[19]	6WQF	77,617	5	1.0
		9	[19]	6WQF	62,200	5	1.0

<sup>a</sup> This structure was determined at pH 7.6. <sup>g</sup> MD simulations were performed for this in our previous study [20].

<sup>b</sup> This structure was determined at pH 6.0. <sup>9</sup>

<sup>c</sup> Protomer A of 1UK3

<sup>d</sup> Protomer B of 1UK3

<sup>e</sup> A homology model (HM) derived from 1UJ1. This model was built by SWISS-MODEL [21],[22]. The query sequence is shown in Table S1 in the supporting information.

<sup>f</sup> The ligand was removed from 6LU7, and thus the systems from 6LU7 were an apo form.

<sup>g</sup> The dyad was ionized (i.e., H41<sup>+</sup>...C145<sup>-</sup>).

## 2. Materials and methods

### 2.1. Preparation of simulation systems

All the target proteases in the present study are shown in Table 1. One simulation system for a dimeric system of SARS-CoV-2 3CL<sup>PRO</sup> was constructed by homology modelling based on the structure of PDBID 1UJ1. The template sequence is shown in Table S1. Since the sequence identity is 96.3 %, the homology modelling approach would provide a reliable structural model. Furthermore, another dimeric system of SARS-CoV-2 3CL<sup>PRO</sup> was constructed from the X-ray structure of PDBID 6LU7. The structural data of the homodimer was downloaded from the "Biological Assembly" of the PDB 6LU7. Note that one protomer's conformation in the dimeric system is identical to the other protomer's because the asymmetric unit of the PDBID 6LU7 is monomeric (i.e., the root mean square displacement between the two protomers is zero). On the basis of the chain IDs, each chain was named "protomer A" and "protomer B".

The protonation state of the histidine H41 would be crucial for the catalytic activity of 3CL<sup>PRO</sup>. For the careful assignment, the catalytic site in a known X-ray structure of 3CL<sup>PRO</sup> (PDBID: 1UJ1) was examined. The surroundings of the catalytic dyad (H41-C145) imply that  $N_{\epsilon}$  of H41 is protonated and  $N_{\delta}$  of H41 is not (Fig. S2). Accordingly,  $N_{\epsilon}$ (H41) was protonated and  $N_{\delta}$  was not in No. 1-7 simulations. Nevertheless, because it is also plausible for 3CL<sup>PRO</sup> to form an ionized dyad H41<sup>+</sup>...C145<sup>-</sup> [17], a dimeric system with two ionized dyad and a monomeric one with an ionized dyad were also constructed (No. 8-9 in Table 1). The computational results of the ionized systems are given in the Supporting Information.

For each structure prepared, the C- and N-terminus were capped by acetyl and amine groups, respectively. Then, each of the structures was immersed into a tetrahedron water box whose margin was 10 Å from each of the planes of the box to the structure. The salt concentration in the box was set to 0.1 M. Energy minimization was executed with all heavy atoms restrained and then without any restraints. Subsequently, five NVT simulations with different velocities were performed during 100 ps, and then five NPT simulations at 1 bar during 100 ps. From the five equilibrated structures, five independent NPT simulations at 1 bar at 310 K were performed. Only for the dimeric system of PDBID 6LU7, the number of simulations was 10. Production runs were performed for  $1 \mu\text{s} \times 10$  for the dimeric system and  $1 \mu\text{s} \times 5$  for the others (Table 1).

Computational setting are as follows: time step was set to 2 fs, an interaction table for cut-off scheme [23] was used whose list was updated every 20 fs. A leap-frog algorithm was used for an integrator, and LINCS [24] was for hydrogen-bond constraints (lincs\_iter=1, lincs\_order=4). V-rescale [25] was for temperature control for two groups, protein and non-protein ( $\tau_T = 0.1$ , compressibility=4.5 × 10<sup>6</sup>). Berendsen coupling [26] was used for the first box equilibration, and Parrinello-Rahman coupling for all production runs to control pressure [27],[28] ( $\tau_p^{\text{ref}} = 1.0$ ,  $\tau_p = 2.0$ ). Short-range electrostatic and van der Waals interaction cut-off values were set to 12 Å. Amber ff99SB-ILDN was employed for protein force field [29] and TIP3P model [30] for water molecules. Reaction-field method [31] for long-range interaction calculation was used with the relative dielectric constant  $\epsilon_{\text{rf}}$  set infinity. Note that, in Gromacs notation, the option epsilon-rf=0 means that  $\epsilon_{\text{rf}}$  is infinity. Snapshots were stored for every 100 ps. Gromacs 2019.2 [32–36] was used for all simulations. The supercomputer TSUBAME 3.0 was used for all simulations. For analysis, MDAnalysis [37],[38] was used for parsing MD trajectory. All pictures for molecules were drawn by PyMOL 2.3.0 [39]. Every MD trajectory from 200 ns to 1 μs was analyzed (see Fig.s S3 and S4).

### 2.2. Principal Component Analysis (PCA)

Cartesian-based PCA (cPCA) was used for description of global pro-

tein dynamics. The cPCA provides dynamic modes, i.e., eigenvectors  $e_1^x, e_2^x, \dots, e_n^x$ , where  $x$  is a system name and  $n$  is the number of coordinates of interest. The eigenvectors  $e_1^x, e_2^x, \dots, e_n^x$  were arranged in a descending order of the eigenvalues. Correlations between  $e_i^x$  and  $e_j^y$  of different proteases were calculated (i.e., inner products  $e_i^x \cdot e_j^y$  ( $i, j = 1, 2, \dots, n$ )) [40]. Here, cPCA was applied to trajectories of the SARS-CoV 3CL<sup>pro</sup> and SARS-CoV-2 3CL<sup>pro</sup> systems. MD snapshot  $i$  of a trajectory was expressed as  $q^i = (q_1^i, q_2^i, \dots, q_{3M}^i)$ , where  $M$  is the number of  $C\alpha$  atoms. Structural superimposition for every snapshot was carried out for all the  $C\alpha$  atoms. The variance-covariance matrix was defined as  $\{C_{mn}\}_{m,n=1,\dots,3M} = \langle q_m q_n \rangle_T - \langle q_m \rangle_T \langle q_n \rangle_T$ , where  $\langle \cdot \rangle_T$  is the ensemble average at temperature  $T$  ( $= 310$  K).  $C_{mn}$  was diagonalized, thereby generating eigenvectors  $e_1^x, e_2^x, \dots, e_M^x$  ( $x = \text{SARSCoV}, \text{SARSCoV2}$ ). Inner products  $e_i^{\text{SARSCoV}} \cdot e_j^{\text{SARSCoV2}}$  ( $i, j = 1, 2, 3$ ) were computed. Global modes were defined as the top-3 eigenvectors  $e_1^x, e_2^x$ , and  $e_3^x$ . The inner products indicate the correlation of global modes between of SARS-CoV-2 3CL<sup>pro</sup> and SARS-CoV 3CL<sup>pro</sup>.

Furthermore, distance-based PCA (dPCA) was also utilized to grasp a structural distribution [41]. In terms of a structural distribution, dPCA shows better performance (e.g., convergence of cumulative variances) than cPCA. [41] The dPCA was thus applied for the active site that was represented as a set of distances between  $C\alpha$  atoms (Table S2). The number of the selected  $C\alpha$  atoms was  $N_{C\alpha} = 25$ . The active site of MD snapshot  $i$  was expressed as a set of distances  $d^i = (d_1^i, d_2^i, \dots, d_K^i)$ , where  $K$  was the number of distance pairs, i.e.,  $K = N_{C\alpha}(N_{C\alpha} - 1) / 2$ . The variance-covariance matrix was defined as  $\{D_{mn}\}_{m,n=1,\dots,K} = \langle d_m d_n \rangle_T - \langle d_m \rangle_T \langle d_n \rangle_T$ . The matrix was diagonalized and then eigenvectors  $a_1^x, a_2^x, \dots, a_K^x$  ( $x = \text{SARSCoV}, \text{SARSCoV2}$ ) were generated. The active site  $d^i$  of each snapshot  $i$  was projected onto every eigenvector  $a_j^x$  by  $x_{ij} = (d^i - \langle d^i \rangle) \cdot a_j^x$ , where  $x_{ij}$  is the position of each snapshot  $i$  on  $j^{\text{th}}$  eigenvector, i.e.,  $j^{\text{th}}$  principal component (PCj).

### 2.3. Statistical analysis of the catalytic dyad

The probability distribution of the catalytic dyad's distance was calculated (Fig. 1). On the basis of the probability distribution, the dyad was judged whether it was broken or not. The following procedure was used for the judgement: (i) The probability distribution of the dyad distance  $d_{\text{dyad}}$  between  $N_\epsilon$ (H41) and  $S_\gamma$ (C145) was calculated. The distribution was expressed as  $P(d_{\text{dyad}})$ . (ii) The mode (the value that appears most frequently) of  $P(d_{\text{dyad}})$  was calculated, which was expressed as  $D_{\text{dyad}}^{\text{mode}}$ . (iii) The deviation of  $D_{\text{dyad}}^{\text{mode}}$  from the reference distance  $D_{\text{dyad}}^{\text{ref}}$  was calculated by

$$\Delta D = D_{\text{dyad}}^{\text{mode}} - D_{\text{dyad}}^{\text{ref}} \quad (1)$$

where  $D_{\text{dyad}}^{\text{ref}}$  was the distance between  $N_\epsilon$ (H41) and  $S_\gamma$ (C145) in the X-ray structure of PDBID 6LU7. The following condition was applied for

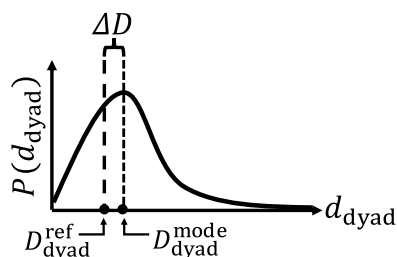


Fig. 1. Probability distribution of the distance  $d_{\text{dyad}}$  between  $N_\epsilon$ (H41) and  $S_\gamma$ (C145).  $D_{\text{dyad}}^{\text{ref}}$  was defined as the distance ( $= 3.9$  Å) between  $N_\epsilon$ (H41) and  $S_\gamma$ (C145) in an X-ray structure (PDBID: 1UK3A).  $D_{\text{dyad}}^{\text{mode}}$  was set to the most frequent value of  $P(d_{\text{dyad}})$ .

the judgement of the dyad's state:

$$\begin{aligned} \Delta D \geq 1.0 \text{ \AA} &\Rightarrow \text{the dyad is broken} \\ \Delta D < 1.0 \text{ \AA} &\Rightarrow \text{the dyad is retained} \end{aligned} \quad (2)$$

Using this condition, a confusion matrix was created (Table 3). This matrix was used for the evaluation of the structural difference of the dyad between the monomeric and the dimeric systems. Two-sided Fisher's exact test of independence was employed for the statistical assessment. The null hypothesis was that oligomeric states (i.e., monomeric or dimeric) are independent of the states of the dyad (i.e., broken or retained). Although this test is similar to  $\chi^2$  test of independence, the use of Fisher's exact test is recommended if sample size is less than 1000 [42]. Matthew's correlation coefficient (MCC) of the matrix was also calculated for the assessment of the correlation between the two categories: monomeric/dimeric states and broken/retained dyad. MCC is considered a reliable statistical measure [43].

### 2.4. Residence time of water molecules

Survival probability (SP) [44] of water molecules within the defined sphere (Fig. 2) was calculated by the following procedure: The average coordinates  $r_A$  of the triad (H41, H164, and D187) is calculated for each snapshot. The number of water molecules at time  $t$  within the sphere of radius  $5$  Å centered at  $r_A$  is expressed as  $N(t)$ . The number of water molecules within the sphere from  $t$  to  $t + \tau$  is represented as  $N(t, t + \tau)$ , where  $\tau$  is called a lag time. The SP is then defined as:

$$SP(\tau) = \frac{1}{T} \sum_{t=1}^T \frac{N(t, t + \tau)}{N(t)} \quad (3)$$

where  $T$  is the entire simulation time of interest and  $0 \leq SP(\tau) \leq 1$ . The SP tells us how fast water molecules within the sphere diffuse. Let us consider two extreme cases: if  $SP(\tau) = 1$  the number of water molecules within the sphere never decreases after time  $\tau$ , i.e., the water molecules stay in the sphere for the time  $\tau$ ; if  $SP(\tau) = 0$ , the number of water molecules within the sphere vanishes for the time  $\tau$ . Residence time  $\tau_{0.5}^{\text{system}}$  was defined as the time  $\tau$  when SP reaches 0.5. A python library, MDAnalysis, was used for the SP calculation [37,38,44].

## 3. Results and discussion

This section presents results of the systems with neutral dyads mainly, whereas results of the systems with ionized dyads are given in the supporting information.

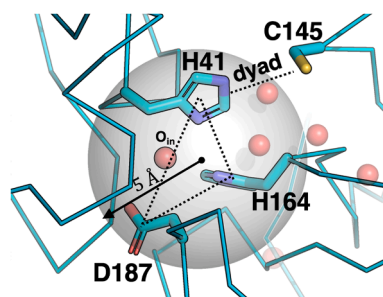


Fig. 2. The definition of the spherical space used in the survival probability calculation. The transparent grey space was defined as the sphere whose center was positioned at the average coordinates of H41, H164, and D187 and whose radius was  $5$  Å. This depicted structure was taken from the PDB ID 6WQF [19]. The symbol  $O_{\text{in}}$  indicates an oxygen atom of a water molecule within the sphere. The water molecule is surrounded by H41, H164, and D187 and corresponds to the crystal water that are considered important for the catalytic activity [19],[45].

### 3.1. Correlation between major dynamic modes of the two proteases

Performing cPCA for the entire protein structure, we produced their dynamic modes (i.e., eigenvectors) for the dimeric SARS-CoV 3CL<sup>pro</sup> and SARS-CoV-2 3CL<sup>pro</sup> systems. Since it is unclear how much the two proteases' global motion is conserved, we examined dynamic modes' correlations between the two proteases. Here, the correlation was defined as cosine similarity between a pair of eigenvectors, which were scaled to be unit vectors. Table 2 shows the correlation for each pair of eigenvectors. The highest correlation was 0.37 between  $e_3^{\text{SARSCoV}}$  and  $e_3^{\text{SARSCoV2}}$ , the second was 0.36 between  $e_2^{\text{SARSCoV}}$  and  $e_3^{\text{SARSCoV2}}$ , and the third was 0.35 between  $e_2^{\text{SARSCoV}}$  and  $e_3^{\text{SARSCoV2}}$ . The second mode  $e_2^{\text{SARSCoV}}$  was distributed into  $e_1^{\text{SARSCoV2}}$  and  $e_3^{\text{SARSCoV2}}$ . These results indicated that the major modes were not highly conserved, while the amino-acid sequence identity between two proteases was 96% that was very high value. The slight mutations on SARS-CoV-2 3CL<sup>pro</sup> altered the structural dynamics of the previous SARS-CoV 3CL<sup>pro</sup>. The alternation implied that the global dynamics were not a fundamental aspect for the proteases' function.

### 3.2. Structural Fluctuations of the Active Site

This section deals with structural fluctuations of the active site. For the analysis, we applied distance-based principal component analysis (dPCA) (Table S2 and Fig. 3E). Recall that the dimer has one active site in each protomer. We executed dPCA for each active site, obtaining the first principal component (PC1).

For the dimeric systems with neutral dyads, the active site in a protomer fluctuated differently from the other protomer's (Fig. 3). For instance, Traj2 in Fig. 3A shows two distinguishable probability distributions of PC1. Although the undistinguished distribution in Traj5 of Fig. 3D also occurred, it appeared rarer than the separated distributions. At the beginning of the simulations for 6LU7, the two active sites were structurally identical. This eliminates the possibility that the separated distribution was attributed to the initial structures. The asymmetric fluctuations may imply that substrate specificity of a protomer differ from that of the other. This asymmetric fluctuations of the active site were supported by several crystal structures whose protomer has conformationally distinct active site (e.g., PDBID: 1UK3) [9] and by earlier computational studies [10],[47]. An earlier study also pointed out the structural heterogeneity of the active site on the basis of known crystal structures [46].

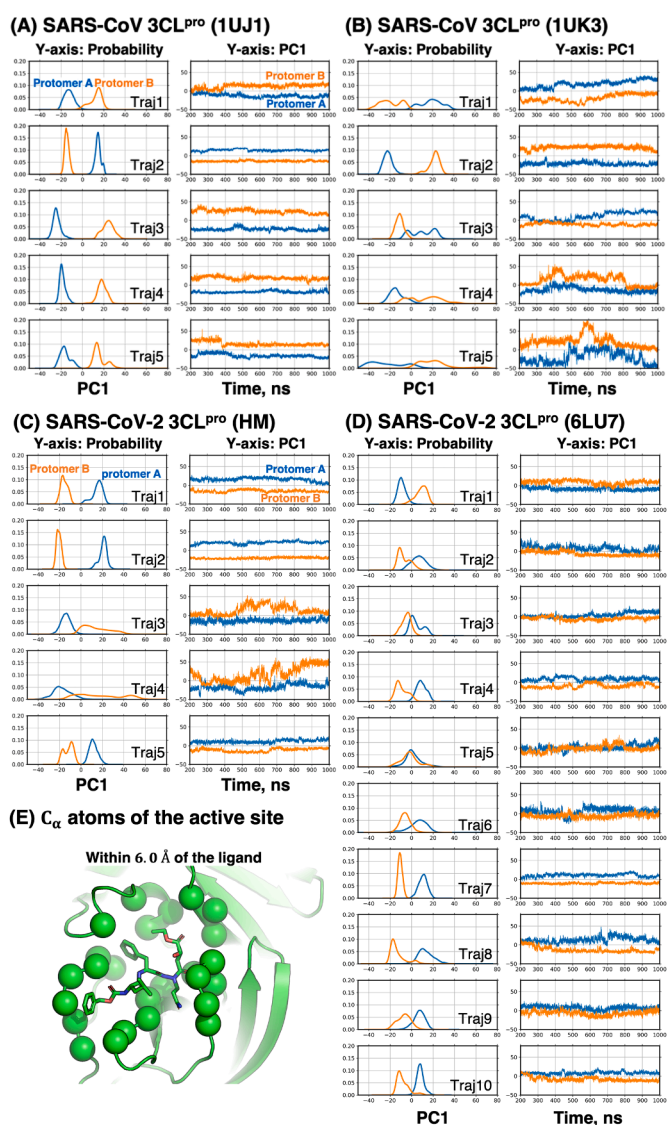
For the monomeric systems with a neutral dyad, we observed multimodal probability distributions along with PC1 (Fig. S5). For example, Traj5 in Fig. S5C shows a bimodal distribution, which indicated that the monomeric systems' active site was unstable more than the dimeric systems. Keeping active site's conformation was thus likely to be a difficult task for the monomer; this can be one explanation of why the monomer is inactive. For both proteases, the dynamic modes were altered by the slight mutations (Table 2), but the distributions of the active site showed the same tendency: each active site fluctuated asymmetrically (Fig. 3).

**Table 2**

Correlation coefficient between major eigenvectors.

	Mode	SARS-CoV-2 3CL <sup>pro</sup> (6LU7)		
		$e_1$	$e_2$	$e_3$
SARS-CoV 3CL <sup>pro</sup> (1UJ1&1UK3) <sup>a</sup>	$e_1$	0.03	0.02	0.07
	$e_2$	0.36	0.11	0.35
	$e_3$	0.05	0.13	0.37

<sup>a</sup> The trajectories were initiated from two different initial structures 1UJ1 and 1UK3. Subsequently, they were merged into a single trajectory, which was analyzed via cPCA.



**Fig. 3.** Probability distribution of the first principal component (PC1) and PC1's trajectories along time evolution. PC1 was obtained from dPCA using  $\alpha$  atoms of the active site. Blue and orange lines are for protomer A and B, respectively. (A) SARS-CoV 3CL<sup>pro</sup> (1UJ1), (B) SARS-CoV 3CL<sup>pro</sup> (1UK3), (C) SARS-CoV-2 CoV 3CL<sup>pro</sup> (HM), and (D) SARS-CoV-2 3CL<sup>pro</sup> (6LU7). (E) The  $\alpha$  atoms of the active site that was used in dPCA. PC1 of each graph was not identical to each other, i.e., we analyzed each Traj  $i$  ( $i=1,2,\dots,10$ ) individually in the dPCA, because merging multiple trajectories results in the expansion of conformational space, leading to the poor capability of dPCA to discriminate conformations.

### 3.3. Dyad stability

We investigated the distance fluctuations of the catalytic dyad in order to connect the dyad's fluctuations with the catalytic activity of the 3CL<sup>pro</sup>. We calculated the distance  $d_{\text{dyad}}$  between  $N_{\epsilon}$ (H41) and  $S_{\gamma}$ (C145). For each trajectory, we computed a probability distribution of  $d_{\text{dyad}}$  and obtained the mode  $D_{\text{dyad}}^{\text{mode}}$  (i.e., the most frequent value). Using the mode and the reference distance  $D_{\text{dyad}}^{\text{ref}}$ , we judged whether the dyad was retained or broken (Table 3). Further details of this procedure are given in Materials and Methods.

We found that the dimeric systems tended to retain the active dyad's conformation more than the monomeric systems did (Fig. 4 and Table 3). To assess the difference between the monomeric and dimeric systems, we conducted Fisher's exact test and calculated Matthew's

**Table 3**

Confusion matrix for the dependency of monomeric/dimeric states on the dyad's conformation. "Monomeric" (first column) and "Dimeric" (second column) data were derived from the trajectories No. 3, 4, and 7 and No. 1, 2, and 6 (see Table 1).

	Monomeric	Dimeric	p-value= 0.0025
Broken dyad	10	8	MCC=0.48
Retained dyad	5	32	

correlation coefficient (MCC). We have shown the statistical significance (p-value = 0.0025) if we employed the significance level of  $\alpha = 0.05$ . Furthermore, MCC was 0.48, showing a positive correlation of broken dyad with the monomeric systems. Thus, from a statistical viewpoint, the dimeric systems are likely to retain the dyad, whereas the monomeric systems are less likely to do so. The dyad instability in the monomer was in agreement with the result that SARS-CoV 3CL<sup>PRO</sup> (not SARS-CoV-2 3CL<sup>PRO</sup>) at low concentration (corresponding to a monomeric state) did not show its enzymatic activity [10].

Table S3 shows that 20% dyad of one protomer in the dimeric systems tended to be broken. It implied that the dyad in each protomer kept its distance asymmetrically like the asymmetric conformation of the active sites in Fig. 3. We thus conjectured that the active site and the dyad in a protomer form conformations different from those in the other protomer and that each protomer has inhomogeneous catalytic activity. Although no study gave the direct evidence to this inhomogeneity, the structural heterogeneity of the active site has been pointed out from crystal structures [46].

It should be noted that the homology modelling structure could not retain the geometry of the dyad even though the template structure 1UJ1 retained its dyad geometry (Table S3). The entire dynamics seemed appropriate, but the probability distribution of the distance for the dyad might have been affected by the modelled initial structure. It suggested that a homology model should be handled carefully by examining not only the overall dynamics but also local structural fluctuations.

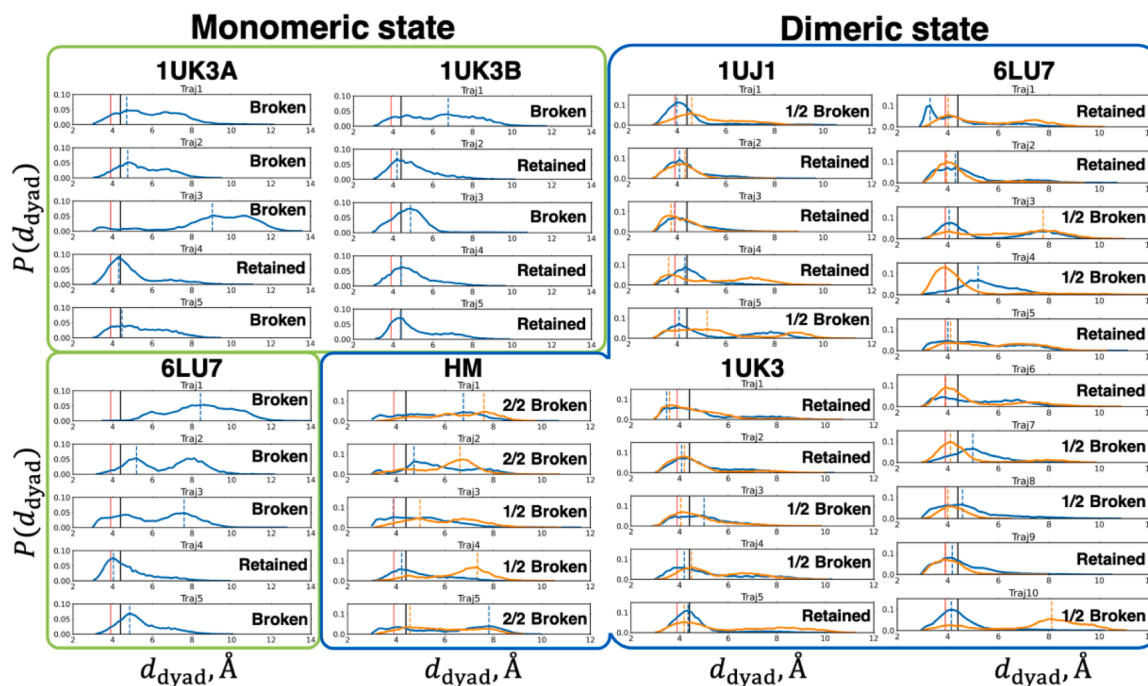
### 3.4. Consensus hydration sites of water molecules

To investigate the role of hydration water in the catalytic activity, we identified sites in which water molecules stayed for more than 500 ns (Fig. S6 and S7). From the sites, we selected and studied three sites (M<sub>1</sub>, M<sub>2</sub>, and D<sub>1</sub>, D<sub>2</sub>, and D<sub>3</sub> in Fig. S6 and S7; M and D stands for Monomeric and Dimeric, respectively). Since these were found in most trajectories, we called the sites "consensus hydration sites".

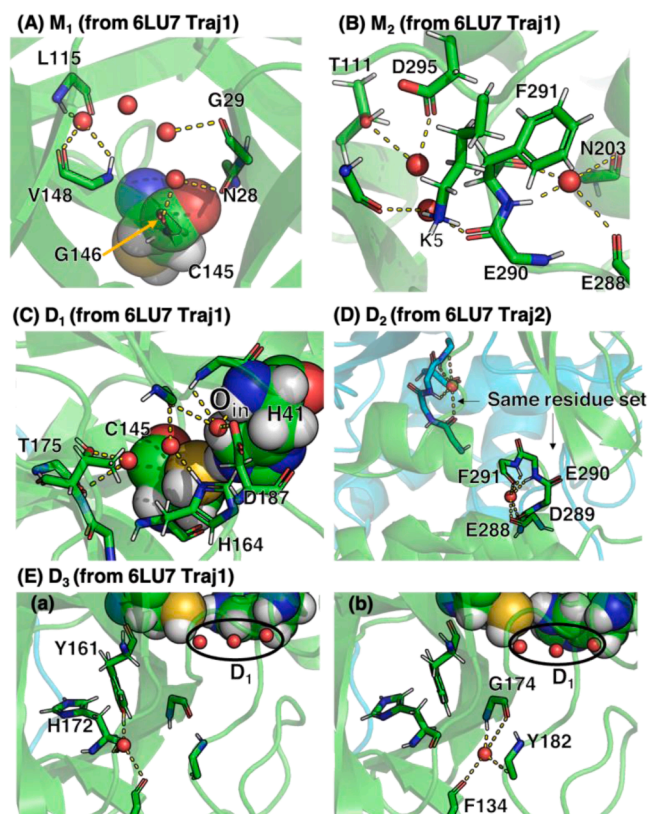
The consensus hydration sites had one noticeable feature: several water molecules were closely located and were stabilized mainly by the hydrogen bonds to N-H or C=O of the backbone (Fig. 5). For example, backbone's polar atoms in the site D<sub>2</sub> of both protomers interacted with a water molecule (Fig. 5D). The water molecule stabilized the loop in the site D<sub>2</sub>. Interestingly, since some residues in the site D<sub>2</sub> (e.g., F291, E290, and E288) were seen even in the site M<sub>2</sub>, this hydration site was likely to be conserved in both the monomeric and dimeric systems. The hydration site D<sub>2</sub> may play a crucial role in the dimerization: An earlier study suggested that the mutation E290A for SARS-CoV 3CL<sup>PRO</sup> led to a complete loss of the activity and dimerization because E290 in a protomer formed a salt bridge with R4 in the other protomer [48]. If the hydration water molecule does not stabilize the site D<sub>2</sub> containing E290, forming an appropriate orientation of E290 towards R4 can be difficult. The hydration water was also observed in crystal structures.

The site D<sub>1</sub> in the dimeric systems contained three water molecules (Fig. 5C). This site accommodated the water molecule (dubbed O<sub>in</sub>) found in most crystal structures that is hydrogen-bonded to N<sub>δ</sub> (H41), N<sub>ε</sub> (Q163), and O<sub>δ</sub> (D187) [19],[45]. The three water molecules were positioned behind the catalytic dyad, whereas water molecules in the site D<sub>1</sub> were not observed in the monomeric systems. Instead, the site M<sub>1</sub> appeared and was adjacent to D<sub>1</sub>. M<sub>1</sub> was likely to be transformed into D<sub>1</sub> in the dimeric systems (Fig. S8).

The site D<sub>1</sub> accommodating the three water molecules corresponds to the one where a single crystal water molecule is bound (e.g., PDB ID 6WQF, 6LU7, 1UK3, and 1UJ1) (Fig. 2). The difference between our MD trajectories and the crystal structures was the number of hydration



**Fig. 4.** Probability distribution of  $d_{\text{dyad}}$  for each system (blue lines are for protomer A, and orange for protomer B). The green rectangle region is for monomeric systems; the blue region is for dimeric systems. Solid black lines indicate the upper bound for the judgement of whether the dyad is broken or retained. Red solid vertical lines are the distance between N<sub>ε</sub>(H41) and S<sub>γ</sub>(C145) in an X-ray structure (PDBID: 1UK3A). Orange and blue dotted lines are the mode (i.e., the most frequent value) of each probability distribution.



**Fig. 5.** Consensus hydration sites obtained from representative snapshots of 6LU7. The sites are, respectively, shown in (A)  $M_1$  (B)  $M_2$ , (C)  $D_1$ , (D)  $D_2$ , and (E)  $D_3$ . In (E), two representative structure (a) and (b) are shown. In (C), the water molecule  $O_{in}$  corresponds to the crystal water that are considered important for the catalytic activity [19],[45]. The sphere representation indicates the dyad H41 and C145. The small red spheres represent oxygen atoms of water molecules. Green and cyan cartoon representations show the protomer A and B, respectively. The dotted-yellow lines indicate hydrogen bonds of water molecules to amino-acid residues. The interactions were mainly from backbone atoms, except for the electrostatic interactions with the tips of following residues: K5 (Fig. 5B); T175, H164 and D187 (Fig. 5C); Y161 (Fig. 5E(a)).

water molecules. The crystal structures show the existence of a single crystal water molecule in the site  $D_1$ , whereas our trajectories indicated several water molecules in the site.

The site  $D_3$  was observed in most of the trajectories, which was located near  $D_1$  (Fig. 5E and the blue circles in Fig. S7). Thus, the site  $D_3$  can be said to be the most agreed hydration site among the dimeric systems. Our result indicated that this site was conserved among the two different proteases (Fig. S7). Since the water molecule in the site  $D_3$  stayed near  $D_1$ , it can contribute to the stability of the site  $D_1$ , and thus to the stability of the dyad. In the site  $D_3$ , a crystal water molecule is also seen in crystal structures (e.g., 6WQF, 6LU7, 1UK3, 1UJ1, 6M03); it corresponds to the water molecule in Fig. 5E(b). Our trajectories demonstrated that the water molecule in the site  $D_3$  fluctuated between the following hydrogen-bond networks: Y182-G174-F134 and Y161-H172 (Fig. 5E(a) and 5E(b)). We noted that  $D_3$  was not observed in all the trajectories of the monomeric systems (Fig. S6).

### 3.5. Residence time of water molecules in the hydration site $D_1$

In the previous section, we discussed the five hydration sites, and the site  $D_1$  contained three water molecules, one of which corresponded to the catalytic water  $O_{in}$ . Here, we investigate the site  $D_1$  further, using residence time  $\tau_{0.5}^{system}$ , which indicates how long water molecules stay in a site on average (see the definition Eqn. 3). We computed survival

probability (SP) of water molecules within the spherical space defined in Fig. 2 (This space was the definition of the site  $D_1$ ). We then estimated residence time  $\tau_{0.5}^{system}$  for each system (Fig. 6 and S9, and Table S4).

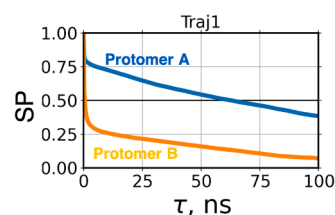
For the monomeric systems, the residence time for 1UK3A and 1UK3B were shorter than the one for 6LU7 (left columns in Table S4). This showed that the ability of the monomeric SARS-CoV-2 3CL<sup>PRO</sup> to keep water molecules in the site  $D_1$  was slightly higher than the monomeric SARS-CoV 3CL<sup>PRO</sup>. Also, since the site  $D_1$  of the monomeric SARS-CoV-2 3CL<sup>PRO</sup> had the two distinctive values of the residence time (e.g.,  $\tau_{0.5}^{6LU7} = 62.1$  ns and  $\tau_{0.5}^{6LU7} = 0.4$  in Table S4). The site  $D_1$  appeared to form at least two water-bound modes. We referred to the two water-bound modes as “water-weakly-bound mode” (if  $\tau_{0.5}^{system} < 30$  ns) and “water-strongly-bound mode” ( $\tau_{0.5}^{system} \geq 30$  ns), respectively. On the other hand, the site  $D_1$  in the monomeric SARS-CoV 3CL<sup>PRO</sup> appeared to form only water-weakly-bound mode, because of the short residence time ( $\tau_{0.5}^s \approx 5.0$  ns,  $s = 1UK3A$  or  $1UK3B$ ).

For dimeric systems, the site  $D_1$  of one protomer was likely to hold water molecules for longer residence time than that of the other protomer (right columns in Table S4). For both SARS-CoV 3CL<sup>PRO</sup> and SARS-CoV-2 3CL<sup>PRO</sup>, the site  $D_1$  of one protomer was capable of forming the water-strongly-bound mode, whereas the site  $D_1$  of the other protomer formed the water-weakly-bound mode. Both protomers were unable to form the water-strongly-bound mode simultaneously (e.g., Fig. 6), but both were able to do the water-weakly-bound mode at the same time.

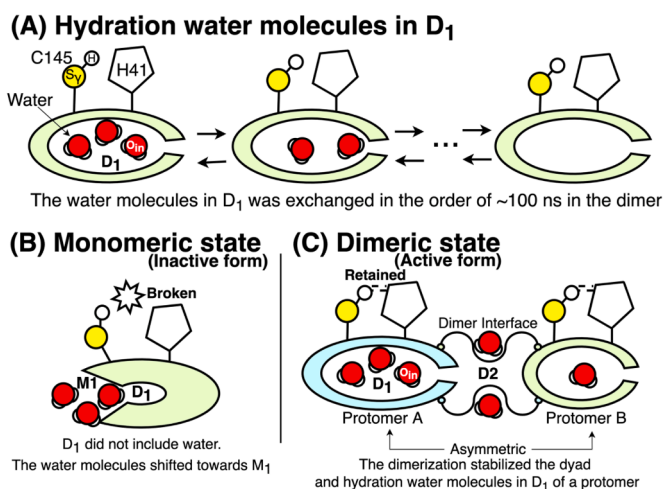
The comparison of the monomeric and dimeric systems demonstrated that the dimerization stabilized water molecules in the site  $D_1$ , because the residence time of the monomeric systems was at most 62.1 ns, whereas that of the dimeric systems was more than 100 ns. Indeed, several water molecules settled in the site  $D_1$  for more than 800 ns (not residence time but simulation time). As further analysis, WaterMap method would also be useful to obtain insight into the hydration sites and drug design [49],[50].

It would be beneficial to reiterate the two features for the dimeric systems: (i) the longer residence time of water molecules in the site  $D_1$  than the monomeric systems; (ii) more stable dyad than the monomeric systems. The two results showed that the dimerization stabilized the dyad and hydration water in the site  $D_1$ . Unexpectedly, the dyad stability appeared irrelevant to the residence time of the hydration water (Eqn S1 and Table S5), i.e., the hydration water was unlikely to stabilize the dyad. The hydration water molecules may have other roles in the catalytic activity, and we speculated the roles in the subsequent section.

Fig. 7 summarizes the relationship between the water molecules, the dyad, and the oligomeric states. Fig. 7A illustrates the mobility of the hydration water molecules in  $D_1$ . The mobility depended on the oligomeric state of the proteases (Table S4 and Fig. S9). In a monomeric state, hydration water molecules did not stay in  $D_1$ , and the dyad tended to be broken (Fig. 7B). The dimerization stabilized the water molecules in  $D_1$  and the dyad; the hydration water in  $D_2$  appeared to contribute the dimerization (Fig. 7C).



**Fig. 6.** A representative of survival probability (SP) for water molecules in the site  $D_1$ , which was taken from Traj1 (1UJ1) in Fig. S9. The horizontal solid line indicates 0.5 of SP, by which  $\tau_{0.5}^{system}$  was defined. All of the results are given in Fig. S9.



**Fig. 7.** The relation between hydration water molecules, the dyad, and oligomeric states. The red spheres with two white arcs indicate hydration water molecules, the red spheres labeled by O<sub>in</sub> are the crystal water molecule observed in many crystal structures; In our simulations, the crystal water O<sub>in</sub> was observed in the site D<sub>1</sub>. The yellow and white spheres labeled by S<sub>γ</sub> and H, respectively, are sulfur atoms and hydrogen atoms of C145. (A) The mobility of hydration water molecules in the site D<sub>1</sub>. (B) Hydration water molecules and the dyad in a monomeric state. (C) Hydration water molecules in the site D<sub>1</sub> and D<sub>2</sub> in a dimeric state. The molecules in D<sub>1</sub> of each protomer (cyan and light green) had different residence time.

### 3.6. Speculative roles of hydration water

For further consideration, we conjectured three roles of the hydration water molecules. The first possibility is that the water molecules stabilize the dyad, thereby keeping the activity. The water molecules, however, seemed irrelevant to the dyad stability because of the conditional probabilities calculated from Table S5. For the dyad stability, the dimerization did matter (Fig. 4 and Table 3).

The second possibility is that some of the water molecules in D<sub>1</sub> may transfer protons to C145 that is covalently bonded to a substrate, allowing the substrate to dissociate. It is thought that the catalytic process involves the hydrolysis of the covalent bond, which is attacked by a water molecule [51]. We speculated that a water molecule in D<sub>1</sub> can donate a proton to the covalent bond easier than bulk water molecules does because of the steric hindrance of the covalently-bonded substrate.

The third possibility is that the hydration water may prevent the active site from being denatured by accepting the heat generated by the catalytic reaction and dissipating it into bulk. The hydration water molecules in the site D<sub>1</sub> might act as a medium that gets and releases the heat into bulk. Earlier studies investigated the heat generated by a catalytic reaction of some enzymes: [52–54] The heat increased the enzyme's diffusivity [52],[53], and the energy of the heat was about 5 kcal/mol [54]. The results raise the question of why the resultant heat does not cause severe denaturation of the active site. We speculated that if the heat dissipates into the hydration water molecules in D<sub>1</sub>, the active site may not be intensely perturbed. Since the water molecules in D<sub>1</sub> were mobile (Fig. 6 and S9), they were readily released into the bulk and stored again. In this way, the active site may be prevented from being overheated. For direct verification, future work will take the propagation of the heat into account. A computational method by Stock Gerhard group would be suitable, which applies local temperature jump and then keeps track of the energy propagation [55–57].

## 4. Conclusion

The investigation of SARS-CoV 3CL<sup>pro</sup> and SARS-CoV-2 3CL<sup>pro</sup> in an

apo form provides fundamental insight into how they sustain their catalytic activity. In the present study, we simulated the two proteases. We revealed that major global modes were not conserved (Table 2). The changes in the modes would originate from the slight mutations on SARS-CoV-2 3CL<sup>pro</sup>. We observed that each active site in the dimeric 3CL<sup>pro</sup> obeyed different conformational distribution (Fig. 3). This may imply that each protomer has different substrate specificity. We found that the catalytic dyad in the dimeric systems tended to be retained more than that in the monomeric systems (Fig. 4 and Table 3), which indicated that the dimeric state is active.

Furthermore, we identified hydration water around the two proteases, the water which might be functionally relevant (Fig. 5). Hydration water molecules in the site D<sub>1</sub> of a dimeric state stayed longer than of a monomeric state (Fig. 6). Their residence time depended on the oligomeric states of the two proteases. Since it is thought that the dimer is an active form, these hydration water molecules might be functionally relevant. These tendencies above were also observed in the systems with ionized dyads.

Lastly, we point out unsolved problems. Although the dimerization is crucial for the catalytic activity, the dimerization interface has a large interaction surface, which causes difficulty in targeting the surface. For future work, we illustrated important residues on the interface in “Contact Residues on Dimer Interface” in the supporting information. Plus, even though the allostery cause by several mutations was investigated,[58] the knowledge of the effects on the protease's dynamics is still not adequate for application. Future work should describe more (e. g., by pointing out allosteric pathways) to allow us to manipulate the protease's function. SARS-CoV-2 can mutate readily because it is an RNA virus that mutates more quickly than DNA viruses.[59] The understanding of the mutation effects on the activity is indispensable to a fight for the viruses' drug-resistance.

### CRedit authorship contribution statement

**Shinji Iida:** Conceptualization, Methodology, Investigation, Validation, Formal analysis, Visualization, Writing – original draft, Data curation. **Yoshifumi Fukunishi:** Conceptualization, Project administration, Supervision, Funding acquisition, Writing – review & editing.

### Declarations of Competing Interest

None.

### Acknowledgements

This work was supported by grants of the Development of core technologies for innovative drug development based upon IT from Japan Agency for Medical Research and Development (AMED). Our simulations were able to be performed on TSUBAME3.0 supercomputers at the Tokyo Institute of Technology via HPCI System Research Project (Project IDs: hp170020) and with TSUBAME Encouragement Program for Young/Female/ Younger Users (Project ID: tge-20LJ0047).

### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.bbadv.2021.100016.

### References

- [1] C. Drosten, et al., Identification of a novel coronavirus in patients with severe acute respiratory syndrome, *N. Engl. J. Med.* 348 (2003) 1967–1976.
- [2] H. Yang, et al., Design of wide-spectrum inhibitors targeting coronavirus main proteases, *PLoS Biol.* 3 (2005).
- [3] Z. Jin, et al., Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (2020) 289–293.

- [4] W. Dai, et al., Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, *Science* (80-) 368 (2020) 1331–1335.
- [5] M.K. Tripathi, et al., Identification of bioactive molecule from *Withania somnifera* (Ashwagandha) as SARS-CoV-2 main protease inhibitor, *J. Biomol. Struct. Dyn.* 0 (2020) 1–14.
- [6] N.M. Tam, et al., Binding of inhibitors to the monomeric and dimeric SARS-CoV-2 Mpro, *RSC Adv.* 11 (2021) 2926–2934.
- [7] S. Ullrich, C. Nitsche, The SARS-CoV-2 main protease as drug target, *Bioorg. Med. Chem. Lett.* 30 (2020), 127377.
- [8] B. Goyal, D. Goyal, Targeting the dimerization of the main protease of coronaviruses: a potential broad-spectrum therapeutic strategy, *ACS Comb. Sci.* 22 (2020) 297–305.
- [9] H. Yang, et al., The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 13190–13195.
- [10] H. Chen, et al., Only one protomer is active in the dimer of SARS 3C-like proteinase, *J. Biol. Chem.* 281 (2006) 13894–13898.
- [11] D.W. Kneller, et al., Room-temperature X-ray crystallography reveals the oxidation and reactivity of cysteine residues in SARS-CoV-2 3CL Mpro: Insights into enzyme mechanism and drug design, *IUCr J.* 7 (2020) 1028–1035.
- [12] L. Zhang, et al., Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors, *Science* 368 (80-) (2020) 409–412.
- [13] T.J. El-Baba, et al., Allosteric inhibition of the SARS-CoV-2 main protease: insights from mass spectrometry based assays\*\*, *Angew. Chemie - Int* 59 (2020) 23544–23548.
- [14] W.C. Hsu, et al., Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease, *J. Biol. Chem.* 280 (2005) 22741–22748.
- [15] P.Y. Lin, C.Y. Chou, H.C. Chang, W.C. Hsu, G.G. Chang, Correlation between dissociation and catalysis of SARS-CoV main protease, *Arch. Biochem. Biophys.* 472 (2008) 34–42.
- [16] J. Liang, et al., Site mapping and small molecule blind docking reveal a possible target site on the SARS-CoV-2 main protease dimer interface, *Comput. Biol. Chem.* 89 (2020), 107372.
- [17] D.W. Kneller, et al., Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography, *J. Biol. Chem.* 295 (2020) 17365–17373.
- [18] Z. Jin, et al., Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (2020) 289–293.
- [19] D.W. Kneller, et al., Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography, *Nat. Commun.* 11 (2020) 7–12.
- [20] S. Iida, H.K. Nakamura, T. Mashimo, Y. Fukunishi, Structural fluctuations of aromatic residues in an apo-form reveal cryptic binding sites: implications for fragment-based drug design, *J. Phys. Chem. B* 124 (2020) 9977–9986.
- [21] M. Biasini, et al., SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Res.* 42 (2014).
- [22] J.M. Hancock, M.J. Zvelebil, R. Dunbrack, *SwissModel. Dictionary of Bioinformatics and Computational Biology, 2004*, <https://doi.org/10.1002/9780471650126.dob0720.pub2>.
- [23] S. Páll, B. Hess, A flexible algorithm for calculating pair interactions on SIMD architectures, *Comput. Phys. Commun.* 184 (2013) 2641–2650.
- [24] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472.
- [25] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* 126 (2007).
- [26] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* 81 (1984) 3684–3690.
- [27] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, *J. Appl. Phys.* 52 (1981) 7182–7190.
- [28] S. Nosé, M.L. Klein, Constant pressure molecular dynamics for molecular systems, *Mol. Phys.* 50 (1983) 1055–1076.
- [29] K. Lindorff-Larsen, et al., Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins Struct. Funct. Bioinforma.* 78 (2010) 1950–1958.
- [30] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (1983) 926–935.
- [31] I.G. Tironi, R. Sperb, P.E. Smith, W.F. Van Gunsteren, A generalized reaction field method for molecular dynamics simulations, *J. Chem. Phys.* 102 (1995) 5451–5459.
- [32] S. Páll, M.J. Abraham, C. Kutzner, B. Hess, E. Lindahl, Tackling exascale software challenges in molecular dynamics simulations with GROMACS, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8759 (2015) 3–27.
- [33] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun.* 91 (1995) 43–56.
- [34] M.J. Abraham, et al., Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1–2 (2015) 19–25.
- [35] E. Lindahl, B. Hess, D. van der Spoel, GROMACS 3.0: a package for molecular simulation and trajectory analysis, *J. Mol. Model.* 7 (2001) 306–317.
- [36] S. Pronk, et al., GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics* 29 (2013) 845–854.
- [37] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, O. Beckstein, MDAnalysis: A toolkit for the analysis of molecular dynamics simulations, *J. Comput. Chem.* 32 (2011) 2319–2327.
- [38] R. Gowers, et al., MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations, *Proc. 15th Python Sci. Conf.* 98–105 (2016), <https://doi.org/10.25080/Majora-629e541a-00e>.
- [39] L DeLano, W. *PyMOL: An Open-Source Molecular Graphics Tool*, Ccp4 Newslett. Protein Crystallogr. (2002), <http://www.ccp4.ac.uk/newsletters/newsletter36.pdf>.
- [40] B.S. Hanson, et al., Continuum mechanical parameterisation of cytoplasmic dynein from atomistic simulation, *Methods* 185 (2020) 39–48.
- [41] M. Ernst, F. Sittel, G. Stock, Contact- and distance-based principal component analysis of protein dynamics, *J. Chem. Phys.* 143 (2015).
- [42] J.H. McDonald, Fisher's exact test of independence, *Handb. Biol. Stat.* 9 (2014), <https://doi.org/10.1002/0470011815.b2a10020>.
- [43] D. Chicco, G. Jurman, The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, *BMC Genomics* 21 (6) (2020).
- [44] P. Liu, E. Harder, B.J. Berne, On the calculation of diffusion coefficients in confined fluids and interfaces with an application to the liquid-vapor interface of water, *J. Phys. Chem. B* 108 (2004) 6595–6602.
- [45] J. Tan, et al., pH-dependent conformational flexibility of the SARS-CoV main proteinase (Mpro) dimer: Molecular dynamics simulations and multiple X-ray structure analyses, *J. Mol. Biol.* 354 (2005) 25–40.
- [46] M.A.M. Behnam, Protein structural heterogeneity: A hypothesis for the basis of proteolytic recognition by the main protease of SARS-CoV and SARS-CoV-2, *Biochimie* 182 (2021) 177–184.
- [47] D. Suárez, N. Díaz, SARS-CoV-2 Main Protease: A Molecular Dynamics Study, *J. Chem. Inf. Model.* (2020), <https://doi.org/10.1021/acs.jcim.0c00575>.
- [48] C.Y. Chou, et al., Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease, *Biochemistry* (2004), <https://doi.org/10.1021/bi0490237>.
- [49] L. Wang, B.J. Berne, R.A. Friesner, Ligand binding to protein-binding pockets with wet and dry regions, *Proc. Natl. Acad. Sci.* 108 (2011) 1326–1330.
- [50] A. Fernandez, WaterMap originates in dehydron-based drug design, *J. Pharmacogenomics Pharmacoproteomics* (2017), <https://doi.org/10.4172/2153-0645.100e156>.
- [51] H. Wang, et al., Comprehensive Insights into the Catalytic Mechanism of Middle East Respiratory Syndrome 3C-Like Protease and Severe Acute Respiratory Syndrome 3C-Like Protease, *ACS Catal* 10 (2020) 5871–5890.
- [52] A.J. Wand, Enzymes surf the heat wave, *Nature* 517 (2015) 149–150.
- [53] C. Riedel, et al., The heat released during catalytic turnover enhances the diffusion of an enzyme, *Nature* 517 (2015) 227–230.
- [54] A. Dobry, J.M. Sturtevant, Heats of hydrolysis of amide and peptide bonds, *J. Biol. Chem.* 195 (1952) 141–147.
- [55] P.H. Nguyen, S.M. Park, G. Stock, Nonequilibrium molecular dynamics simulation of the energy transport through a peptide helix, *J. Chem. Phys.* 132 (2010).
- [56] S.M. Park, P.H. Nguyen, G. Stock, Molecular dynamics simulation of cooling: heat transfer from a photoexcited peptide to the solvent, *J. Chem. Phys.* 131 (2009) 37–40.
- [57] A. Gulzar, L. Valiño Borau, S. Buchenberg, S. Wolf, G. Stock, Energy transport pathways in proteins: a non-equilibrium molecular dynamics simulation study, *J. Chem. Theory Comput.* 15 (2019) 5750–5757.
- [58] L. Lim, J. Shi, Y. Mu, J. Song, Dynamically-driven enhancement of the catalytic machinery of the SARS 3C-Like Protease by the S284-T285-I286/A mutations on the extra domain, *PLoS One* 9 (2014), e101941.
- [59] K.M. Peck, A.S. Lauring, Complexities of Viral Mutation Rates, *J. Virol.* (2018), <https://doi.org/10.1128/jvi.01031-17>.