



OPEN Affinity of drug candidates binding to SARS-CoV-2 PLpro assessed using a generalized-ensemble method

Masashi Muramoto¹, Simon Hikiri², Suzuka Saito¹, Qilin Xie¹, Kota Kasahara^{2,3}, Junichi Higo^{4,5} & Takuya Takahashi²✉

To establish a procedure for screening compounds that inhibit ligand–receptor binding, we employed a multidimensional virtual-system coupled molecular dynamics (mD-VcMD), a generalized ensemble method recently developed by our group. In this approach, each compound is initially placed far from the receptor. Both receptor and compound were fully flexible in explicit solvent during sampling. The mD-VcMD generated a free-energy landscape of the compound–receptor interactions, in which a probability of existence was assigned to each sampled conformation. We examined four compounds that bind to the papain-like protease (PLpro) of SARS-CoV-2. The resultant free-energy landscapes were funnel-like for all compounds. The probabilities assigned to the free-energy basins correlated well with the measured dissociation constants. Furthermore, structural clustering revealed two types of binding modes within the free-energy basin. The probabilities assigned to the binding modes correlated well with the measured enzyme inhibitory activity. These results suggest that the proposed procedure is effective for selecting candidate inhibitors among the examined compounds.

Keywords Generalized ensemble, mD-VcMD, SARS-CoV-2 PLpro, Drug discovery, Binding affinity, Molecular binding

Since the 1990s, drug discovery research using computers has progressed rapidly. This trend coincides with significant improvements in computer performance. Simultaneously, the development of molecular dynamics simulations (MD) and the advancement of related application technologies have accelerated dramatically. For example, multicanonical MD-based dynamic docking simulations have elucidated the binding mechanism of a receptor and a ligand (i.e., a compound binding to the receptor). Moreover, the calculated binding constants are consistent with experimentally measured values^{1,2}.

Progress has advanced rapidly in the development of classical in-silico screening methods by which, after various ligands are arranged on the surface of the receptor, the plausibility of these complexes is judged using an empirical evaluation function without performing the heavy calculations necessary for MD simulation³. Nevertheless, it remains difficult to calculate the multidimensional free-energy landscape in a wide conformational space that includes the binding and dissociation states of two flexible molecules in an explicit solvent. Efficient calculation and utilization of the free-energy landscape are expected to be useful for elucidating the binding process. Classical screening methods are often inadequate when molecules (receptors and/or ligands) are in high motion and when the solvent effects on the complex formation are not negligible.

Various enhanced sampling methods have been used widely for such efficient computational sampling^{4–20}. In fact, the multi-dimensional virtual-system coupled molecular dynamics (mD-VcMD) method used for the present study is one such method^{7,8,19}. As a concrete example, we investigated the free energy landscape of complex formation and the most stable complex structure between the membrane protein GPCR and a drug molecule: bosentan²¹. The simulation was initiated from an unbound conformation, with GPCR and bosentan

¹Graduate School of Life Sciences, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan.

²College of Life Sciences, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan. ³Central Pharmaceutical Research Institute, Japan Tobacco Inc., 1-1 Murasaki-cho, Takatsuki, Osaka 569-1125, Japan.

⁴Graduate School of Information Science, University of Hyogo, 7-1-28 Minatojima Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. ⁵Research Organization of Science and Technology, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan. ✉email: tkhs@sk.ritsumei.ac.jp

spatially separated. As a result of an extensive structural search, the state with the lowest free energy (the most thermodynamically stable state) coincided with the experimentally

determined complex structure. By analyzing the free energy landscape, we were able to elucidate the binding mechanism in the complex formation²¹. We obtained a similar result from an mD-VcMD simulation of binding between globular protein 14-3-3 and its associated intrinsically disordered protein (MLF1)⁹.

The mD-VcMD method has also been applied for drug discovery. It has been demonstrated by experimentation that a compound that inhibits the complex formation of the mammalian SIN3 transcription corepressor isoform B (mSin3B) and repressor element 1 silencing transcription factor (REST) suppresses germination²². mD-VcMD simulation⁴ of several compounds binding to mSin3B has reported that a compound that binds firmly to the cleft of mSin3B in the simulation has greater inhibitory activity, measured by experimentation, than those of other compounds²². This finding suggests that mD-VcMD is applicable to drug discovery.

In addition, the free-energy landscape obtained from the mD-VcMD method is useful not only for finding the most stable complex structure, but also for identifying the following quantities: the positions of the most stable and metastable complexes (e.g., encounter complexes) in the landscape, the paths of structural changes among these stable complexes, and the free-energy barriers among the complexes^{4,6,9,21}. Consequently, the free-energy landscape provides comprehensive understanding of the complex formation process. When mD-VcMD is applied to different compounds, the obtained free-energy landscapes provide differences in complex structure formation among the compounds^{4,6,9,21}.

We note that the mD-VcMD method was developed based on statistical mechanics and physical chemistry, as explained in Refs. 4, 6, 8, 9, 12, 15, 17, 18, 21, 43, and 44. This method assigns thermodynamic weights to all sampled snapshots, enabling the construction of a free energy landscape that includes both bound and dissociated states. By sampling diverse conformations, we account for entropic contributions from molecular motions and conformational flexibility, as well as enthalpic contributions from inter- and intra-molecular interactions. Accurate evaluation of complex stability requires consideration of both states and their statistical weights.

The study described herein targets SARS-CoV-2 PLpro, the papain-like protease of SARS-CoV-2 which has caused a global pandemic. This protease was discovered next to SARS-CoV-2 3CLpro. For this study, SARS-CoV-2 PLpro is designated simply as PLpro. Below is a biological explanation of PLpro. The papain-like protease, including PLpro, is an essential enzyme of coronaviruses that is required for the processing of viral polyproteins to form functional replicase complexes and to enable viral spread. Also, PLpro is involved in cleaving proteinaceous post-translational modifications to host proteins as an evasion mechanism against the host's antiviral immune response. PLpro selectively cleaves ubiquitin-like interferon-stimulated gene 15 (ISG15) proteins. Crystallographic analysis of PLpro in complex with ISG15 shows unique interaction with the amino-terminal ubiquitin-like domain of ISG15 with high affinity and specificity. Furthermore, PLpro is involved in the cleavage of ISG15 from interferon responsive factor 3 (IRF3) during infection, attenuating the type I interferon response. Specifically, inhibition of PLpro with GRL0617 in infected cells suppresses virus-induced cytopathic effects, maintains the antiviral interferon pathway, and reduces viral replication^{23,24}. Consequently, targeting PLpro is a promising therapeutic strategy that combines the two aspects of suppressing SARS-CoV-2 infection and promoting antiviral immunity, enabling more efficient drug discovery.

For drug discovery targeting the proteases 3CLpro and PLpro, experimentation has already been conducted with SARS-CoV-2, such as structural analyses of proteases^{25,26}. Moreover, computational studies of proteases, including in silico screening using machine learning²⁷, in silico screening using MD²⁸, functional dynamics research covering homologous proteins²⁹, inhibitor binding research on pathway dynamics^{30,31} and in silico screening that combines information spectral method ISM and MD calculations³² are being conducted actively. Understanding of these important topics is deepening.

In the present study, we selected four complexes of PLpro with compounds from the ligand database designed by Shen et al.: GRL0617, XR8-24, XR8-23, and ZN-3-56 (Fig. 1a)³³. By applying the mD-VcMD method to these compounds binding to PLpro, we went beyond simple screening to elucidate details of the inhibition mechanism. Furthermore, we investigated the correlation between the computed stabilities of the compound-PLpro complexes and the experimentally obtained dissociation constants and IC_{50} values³³. We also applied clustering methods of two types^{34,35} to classify the numerous complex structures obtained from the simulations.

Materials and methods

To study binding of compounds to the receptor PLpro, we conducted the mD-VcMD simulation starting from an unbound conformation. Each system consists of PLpro and one of four compounds: GRL0617, XR8-24, XR8-23, or ZN-3-56 (Fig. 1a). We respectively designated the systems as PLpro-GRL0617, PLpro-XR8-24, PLpro-XR8-23, and PLpro-ZN-3-56.

The mD-VcMD simulation is a generalized ensemble method. For it, three variants have been developed to date: the original, subzone-based, and GA-guided mD-VcMD methods⁸. For this study, we used the subzone-based mD-VcMD, although we call this variant simply “mD-VcMD”. The mD-VcMD enhances the conformational sampling in a space (RC space) constructed by multiple reaction coordinates (RCs), which are set in advance. The mD-VcMD simulation generates an ensemble of conformations (i.e., snapshots) in the RC space. After the simulation, a statistical weight (i.e., a thermodynamically equilibrated weight at the simulation temperature) is assigned to each snapshot stored in the ensemble⁸. Therefore, this ensemble is regarded as a canonical ensemble (i.e., thermodynamically equilibrated ensemble), from which various physical quantities at equilibrium are computed.

Below, after first explaining the simulation systems, we introduce two RCs. Therefore, mD-VcMD is “2D”-VcMD in this study. Third, the mD-VcMD procedure is explained briefly. Fourth, we present a method to compute a physical quantity from the resultant ensemble. Subsequently, the principal component analysis (PCA)

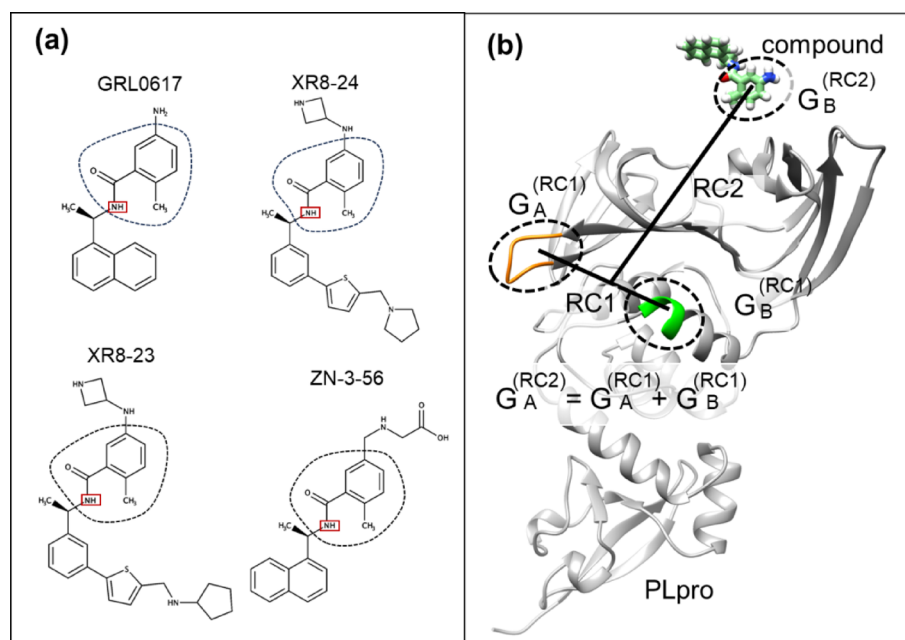


Fig. 1. (a) Chemical structures of four compounds used for mD-VcMD. Broken line circles are “common regions” that are common among the four compounds. This region is used for setting the atom group $G_B^{(RC2)}$ to define $RC2$. (b) Two atom groups $G_A^{(RC1)}$ (orange segment) and $G_B^{(RC1)}$ (green segment) define $RC1$; $G_A^{(RC2)}$ and $G_B^{(RC2)}$ define $RC2$. Table S2 presents the atom groups explicitly. Supplementary Sect. 2 presents the procedure for setting reaction coordinates $RC1$ and $RC2$.

method is explained, by which the receptor–ligand binding affinity is computed. Finally, we present a procedure to compute structural clusters from the ensemble.

Molecular systems and initial conformations of the simulation

The simulation systems were generated as explained hereinafter. First, the crystallographic structure of the PLpro–GRL0617 complex was taken from PDB (PDB ID: 7CJM; 3.2 Å resolution)³⁶. Then, we removed the compound GRL0617 from the structure. The 111-th amino-acid residue of PLpro in the crystallographic structure is serine, whereas it is cysteine in the wild-type PLpro. Therefore, we back-mutated it to cysteine. Acetyl and N-methyl groups were put respectively to the N-terminal and C-termini of PLpro.

PLpro was put at the center of a periodic boundary box filled by water molecules: Box size = $100 \text{ \AA} \times 100 \text{ \AA} \times 100 \text{ \AA}$. One can recall that GRL0617 was removed from the crystallographic structure described above. Then, we put one of the four compounds (GRL0617, XR8-24, XR8-23, and ZN-3-56) at a position distant from the ligand-binding site of PLpro. In this stage, therefore, the system comprises PLpro, one of four compounds, and water molecules. When putting PLpro and the compound in the box, water molecules overlapping to PLpro or the compound were removed. Furthermore, Na^+ and Cl^- ions were put randomly in the solvent region. Water molecules overlapping the introduced ions were removed. The quantities of ions were set so that the NaCl concentration was 150 mM and so that the net charge of the system was neutralized. Table S1 of Supporting Information (SI) presents the quantities of constituent atoms of the generated systems. Before initiating the mD-VcMD calculations, we examined various box sizes. Larger boxes require longer computation times, and if the ligand and receptor are too far apart, informative snapshots cannot be obtained. Notably, two spatially distant molecules contribute to the system’s entropy, but such entropic contributions cancel out when comparing different systems. Conversely, smaller boxes reduce computation time but fail to capture structures in which the ligand and receptor are fully dissociated. To obtain the free-energy landscape of complex formation, it is essential to sufficiently incorporate dissociated states into the ensemble. Ultimately, we found that a $100 \times 100 \times 100 \text{ \AA}$ box allows complete dissociation of the ligand from the receptor in all systems, and thus we adopted this box size.

The conformation of each system generated above was energy-minimized. An NVT simulation (300 K temperature) was done to relax the structural frustrations originating from the PDB structure. Then, an NPT simulation (1 atm pressure; 300 K temperature) was conducted to relax the box size at the pressure. Table S1 of SI presents the resultant box size for each system. The resultant conformations are portrayed in Figure S1 of SI, which indicates that the compound and PLpro are mutually separated in solvent for all systems. We used these unbound conformations for the initial conformations of the mD-VcMD simulations.

The four systems generated above did not involve a zinc (Zn) ion. By contrast, the crystallographic complex structures for PLpro–GRL0617 (PDB ID: 7CJM; Figures S2a of SI) and PLpro–XR8-24 (PDB ID: 7LBS; Figures S2b of SI) respectively contain three and one Zn ions. No crystallographic structure is reported for the other two

systems, PLpro–XR8-23 and PLpro–ZN-3-56, in PDB (Aug. 2023). In the structure of 7CJM (Figure S2a) the distances from the N2 atom of GRL0617 to the three Zn ions are 22.1 Å, 30.9 Å, and 39.4 Å. Similarly, the distance from the N03 atom of XR8-24 to the Zn ion is 30.9 Å in 7LBS (Figure S2b). Figure S2c of SI presents a ligand-free (apo-form) of PLpro (PDB ID: 6XG3), where the Zn ion is also distant from the ligand-binding site of PLpro. Furthermore, we searched other crystallographic structures containing PLpro in PDB, which revealed that most Zn ions are far from the ligand-binding site, as shown in Supplementary Sect. 1 of SI. Therefore, it is likely that the direct effect of Zn ions on the compound-binding is small. For that reason, no Zn ion was involved in the systems.

The force field parameters for PLpro, and the ions (Na^+ and Cl^-) were obtained respectively from AMBER99SB-ILDN³⁷ and the Joung–Cheatham parameters³⁸. The TIP3P potential³⁹ was used for water. Force field parameters for the four compounds were modeled by us. First, the atomic partial charges of the compounds were calculated quantum-chemically using Gaussian16⁴⁰ at the HF/6-31G* level, with subsequent RESP fitting⁴¹. Then, the obtained atomic partial charges were incorporated into a general amber force field 2 (GAFF2)⁴², which was designed to be compatible with conventional AMBER force-fields.

Two RCs introduced to control system motion

The mD-VcMD simulation enhances the conformational sampling in a space constructed by multiple reaction coordinates (RCs)⁸. The RCs are usually set before the simulation. The i -th RC, denoted herein as RC_i , was defined by the inter-centroid distance between two atom groups introduced into the system: $G_A^{(RC_i)}$ and $G_B^{(RC_i)}$ (Supplementary Sect. 2 of SI, Figure S3). As explained earlier, the space constructed by the multiple RCs is designated as an RC space. Because we introduced two RCs in this study, the RC space is two-dimensional.

Figure 1b shows $RC1$ and $RC2$ for the PLpro–GRL0617 system. $RC1$ is the inter-centroid distance between the atom groups $G_A^{(RC1)}$ and $G_B^{(RC1)}$, both of which are set in PLpro. The atom group $G_A^{(RC1)}$ is the orange segment in the figure, which is a loop region (residues G266–N267–Y268–Q269–C270–G271). Also, $G_B^{(RC1)}$ is the green segment, which is a part of a helix (residues V165–R166–E167–T168). The side-chains of the segments are eliminated from the definition of the atom groups. Also, $G_A^{(RC1)}$ and $G_B^{(RC1)}$ form the cleft of the ligand-binding site of PLpro. $RC2$ is the inter-centroid distance separating $G_A^{(RC2)}$ from $G_B^{(RC2)}$. Atom group $G_A^{(RC2)}$ is set to the combined region of $G_A^{(RC1)}$ and $G_B^{(RC1)}$ (i.e., $G_A^{(RC2)} = G_A^{(RC1)} + G_B^{(RC1)}$). Atom groups $G_A^{(RC1)}$, $G_B^{(RC1)}$, and $G_A^{(RC2)}$ are the same for all systems. Atom group $G_B^{(RC2)}$ is set to a region of the compound GRL0617. One might consider that $G_B^{(RC2)}$ depends on the system because $G_B^{(RC2)}$ is set to the different compounds. However, the compounds share a common region (Fig. 1a). Then we set $G_B^{(RC2)}$ to the common region. Consequently, $G_B^{(RC2)}$ is the same among the four systems. Table S2 of SI presents details for the atom groups. The benefit of using the same atom groups for all systems is that the free-energy landscape can be investigated from universal points of view.

Figure 1b shows that $RC1$ quantifies the cleft width of the ligand-binding site of PLpro and that $RC2$ quantifies the distance from the ligand-binding site to the compound. Therefore, $RC1$ controls opening (or closing) of the cleft, and $RC2$ controls the compound approach to (or departure from) the binding site. We have used similar RCs to study ligand-binding to the binding site of a receptor^{4,6,21,43,44}.

During an mD-VcMD simulation, the conformation is confined to the range $[RC_{i_{min}}, RC_{i_{max}}]$ along each of the RC_i axes, where $RC_{i_{min}}$ and $RC_{i_{max}}$ respectively represent the minimum and maximum boundaries for RC_i . To ensure that mD-VcMD samples the 2D-RC space widely, we set the 2D-RC range so that the conformation can adopt various structures. So that RC_i does not protrude from the range, the confinement is achieved by application of a restoring force to the conformation¹² only when RC_i is flying outside the range. For the present study, these 2D ranges involve both bound and unbound conformations. Table S3 of SI presents the actual values for the ranges. The smallest RCs are set as $RC_{i_{min}} = 0.0 \text{ \AA}$ ($i = 1$ and 2), which is the smallest value in principle because a negative distance does not exist in the real space. The value of 25.4890 Å for $RC1_{max}$ is sufficiently large to open the binding cleft. The value of 49.2801 for $RC2_{max}$ is also sufficiently large to allow the compound to move in the unbound state. We show in *Results and Discussion* that various complex and unbound conformations are sampled during the simulation.

Division of RC axes into small zones and execution of mD-VcMD

With mD-VcMD, each RC axis is divided into small zones, which are set in the range of $[RC_{i_{min}}, RC_{i_{max}}]$. The conformation moves in a zone (intra-zone movement) for a time interval and transitions to another zone (inter-zone transition) occasionally, as explained later. The numbers of zones set along the RC_i axis are set as $n_{vs}^{(RC_i)}$: $n_{vs}^{(RC1)} = 34$ and $n_{vs}^{(RC2)} = 48$ (Table S3 of SI). The lower and upper boundaries for the k -th zone of the axis RC_i are denoted respectively as $RC_{i_{low}}(k)$ and $RC_{i_{up}}(k)$. Their actual values are given in Table S4 of SI.

The simulation is assumed to be at time t at present. The conformation at this time is assumed as in the l -th and m -th zones, respectively, along the $RC1$ the $RC2$ axes. In other words, the conformation is in the 2D zone (l, m) at t . We designated this 2D zone as the “current zone”. In mD-VcMD, the conformation is confined in the current zone for a short interval $\Delta \tau_{trns}$ where the restoring force¹² towards the current zone is applied to the system only when the conformation is flying to outside the current zone. Alternatively, the conformation moves without the restoring force when it is inside the current zone. At a time-passage of $\Delta \tau_{trns}$ (time = $t + \Delta \tau_{trns}$), the conformation transitions from (l, m) to another zone (l', m') using

an inter-zone transition probability. The protocol to execute the transition and a method to set the transition probability are explained in Supplementary Sect. 3 and Figure S4 of SI. The zone (l', m') is one of zones surrounding the current zone (l, m) or the current zone itself (self-transition). The current zone is updated from (l, m) to (l', m') . We set $\Delta \tau_{trns} = 2 ps$, which corresponds to 1×10^3 simulation steps in this study (simulation time-step = 2 fs).

Repeating this procedure, the conformation moves among many 2D zones during the simulation. However, because the RC space $[RC_{imin}, RC_{imax}]$ is too large to be sampled in a single run, the mD-VcMD simulation is conducted iteratively. When the J -th iteration has finished, we calculate a probability distribution function $Q_{cano}^{[J]}(l, m)$. The probability of existence of the system assigned to a 2D RC zone (l, m) is calculated using simulation data from the fourth to J -th iterations⁸. As explained later, we conducted 20 iterations for each system in this study. Data from the first to third iterations were discarded to eliminate an error resulted from a structural distortion in the initial conformation of simulation, although we presume that such error was small in this study because the simulation starts from the unbound conformation (Figure S1).

When convergence of $Q_{cano}^{[J]}$ has been realized in the 2D-RC space, $Q_{cano}^{[J]}(l, m)$ is regarded as the system's canonical distribution (i.e., a thermally equilibrated distribution at the 300 K simulation temperature). Then, we simply denote $Q_{cano}^{[J]}(l, m)$ as $Q_{cano}(l, m)$ and assign a thermodynamic weight at T to each snapshot⁸. The convergence check is explained in Supplementary Sect. 4. If the convergence has not yet been obtained in iteration J , then the inter-zone transition probability is calculated; also, iteration $J + 1$ is started.

To increase the sampling efficiency further, we conducted N_{run} (= 1,000) runs of the mD-VcMD simulation in parallel in each iteration, where the runs started from different conformations. Conformations from the N_{run} runs construct a canonical ensemble, in theory^{20,45}. If assuming that the length of each run is $\Delta \tau_{run}$, and that the runs are initiated from various conformations, then the N_{run} trajectories cover a conformational space more widely than a long single run of length $N_{run} \times \Delta \tau_{run}$ does²⁰. Exceptionally, the N_{run} runs for the first iteration start from the single conformation (Figure S1). Then, the sampled region increases with progress of the simulation to the second and third iterations. The N_{run} initial conformations for the fourth iteration are the last conformations of the third iteration. When the J -th iteration has finished ($J \geq 4$), we have snapshots sampled from the fourth to J -th iterations. Then, the N_{run} initial conformations for the $(J + 1)$ -th iteration are selected from those snapshots so that they are distributed as evenly as possible in the 2D-RC space.

We set the simulation length $\Delta \tau_{run}$ of each of the N_{run} runs to 2.0 ns (1×10^6 steps) for each of the four systems. Consequently, the total length for an iteration was 2.0 μs (2.0 ns \times 1,000 runs). As we described above, after doing 20 iterations, we used the trajectories from the fourth to the twentieth iterations for analyses, for which the simulation length was 2.0 ns \times 1,000 \times 17 iterations = 34.0 μs . We saved a snapshot at every 5×10^4 steps (0.1 ns) in each run. The number of snapshots saved for the analyses was therefore 3.4×10^5 (= 34 μs / 0.1 ns) for each system. The integrated trajectories produce a canonical ensemble when the N_{run} simulations have been performed repeatedly, starting from different initial conformations. These N_{run} runs are executed in parallel^{20,45}. Therefore, the total trajectory length does not represent a continuous 34.0 μs simulation, but rather an ensemble-based sampling approach^{20,45}.

The mD-VcMD algorithm was first implemented to an MD simulation program omegagene/myPresto¹⁹ and next to GROMACS⁴⁶ with the PLUMED plug-in⁴⁷. The present simulation was done using the GROMACS version. The simulation conditions were the following: LINCS⁴⁸ to constrain the covalent-bond lengths related to hydrogen atoms, the Nosé-Hoover thermostat^{49,50} to control the simulation temperature, and the zero-dipole summation method^{51,52} to compute the long-range electrostatic interactions. No structural restraint was applied to the system: All molecules were flexible in the simulation except for the hydrogen-related covalent bonds constrained by LINCS, as explained above. We set the time-step of simulation to 2 fs ($\Delta t = 2$ fs) and the simulation temperature to 300 K. All simulations were conducted using the TSUBAME4.0 supercomputer at the Institute of Science Tokyo using GP-GPU. Other analyses were done at Ritsumeikan University and the Research Center for Computational Science, Okazaki, Japan. The force field parameters used for the present study were explained earlier herein.

One might infer that confinement of the system's motions in a zone for an interval of $\Delta \tau_{trns}$ might affect the resultant conformational ensemble. This criticism is valid for a conventional MD simulation that uses an artificial potential. For mD-VcMD, however, the bias effect is removed completely from the resultant ensemble by a reweighting technique^{8,21}. The resultant thermodynamic weight assigned to a snapshot is a quantity without the conformational confinement in zones. In fact, mD-VcMD differs from umbrella sampling⁵³, whereas the conformation is confined in a 2D-RC zone for the short time interval $\Delta \tau_{trns}$, as explained in our previous study⁴⁴.

Some physical quantities to study binding mechanisms

One can calculate the thermodynamic average (expected value) of a physical quantity Ω at the simulation temperature T (300 K) using the canonical ensemble of snapshots:

$$\langle \Omega \rangle = \frac{\sum_{i=1}^{N_{snap}} w_i \Omega_i}{\sum_{i=1}^{N_{snap}} w_i}, \quad (1)$$

where w_i denotes the statistical weight (thermodynamic weight) assigned to snapshot i , Ω_i represents the value of Ω for the snapshot, and N_{snap} (= 3.4×10^5) stands for the number of snapshots stored in the ensemble of a single system. Here, Ω can be either a scalar or vector. If one aims to restrict the ensemble so that the snapshots satisfy a condition, then Eq. 1 is redefined as

$$\langle \Omega \rangle = \frac{\sum_{i=1}^{N_{snap}} w_i \Omega_i \times D_i}{\sum_{i=1}^{N_{snap}} w_i \times D_i}, \quad (2)$$

where

$$D_i = \begin{cases} 1 & \text{(if snapshot } i \text{ satisfies the condition)} \\ 0 & \text{(else)} \end{cases} \quad (3)$$

The role of D_i is simple: If a snapshot does not satisfy the condition, then the snapshot is eliminated from the analysis. The standard deviation of Ω is defined as

$$SD_{\Omega} = \left[\frac{\sum_{i=1}^{N_{snap}} w_i \Omega_i^2 \times D_i}{\sum_{i=1}^{N_{snap}} w_i \times D_i} - \langle \Omega \rangle^2 \right]^{1/2}. \quad (4)$$

Principal component analysis (PCA)

Because a molecular conformation is specified by many atomic coordinates, the snapshot conformation is represented by a point in a high-dimensional space, and many snapshots obtained from a simulation distribute in the high-dimensional space. PCA has been used to visualize the distribution by projecting the points to a low-dimensional space^{54–57}. We applied PCA to the ensembles of the snapshots according to our earlier study⁵⁷, the methodological details of which are explained in Supplementary Sect. 5. Below we briefly present an outline of the PCA done for this study.

The variance–covariance matrix A is calculated from atom–pair distances between the whole Ca atoms of PLpro (316 Ca atoms) and the heavy atoms (eight atoms) in a core region of the compound (Figure S5 of SI). The element $A_{i,j}$ of A is defined in Equation S7 of SI. Adoption of all Ca atoms is done to detect contacts between the compound and the entire surface of PLpro. Alternatively, selection of only the core region is done to compare the PCA results among the four systems from the same perspective: the atoms taken for PCA are the same when the core region is used. The core region is a ring part of the common region of the compounds (Fig. 1a). Both the crystallographic structures for the PLpro–GRL0617 complex (PDB ID: 7CJM) and the PLpro–XR8-24 complex (PDB ID: 7LBS) demonstrate that this ring binds to the binding site of PLpro firmly with the same orientation. This point is explained further in the last paragraph of Supplementary Sect. 5.

We denote two eigenvectors with the largest and second largest eigenvalues respectively as $PC1$ and $PC2$. The contribution ratios CR_1 and CR_2 (Equation S11) to the standard deviation of the entire distribution were $CR_1 = 0.647$ and $CR_2 = 0.190$ and $CR(1;2) = 0.837$ (Equation S12). This large $CR(1;2)$ indicates that the 2D-PC space constructed by $PC1$ and $PC2$ is a conformational space that can capture important structural features of the system's distribution. Then, we computed a distribution function $\rho(PC1, PC2)$ for each system in the 2D-PC space. Because $\rho(PC1, PC2)$ is computed from the ensemble at 300 K, this function is the thermally equilibrated distribution at 300 K. A free-energy landscape (potential of mean force) at 300 K in 2D-PC space is defined as

$$PMF(PC1, PC2) = -RT \ln[\rho(PC1, PC2)], \quad (5)$$

where R stands for the gas constant, and where the unit of PMF is kilocalories per mole. Because we intend to compare PMF with an equal weight among the four systems, $\rho(PC1, PC2)$ is normalized for each system as $\iint \rho(PC1, PC2) dPC1 \times dPC2 = 1.0$.

Clustering of complex conformations

PCA is a convenient method to view the conformational distribution (the free-energy landscape; $PMF(PC1, PC2)$) and to identify a region occupied by thermodynamically stable conformations (i.e., the region with a low PMF) in the conformational space. However, it does not mean that PCA can pick the compound's conformations which have a particular structural feature merely by viewing the free-energy landscape. Therefore, structural clustering is important to classify the sampled conformations and to assign a binding mode to each of the resultant clusters.

Then, to elucidate the structural features of the sampled conformations in the entire conformational ensemble, we used the t-distributed stochastic neighbor embedding (t-SNE) method³⁴ and the density-based spatial clustering of applications with noise (DBSCAN) method³⁵ sequentially. We explain those procedures below.

A quantity that characterizes the compound–receptor complex structure is the pattern of atomic contacts between the two molecules. Then we quantified the heavy-atomic contacts between PLpro and the core region of each compound. As described earlier, the core region was also used for PCA. Discarding of the non-core region is done to eliminate unimportant inter-molecular contacts from the analysis. As discussed in the preceding subsection, the crystallographic structures for the PLpro–GRL0617 and the PLpro–XR8-24 complexes demonstrate that the core region binds firmly to the ligand-binding site of PLpro with the same orientation.

We represented the atomic distance between the i -th $C\alpha$ atom of PLpro and the j -th heavy-atom of the compound's core region of snapshot k as $l_{i,j}^{(k)}$, where the ordering of i and j is arbitrary. Then, we digitized

$$l_{i,j}^{(k)} \text{ as}$$

$$c_{i,j}^{(k)} = \begin{cases} 1 & (\text{if } l_{i,j}^{(k)} \leq l_{lim}^{(cnt)}) \\ 0 & (\text{if } l_{i,j}^{(k)} > l_{lim}^{(cnt)}) \end{cases} \quad (6)$$

where $l_{lim}^{(cnt)}$ is a contact-limit parameter. Atoms i and j are contacting when $c_{i,j}^{(k)} = 1$. Also, $l_{lim}^{(cnt)}$ is set as 8 Å for this study. To define the contact between two heavy atoms more rigorously, the criteria $l_{lim}^{(cnt)} = 7 \text{ \AA}$ might be better than 8 Å: Assuming that the radius of a heavy atom and the diameter of a water molecule are, respectively, approximately 2 Å and 3 Å, the water molecule only slightly passes through the space between two heavy atoms when the distance between the two heavy atoms is 7 Å (= 2 Å + 2 Å + 3 Å). However, to conserve computer memory resources, we used only a C α atom as a representative atom for each amino-acid residue. Therefore, we increased $l_{lim}^{(cnt)}$ from 7 Å to 8 Å in this study.

PLpro and the core region of a compound respectively contain 316 residues and 8 heavy atoms. Therefore, a set of the intermolecular contacts for the k -th snapshot is expressed as a vector of 2,528 (= 316 × 8) elements:

$$\mathbf{c}_k = (c_{1,1}^{(k)}, \dots, c_{316,1}^{(k)}, c_{1,2}^{(k)}, \dots, c_{316,8}^{(k)}) \quad (7)$$

One can imagine that the N_{snap} snapshots $\{\mathbf{c}_1, \dots, \mathbf{c}_{N_{snap}}\}$ for each system are distributed in the 2,528-dimensional space, where \mathbf{c}_k denotes the position of the k -th snapshot in the space. As stated earlier, N_{snap} is the number of snapshots saved for each system. We designate this space as a “contact space”. Because all C α atoms of PLpro are involved in Eq. 7, not only the intermolecular native contacts but also non-native ones are counted in this equation. Next, we reduced the dimensionality of the contact space from 2,528 to 2 using t-SNE. Accordingly, the N_{snap} data points were translated as

$$\{\mathbf{c}_1, \dots, \mathbf{c}_{N_{snap}}\} \rightarrow \{\mathbf{c}_1^{(2D)}, \dots, \mathbf{c}_{N_{snap}}^{(2D)}\}, \quad (8)$$

where $\mathbf{c}_k^{(2D)}$ represents the position of the k -th snapshot in the 2D contact space. Clustering of the 2D data points was done using DBSCAN.

We denote the m -th cluster for the system α ($\alpha = \text{PLpro-GRL0617, PLpro-XR8-24, PLpro-XR8-23 or PLpro-ZN-3-56}$) as $Cl_s_m^{(\alpha)}$. Then, the probability assigned to $Cl_s_m^{(\alpha)}$ for each system is computed as

$$P_{Cl_s_m^{(\alpha)}} = \sum_{i=1}^{N_{snap}} w_i \times D_i^{(Cl_s_m^{(\alpha)})} \quad (9)$$

where $D_i^{(Cl_s_m^{(\alpha)})}$ is a specific case of D_i (Eq. 3) defined as

$$D_i^{(Cl_s_m^{(\alpha)})} = \begin{cases} 1 & (\text{if snapshot } i \text{ of system } \alpha \text{ is in } Cl_s_m^{(\alpha)}) \\ 0 & (\text{else}) \end{cases} \quad (10)$$

Results and discussion

Conformational distribution in the 2D-RC space

We performed 20 iterations of the mD-VcMD simulation for each of the four systems starting from the unbound conformations (Figure S1). Figure 2 presents the probability distribution $Q_{cano}(l, m)$ computed from iterations 4–20 with discarding of iterations 1–3 as explained in the *Materials and Methods* section. Apparently, a small-RC2 region was characterized by a large probability for all systems, which means that the compound binding to the ligand-binding site of PLpro was more stable than those in the other regions. The eight sites of c_1, \dots, c_8 are marked in Fig. 2. Later, the conformations at these eight sites are shown. First, however, we assess the simulation data quality.

Convergence of simulation data was checked using the correlation coefficient $C_{cor}^{[J,20]}$ of the potential of mean force $F_J(l, m)$ (Equation S5) between the J -th and 20-th iterations (Equation S6.1 of SI). Figure S6 of SI demonstrates that $F_J(l, m)$ approached $F_{20}(l, m)$ quickly in an iteration range of $1 \leq J \leq 10$ and slowly in $11 < J \leq 16$. Finally, the coefficient almost reached a plateau for $J \geq 17$ in all systems. Therefore, we judged that the simulation provided an equilibrated ensemble approximately. As described in *Materials and Methods*, we discarded iterations 1–3 from the analyses. Figure S6 shows that the correlation coefficient increased quickly in the three iterations. Therefore, the $N_{run} (= 1,000)$ trajectories spread quickly from the single conformation (Figure S1) in the 2D-RC space during the three iterations.

Although the convergence was obtained well in the iterations, Fig. 2 depicts that the data scattering along the RC1 direction (i.e., the degree of the binding-site cleft opening) differs among the systems. To check this difference of distribution quantitatively, we calculated the fluctuation of RC1 (i.e., SD_{RC1}) at various values of RC2. Details for SD_{RC1} are presented in Supplementary Sect. 6 of SI. Figure S7 of SI shows that SD_{RC1} values were similar among the four systems in the large RC2 range. In fact, snapshots with a very small thermodynamic weight w_i contribute less to SD_{RC1} (Equation S13). Figure 2 is depicted with a log scale of Q_{cano} .

As explained earlier, RC1 was introduced to enhance the motion of the binding-site cleft of PLpro (Fig. 1b). In fact, Figure S7 shows that RC1 moved greatly. By contrast, Figure S7 shows that SD_{RC1} in a small-RC2

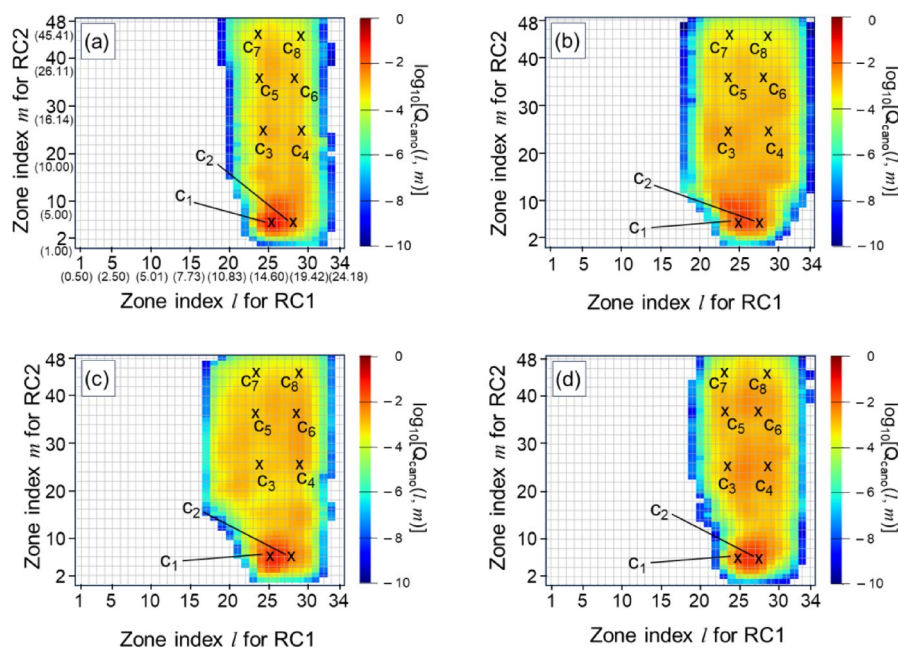


Fig. 2. Spatial patterns of the distribution $Q_{cano}(l, m)$ for (a) PLpro–GRL0617, (b) PLpro–XR8-24, (c) PLpro–XR8-23, and (d) PLpro–ZN-3-56. The X-axis represents zone index l for $RC1$. The Y-axis shows the zone index m for $RC2$. The main values of panel (a) are zone indices l and m . Values in parentheses $RC1$ and $RC2$ respectively correspond to l and m . Correspondence between zone index k ($k = l$ or m) and RCi ($i = 1$ or 2) is given as $RCi = [RCi_{low}(k) + RCi_{up}(k)]/2$. The RC values for the other panels are omitted. Table S4 of SI presents the values of $RCi_{low}(k)$ and $RCi_{up}(k)$. The value of $\log_{10}[Q_{cano}]$ at 2D zone (l, m) is shown by the color scale.

range has some variety, which is natural because direct interaction between PLpro and the component differs among the systems when the compound and PLpro mutually contact (i.e., when $RC1$ is small).

Figure 3 displays the conformations of PLpro–GRL0617 picked from the sites c_1, \dots, c_8 marked in Fig. 2a. The conformations c_1 and c_2 , which are from the lowest free-energy region of the PLpro–GRL0617 system, were complex structures (Fig. 3 c_1 and c_2), where GRL0617 bound to the ligand-binding site of PLpro. Conformations c_3 and c_4 were those for which the compound bound to a surface of PLpro deviating from the ligand-binding site (Fig. 3 c_3 and c_4). Conformations c_5 and c_6 showed that the compound contacted to various surfaces of PLpro (Fig. 3 c_5 is an example of the compound binding to PLpro) or that it was rarely dissociated from PLpro as in Fig. 3 c_6 . Conformations in a large- $RC2$ region (c_7 and c_8 in Fig. 2a) were unbound conformations (Fig. 3 c_7 and c_8). This positioning of the compound in the large- $RC2$ region is natural because the upper bound of $RC2$ ($RC2_{max} = 49.2801 \text{ \AA}$; Table S3) was set so that the compound could move freely without touching PLpro. A similar situation was found in the other systems, as shown in Figures S8–10 of SI.

It is noteworthy that Fig. 2 depicts a high free energy (i.e., a low probability) assigned to unbound conformations or non-native complex structures. Therefore, we infer that the free-energy landscape of the four systems is funnel-like in 2D-RC space.

PCA

PCA was applied to the resultant conformational ensemble of the four systems, for which details are given in Supplementary Sect. 5 of SI. Figure 4 presents $PMF(PC1, PC2)$ at 300 K in the 2D-PC space (Eq. 5). The spatial patterns of PMF were similar among the four systems: a deep free-energy basin existed at nearly the same position in the space; PMF increased with departure from the basin. In other words, the free-energy landscape is funnel like. As explained in *Materials And Methods*, the contribution $CR(1; 2)$ from $PC1$ and $PC2$ (Equation S12) to the entire conformational variety was large: $CR(1; 2) = 0.837$. Therefore, this funnel-like feature is likely to be the main feature in the full-dimensional space.

Figure 5 portrays conformations of the four systems picked randomly from the free-energy basin (i.e., region of $PMF(PC1, PC2) \leq 1.0 \text{ kcal/mol}$) of Fig. 4. All the conformations were complex structures in that the compound bound to the ligand-binding site of PLpro, whereas the conformation of the compound showed some variation in the binding site. Because these conformations were characterized by the low PMF value, it is likely that they are related to the dissociation constant K_d .

However, the existence of a low PMF region does not necessarily mean that the complex is thermodynamically stable. If the low PMF region is narrow in the PC space, then the complex might be unstable. Therefore, we computed the probability assigned to the low PMF region as

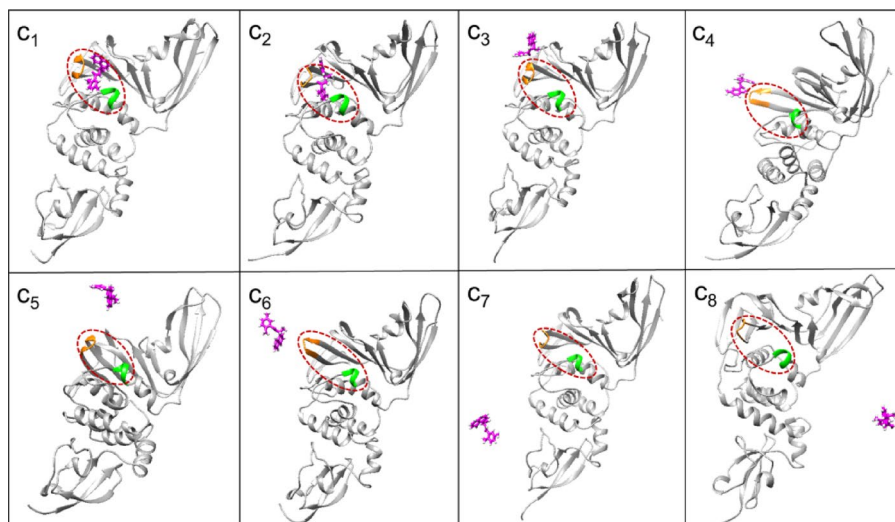


Fig. 3. Conformations of the PLpro–GRL0617 system selected from sites c_1, \dots, c_8 defined in (Fig. 2a). Panel names of this figure correspond to site names c_1, \dots, c_8 of (Fig. 2a). Purple molecules are GRL0617. Orange and green segments are atom groups used to define reaction coordinate $RC1$ (Fig. 1b): Residues 266–271 and 165–168. The brown broken line represents the position of the ligand-binding site of PLpro. Molecules in panels are drawn from different directions to clarify the positions of the ligand-binding site and compound.

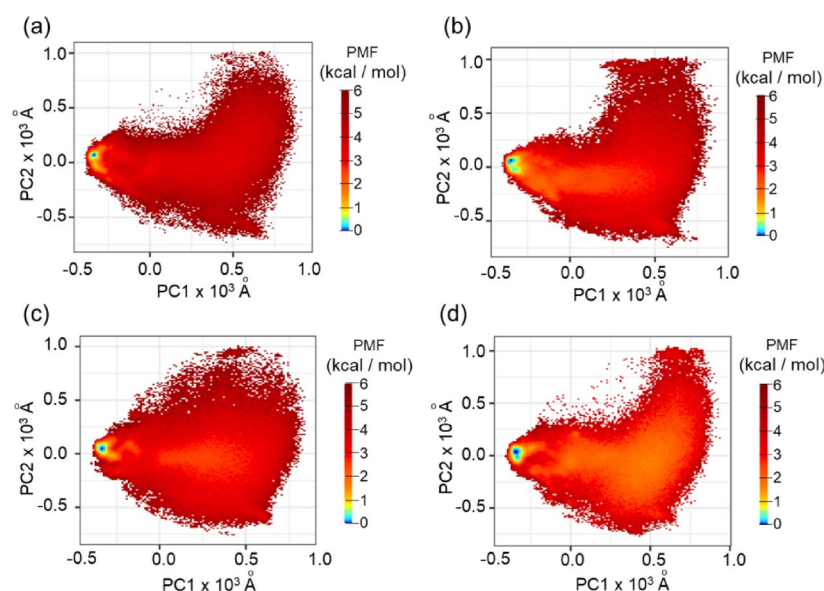


Fig. 4. Spatial pattern of PMF ($PC1, PC2$) in 2D-PC space for (a) PLpro–GRL0617, (b) PLpro–XR8-24, (c) PLpro–XR8-23, and (d) PLpro–ZN-3-56.

$$Q_{PC}(\epsilon) = \iint_{PMF \leq \epsilon} \rho(PC1, PC2) dPC1 \times dPC2 \quad (11)$$

Therein, ϵ is a threshold specifying the low PMF region. We examined not only 1.0 kcal/mol but also 0.3 kcal/mol and 0.5 kcal/mol for ϵ .

Table 1 presents $Q_{PC}(\epsilon)$ of the four systems together with their dissociation constants K_d . The dissociation constant K_d was reported from a study conducted by experimentation³³. Table 2 presents the correlation coefficient between $Q_{PC}(\epsilon)$ and K_d at each value of ϵ . It is particularly interesting that the correlation coefficient increased concomitantly with increasing ϵ and that the strongest correlation was -0.93 from $\epsilon = 1.0$ kcal/mol. In general, the structural variety of the compound in the complex increases concomitantly with increasing ϵ . Consequently, the conformations with $\epsilon = 1.0$ kcal/mol have the greatest structural diversity among the three values of ϵ examined. Those with $\epsilon = 0.3$ kcal/mol have the least diversity.

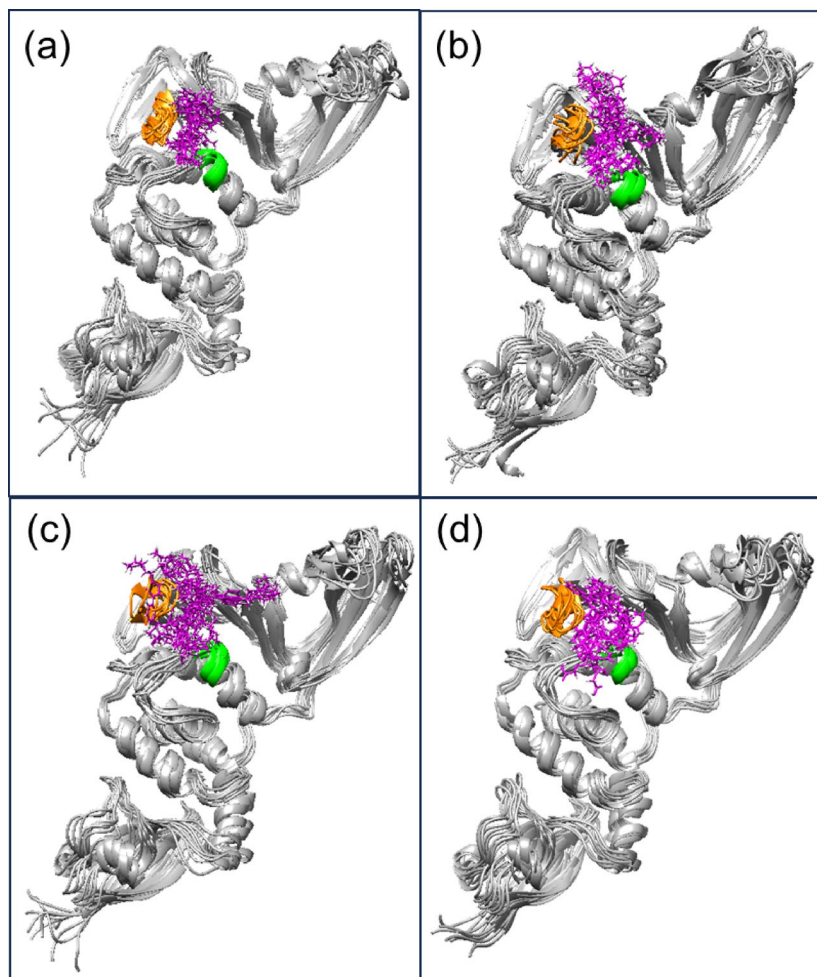


Fig. 5. Conformations picked randomly from free-energy basin (i.e., region of with $PMF(PC1, PC2) \leq 1.0$ kcal/mol of Fig. 4 for (a) PLpro-GRL0617, (b) PLpro-XR8-24, (c) PLpro-XR8-23, and (d) PLpro-ZN-3-56.

System	ϵ (kcal/mol)			k_d (μ M)
	0.3	0.5	1.0	
PLpro-GRL0617	0.170	0.260	0.380	2.7
PLpro-XR8-24	0.131	0.219	0.383	0.372
PLpro-XR8-23	0.178	0.197	0.329	0.235
PLpro-ZN-3-56	0.119	0.162	0.212	26.5

Table 1. $Q_{PC}(\epsilon)$ Assigned to region of $PMF \leq \epsilon$ and K_d for each system.

	ϵ (kcal/mol)		
	0.3	0.5	1.0
Correlation	-0.68	-0.71	-0.93

Table 2. Correlation coefficient between $Q_{PC}(\epsilon)$ and K_d .

Therefore, we infer that the reported dissociation constant found through experimentation is contributed by stably formed complex conformations, even when those conformations involve some structural diversity.

The correlation coefficient of -0.93 found for $Q_{PC}(\epsilon = 1.0$ kcal/mol) does not mean that the current study reproduced the absolute experimentally obtained value for K_d . Instead, the current study shows that the

order of K_d for the four compounds was reproduced well. Therefore, we infer that the current method is useful to select good candidate compounds among the computed compounds.

In fact, the periodic boundary box of the current study is smaller than that to reproduce the absolute value of K_d . The box–volume is expected to involve a sufficient margin where the compound fluctuates without any effect from PLpro. A simple way to define the thermodynamic stability of a complex is to calculate the ratio $\Delta Q_{bound}/\Delta Q_{unbound}$, where ΔQ_{bound} is the probability assigned to the bound state (the stable complex state) and $\Delta Q_{unbound}$ is that assigned to the unbound state. If the margin–volume is not counted accurately, then $\Delta Q_{unbound}$ cannot be estimated accurately. It is noteworthy that this argument omits correction by a huge volume of bulk solvent, which exists outside the simulation box in a real system, and omits correction by the ligand and receptor concentrations in the solution. We emphasize that the high correlation (-0.93) between Q_{PC} (1.0 kcal/mol) and K_d suggests that the error resulted from the margin region was cancelled among the four systems. Finally, we conclude that the current method is useful to screen good candidates for use as inhibitors.

Structural clusters

As shown in the preceding subsection, the complex structures from the free-energy basin exhibited some structural variety (Fig. 5). This variety suggests that the stable conformations are mixtures of some binding modes. Consequently, the structural clustering might help to identify the binding modes of the compound and to discuss whether these modes have inhibitory activity to PLpro, or not.

We applied t-SNE and DBSCAN to the conformational ensemble for the structural clustering, which is based on the atom-contact patterns between the compounds and PLpro (Eq. 6). Because it is difficult to set the optimal parameters for t-SNE and DBSCAN in general, we conducted clustering with three sets of input parameters to search the conformational clusters comprehensively: Set 1, Set 2, and Set 3.^{58–60} Parameters of Set 1 are *perplexity* = 100, *init* = *pca*, and *learning_rate* = 800 for t-SNE, and *eps* = 2.0 and *minpts* = 14000 for DBSCAN. Those of Set 2 are *perplexity* = 150, *init* = *pca*, and *learning_rate* = 800 for t-SNE, and *eps* = 1.5 and *minpts* = 15000 for DBSCAN. Those of Set 3 are *perplexity* = 150, *init* = *pca*, and *learning_rate* = 800 for t-SNE, and *eps* = 2.5 and *minpts* = 15000 for DBSCAN. Table 3 presents the resultant clusters $Cls_m^{(\alpha)}$ and their probabilities $P_{Cls_m^{(\alpha)}}$ (Eq. 9) for each system: α = PLpro–GRL0617, PLpro–XR8-24, PLpro–XR8-23 or PLpro–ZN-3-56. To specify that a cluster $Cls_m^{(\alpha)}$ resulted from parameter Set *i* (*i* = 1, 2 or 3), we designate the cluster as $Cls_m^{(\alpha)}$ (Set *i*) if necessary. In Table 3, snapshots in which the compound did not contact to PLpro were classified in a single cluster $Cls_D^{(\alpha)}$. Those in which only the non-core region of the compound contacted to PLpro were also assigned to $Cls_D^{(\alpha)}$ because such a contact was not counted in Eq. 6 ($c_{i,j}^{(k)} = 0$). “Outliers” in Table 3 are conformations that were not assigned to a cluster (i.e., singletons). A large majority of the outliers were conformations for which the core region contacted to a surface of PLpro other than the ligand-binding site. Because these contacts were less stable than the contacts formed between the compound’s core region and the PLpro’s ligand-binding site, no cluster was formed by the conformations in the outliers. Finally, a few (1, 2 or 3) clusters were identified from the clustering.

The intermolecular contact pattern in each cluster from parameter Set 1 is shown in Figs. 6, 7, 8 and 9. Those from Set 2 and Set 3 are presented respectively in Supplementary Figures S11 and S12. In all clusters, the compounds contacted to the ligand-binding site ($G_A^{(RC1)}$ and $G_B^{(RC1)}$; Fig. 1b) of PLpro, irrespective of the compound type. Although Figs. 6, 7, 8 and 9, S11 and S12 clearly characterize the contact patterns of the clusters, we cannot imagine the tertiary structures of the clusters from the figures.

The complex structures from the clusters of the four systems from parameter Set 1 are displayed in Figs. 10, 11, 12 and 13. Those from Set 2 and Set 3 are shown respectively in Supplementary Figures S13 and S14. These complex structures were typical of the clusters to which the structures belonged. The figures again show that the core-region of the compounds bound to the ligand-binding site of PLpro.

For α = PLpro–GRL0617, three clusters were obtained from Set 1. The first cluster $Cls_1^{(\alpha)}$ (Set 1) was similar to the ligand orientations of the native-like complex structures (PDB ID: 7CJM), but the side-chain of Y268 was standing on the surface of PLpro and the side-chain of Q269 covered the core region of GRL0617 (Fig. 10a). The second cluster $Cls_2^{(\alpha)}$ (Set 1) consisted of the native-like complex structures where the side-chains of Y268 and Q269 covered the core region of GRL0617 (Fig. 10b). In the third cluster $Cls_3^{(\alpha)}$ (Set 1), by contrast, the orientation of GRL0617 was greatly different from that in $Cls_2^{(\alpha)}$ (Set 1) (Fig. 10b) where the side-chains of Y268 and Q269 were standing on the PLpro surface and were in contact with the double-ring part (non-core region) of GRL0617 (Fig. 10c). In parameter Set 2, we again obtained three clusters: $Cls_1^{(\alpha)}$ (Set 2) (Figure S13a), $Cls_2^{(\alpha)}$ (Set 2) (Figure S13b) and $Cls_3^{(\alpha)}$ (Set 2) (Figure S13c), with binding modes that were respectively similar with those of $Cls_1^{(\alpha)}$ (Set 1), $Cls_2^{(\alpha)}$ (Set 1) and $Cls_3^{(\alpha)}$ (Set 1). Also, two clusters were identified from Set 3: $Cls_1^{(\alpha)}$ (Set 3), and $Cls_2^{(\alpha)}$ (Set 3). The cluster $Cls_1^{(\alpha)}$ (Set 3) was similar to $Cls_2^{(\alpha)}$ (Set 1) and $Cls_2^{(\alpha)}$ (Set 2) (Figure S14a), and $Cls_2^{(\alpha)}$ (Set 3) was similar to $Cls_3^{(\alpha)}$ (Set 1) and $Cls_3^{(\alpha)}$ (Set 2) (Figure S14b).

For α = PLpro–XR8-24, only one cluster $Cls_1^{(\alpha)}$ (Set 1) was obtained from parameter Set 1. XR8-24 had a native-like binding mode (PDB ID: 7LBS) and the side-chains Y268 and Q269 covered XR8-24 (Fig. 11). It is particularly interesting that $Cls_1^{(\alpha)}$ (Set 1) was divided into two clusters in Set 2: $Cls_1^{(\alpha)}$ (Set 2) and $Cls_2^{(\alpha)}$ (Set 2). In $Cls_1^{(\alpha)}$ (Set 2), the core region of XR8-24 did not contact to C111 of PLpro (Figure S11d), although the core region contacted to C111 of that in $Cls_2^{(\alpha)}$ (Set 2) (Figure S11e). In Set 3, we found only one cluster $Cls_1^{(\alpha)}$ (Set 3) (Figure S14c), for which the binding modes were similar, respectively, with those of $Cls_1^{(\alpha)}$ (Set 1).

Cluster	$P_{Cls_m}^{(\alpha)}$		
	Set 1	Set 2	Set 3
$\alpha = \text{PLpro} - \text{GRL0617}$			
$Cls_1^{(\alpha)}$	0.132	0.131	0.355
$Cls_2^{(\alpha)}$	0.121	0.097	0.181
$Cls_3^{(\alpha)}$	0.096	0.083	--
$Cls_D^{(\alpha)}$	0.191	0.191	0.191
Outliers	0.460	0.497	0.273
$\alpha = \text{PLpro-XR8-24}$			
$Cls_1^{(\alpha)}$	0.303	0.153	0.355
$Cls_2^{(\alpha)}$	--	0.062	0.123
$Cls_3^{(\alpha)}$	--	--	--
$Cls_D^{(\alpha)}$	0.204	0.204	0.204
Outliers	0.493	0.581	0.461
$\alpha = \text{PLpro-XR8-23}$			
$Cls_1^{(\alpha)}$	0.114	0.090	0.175
$Cls_2^{(\alpha)}$	0.073	0.081	0.107
$Cls_3^{(\alpha)}$	0.046	--	--
$Cls_D^{(\alpha)}$	0.329	0.329	0.329
Outliers	0.438	0.501	0.390
$\alpha = \text{PLpro-ZN-3-56}$			
$Cls_1^{(\alpha)}$	0.129	0.113	0.159
$Cls_2^{(\alpha)}$	0.033	--	0.041
$Cls_3^{(\alpha)}$	0.028	--	--
$Cls_D^{(\alpha)}$	0.538	0.538	0.538
Outliers	0.272	0.349	0.262

Table 3. Probabilities $P_{Cls_m}^{(\alpha)}$ of clusters Cls_m calculated with three sets of parameters.

For $\alpha = \text{PLpro-XR8-23}$, three clusters were obtained from Set 1. In $Cls_1^{(\alpha)}$ (Set 1), the side-chains of Y268 and Q269 covered the core region (Fig. 12a). In $Cls_2^{(\alpha)}$ (Set 1), however, the side-chains of Y268 and Q269 were standing on the PLpro surface and the tail part (non-core region) contacted to the side-chains (Fig. 12b). Furthermore, in $Cls_3^{(\alpha)}$ (Set 1), the core region of XR8-23 was covered by the side-chains of Y268 and Q269, as in $Cls_1^{(\alpha)}$ (Set 1), and also contacted C111 (Figs. 8c and 12c). The parameters Set 2 and Set 3 obtained two clusters each. For Set 2, the binding mode of $Cls_1^{(\alpha)}$ (Set 2) (Figure S13f) was similar to that of $Cls_2^{(\alpha)}$ (Set 1). That of $Cls_2^{(\alpha)}$ (Set 2) (Figure S13g) was similar to that of $Cls_1^{(\alpha)}$ (Set 1). This situation was the same for Set 3. The binding modes of $Cls_1^{(\alpha)}$ (Set 3) and $Cls_1^{(\alpha)}$ (Set 1) were similar. Those of $Cls_2^{(\alpha)}$ (Set 3) and $Cls_2^{(\alpha)}$ (Set 1) were similar.

For $\alpha = \text{PLpro-ZN-3-56}$, three clusters were obtained from Set 1. In the cluster $Cls_1^{(\alpha)}$ (Set 1), the side-chain of Y268 was standing on the surface of PLpro and the side chain of Q269 covered the core region of ZN-3-56 (Fig. 13a). In both $Cls_2^{(\alpha)}$ (Set 1) (Fig. 13b) and $Cls_3^{(\alpha)}$ (Set 1) (Fig. 13c), the core region of ZN-3-56 was covered by the side chains of Y268 and Q269. In this respect, $Cls_2^{(\alpha)}$ (Set 1) and $Cls_3^{(\alpha)}$ (Set 1) were similar. However, in $Cls_2^{(\alpha)}$ (Set 1), the ZN-3-56 core region did not contact to C111 of PLpro (Fig. 9b), whereas in $Cls_3^{(\alpha)}$ (Set 1), the core region did contact to C111 of that (Fig. 9c). In Set 2, by contrast, we found only one cluster: $Cls_1^{(\alpha)}$ (Set 2). The binding mode of $Cls_1^{(\alpha)}$ (Set 2) was similar to that of $Cls_1^{(\alpha)}$ (Set 1) (Figure S13h). This point is argued in the next paragraph. In Set 3, two clusters $Cls_1^{(\alpha)}$ (Set 3) and $Cls_2^{(\alpha)}$ (Set 3) were obtained. The first cluster $Cls_1^{(\alpha)}$ (Set 3) corresponded to $Cls_1^{(\alpha)}$ (Set 1) and $Cls_1^{(\alpha)}$ (Set 2). Also, the second cluster $Cls_2^{(\alpha)}$ (Set 3) was similar to $Cls_3^{(\alpha)}$ (Set 1).

The results of clustering with three parameter sets showed that even though the ligand orientations are similar, they are classified as different clusters due to the different orientations of the Y268 and Q269 side-chains in the ligand binding site. For example, $Cls_1^{(\alpha)}$ (Set 1) and $Cls_2^{(\alpha)}$ (Set 1) of $\alpha = \text{PLpro-GRL0617}$ had the same ligand orientation (Fig. 10a and b), but the orientation of the side-chains of Y268 and Q269 were different. In $Cls_1^{(\alpha)}$ (Set 1), the side-chain of Y268 was perpendicular to the surface of PLpro, while

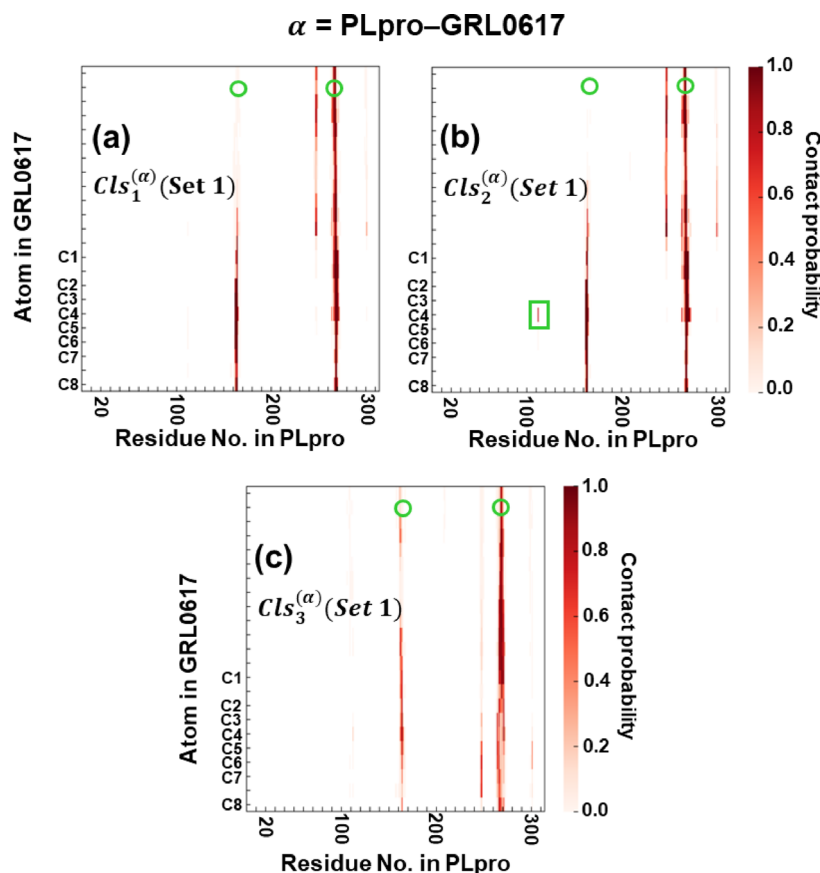


Fig. 6. Contact map between PLpro (X-axis) and GRL0617 (Y axis) for each cluster with clustering parameter Set 1 (Table 3). The contact probability, which is normalized in each cluster, is presented by color scale. Panels (a–c) are for clusters $Cls_1^{(\alpha)}$ (Set 1), $Cls_2^{(\alpha)}$ (Set 1) and $Cls_3^{(\alpha)}$ (Set 1) of $\alpha = \text{PLpro-GRL0617}$ system. Only the names of atoms in the core region are displayed on the Y axis (Fig. S5 of SI). Small green rings denote the position of atom group $G_A^{(RC1)}$ (residues 266–271) and $G_B^{(RC1)}$ (residues 165–168) (Fig. 1b). Space between $G_A^{(RC1)}$ and $G_B^{(RC1)}$ corresponds to the ligand-binding site. The green rectangle represents C111 of PLpro contacting the core-region of the compounds.

the side-chain of Q269 was parallel to the surface of PLpro and covered the core region of GRL0617 (Fig. 10a). However, in $Cls_2^{(\alpha)}$ (Set 1), both the side-chains of Y268 and Q269 covered the core region of GRL0617 (Fig. 10b). The classification of those binding modes was also similar to the parameter Set 2 ($Cls_1^{(\alpha)}$ (Set 2) and $Cls_2^{(\alpha)}$ (Set 2)) for $\alpha = \text{PLpro-GRL0617}$, Set 1 ($Cls_1^{(\alpha)}$ (Set 1) and $Cls_2^{(\alpha)}$ (Set 1)) and Set 3 ($Cls_1^{(\alpha)}$ (Set 3) and $Cls_3^{(\alpha)}$ (Set 3)) for $\alpha = \text{PLpro-ZN-3-56}$. As described above, even though the orientation of the ligands were similar, they were classified as different clusters due to the different side-chain orientations of Y268 and Q269 ($G_A^{(RC1)}$) at the ligand-binding site. Furthermore, the clusters similar to $Cls_1^{(\alpha)}$ (Set 1) and $Cls_1^{(\alpha)}$ (Set 2) for $\alpha = \text{PLpro-GRL0617}$ were obtained from parameter Set 3. Therefore, this is a concrete case in which the clustering result depends on the parameter set. We infer that it is important to examine some parameter sets to analyze the binding modes comprehensively. Also, depending on the orientation of the ligands, we were able to detect clusters of structures in which the side chains of Y268 and Q269 were parallel or perpendicular to the surface of PLpro. These structural features are used to assign the clusters obtained from the structural clustering to a binding mode below.

The position of the core region relative to the ligand-binding site of PLpro was common for all the clusters. Consequently, the core-region position is not a good identifier for analyzing the binding mode. By contrast, the molecular orientation of the compound relative to the PLpro's surface can be a good identifier. For instance, the compound's orientation in $Cls_1^{(\alpha)}$ (Set 1) and $Cls_2^{(\alpha)}$ (Set 1) for $\alpha = \text{PLpro-GRL0617}$ was approximately parallel to the PLpro surface (Fig. 10a and b). By contrast, in $Cls_3^{(\alpha)}$ (Set 1), a part of the compound (usually non-core region) was perpendicular to the surface (Fig. 10c). As another identifier, the side-chain orientation of Y268 and Q269 of PLpro relative to the surface of PLpro exhibited clear differences among the clusters. The side-

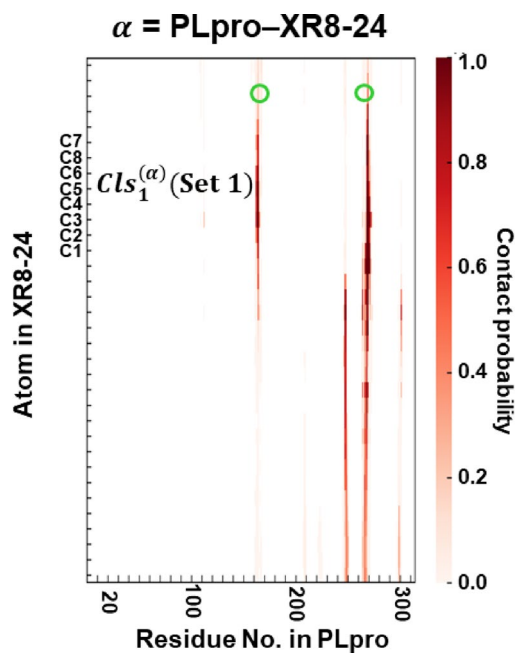


Fig. 7. Contact map between PLpro (X-axis) and XR8-24 (Y axis) for each cluster with clustering parameter Set 1 (Table 3). The contact probability, which is normalized in each cluster, is presented by color scale. This figure is for clusters $Cls_1^{(\alpha)}$ (Set 1) of $\alpha = \text{PLpro-XR8-24}$ system. Only the names of atoms in the core region are displayed on the Y axis (Fig. S5 of SI). Small green rings denote the position of atom $G_A^{\text{group}}(RC1)$ (residues 266–271) and $G_B^{\text{group}}(RC1)$ (residues 165–168) (Fig. 1b). Space between $G_A^{\text{group}}(RC1)$ and $G_B^{\text{group}}(RC1)$ corresponds to the ligand-binding site.

chains in $Cls_2^{(\alpha)}$ (Set 1) covered the core region of the compound, whereas those in $Cls_3^{(\alpha)}$ (Set 1) contacted to a part (usually non-core region) of the compound. Therefore, we classified the complex structures into two binding modes using the identifiers above: a parallel mode (Fig. 14a) and a perpendicular mode (Fig. 14b). In the parallel mode, the compound is approximately parallel to the PLpro surface. And, the two side-chains of Y268 and Q269 or the side-chains of Q269 cover the core region of the compound. In the perpendicular mode, a part (mostly the non-core region) of the compound took an orientation approximately perpendicular to the PLpro surface. This part contacts to the side-chains of Y268 and/or Q269, which are also approximately perpendicular to the PLpro surface. Tables 4 and 5 respectively assign clusters to the parallel mode and the perpendicular mode.

In all clusters, the core region bound to the ligand-binding site of PLpro (Figs. 10, 11, 12 and 13 and S13 and S14). Therefore, we inferred that all clusters obtained here have an enzyme-inhibitory effect. We defined the fraction for these clusters to the entire conformational ensemble as

$$Q_{inhib}^{(\alpha)} = \sum_m P_{Cls_m^{(\alpha)}} \quad (12)$$

In this equation, the probabilities for $Cls_D^{(\alpha)}$ and outliers were eliminated from the summation. Table 6 presents the resultant $Q_{inhib}^{(\alpha)}$ and the enzyme inhibition IC_{50}^{33} for each system: IC_{50} is the concentration of the compound in the solution at which 50% of the enzymatic function of PLpro is inhibited. Table 7 presents the correlation coefficient between $Q_{inhib}^{(\alpha)}$ and IC_{50} . Also, Table 8 presents the correlation coefficient between $Q_{inhib}^{(\alpha)}$ and K_d . Those shows that the coefficient is anti-correlative, irrespective of the parameter set. In fact, lower values of IC_{50} and K_d are generally associated with stronger inhibition. Therefore, the anti-correlation reflects that the computed $Q_{inhib}^{(\alpha)}$ supports the experimental IC_{50} and K_d for all parameter sets. Findings from the present study suggest that the mD-VcMD simulation can be used effectively for screening promising compounds from others with using $Q_{inhib}^{(\alpha)}$. Consequently, GRL0617, XR8-24 and XR8-23 are more promising than and ZN-3-56.

As described earlier, correlation between $Q_{PC}(\epsilon)$ (Eq. 11) and K_d increased concomitantly with increase of threshold ϵ . The best correlation was from $\epsilon = 1.0$ kcal/mol (-0.93; Table 2) among the three examined ϵ

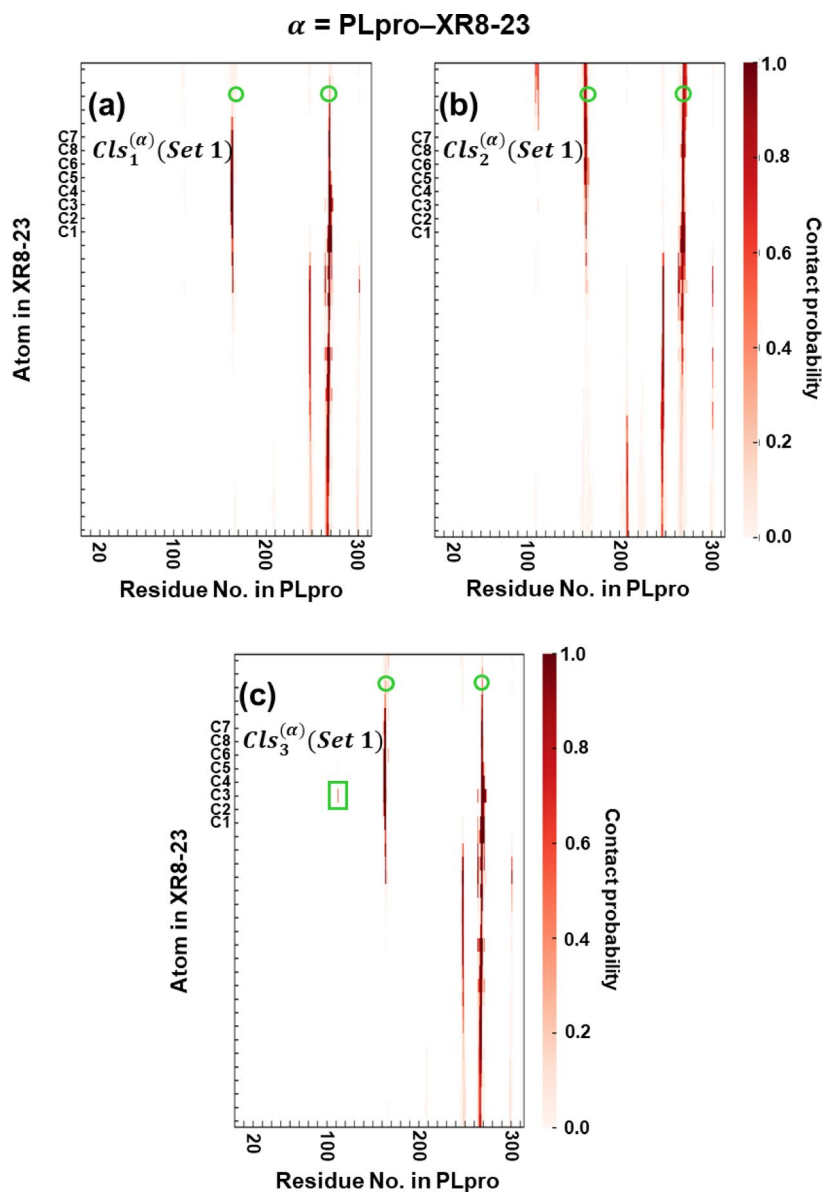


Fig. 8. Contact map between PLpro (X-axis) and XR8-23 (Y axis) for each cluster with clustering parameter Set 1 (Table 3). The contact probability, which is normalized in each cluster, is presented by color scale. Panels (a–c) are for clusters $Cls_1^{(\alpha)}$ (Set 1), $Cls_2^{(\alpha)}$ (Set 1) and $Cls_3^{(\alpha)}$ (Set 1) of $\alpha = \text{PLpro-XR8-23}$ system. Only the names of atoms in the core region are displayed on the Y axis (Fig. S5 of SI). Small green rings denote the position of atom group $G_A^{(RC1)}$ (residues 266–271) and $G_B^{(RC1)}$ (residues 165–168) (Fig. 1b). Space between $G_A^{(RC1)}$ and $G_B^{(RC1)}$ corresponds to the ligand-binding site. The green rectangle represents C111 of PLpro contacting the core-region of the compounds.

values. Therefore, the complex state involves some structural variety. Furthermore, the complex state, which has inhibitory activity, consists of structurally distinct clusters (Fig. 14). These results might represent a concrete case in which both the receptor and the ligand (i.e., the compound in the present study) vary their tertiary structures to search stable complex structures⁶¹. Furthermore, the existence of two binding modes suggests that the binding mechanism of the current system is a fuzzy complex formation⁶² or a multimodal binding mechanism^{10,16}.

We examined four compounds to verify our screening procedure. One suggestion is to apply the procedure to more compounds. We consider that suggestion to be valid. An earlier study³³ conducted by experimentation found k_d and IC_{50} values measured for many compounds, which bind to PLpro. We consider that future

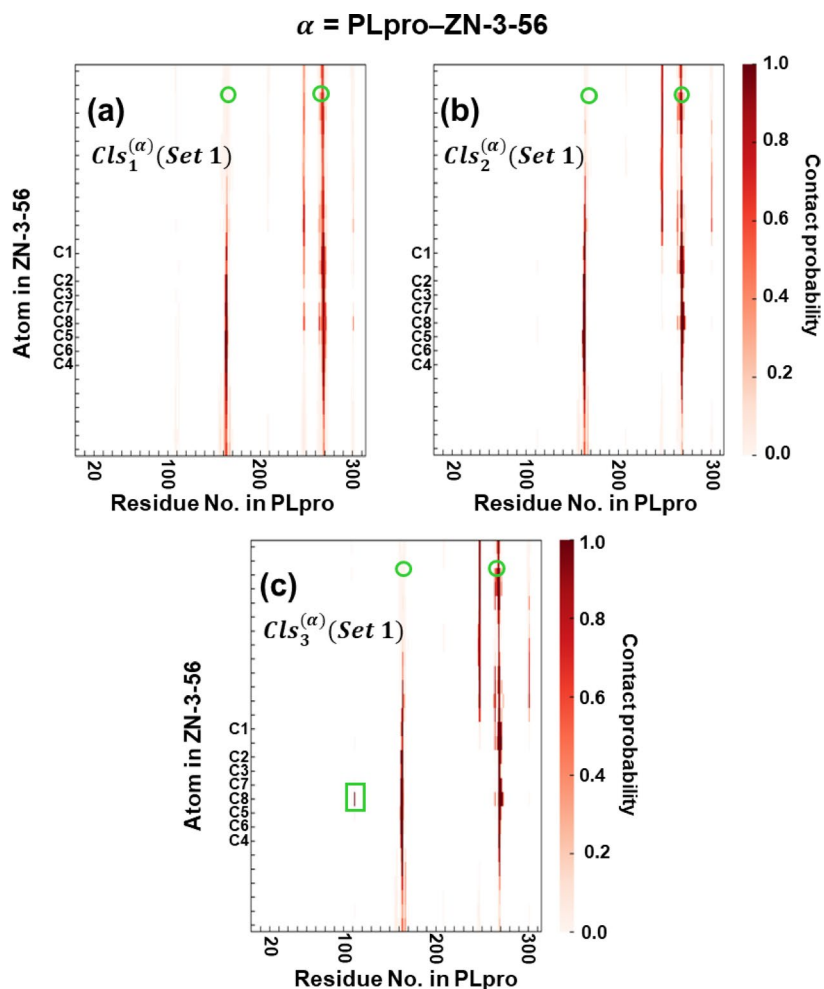


Fig. 9. Contact map between PLpro (X-axis) and ZN-3-56 (Y axis) for each cluster with clustering parameter Set 1 (Table 3). The contact probability, which is normalized in each cluster, is presented by color scale. Panels (a–c) are for clusters $Cls_1^{(\alpha)}$ (Set 1), $Cls_2^{(\alpha)}$ (Set 1) and $Cls_3^{(\alpha)}$ (Set 1) of $\alpha = \text{PLpro-ZN-3-56}$ system. Only the names of atoms in the core region are displayed on the Y axis (Fig. S5 of SI). Small green rings denote the position of atom group $G_A^{(RC1)}$ (residues 266–271) and $G_B^{(RC1)}$ (residues 165–168) (Fig. 1b). Space between $G_A^{(RC1)}$ and $G_B^{(RC1)}$ corresponds to the ligand-binding site. The green rectangle represents C111 of PLpro contacting the core-region of the compounds.

studies should be done for those compounds to clarify the correlation between computational probabilities ($Q_{PC}(\epsilon)$ or $Q_{inhib}^{(\alpha)}$) and experimentally obtained K_d or IC_{50} values.

Determination of atom groups $G_A^{(RCi)}$ and $G_B^{(RCi)}$ is important to set RCs effectively. We presume that an inappropriate RC set requires more computational time than an appropriate one, although any set of RCs can produce a canonical ensemble of the system, in theory. For the present study, the crystallographic structures of the PLpro–GRL0617 complex (PDB ID: 7CJM) and the PLpro–XR8-24 complex (PDB ID: 7LBS) helped the setting of RCs. However, if an essential region for the target compound and the binding site of the receptor is unknown, then another approach is required. Moreover, such an approach must be free from the setting of atom groups. Such a version free from atom-group setting is now under development.

We emphasize that mD-VcMD is a method to compute a free-energy landscape to study biomolecular binding^{4,6,21,43,44}. The study described herein specifically addressed only the free-energy basin of the conformational space because the aim of this study is verification of the correlation between the computational $Q_{PC}(\epsilon)$ (or $Q_{inhib}^{(\alpha)}$) and the experimentally obtained values of K_d (or IC_{50}). In this last paragraph, it must be stated that the free-energy landscape and the conformational ensemble from mD-VcMD can provide various intermediates (encounter complexes), which might be far from the free-energy basin in the conformational space, and pathways among the various complexes. Such insights from the free-energy landscape might provide important hints supporting drug discovery, along with knowledge related to the basic study of biology.

As Zn ions in PLpro are known to be important for the structure and function^{61–63}, we also performed preliminary mD-VcMD simulations including zinc ions. We then calculated the RMSD (root-mean-squared-

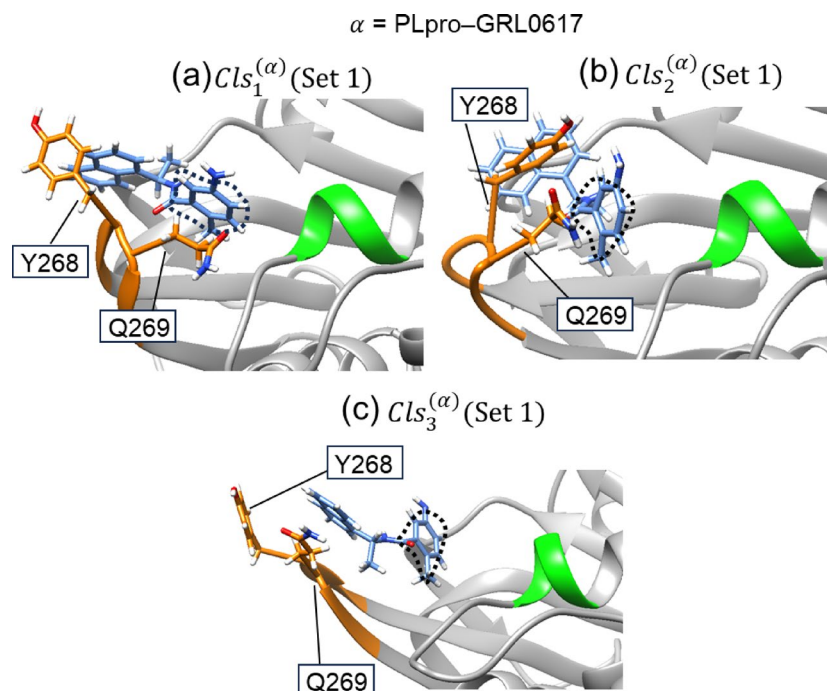


Fig. 10. Complex structure picked from each cluster obtained from clustering parameter Set 1 of $\alpha = \text{PLpro-GRL0617}$ system (Table 3). Panels (a), (b) and (c) are respectively for clusters $Cl_1^{(\alpha)}$ (Set 1), $Cl_2^{(\alpha)}$ (Set 1) and $Cl_3^{(\alpha)}$ (Set 1). The orange segment is the loop region ($G_A^{(RC1)}$; residues 266–271). The green one is a part of helix ($G_B^{(RC1)}$; residues 165–168); Fig. 1b shows details for these segments. Side-chains of two amino-acid residues Y268 and Q269 are labeled. Cornflower blue molecules are compounds, with broken-line circles representing their core regions (Fig. S5 of SI).

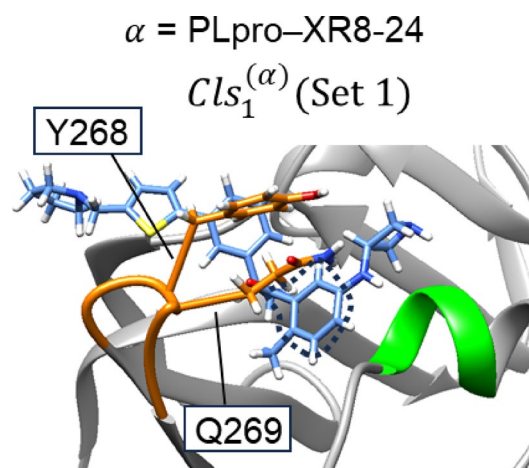


Fig. 11. Complex structure picked from each cluster obtained from clustering parameter Set 1 of $\alpha = \text{PLpro-XR8-24}$ system (Table 3). This figure is respectively for clusters $Cl_1^{(\alpha)}$ (Set 1). The orange segment is the loop region ($G_A^{(RC1)}$; residues 266–271). The green one is a part of helix ($G_B^{(RC1)}$; residues 165–168); Fig. 1b shows details for these segments. Side-chains of two amino-acid residues Y268 and Q269 are labeled. Cornflower blue molecules are compounds, with broken-line circles representing their core regions (Fig. S5 of SI).

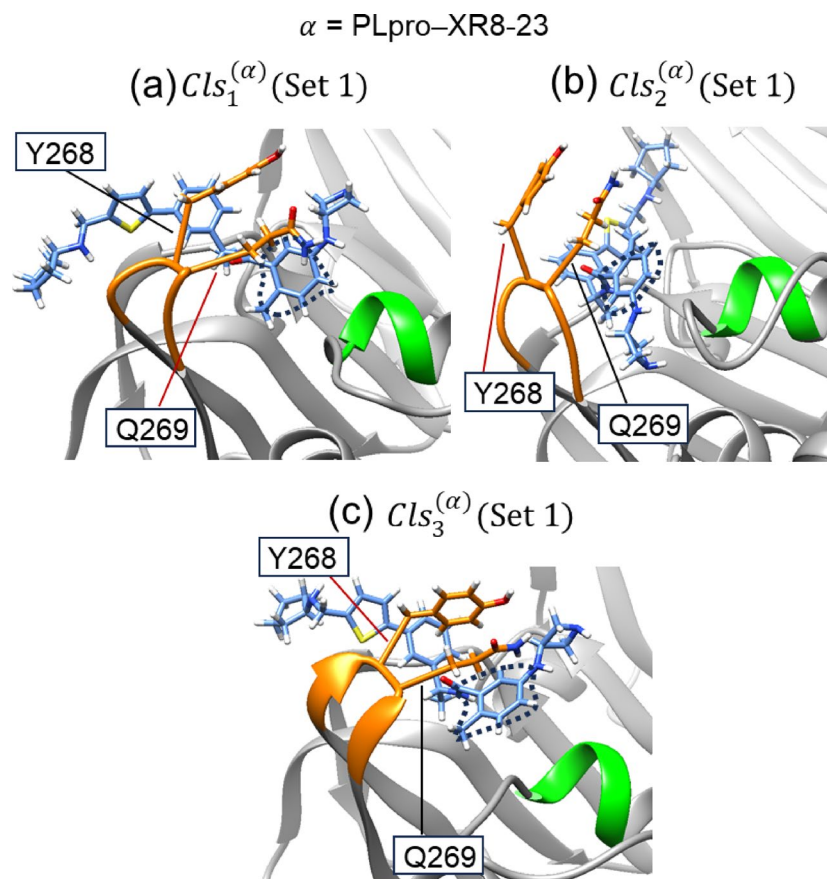


Fig. 12. Complex structure picked from each cluster obtained from clustering parameter Set 1 of $\alpha = \text{PLpro-XR8-23}$ system (Table 3). Panels (a–c) are respectively for clusters $Cl_s_1^{(\alpha)}$ (Set 1), $Cl_s_2^{(\alpha)}$ (Set 1) and $Cl_s_3^{(\alpha)}$ (Set 1). The orange segment is the loop region ($G_A^{(RC1)}$; residues 266–271). The green one is a part of helix ($G_B^{(RC1)}$; residues 165–168); Fig. 1b shows details for these segments. Side-chains of two amino-acid residues Y268 and Q269 are labeled. Cornflower blue molecules are compounds, with broken-line circles representing their core regions (Fig. S5 of SI).

displacement among conformations) for the entire PLpro and for the binding site residues. The resulting RMSD values ranged between 2 and 3 Å in almost all calculations, regardless of the presence or absence of Zn ions, using the experimental structure (PDB ID: 7CJM) as the reference. Importantly, the experimental binding pose was successfully reproduced, likely because the Zn ions are located far from the ligand-binding site and the overall PLpro structure remained unchanged. The structural integrity of PLpro during the simulation was maintained by setting an upper limit for reaction coordinate RC1 (25.49 Å; see Table S4), which was introduced to effectively enhance ligand–receptor binding. We presume that this upper limit helped prevent structural collapse. Therefore, we conducted detailed mD-VcMD simulations excluding zinc ions in this study and confirmed that the PLpro structure was preserved across all systems. Notably, the binding site of PLpro underwent substantial movement due to the enhancement of RC1 and RC2, without structural collapse, and the conformational space was extensively sampled, as shown in Fig. 2.

Our simulation results are consistent with the experimental results reported by Shen et al.³³, as shown in Tables 1, 2 and 6–8, confirming the validity of our computational method.

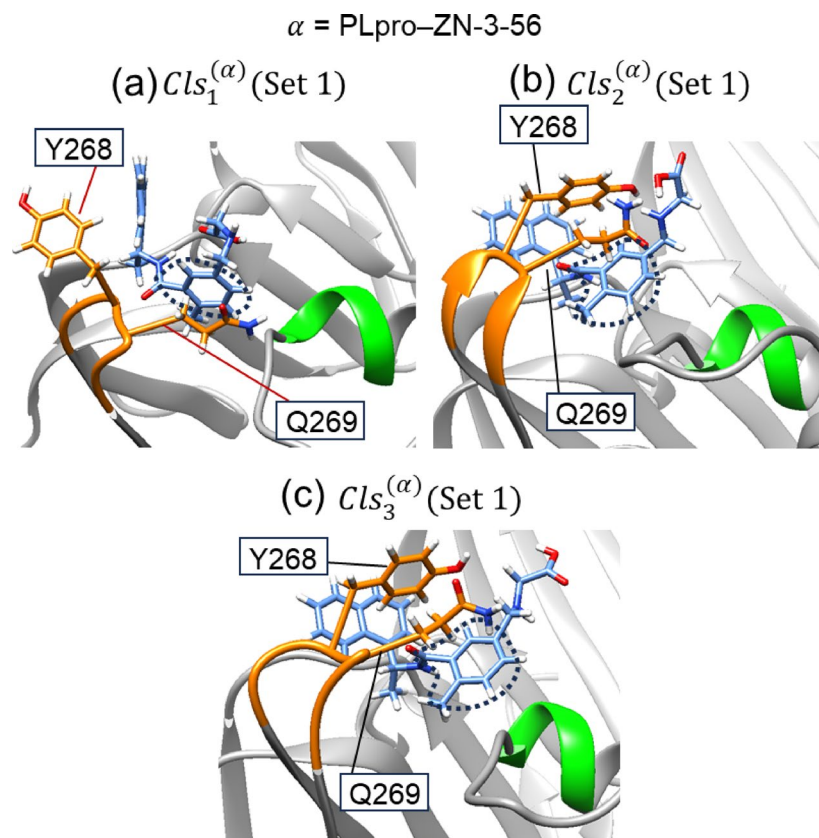


Fig. 13. Complex structure picked from each cluster obtained from clustering parameter Set 1 of $\alpha = \text{PLpro-ZN-3-56}$ system (Table 3). Panels (a–c) are respectively for clusters $Cl_s_1^{(\alpha)}$ (Set 1), $Cl_s_2^{(\alpha)}$ (Set 1) and $Cl_s_3^{(\alpha)}$ (Set 1). The orange segment is the loop region ($G_A^{(RC1)}$; residues 266–271). The green one is a part of helix ($G_B^{(RC1)}$; residues 165–168); Fig. 1b shows details for these segments. Side-chains of two amino-acid residues Y268 and Q269 are labeled. Cornflower blue molecules are compounds, with broken-line circles representing their core regions (Fig. S5 of SI).

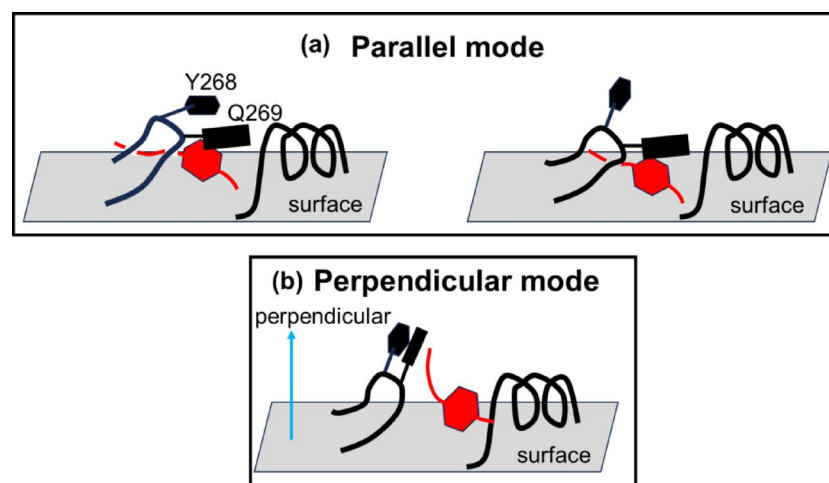


Fig. 14. Schematical representation of two binding modes: (a) parallel mode and (b) perpendicular mode. The red model is compound, with a hexagon as its core region. The blue vector with the label “perpendicular” in panel (b) shows that the side chains of Y268 and Q269 are standing on the PLpro surface. The main text explains more details related to the modes.

System (α)	Cluster (m)
$\alpha = \text{PLpro-GRL0617}$	$Cls_1^{(\alpha)}$ (Set 1), $Cls_2^{(\alpha)}$ (Set 1), $Cls_1^{(\alpha)}$ (Set 2), $Cls_2^{(\alpha)}$ (Set 2), $Cls_1^{(\alpha)}$ (Set 3)
$\alpha = \text{PLpro-XR8-24}$	$Cls_1^{(\alpha)}$ (Set 1), $Cls_2^{(\alpha)}$ (Set 2), $Cls_2^{(\alpha)}$ (Set 2), $Cls_1^{(\alpha)}$ (Set 3)
$\alpha = \text{PLpro-XR8-23}$	$Cls_1^{(\alpha)}$ (Set 1), $Cls_3^{(\alpha)}$ (Set 1), $Cls_2^{(\alpha)}$ (Set 2)
$\alpha = \text{PLpro-ZN-3-56}$	$Cls_1^{(\alpha)}$ (Set 1), $Cls_2^{(\alpha)}$ (Set 1), $Cls_2^{(\alpha)}$ (Set 1), $Cls_1^{(\alpha)}$ (Set 2), $Cls_1^{(\alpha)}$ (Set 3), $Cls_2^{(\alpha)}$ (Set 3)

Table 4. Assignment of clusters $Cls_m^{(\alpha)}$ (Set i) to parallel mode.

System (α)	Cluster (m)
$\alpha = \text{PLpro-GRL0617}$	$Cls_2^{(\alpha)}$ (Set 1), $Cls_3^{(\alpha)}$ (Set 2), $Cls_2^{(\alpha)}$ (Set 3)
$\alpha = \text{PLpro-XR8-24}$	No cluster detected
$\alpha = \text{PLpro-XR8-23}$	$Cls_2^{(\alpha)}$ (Set 1), $Cls_1^{(\alpha)}$ (Set 2), $Cls_2^{(\alpha)}$ (Set 3)
$\alpha = \text{PLpro-ZN-3-56}$	No cluster detected

Table 5. Assignment of clusters $Cls_m^{(\alpha)}$ (Set i) to perpendicular mode.

System (α)	$Q_{inhib}^{(\alpha)}$			IC_{50} (μM)
	Set 1	Set 2	Set 3	
PLpro-GRL0617	0.349	0.311	0.536	1.61
PLpro-XR8-24	0.303	0.215	0.335	0.56
PLpro-XR8-23	0.233	0.171	0.282	0.39
PLpro-ZN-3-56	0.19	0.113	0.200	3.9

Table 6. $Q_{inhib}^{(\alpha)}$ For each system and enzyme Inhibition IC_{50} For each compound.

	Set 1	Set 2	Set 3
Correlation	-0.50	-0.44	-0.35

Table 7. Correlation coefficient between $Q_{inhib}^{(\alpha)}$ and enzyme Inhibition IC_{50} .

	Set 1	Set 2	Set 3
Correlation	-0.69	-0.65	-0.57

Table 8. Correlation coefficient between $Q_{inhib}^{(\alpha)}$ and enzyme Inhibition K_d .

Data availability

All data generated or examined throughout this study are included in this published article and its Supplementary Information.

Received: 26 July 2024; Accepted: 27 December 2025

Published online: 11 January 2026

References

1. Bekker, G. J., Fukuda, I., Higo, J. & Kamiya, N. Cryptic-site binding mechanism of medium-sized Bcl-xL inhibiting compounds elucidated by McMD-based dynamic Docking simulations. *Sci. Rep.* **11**, 5046. <https://doi.org/10.1038/s41598-021-84488-z> (2021).

2. Bekker, G. J., Fukuda, I., Higo, J. & Kamiya, N. Mutual population-shift driven antibody-peptide binding elucidated by molecular dynamics simulations. *Sci. Rep.* **10**, 1406. <https://doi.org/10.1038/s41598-020-58320-z> (2020).
3. Fukunishi, Y., Higo, J. & Kasahara, K. Computer simulation of molecular recognition in biomolecular system: from in-silico screening to generalized ensembles. *Biophys. Rev.* **14**, 1423–1447. <https://doi.org/10.1007/s12551-022-01015-8> (2022).
4. Hayami, T. et al. Difference of binding modes among three ligands to a receptor mSin3B corresponding to their inhibitory activities. *Sci. Rep.* **11**, 6178. <https://doi.org/10.1038/s41598-021-85612-9> (2021).
5. Moritsugu, K. et al. Flexibility and cell permeability of Cyclic Ras-Inhibitor peptides revealed by the coupled Nosé–Hoover equation. *J. Chem. Inf. Model.* **61**, 1921–1930. <https://doi.org/10.1021/acs.jcim.0c01427> (2021).
6. Higo, J., Takashima, H., Fukunishi, Y. & Yoshimori, A. Generalized-ensemble method study: A helix-mimetic compound inhibits protein–protein interaction by long-range and short-range intermolecular interactions. *J. Comput. Chem.* **42**, 956–969. <https://doi.org/10.1002/jcc.26516> (2021).
7. Kasahara, K. et al. myPresto/omegagene 2020: a molecular dynamics simulation engine for virtual-system coupled sampling. *Biophys. Physicobiol.* **17**, 140–146. <https://doi.org/10.2142/biophysico.BSJ-2020013> (2020).
8. Higo, J. et al. GA-guided mD-VcMD: A genetic-algorithm-guided method for multi-dimensional virtual-system coupled molecular dynamics. *Biophys. Physicobiology.* **17**, 161–176. <https://doi.org/10.2142/biophysico.BSJ-2020008> (2020).
9. Higo, J. et al. Molecular interaction mechanism of a 14-3-3 protein with a phosphorylated peptide elucidated by enhanced conformational sampling. *J. Chem. Inf. Model.* **60**, 4867–4880. <https://doi.org/10.1021/acs.jcim.0c00551> (2020).
10. Iida, S. et al. Multimodal structural distribution of the p53 C-terminal domain upon binding to S100B via a generalised ensemble method: from disorder to Extra-Disorder. *J. Chem. Theory Comput.* **15**, 2597–2607. <https://doi.org/10.1021/acs.jctc.8b01042> (2019).
11. Shimato, T., Kasahara, K., Higo, J. & Takahashi, T. Effects of number of parallel runs and frequency of bias-strength replacement in generalized ensemble molecular dynamics simulations. *PeerJ Phys. Chem.* **1**, e4. <https://doi.org/10.7717/peerj-pchem.4> (2019).
12. Hayami, T., Higo, J., Nakamura, H. & Kasahara, K. Multidimensional virtual-system coupled canonical molecular dynamics to compute free-energy landscapes of peptide multimer assembly. *J. Comput. Chem.* **40**, 2453–2463. <https://doi.org/10.1002/jcc.26020> (2019).
13. Kasahara, K., Terazawa, H., Takahashi, T. & Higo, J. Studies on molecular dynamics of intrinsically disordered proteins and their fuzzy complexes: a mini-review. *Comput. Struct. Biotech. J.* **17**, 712–720. <https://doi.org/10.1016/j.csbj.2019.06.009> (2019).
14. Iwai, R., Kasahara, K. & Takahashi, T. Influence of various parameters in the replica-exchange molecular dynamics method: number of replicas, replica-exchange frequency, and thermostat coupling time constant. *Biophys. Physicobiol.* **15**, 165–172. https://doi.org/10.2142/biophysico.15.0_165 (2018).
15. Hayami, T., Kasahara, K., Nakamura, H. & Higo, J. Molecular dynamics coupled with a virtual system for effective conformational sampling. *J. Comput. Chem.* **39**, 1291–1299. <https://doi.org/10.1002/jcc.25196> (2018).
16. Kasahara, K. et al. Phosphorylation of an intrinsically disordered region of Ets1 shifts a multi-modal interaction ensemble to an out-inhibitory state. *Nucleic Acids Res.* **46**, 2243–2251. <https://doi.org/10.1093/nar/gkx1297>
17. Higo, J., Kasahara, K. & Nakamura, H. Multi-dimensional virtual system introduced to enhance canonical sampling. *J. Chem. Phys.* **147**, 134102. <https://doi.org/10.1063/1.4986129> (2017).
18. Higo, J., Kasahara, K., Dasgupta, B. & Nakamura, H. Enhancement of canonical sampling by virtual-state transitions. *J. Chem. Phys.* **146**, 044104. <https://doi.org/10.1063/1.4974087> (2017).
19. Kasahara, K. et al. myPresto/omegagene: a GPU-accelerated molecular dynamics simulator tailored for enhanced conformational sampling methods with a non-Ewald electrostatic scheme. *Biophys. Physicobiology.* **13**, 209–216. https://doi.org/10.2142/biophysico.13.0_209 (2016).
20. Ikebe, J. et al. Theory for trivial trajectory parallelization of multicanonical molecular dynamics and application to a polypeptide in water. *J. Comput. Chem.* **32**, 1286–1297. <https://doi.org/10.1002/jcc.21710> (2011).
21. Higo, J. et al. Fly casting with ligand sliding and orientational selection supporting complex formation of a GPCR and a middle sized flexible molecule. *Sci. Rep.* **12**, 13792. <https://doi.org/10.1038/s41598-022-17920-7> (2022).
22. Kurita, J. et al. Sertraline, chlorprothixene, and chlorpromazine characteristically interact with the REST-binding site of the corepressor mSin3, showing Medulloblastoma cell growth inhibitory activities. *Sci. Rep.* **8**, 13763. <https://doi.org/10.1038/s41598-018-31852-1> (2018).
23. Shin, D. et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587**, 657–662. <https://doi.org/10.1038/s41586-020-2601-5> (2020).
24. Narayanan, A. et al. Identification of SARS-CoV-2 inhibitors targeting Mpro and PLpro using in-cell-protease assay. *Commun. Biol.* **5**, 169. <https://doi.org/10.1038/s42003-022-03090-9> (2022).
25. Zhao, Y. et al. High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors. *Protein Cell.* **12**, 877–888. <https://doi.org/10.1007/s13238-021-00836-9> (2021).
26. Calleja, D. J., Lessene, G. & Komander, D. Inhibitors of SARS-CoV-2 PLpro. *Front. Chem.* **10**, 876212. <https://doi.org/10.3389/fchem.2022.876212> (2022).
27. Zhang, L. et al. Design of SARS-CoV-2 Mpro, PLpro dual-target inhibitors based on deep reinforcement learning and virtual screening. *Future Med. Chem.* **14**, 393–405. <https://doi.org/10.4155/fmc-2021-0269> (2022).
28. Negru, P. A. et al. Virtual screening of substances used in the treatment of SARS-CoV-2 infection and analysis of compounds with known action on structurally similar proteins from other viruses. *Biomed. Pharmacother.* **153**, 113432. <https://doi.org/10.1016/j.biopha.2022.113432> (2022).
29. Kidera, A., Moritsugu, K., Ekimoto, T. & Ikeguchi, M. Allosteric regulation of 3CL protease of SARS-CoV-2 and SARS-CoV observed in the crystal structure ensemble. *J. Mol. Biol.* **443**, 167324. <https://doi.org/10.1016/j.jmb.2021.167324> (2021).
30. Moritsugu, K., Ekimoto, T., Ikeguchi, M. & Kidera, A. Binding and unbinding pathways of peptide substrates on the SARS-CoV-2 3CL protease. *J. Chem. Inf. Model.* **63**, 240–250. <https://doi.org/10.1021/acs.jcim.2c00946> (2023).
31. Hengphasatporn, K. et al. Promising SARS-CoV-2 main protease inhibitor ligand-binding modes evaluated using LB-PaCS-MD/FMO. *Sci. Rep.* **12**, 17984. <https://doi.org/10.1038/s41598-022-22703-1> (2022).
32. Sencanski, M. et al. Identification of SARS-CoV-2 Papain-like protease (PLpro) inhibitors using combined computational approach. *ChemistryOpen* **11**, e202100248. <https://doi.org/10.1002/open.202100248> (2022).
33. Shen, Z. et al. Design of SARS-CoV-2 PLpro inhibitors for COVID-19 antiviral therapy leveraging binding cooperativity. *J. Med. Chem.* **65**, 2940–2955. <https://doi.org/10.1021/acs.jmedchem.1c01307> (2022).
34. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
35. Ester, M., Kriegel, H. P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)*. 226–231. (AAAI, 1996).
36. Fu, Z. et al. The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nat. Commun.* **12**, 488. <https://doi.org/10.1038/s41467-020-20718-8> (2021).
37. Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the amber ff99SB protein force field. *Proteins: Struct. Funct. Genet.* **78**, 1950–1958. <https://doi.org/10.1002/prot.22711> (2010).
38. Joung, I. S. & Cheatham, T. E. III Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B.* **112**, 9020–9041. <https://doi.org/10.1021/jp8001614> (2008).
39. Jorgensen, W. L. et al. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–936. <https://doi.org/10.1063/1.445869> (1983).

40. Frisch, M. J. et al. Gaussian 16, Revision C.01 Gaussian, Inc., <https://gaussian.com/gaussian16/> (2016).
41. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280. <https://doi.org/10.1021/j100142a004> (1993).
42. Wang, J. M. et al. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174. <https://doi.org/10.1002/jcc.20035> (2004).
43. Higo, J. et al. Free-energy landscape of molecular interactions between endothelin 1 and human endothelin type B receptor: Fly-casting mechanism. *Protein Eng. Des. Sel.* **32**, 297–308. <https://doi.org/10.1093/protein/gzz029> (2019).
44. Higo, J. et al. Binding free-energy landscapes of small molecule binder and non-binder to FMN riboswitch: All-atom molecular dynamics. *Biophys. Physicobiology.* **20**, e200047. <https://doi.org/10.2142/biophysico.bppb-v20.0047> (2023).
45. Higo, J. et al. Verifying trivial parallelization of multicanonical molecular dynamics for conformational sampling of a polypeptide in explicit water. *Chem. Phys. Lett.* **473**, 326–329. <https://doi.org/10.1016/j.cplett.2009.03.077> (2009).
46. Kutzner, C. et al. More Bang for your buck: improved use of GPU nodes for GROMACS 2018. *J. Comput. Chem.* **40**, 2418–2431. <https://doi.org/10.1002/jcc.26011> (2019).
47. Bonomi, M. et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comp. Phys. Commun.* **180**, 1961–1972. <https://doi.org/10.1016/j.cpc.2009.05.011> (2009).
48. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
49. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519. <https://doi.org/10.1063/1.447334> (1984).
50. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A.* **31**, 1695–1697. <https://doi.org/10.1103/PhysRevA.31.1695> (1985).
51. Kamiya, N., Fukuda, I. & Nakamura, H. Application of zero-dipole summation method to molecular dynamics simulations of a membrane protein system. *Chem. Phys. Lett.* **568–569**, 26–32. <https://doi.org/10.1016/j.cplett.2013.03.014> (2013).
52. Fukuda, I., Yonezawa, Y. & Nakamura, H. Molecular dynamics scheme for precise Estimation of electrostatic interaction via zero-dipole summation principle. *J. Chem. Phys.* **134**, 164107. <https://doi.org/10.1063/1.3582791> (2011).
53. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* **23**, 187–199. [https://doi.org/10.1016/0021-9991\(77\)90121-8](https://doi.org/10.1016/0021-9991(77)90121-8) (1977).
54. Pearson, K. On lines and planes of closest fit to systems of points in space. *Phil. Mag.* **2**, 559–572. <https://doi.org/10.1080/14786440109462720> (1901).
55. Higo, J., Galzitskaya, O. V., Ono, S. & Nakamura, H. Energy landscape of a β -hairpin peptide in explicit water studied by multicanonical molecular dynamics. *Chem. Phys. Lett.* **337**, 169–175. [https://doi.org/10.1016/S0009-2614\(01\)00118-X](https://doi.org/10.1016/S0009-2614(01)00118-X) (2001).
56. Higo, J. et al. Energy landscape of a peptide consisting of α -helix, 3_{10} -helix, β -turn, β -hairpin, and other disordered conformations. *Protein Sci.* **10**, 1160–1171. <https://doi.org/10.1110/ps.44901> (2008).
57. Xie, Q., Kasahara, K., Higo, J. & Takahashi, T. Molecular mechanisms of functional modulation of transcriptional coactivator PCA4 via phosphorylation on its intrinsically disordered region. *ACS Omega.* **8**, 14572–14582. <https://doi.org/10.1021/acsomega.3c00364> (2023).
58. Li, M. et al. A novel Traveling-Wave-Based method improved by unsupervised learning for fault location of power cables via sheath current monitoring. *Sensors* **19**, 2083. <https://doi.org/10.3390/s19092083> (2019).
59. Sonnewald, M., Dutkiewicz, S., Hill, C. & Forget, G. Elucidating ecological complexity: unsupervised learning determines global marine eco-provinces. *Sci. Adv.* **6**, eaay4740. <https://doi.org/10.1126/sciadv.aay4740> (2020).
60. Yigin, B. O. & Saygili, G. Effect of distance measures on confidences of t-SNE embeddings and its implications on clustering for scRNA-seq data. *Sci. Rep.* **13**, 6567. <https://doi.org/10.1038/s41598-023-32966-x> (2023).
61. Ratia, K. et al. Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. *Proc. Natl. Acad. Sci. U S A.* **103**, 5717–5722. <https://doi.org/10.1073/pnas.0510851103> (2006).
62. Barretto, N. et al. The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *J. Virol.* **79**, 15189–15198. <https://doi.org/10.1128/jvi.79.24.15189-15198.2005> (2005).
63. Arya, R., Prashar, V. & Kumar, M. Evaluating stability and activity of SARS-CoV-2 PLpro for high-throughput screening of inhibitors. *Mol. Biotechnol.* **64**, 1–8. <https://doi.org/10.1007/s12033-021-00383-y> (2022).

Acknowledgements

This work was supported by JSPS KAKENHI Grants No. 21K06052 (J.H.) and was conducted in part under the Cooperative Research Program of the Institute for Protein Research, Osaka University, CR22-02 and CR-23-02. We acknowledge support by the HPCI System Research Project (Project IDs: hp220002, hp220015, hp220022, hp230003, hp230011, hp230013, hp230109, hp240032, hp240167, hp250017, and hp250061). J.H. is supported by the Project Focused on Developing Key Technology for Discovering and Manufacturing Drugs for Next-Generation Treatment and Diagnosis (project ID: 23ae0121028h0003; 2018–2021 and 2021–) from AMED and the Japan Biological Informatics Consortium (JBIC). The mD-VcMD was performed on TSUBAME4.0 supercomputers at the Institute of Science Tokyo. Other analyses were done at Ritsumeikan University and Research Center for Computational Science, Okazaki, Japan (Project: 25-IMS-C175). We used UCSF Chimera ver. 1.15 for drawing molecular structures and gnuplot for graphs. Other analyses were done using programs produced in-house.

Author contributions

Research project was proposed by Masashi Muramoto and Takuya Takahashi. Computer program preparation was done by Masashi Muramoto, Qilin Xie and Kota Kasahara; MD simulation was performed by Masashi Muramoto. Analyses were performed by Muramoto Masashi, Simon Hikiri, Suzuka Saito and Junichi Higo. Manuscript writing was done by Masashi Muramoto, Simon Hikiri, Suzuka Saito, Junichi Higo and Takuya Takahashi. The entire process of this research was supervised by Takuya Takahashi.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-34319-2>.

Correspondence and requests for materials should be addressed to T.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026