The Conformational Space of the SARS-CoV-2 Main Protease Active Site Loops is Determined by Ligand Binding and Interprotomer Allostery

Ethan Lee[†,‡] and Sarah Rauscher[*,†,‡,¶]

†*Department of Chemical and Physical Sciences, University of Toronto Mississauga*

‡*Department of Chemistry, University of Toronto*

¶*Department of Physics, University of Toronto*

E-mail: sarah.rauscher@utoronto.ca

Abstract

The main protease (M^{pro}) of SARS-CoV-2 is essential for viral replication and is, therefore, an important drug target. Here, we investigate two flexible loops in M^{pro} that play a role in catalysis. Using all-atom molecular dynamics simulations, we analyze the structural ensemble of M^{pro} in an apo state and substrate-bound state. We find that the flexible loops can adopt open, intermediate (partly open) and closed conformations in solution, which differs from the partially closed state observed in crystal structures of M^{pro}. When the loops are in closed or intermediate states, the catalytic residues are more likely to be in close proximity, which is crucial for catalysis. Additionally, we find that substrate binding to one protomer of the homodimer increases the frequency of intermediate states in the bound protomer, while also affecting the structural propensity of the apo protomer's flexible loops. Using dynamic network analysis, we identify multiple allosteric pathways connecting the two active sites of the homodimer. Common to these pathways is an allosteric hotspot involving the N-terminus, a critical

region that comprises part of the binding pocket. Taken together, the results of our simulation study provide detailed insight into the relationships between the flexible loops and substrate binding in a prime drug target for COVID-19.

1 Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a betacoronavirus from the *Coronaviridae* family, of which SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) are also members.[1] SARS-CoV-2 spreads through human-to-human transmission and is responsible for the COVID-19 pandemic. During viral replication, two-thirds of the genome is translated into two polyproteins containing 16 nonstructural proteins (nsps).[2] These polyproteins are cleaved by the papain-like protease and the main protease (M^{pro}). M^{pro} cleaves the polyproteins at 11 sites, including two sites liberating itself.[3,4] Dimerization of M^{pro} is essential for its catalytic activity,[5,6] as the N-terminal region of one monomer forms part of the substrate binding site of the other monomer.[7] Each M^{pro} monomer is 306 residues in length and comprises three domains.[7,8] Domains I and II are six-stranded antiparallel β-barrels that enclose the active site between them, while Domain III contains five helices and plays a role in dimerization.[7,8] The catalytic Cys-His dyad located in the active site of M^{pro} cleaves the polyprotein substrate. The first step in the reaction mechanism involves His_{41} accepting a proton from Cys_{145}, which results in a charged thiolate and is followed by a nucleophilic attack on the carbonyl carbon atom of the peptide bond in the substrate by the $S\gamma$ atom of Cys_{145}.[9]

M^{pro} is a prime target for inhibition due to its importance in viral replication.[7,10] There are no human proteases that target similar cleavage sequences to M^{pro}, which suggests that inhibitors would have a lower potential for toxicity.[7] To date, over 100 inhibitors of M^{pro} have been identified,[10] and the Protein Data Bank (PDB) contains over 1200 structures of M^{pro}. Furthermore, the FDA has approved a ritonavir-boosted nirmatrelvir protease inhibitor known as Paxlovid.[11] Continued investigation of M^{pro} is crucial not only for enhancing

the effectiveness and specificity of antiviral drugs, but also because several SARS-CoV-2 variants of concern have mutations in M^{pro},[12] and mutations in M^{pro} can confer resistance to nirmatrelvir.[13] Furthermore, a detailed understanding of the structure and function of M^{pro} will strengthen global preparedness against potential future pandemics where related viral proteases may be targets.

Most inhibitors of M^{pro} bind in or near its active site,[10] which is surrounded by two mobile loops.[14,15] Flexibility of the active site and the surrounding loops is functionally important because it allows for the accommodation of diverse substrates.[4] Other viral proteases and enzymes have mobile active site loops that play critical roles in their function and dynamics.[16] For instance, a study by Jupin et al. reported the switch action of an active site loop in an ovarian tumour-like protease/deubiquitinase. This loop samples either an open, intermediate, or closed state. These loop states are related to the protein's function: the state of the loop determines whether the protein acts as a protease or a deubiquitinase.[17] Similarly, studies on HIV-1 protease showed two symmetric active site loops that populate open, intermediate, and closed states depending on the occupancy of the active site, as revealed by nuclear magnetic resonance (NMR) experiments and molecular dynamics (MD) simulations.[18,19]

MD simulations are a powerful tool for studying viral proteases, as well as mobile loops in enzymes more generally, providing complementary information to crystallographic and biophysical studies.[16] Simulation studies of M^{pro} have advanced our understanding of the structure and dynamics of this enzyme in order to provide useful information for the development of viable inhibitors.[15,20–32] The simulation study carried out in this work builds on the growing body of research into the functional motions of M^{pro}. Specifically, we aim to describe specific conformational states of the active site in both protomers of the homodimer and quantify the extent to which information is transmitted between the active sites through interprotomer allostery.

Here, we focus on M^{pro}'s flexible active site loops, presenting a combined 92.5 µs of

all-atom simulation data on M^{pro} in different conditions and from multiple coronaviruses. With this extensive simulation data set, we demonstrate that the active site loops sample open, closed, and intermediate states, and the frequency of these states depends on the occupancy of the active site. We find evidence for interprotomer allostery: the frequency of the loop states in one protomer depends on the bound state of the opposite protomer. To further explore this allostery, we used dynamic network analysis to identify residues involved in interprotomer communication. Additionally, we find that the isolated monomer has a higher frequency of open states compared to the dimer, which may explain its weaker ligand binding.[6,26] Finally, we compare M^{pro} of SARS-CoV-2, SARS-CoV, and MERS-CoV and find similarities between the three proteases in the dynamics of the flexible loops and the allosteric paths connecting the two active sites.

2 Results and Discussion

2.1 Assessment of Sampling, Equilibration, and Accuracy

In this study, we conducted MD simulations of the M^{pro} dimer with either one, two, or zero (apo) substrates bound. The complete list of all simulations performed is provided in Table 1. We define the bound systems that were simulated as follows: an M^{pro} dimer with two substrates bound (one per active site) is referred to as the *two-substrate system*. A complex with one substrate bound (where the dimer has one bound and one apo protomer) is referred to as the *one-substrate system*. Additionally, simulations were performed on M^{pro} as a monomer, with and without substrate, as well as on apo SARS-CoV and MERS-CoV M^{pro} dimers. To mimic the polyprotein cleavage site between nsp4 and the N-terminus of M^{pro}, we used the same pentapeptide substrate as Suárez and Díaz.[23] This substrate contains a –P4–P3–P2–P1–P1'- sequence of -Ala–Val–Leu–Gln–Ser–, where cleavage would occur at the P1–P1' site (Figure 1B and C, see SI Section 1.1 for more details).[23] Equilibrated sections of the trajectories were determined using four metrics: root-mean-square deviation (RMSD),

radius of gyration (R_{gyr}), number of hydrogen bonds, and solvent accessible surface area (SASA). Convergence of these four metrics occurred at ∼0.5 µs (Figure 1D). Only the data after t = 0.5 µs were included in further analysis (see SI Section 1.2).

Table 1: **Summary of M^{pro} MD simulations.** The setup of substrate-bound systems is described in SI Section 1.1.

Virus	Oligomeric State	Trajectory Length (µs)	Replicates	Bound State
SARS-CoV-2	Dimer	1.5	5	Apo
SARS-CoV-2	Dimer	4.0	5	2 Substrates
SARS-CoV-2	Dimer	4.0	5	1 Substrate
SARS-CoV-2	Monomer	3.0	5	Apo
SARS-CoV-2	Monomer	3.0	5	Substrate
SARS-CoV	Dimer	1.5	5	Apo
MERS-CoV	Dimer	1.5	5	Apo

To evaluate the accuracy of our simulations, we computed residual dipolar couplings (RDCs) using the M^{pro} apo dimer simulations. We then compared the computed RDCs to those obtained using NMR spectroscopy by Robertson et al.[34] (Figure 1E and F). We observe relative agreement between the two, as shown through a mean Q-factor of 0.398 ± 0.007 comparing $^1D_{NH}$ and $^2D_{C'H}$ RDC values. To put this comparison into context, we note that Robertson et al. report Q-factors of 0.2-0.53 between their measured RDCs and RDCs computed using X-ray crystal structures of M^{pro}.[34] Q-factors of 0.2 to 0.7 have been reported when comparing simulated ensembles using multiple force fields and experimental RDCs in a study evaluating the accuracy of MD simulations.[35] Differences between experimental RDCs and RDCs predicted from the simulation ensembles may stem from using a C145A active site mutant and an alignment medium in the NMR experiments,[34] as well as limitations in simulation accuracy and conformational sampling. Based on this comparison to the available NMR data for M^{pro}, as well as additional qualitative comparisons to other structural and biophysical studies described below, we proceed with the analysis of the conformational ensembles of M^{pro} in different conditions.

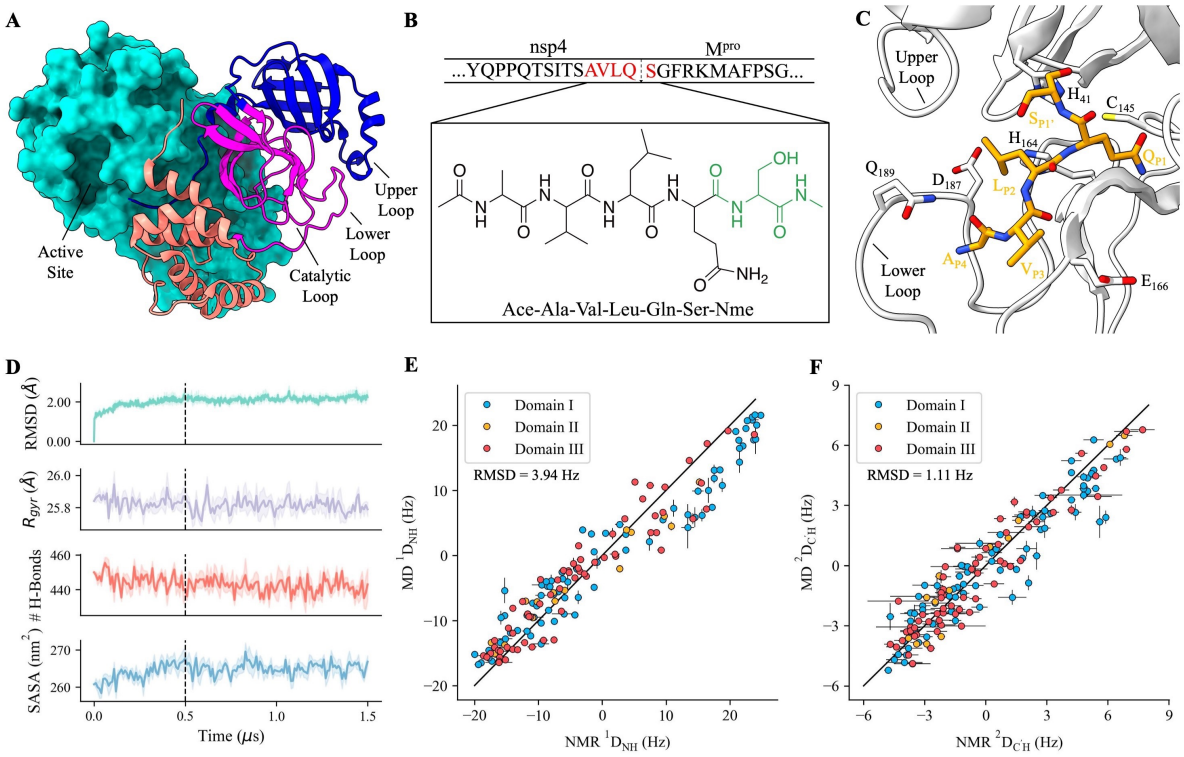

Figure 1: **Simulation setup and validation.** (A) Surface area and ribbon models show two M^{pro} protomers (PDB ID: 6XHU)[33] The ribbon model is split into three domains: Domain I (blue, residues 8-101), Domain II (magenta, 102-184), and Domain III (orange, 201-303). The linker connecting Domains II and III (185-200) is also coloured magenta. We define three loops: the upper loop (residues 44-54), lower loop (184-194), and a loop in the active site that we refer to as the catalytic loop (162-174). (B) Pentapeptide substrate mimetic used in all simulations of the bound systems. The upper row shows the –P4–P3–P2–P1–P1' sequence of -Ala-Val-Leu-Gln-Ser- in red, where cleavage occurs at the P1–P1' site (arrow) between the sequence of nsp4 and M^{pro}.[23] The black portion of the peptide was taken from the N3 inhibitor in PDB 6LU7,[8] while the green portion was constructed along the inhibitor backbone by Suárez and Díaz.[23] (C) The substrate mimetic in the active site pocket. The initial structure used in simulations, obtained from Suárez and Díaz.,[23] is shown. The substrate is coloured in orange, with nitrogen atoms and oxygen atoms coloured blue and red, respectively. M^{pro} side chains coloured in white are those that interact directly with the substrate. Acetyl and N-methyl groups capping the N- and C-termini are omitted for clarity. (D) Metrics used to determine equilibration in the apo dimer simulations. Based on root mean square deviation (RMSD), radius of gyration (R_{gyr}), number of hydrogen bonds, and solvent accessible surface area (SASA), the first 0.5 µs was excluded as equilibration. The dashed line delineates the equilibration from the part of the trajectories used for analysis. Shaded regions denote the standard error of the mean from five independent trajectories. (E), (F) Comparison of mean RDCs computed from apo dimer simulations and experimental values.[34] Points are coloured by the domain to which each residue belongs. The root-mean-square deviation (RMSD) between experimental values and RDCs computed from simulation data are shown on each plot. The Pearson correlation, r, for $^1D_{NH}$ and $^2D_{C'H}$ RDC values is $r = 0.964 \pm 0.006$ and 0.941 ± 0.009, respectively.

2.2 The Flexibility and States of the Two Active Site Loops

Surrounding the active site of each protomer are two loops, the upper loop (residues 44-54) and the lower loop (residues 184-194), with loop boundaries defined consistently with previous studies.[23,25] Mutations in both loops occur in SARS-CoV-2 variants of concern.[12] These mutations increase (L50F and E47K) or decrease (E47N) the enzyme's activity due to changes in charge distribution and the orientations of side chains in the substrate binding site.[12] Mutations in both the upper (L50F) and lower loops (V186V, A193P) confer resistance to nirmatrelvir,[13] suggesting the importance of these loops to the function of M^{pro} and the utility of studying their structure and dynamics in detail. An additional loop of interest due to its interactions with bound ligands comprises residues 162-174; we refer to this loop as the catalytic loop (Figure 1A).

Both the upper and lower loops exhibit high flexibility, as evidenced by their higher root-mean-square fluctuation (RMSF) compared to the rest of the protein (Figure 2C and 3B). This observed flexibility is consistent with other MD simulation studies of M^{pro}[15,25,26] and is likely crucial in accommodating different types of residues in the active site.[4] Increased flexibility in the upper and lower loops is not evident in crystallographic B-factors (Figure S1). However, both loops are significantly more flexible in solution simulations compared to what would be expected based on the B-factors. An analysis of over 200 crystal structures of M^{pro} reveals that each loop adopts a distinct orientation corresponding to a partially closed conformation in all of these structures (with the upper loop in a closed state and the lower loop in an intermediate state, see section "Substrate Binding Favours Intermediate and Closed Loop States" below), regardless of whether an inhibitor is bound (Figure S2). Furthermore, an analysis of crystal contacts in crystal structures of apo M^{pro} (Figure S1B) indicates that both the upper and lower loops have many crystal contacts. These contacts are likely stabilizing more compact conformations of both loops. The crystallographic structures may, therefore, provide a limited description of the flexibility of the upper and lower loops. These observations are comparable to HIV-1 protease, which similarly exhibits closed or

semi-open active site loops in crystal structures and heterogeneous loop conformations in solution.[19]

To learn more about the structure and dynamics of the active site loops, we carried out dimensionality reduction using uniform manifold approximation and projection (UMAP),[36] a general nonlinear reduction technique previously used to study protein folding intermediates and loop motions.[37,38] We used UMAP because of its scalability with large datasets, increased computational efficiency, better preservation of global structure, and ability to process non-linear data compared to other dimensionality reduction methods.[36] UMAP constructs a high-dimensional graph representation of a dataset before optimizing a lower-dimensional graph while conserving as much topological similarity as possible. After dimensionality reduction, we used hierarchical density-based spatial clustering of applications with noise (HDBSCAN), whose main advantage over other algorithms is its ability to detect clusters of varying densities.[39]

Because both the upper and lower loops exhibit high flexibility (Figure 2C), we wanted to obtain a quantitative description of the structurally heterogeneous ensembles of each loop. For this purpose, we used UMAP and HBDSCAN to cluster the structural ensemble of the region surrounding the active site. The UMAP projection of the apo dimer's upper loop yields 13 clusters (Figure 2A), each of which corresponds to a closed or open state of this loop. The clusters are categorized as closed or open based on the distance between the loop tip and residue Pro_9 in the N-terminus of the protein, which is used as a reference point due to its stability in the structure (see methods, section "Dimensionality Reduction and Clustering"). The loop behaves similarly in all open state clusters — it is free to move around in space with high flexibility. The high population of the open state (79 ± 8%) compared to the closed state (21 ± 8%) indicates that it is the more favourable conformation in the apo dimer (Table 2).

Applying UMAP and HDBSCAN on the apo dimer's lower loop reveals 12 clusters corresponding to open, closed, and intermediate states (Figure 2A, lower panel). In the closed

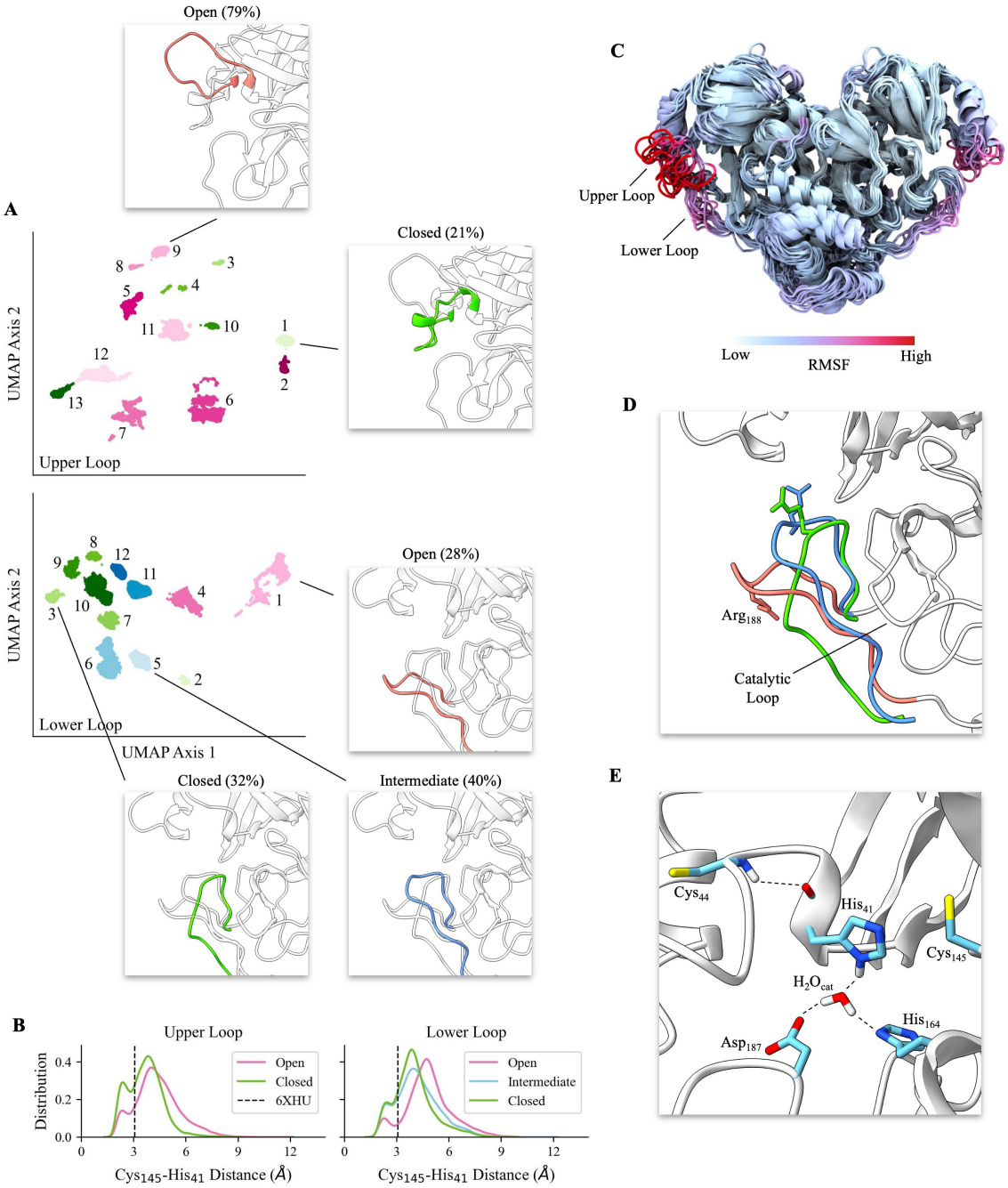

Figure 2: **Flexibility and conformational states of the upper and lower loops in the apo dimer.** (A) UMAP projections and HDBSCAN clusters of the upper and lower loops, shown in the upper and lower panels, respectively. Numbers within the UMAP plots signify each cluster. Clusters are coloured individually. Pink shades denote open states (upper loop: clusters 2, 5, 6, 7, 8, 9, 11, and 12 and lower loop: clusters 1 and 4). Green shades denote closed states (upper loop: clusters 1, 3, 4, 10, and 13 and lower loop: clusters 2, 3, 7, 8, 9, and 10). Blue indicates an intermediate state (lower loop: clusters 5, 6, 11, and 12). Representative structures and the population of each state are provided in insets.

Figure 2: (Continued) (B) Histograms of the catalytic dyad distance (distance between the Cys_{145} thiol hydrogen and the His_{41} imidazole nitrogen) in the different conformational states of the upper (left panel) and lower (right panel) loops. This distance in the crystal structure (PDB 6XHU) is indicated as a dashed line in both panels. (C) Apo dimer coloured by RMSF with a range of 0 (low) to 6 Å (high). Fifteen structures from the simulation are shown (selected from the trajectory at evenly spaced intervals). The data is shown as a line plot in Figure 3B. (D) Representative structures of open, intermediate, and closed lower loop states are coloured in pink, blue, and green, respectively, with the side chain of Arg_{188} shown. (E) View of the active site taken from the apo dimer simulation with both loops in a closed state. Key hydrogen bonds stabilizing the closed state of both loops are shown as dashed lines. Cys_{145} and His_{41} are the catalytic dyad residues, Asp_{187} and His_{41} are catalytic helper residues, and Cys_{44} and His_{41} form a hydrogen bond that orients His_{41} towards Cys_{145}.

Table 2: **Summary of the populations of the upper and lower loop states based on UMAP and clustering with HDBSCAN.** The number in brackets indicates the standard error of the mean for each population (based on n = 5 independent simulations). Sub refers to substrate in the description of each system.

System	**Loop**	**Open %**	**Intermediate %**	**Closed %**
Apo Dimer		79 (8)	-	21 (8)
Apo Protomer from Dimer (1 Sub)		90 (10)	-	10 (10)
Apo Monomer	Upper	100 (0)	-	0 (0)
Bound Dimer (2 Sub)		56 (14)	-	44 (14)
Bound Protomer from Dimer (1 Sub)		62 (15)	-	38 (15)
Apo Dimer		28 (13)	40 (11)	32 (11)
Apo Protomer from Dimer (1 Sub)		6 (6)	44 (14)	50 (16)
Apo Monomer	Lower	75 (11)	25 (11)	0 (0)
Bound Dimer (2 Sub)		13 (8)	87 (8)	0 (0)
Bound Protomer from Dimer (1 Sub)		30 (10)	70 (10)	0 (0)

states, the catalytic loop shifts downwards, enabling the lower loop to curl inward into the active site. This allows for a more compact conformation in solution compared to what is observed in crystal structures. The lower loop shifts further away from the catalytic loop as it moves from closed to intermediate to open (Figure 2D), suggesting that the state of the lower loop may play a role in substrate binding and catalysis.

The flexibility of the two loops observed in our simulations is consistent with the work of Gossen et al., who described the flexibility and conformational changes of M^{pro}'s active site through Markov state model analysis.[14] While they did not identify distinct loop states as we do in the present study, they noted that the volume of the binding site is affected by the flexibility of the loops, suggesting that the movement of loops, and possibly different loop states, may contribute to creating more or less favourable binding conformations. Lee et al. have obtained several crystallographic structures of M^{pro} bound to different substrates, which reveal slight structural differences in the upper and lower loops depending on the bound substrate. These findings provide strong evidence for loop flexibility, which is likely functionally important in accommodating various substrates.[3]

2.3 Catalytic Residues are Stabilized in Closed Loop States

Next, we investigated whether the conformational states of the upper and lower loops correlate with the active site's conformation. The distance between the catalytic dyad residues (Cys_{145} and His_{41}) in the apo dimer is $\leq$ 4.0 Å on average when either loop is in a closed state (Table 3). The Cys_{145}–His_{41} distance is also within error of the closed state distance when the lower loop is in an intermediate state, suggesting some similarity between these two states. In contrast, the Cys_{145}–His_{41} distance rises to an average range of 4.4-4.8 Å when either loop is in an open state, with wider distributions of this distance compared to closed and intermediate states (Figure 2B). This variation in the distance between the catalytic dyad residues likely affects the catalytic efficiency. Additionally, when the upper loop is closed, the side chain of His_{41} and the Cys_{145} thiol become more rigid compared to

when the loop is open, based on an analysis of the RMSF of these residues (Figure S3). This increased rigidity may contribute to the reduced variance and shorter distance between the two catalytic dyad residues. The cause of the increased rigidity may be the closer proximity of the backbone nitrogen of Cys_{44} and the carbonyl oxygen of His_{41} in the closed state (Table 3), where a hydrogen bond is typically present (Figure 2E). The shorter distance may strengthen the hydrogen bond, stabilizing the His_{41} backbone, which in turn stabilizes both its side chain and the interacting Cys_{145} residue. Taken together, these results indicate that Cys_{145} and His_{41} are more likely to be close enough for proton transfer to occur when at least one of the loops is in a closed or intermediate state, suggesting that closed and intermediate states are more favourable for catalysis.

Table 3: **Catalytic site distances in different states of the upper and lower loops.** Values in brackets denote the standard error of the mean for each distance. Residue Cys_{44} is in the upper loop and residue Asp_{187} is in the lower loop. For clarity, only distances that are relevant for comparison between different loop states are shown, and distances that are not relevant for comparison are indicated by dashes.

Loop	**Loop State**	**Cys_{145}-His_{41}**	**Asp_{187}-His_{41}**	**Cys_{44}-His_{41}**
Mean Distance (Å) in the Apo Dimer				
Upper	Open	4.4 (0.1)	-	3.1 (0.3)
	Closed	3.7 (0.2)	-	2.6 (0.2)
Lower	Open	4.8 (0.1)	7.5 (0.8)	-
	Intermediate	4.2 (0.3)	5.1 (0.1)	-
	Closed	4.0 (0.1)	4.90 (0.09)	-
Mean Distance (Å) in the Bound Dimer				
Upper	Open	4.3 (0.2)	-	3.4 (0.3)
	Closed	3.7 (0.2)	-	2.35 (0.01)
Lower	Open	5.6 (0.3)	8.8 (0.4)	-
	Intermediate	3.8 (0.1)	6.9 (0.1)	-
	Closed	-	-	-

We observe additional effects on the conformation of the active site in different states of the upper and lower loops. Similarly to the shorter Cys_{145}–His_{41} distance in closed loop states, a shorter distance is also observed between Asp_{187} and His_{41} when the lower loop is in an intermediate or closed state (Table 3). The interaction between these two residues is

thought to be mediated by a water molecule (H_2O_{cat}, Figure 2E) that plays an essential role in catalysis.[4,40,41] It stabilizes the positive charge of protonated His_{41} through electrostatic interactions with the negatively charged Asp_{187} during cleavage.[3,33] In crystal structures,[4,40,41] as well as in our simulations, this water molecule forms hydrogen bonds with His_{41}, His_{164}, and Asp_{187}. When the lower loop is in an intermediate or closed state, the mean Asp_{187}-His_{41} distance is 5.1 ± 0.1 or 4.90 ± 0.09 Å, respectively, and water-bridged interactions between His_{41}, His_{164}, and Asp_{187} are formed 64 ± 16 and 85 ± 9% of the time, respectively. The open state has a water molecule bridging these three residues only 7 ± 3% of the time, as the mean Asp_{187}-His_{41} distance increases to 7.5 ± 0.8 when open. These results provide additional evidence that the closed and intermediate lower loop states may be more conducive to catalysis than the open state.

Next, we wanted to determine the extent to which the state of one loop influences the state of the opposite loop. We observe that when the upper loop is closed, the lower loop is in an intermediate or closed state 17.5% more often than average. When the upper loop is open, the lower loop is open 4.5% more often than average. Additionally, when the lower loop is open, the upper loop is found to be open 14.5% more than average (Table S1). These results may indicate a cooperative anchoring of closed states. This cooperativity may arise from the orientation of the Arg_{188} side chain on the lower loop in different loop states. The positively-charged Arg side chain is oriented upwards toward the upper loop in closed and intermediate states, allowing it to form inter-loop interactions that bridge the two loops. In contrast, the Arg_{188} side chain often faces downwards in an open state (Figure 2D), resulting in fewer interactions with the upper loop. While Arg_{188} is moderately conserved in coronaviruses (Figure S4), other residues in the lower loop may play similar roles in stabilizing closed loop states when the Arg is not present, such as the lysine residue in the same position in M^{pro} of MERS-CoV. As the closure of either loop stabilizes catalytic residues (Table 3), cooperativity in the closure of loops is likely key to efficient cleavage of the substrate. Inhibitors designed to restrict the closure of one loop may indirectly limit the closure of the other, providing a

unique mechanism for inhibiting the protease.

2.4 Substrate Binding Favours Intermediate and Closed Loop States

To determine the effects of substrate binding on the loops, we carried out the same dimensionality reduction and clustering analysis on the two-substrate system (Figure 3A) and the one-substrate system (Figure S5). When comparing the two-substrate system to the apo dimer system, we find that substrate binding increases the probability of the intermediate state in the lower loop while decreasing the probability of the open state (Table 2). Due to steric hindrance, the lower loop cannot sample a maximally closed state when a substrate is present in the active site. However, the mean distance between His_{41} and Cys_{145} in intermediate and closed states are the same within error, suggesting that either state may be suitable for catalysis (Table 3). The frequency of the closed state in the upper loop is also increased upon substrate binding, further reducing the His_{41}-Cys_{145} distance and rigidifying the catalytic dyad through stronger Cys_{44}-His_{41} interactions, promoting configurations suitable for catalysis.

Upon substrate binding, several regions in the dimer exhibit differences in structure. We assessed these changes in the structural ensemble quantitatively using the Jensen-Shannon distance (JSD) to compare the apo and bound ensembles in terms of inter-residue distances (Figure 3E) and backbone dihedral angles (Figure 3F). The JSD is a symmetrized measure of the similarity of probability distributions, with values closer to zero indicating similarity and values closer to one indicating dissimilarity.[42,43] Computing the JSD between distributions of Cα distances and distributions of dihedral angles reveals that the most significant differences upon binding occur in the active site loops and the catalytic loop, with moderate changes occurring at the C-terminus and the dimerization interface (Figure 3E-F). In contrast, Domain III exhibits only minor structural differences between the apo and bound ensembles. The fact that the most prominent differences occur in the vicinity of the active site is expected because of the increased frequency of closed/intermediate loop states in the

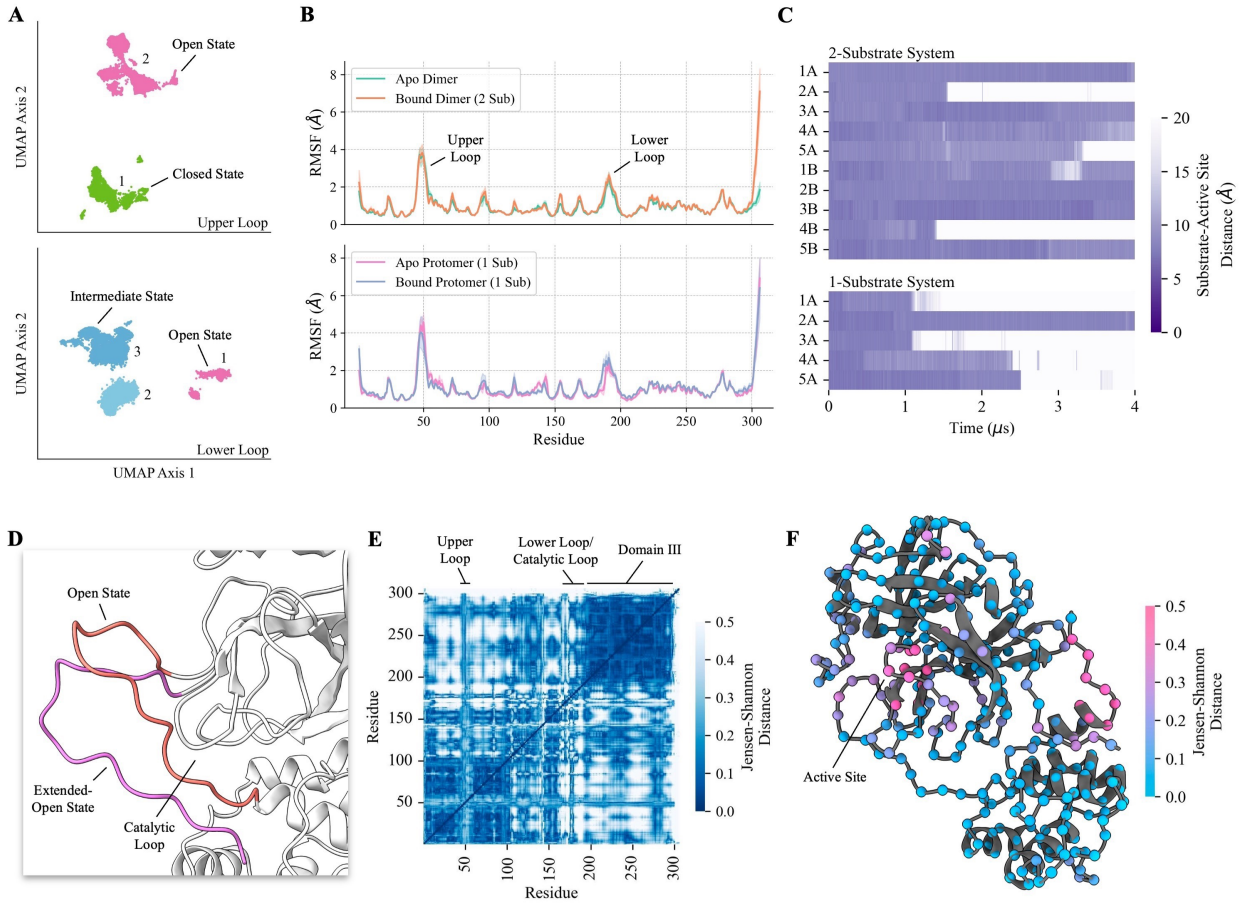

Figure 3: **Effects of substrate binding on M^{pro}.** (A) UMAP projections and HDBSCAN clusters of M^{pro} upper and lower loops in the two-substrate system. Numbers within the UMAP plots label each cluster. Pink, blue, and green shades denote open, intermediate, and closed loop states, respectively. Shades of each colour are assigned arbitrarily. Populations of each state are provided in Table 2. (B) Mean RMSF per residue for the apo dimer and two-substrate system (top) and the bound and apo protomers of the one-substrate system (bottom). Shaded regions denote the standard error of the mean. (C) Distance between the substrate and active site (defined as the distance between Cys_{145} and the closest substrate residue) over the length of the simulations for the two- and one-substrate systems, numbered by trajectory and protomer (e.g. 1A is protomer A of the 1st trajectory). Purple regions indicate that the substrate is in the active site, while white regions show when it has dissociated. (D) Representative structures illustrating the open and extended-open states of the lower loop. (E) Jensen-Shannon distance (JSD) between distributions of Cα distances, comparing the apo dimer and two-substrate ensembles. Pairs of residues with low JSD (blue) sample similar states between systems, whereas high JSD (white) describes significant ensemble differences. (F) The M^{pro} monomer coloured by the JSD between the apo dimer and two-substrate systems using backbone dihedral angles as features. The spheres represent the Cα of each residue and are coloured by JSD.

bound system and the interactions between the substrate and active site.

The observed states of M^{pro}'s active site loops are analogous to other enzymes with flexible active site loops that have distinct apo and bound conformations. For example, Du et al. found that fructose-1,6-bisphosphate aldolase/phosphatase has three active site loops that changed conformation when either a substrate, reaction intermediate, or product is bound.[44] They observed that the loops in the aldolase/phosphatase reorient to bring key side chains and Mg^{2+} cofactors into place. This functional role is comparable to the lower loop of M^{pro}, which brings Asp_{187} closer to the catalytic dyad in a water-bridged interaction. The behaviour of the active site loops of M^{pro} also resembles HIV-1 protease, which has different frequencies of loop conformational states based on the presence or absence of a substrate.[18,19] The loop flexibility in HIV-1 protease has been proposed to have a functional role: product release may be facilitated by the higher conformational entropy of the loops surrounding the active site in the apo compared to the bound state.[19] A similar mechanism may be important in the function of M^{pro}.

2.5 Substrate Binding in One Protomer Correlates with a Bound-Like Conformation in the Apo Partner

To determine the effect of substrate binding on the rest of the protein, we conducted a detailed analysis of the one-substrate system. We used UMAP and HDBSCAN to analyze both protomers in the one-substrate system (Figure S5), following the same approach we used to analyze the loop states in the apo dimer (Figure 2). We find that the apo protomer's upper loop samples open and closed states at frequencies comparable to the apo dimer (Table 2). The lower loop, however, samples the intermediate and closed states significantly more than the open state, comparable to the two-substrate bound dimer despite the absence of substrate (Table 2). The similarity in loop state frequencies between the two-substrate system and the apo protomer from the one-substrate system suggests that binding in one protomer induces a bound-like ensemble in its apo partner.

In addition to the differences in loop states, the one- and two-substrate systems also differ in their retention of the substrate. The substrate dissociates from the active site within 4 µs in 80% of the trajectories of the one-substrate system, compared to only 30% of the two-substrate system (Figure 3C). As we started both simulation systems from the same initial structure and under the same starting conditions (except for the presence of the substrate in one of the protomers), we do not expect the observed difference in dissociation to be a simulation artifact. As shown above, binding in one protomer in the one-substrate system contributes to the increased sampling of intermediate/closed states of its partner's active site. However, while the bound protomer contributes to the closing of the apo protomer's active site, the apo protomer does not reciprocate this effect in the bound protomer. This can be seen in the increased frequency of the open lower loop state in the bound protomer ($30 \pm 10\%$) compared to the apo protomer ($6 \pm 6\%$), which is likely what causes the quicker substrate dissociation in the one-substrate system (Figure 3C). Overall, the observed dependence of the conformational ensemble on the presence or absence of substrate in the opposite protomer suggests that there is allosteric communication between protomers.

2.6 Dynamic Network Analysis

The similarity of the behaviour of the lower loop in the apo protomer and its bound partner in the one-substrate system (Table 2) is suggestive of allosteric communication between protomers. To describe potential allosteric pathways connecting the two protomers and associated correlated motions, we used dynamic network analysis (DNA). DNA uses cross-correlation to reduce trajectories into a "nodes and edges" representation, which can then be analyzed as a network.[45,46] In this analysis, each residue is classified as a node, and nodes are linked to each other through edges, each with an associated correlation coefficient. The network can then be partitioned into "communities" of dense connections, and optimal paths can be determined by computing the shortest pathways between all nodes.[46,47]

The cross-correlation coefficients computed from the simulations of the one-substrate

dimer (Figure 4A) reveal the correlated motion of residues within each domain, consistent with other studies using cross-correlation[30,32] or linear mutual information.[27] The fact that there is moderate correlation between the two active sites (box 1 in Figure 4A) provides further support for allosteric communication occurring between the two protomers. The motion of the N-terminus of one protomer is correlated with the opposite protomer's active site (box 2 in Figure 4A), consistent with the fact that the N-terminus contributes to shaping the active site of the opposite protomer.[7] Using these correlation coefficients, the system can be decomposed into nine communities of high information transfer (Figure 4B), which match well with the defined boundaries of the three domains of M^{pro} (Figure 1A). The system can also be represented as a network with edges weighted by their "betweenness" (Figure 4E), which is defined as the number of shortest paths from all nodes to all others that pass through that edge. In other words, a high betweenness (represented as a thicker red line) corresponds to an avenue of high information transfer.[45] From this network view (Figure 4E), it can be seen that much of the protein's information transfer travels directly through the middle of the protein, strongly suggesting that the dimer interface is important in allosteric communication.

In dynamic network analysis, the shortest distance between two sites corresponds to the optimal path of communication, as shorter distances result in enhanced allostery compared to suboptimal paths.[46] Analysis of the M^{pro} dimer simulations reveals that the shortest pathway between the active sites travels through Domain II of both protomers, which are bridged by one protomer's N-terminal region (specifically, through residue Met_6) (Figure 4F). Computing the 50 shortest paths between the active sites shows that pathways involving several residues in the N-terminal region and Domain II are involved in communication (Figure 4C). The shortest paths consistently pass directly through the center of the protein, including the dimer interface, and through the N-terminus of either protomer (Figure S6), suggesting that any one of these regions may be a target for obstructing allosteric communication. If this communication between protomers is essential to the function of M^{pro}, allosteric inhibitors

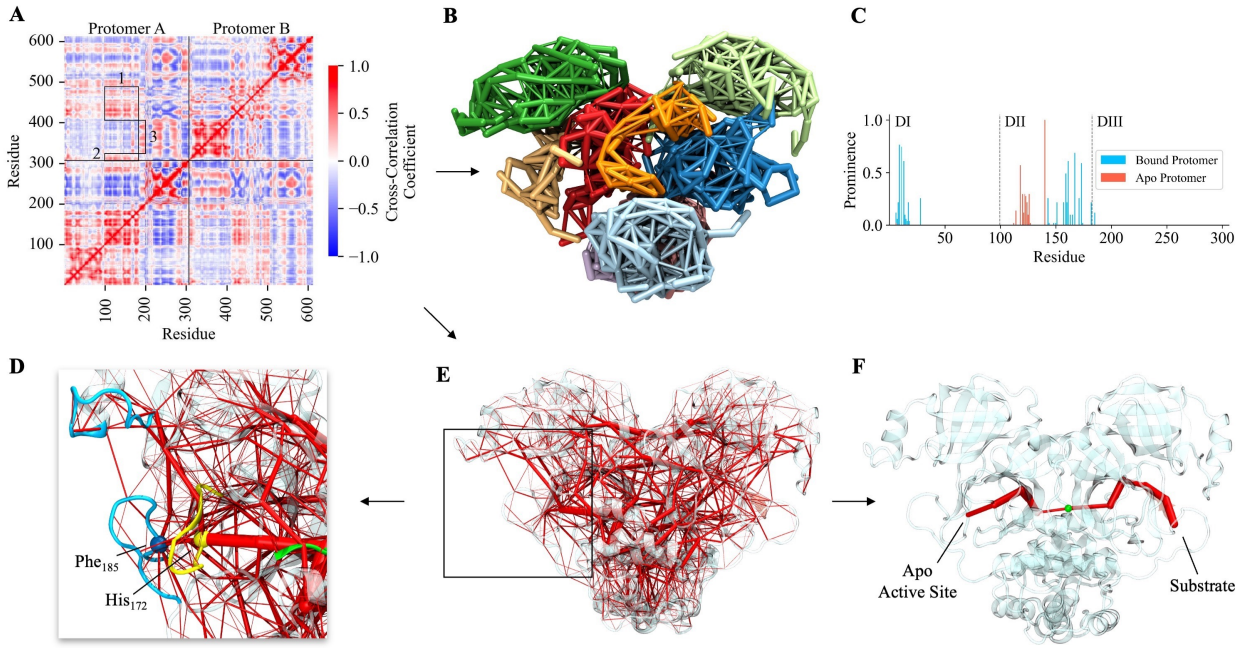

Figure 4: **Dynamic network analysis of the one-substrate system.** (A) Cross-correlation coefficient matrix. Values close to 1, -1, and 0 show correlation, anti-correlation, and no correlation, respectively. Boxed areas show a notable correlation between protomers, which are split by horizontal and vertical black lines. Box 1 shows correlation between Domain IIs, box 2 between the N-terminus and the opposing active site, and box 3 the lower loop and Domain I of the opposite protomer. (B) M^{pro} clustered into nine communities (note that some of the communities are on the opposite side and not visible in this orientation of the dimer). (C) Prominence of each residue, i.e. how often each residue is involved in the 50 shortest paths from one active site to the other. Boundaries between the three domains are indicated by dashed lines. (D) Enlarged view of the apo protomer's active site. The loops are coloured blue, the N-terminus green, and the catalytic loop yellow. Phe_{185} and His_{172} serve as hub nodes (spheres) with several connections to other parts of the active site. (E) Network representation of M^{pro} with edges weighted by betweenness, defined as the number of shortest paths from all nodes to all others that pass through that edge; the thicker the edge, the higher the betweenness. (F) Shortest path from the substrate to the apo protomer's active site. The optimal (shortest) path (red) travels through Met_6 in the N-terminus (green sphere).

that bind at the dimer interface and disrupt this communication may disrupt activity and provide an alternative to competitive inhibition. Such an approach to allosteric modulation of M^{pro} by targeting its dimer interface has been suggested in prior studies.[6,31,48,49] For instance, Douangamath et al. identified three fragments that bind M^{pro} at the interface near the N-terminus,[48] while Cantrelle et al. identified an allosteric hot spot at the interface using NMR spectroscopy.[49] Douangamath et al. identified multiple non-covalent binders that bind at the dimer interface,[48] one of which has been shown to promote dissociation of the dimer.[6]

In our analysis, we find that the N-terminus has a path of high betweenness to His_{172}, which acts as a hub node (Figure 4D-F). His_{172} is closely connected to both loops, as well as the active site (Figure 4D). His_{172}'s high betweenness with Phe_{185}, a hub node on the lower loop, suggests that it may play a role in the conformational change of the loops upon binding. The upper loop is also closely connected to the lower loop through short pathways from either His_{172} or Phe_{185}. The results from dynamic network analysis are supported by prior studies demonstrating the functional importance of His_{172} and Phe_{185}. In particular, protonation of His_{172} has been shown to result in a collapse of the dimer interface[29,30] and a partial collapse in the catalytic site,[20,29] suggesting its importance in the structure and allostery of M^{pro}. Phe_{185} was identified as a mutational coldspot through an analysis of missense mutations reported in the Global Initiative on Sharing All Influenza Data (GISAID).[50] The lack of mutations in key catalytic residues suggests that coldspots are likely functionally important, which agrees with our identification of Phe_{185} as a functionally important residue due to its role as a communication hub. Additionally, Chen et al. identified a mutation at this site, F185S, that exhibited a 30-fold decrease in activity, providing further evidence of the importance of Phe_{185}.[12] Furthermore, both His_{172} and Phe_{185} are highly conserved among coronaviruses (Figure S4).

N-terminal residues are involved in six of the top ten edges of highest betweenness between protomers (Table S2), signifying the importance of the N-terminus in interprotomer communication. N-terminal residues also populate the list of residues with the most connec-

tions between protomers (Table S3). Specifically, Arg_4 acts as a communication hub, with several total edges and edges of high betweenness connecting the two active sites (Table S2 and Table S3). Arg_4 has also been reported to play a role in dimerization, forming a salt bridge with Glu_{290} on Domain III of the opposite protomer[7] and providing an allosteric path between Domain II, the opposite N-terminus, and Domain III. El Ahdab et al. also showed that the distance between the catalytic dyad may be allosterically linked to the distance between Arg_4 and Glu_{290}.[30]

Previous computational studies have reported allostery in M^{pro}.[32,48,51–53] Elastic network modelling by Dubanevics et al. revealed allostery between active sites, which is consistent with the results of our network analysis.[52] Sencanski et al. and Verma et al. discovered allosteric inhibition pockets and noted that inhibitory constants at these sites were significantly stronger than competitive inhibition pockets, with the added benefit that inhibitors need not compete for the active site.[54,55] Campitelli et al. used a dynamic coupling index to quantify allostery between the catalytic Cys_{145}-His_{41} dyad and Glu_{55}, Ile_{59}, and Arg_{60} residues on the opposite protomer, all of which are located in the upper loop.[53] Lastly, Flynn et al. provide evidence of a chain of mutationally-sensitive residues that connect the active site to the dimerization interface, and each of these residues is crucial to protein function.[56] Using the one-substrate system, we have shown that M^{pro}'s active sites are allosterically connected, and using DNA, we find that the optimal allosteric path travels through the middle of Domain II of both protomers in a pathway bridged by the N-terminus. The N-terminus is a key part of the allosteric pathways that facilitate communication with the partner protomer's active site, dimerization domain, and both the upper and lower loops.

2.7 Stability of the M^{pro} Monomer

To expand on the importance of dimerization in the allosteric communication and the dynamics of M^{pro}, we evaluated the differences between monomeric and dimeric M^{pro} assemblies by comparing our monomer and dimer simulations. Based on a comparison of the RMSF

profiles, the monomer is significantly more flexible than the dimer throughout the entire protein, but especially in the N-terminus, the dimer interface (including residues 120-170), and the lower loop (Figure 5A). This difference in flexibility is consistent with other studies,[25,26] and is likely because these are the regions of M^{pro} that have stabilizing interactions in the dimer.

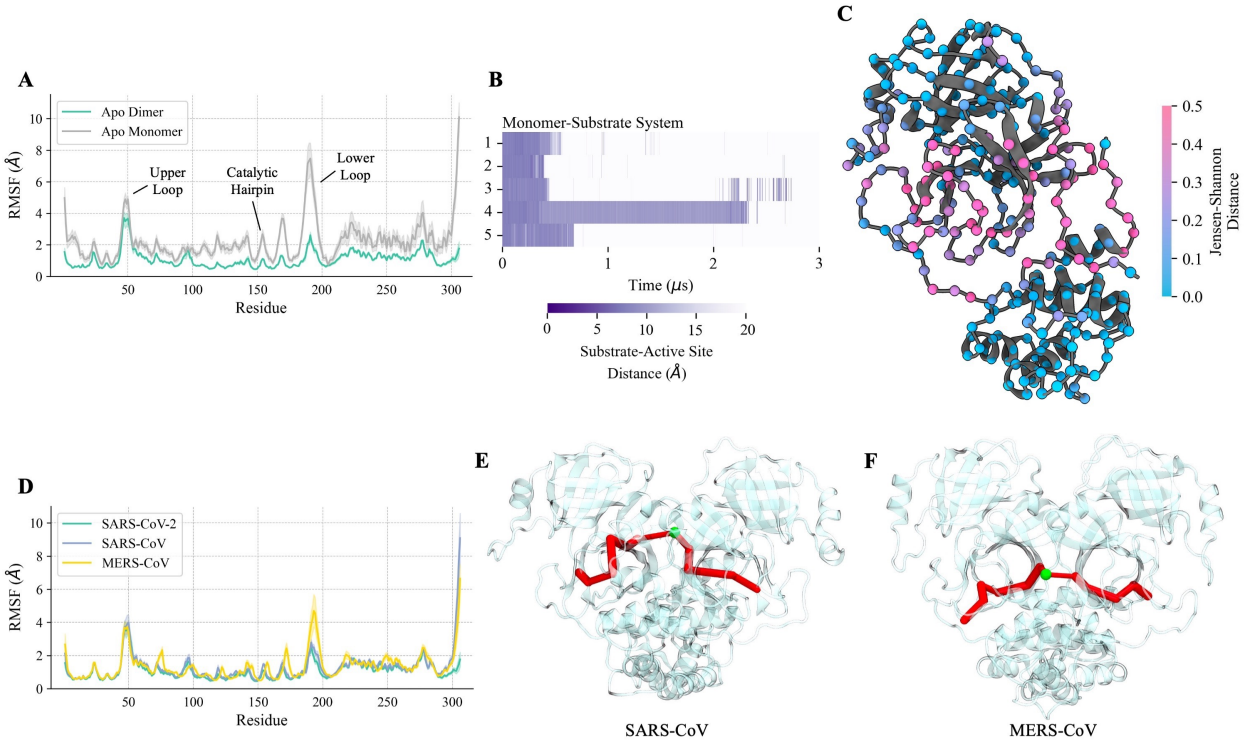

Figure 5: **A comparison of the SARS-CoV-2 apo dimer with the apo monomer, as well as with the SARS-CoV and MERS-CoV dimers.** (A) Mean RMSF per residue for the SARS-CoV-2 apo dimer and apo monomer. Shaded regions denote standard error of the mean. (B) Distance between the substrate and active site (defined as the distance between Cys_{145} and the closest substrate residue) over the length of the simulations for the SARS-CoV-2 monomer-substrate complex, numbered by trajectory. Purple regions indicate that the substrate is in the active site, while white regions show when it has dissociated. (C) The M^{pro} monomer coloured by the JSD between the SARS-CoV-2 apo dimer and apo monomer using backbone dihedral angles as features. The spheres represent the Cα of each residue and are coloured by JSD. The largest differences between the apo monomer and apo dimer ensembles occur in Domain II. (D) Mean RMSF per residue for SARS-CoV-2, SARS-CoV, and MERS-CoV apo M^{pro} dimers. Shaded regions denote standard error of the mean. (E), (F) Shortest path between active sites in SARS-CoV and MERS-CoV M^{pro}s. The optimal (shortest) path (red) travels through the N-terminus (green sphere).

UMAP projections and HDBSCAN clustering of the apo monomer simulations indicate that both the upper and lower loops of the monomer adopt open states more frequently than in the dimer (Figure S7A, Table 2). In the apo monomer, the lower loop no longer samples the closed state. Analysis of the structural ensembles of the monomer and dimer suggests that the major structural difference preventing the lower loop from adopting the closed state is the absence of a hydrogen bond between Glu_{166} of one protomer and Ser_1 of the opposite protomer. This interaction, which cannot occur in the monomer, stabilizes the catalytic loop in the dimer, and its absence results in a more flexible catalytic loop and a change in the structure of the active site, as is evident in the JSD analysis (Figure 5C). The increased flexibility of the catalytic loop results in fewer stabilizing hydrophobic interactions with the lower loop, leading to frequent sampling of an "extended-open" conformation in which the lower loop is dislodged from the catalytic loop (Figure 3D). The frequent sampling of the open state by the lower loop causes dissociation of the substrate within 1 µs in all monomer-substrate complex simulations (Figure 5B). In contrast, the dimer-substrate complex exhibits less dissociation (Figure 3C). Taken together, these results suggest that the rigidity of the lower loop upon dimerization is crucial for retaining the substrate, which may contribute to the experimentally observed inactivity of the M^{pro} monomer.[5,6] Our results are also consistent with a mass spectrometry study by El-Baba et al., which found that the monomer is not only inactive but also does not bind substrate with high affinity.[6]

Dimerization of M^{pro} imparts stability to the active site, as the partner protomer's N-terminus helps form the catalytic pocket through hydrogen bonding.[5,57] In our simulations, we also see a positional shift in Domain III upon dimerization (Figure S8A), as shown through high JSDs between Domain III and the rest of the protein when comparing the dimer and monomer ensembles (Figure S8B). In related coronaviruses (HCoV 229E and TGEV), Domain III has been shown to hold the lower loop in a catalytically competent position (in TGEV),[58] and truncation of this domain in *in vitro* experiments reduced the activity of the protease (in HCoV 229E, TGEV).[58,59] Truncation of Domain III in SARS-

CoV-2 M^{pro} was similarly found to decrease activity due to a shift in the monomer-dimer equilibrium towards the monomer.[60,61] The shift seen in Domain III upon dimerization likely reorients the domain into a more catalytically favourable position that stabilizes the lower loop. These factors explain the instability of the monomer's active site and the dissociation of the substrate observed in simulations of the monomer.

2.8 SARS-CoV and MERS-CoV M^{pro} dynamics and allostery are similar to SARS-CoV-2

To determine the extent to which our findings can be generalized to M^{pro} of other coronaviruses, we carried out MD simulations of SARS-CoV and MERS-CoV M^{pro} dimers in the apo state. Both coronavirus M^{pro}s show strong similarities to SARS-CoV-2 M^{pro} with respect to dynamics and allostery. The RMSF profiles indicate similar flexibility throughout most of the protein (Figure 5D), which is unsurprising considering their high sequence similarity (96.1% and 50.8% for SARS-CoV and MERS-CoV M^{pro} compared to SARS-CoV-2 M^{pro}; sequence alignment provided in Figure S4). As we observe for SARS-CoV-2 M^{pro}, the upper and lower loops exhibit high flexibility.

Network analyses of SARS-CoV and MERS-CoV M^{pro}s (Figure S9) show comparable outcomes to SARS-CoV-2, including similar allosteric pathways between active sites through Domain II of both protomers (Figure 5E and F). Slight differences in optimal paths may be attributed to sampling distinct regions of conformational space among coronavirus M^{pro}s, as demonstrated in a Markov-state modelling study of M^{pro} from all three coronaviruses.[15] In SARS-CoV-2 M^{pro}, hydrophobic interactions between Ala_{285} and Leu_{286} from opposite protomers are known to contribute to the stability of the dimer interface.[57] We find edges of high betweenness connecting these two residues in the network obtained using DNA (Figure S10A). In M^{pro} from SARS-CoV, the residues at these positions are Thr_{285} and Ile_{286}. Instead of hydrophobic interactions, these residues establish only a few contacts, resulting in a weaker bridge between the two protomers. Mutation of these two residues (along with Ser_{284}) to

Ala in M^{pro} from SARS-CoV resulted in enhanced dimerization due to tighter packing and increased activity by a factor of 3.6.[62] The tighter packing may be directly related to the increased activity, as a shorter distance between Arg_4 and Glu_{290} at the dimerization interface tends to correlate with a shorter catalytic distance through allosteric communication.[30] The enhanced packing increases the stability of the N-terminus and the active site, as position 286 (along with 277 and 279) is strongly coupled to the active site, and in SARS-CoV, much of this coupling is lost.[53] The significance of having different residues at these sites in different coronaviruses is reinforced by the presence of fewer edges between Thr_{285} and Ile_{286} in our DNA model of the SARS-CoV M^{pro} compared to the multiple edges of high betweenness between Ala_{285} and Leu_{286} in the SARS-CoV-2 model (Figure S10).

Experimental studies on SARS-CoV M^{pro} have shown that alanine point mutations at the dimer interface in Domain II disrupt catalytic activity while retaining the global structure of the dimer.[63,64] This reduction of activity, despite a lack of observed change in structure, suggests that communication between active sites may play a role in catalysis in SARS-CoV M^{pro}. Similarly, mutational screens have revealed several residues at the dimer interface that affect the activity of SARS-CoV-2 M^{pro}.[56] As we have shown that allosteric paths between active sites travel through the dimer interface, mutations of interface residues could disrupt communication between protomers. The dimer interface provides a hot spot for several potential inhibitors, as revealed by both experimental (mass spectrometry, X-ray crystallography) and computational studies (docking, algebraic topology).[6,32,48,49] We hypothesize that inhibition at this interface may act similarly to the point mutations in reducing catalytic activity by disrupting interprotomer communication.

3 Conclusions

Here, we present a combined 92.5 µs of all-atom simulation data on M^{pro}s from multiple coronaviruses in different forms. First, we find that M^{pro}'s flexible active site loops sample

distinct open, intermediate, and closed conformational states and that closed states of the loops stabilize catalytic residues. We also show that substrate binding increases the frequency of the catalytically favourable closed state. Analysis of loop states provides evidence for allosteric communication between the two protomers. To investigate allosteric transmission in M^{pro}, we use dynamic network analysis, finding potential allosteric paths between the active sites in opposite protomers. We observe the instability of the monomer compared to the dimer due to an increased flexibility, which results in its inability to retain the substrate. Finally, we find similarities between M^{pro}s from different coronaviruses in terms of loop states, flexibility and allosteric paths, which makes it possible to generalize our findings to M^{pro} from related viruses. The flexibility of the upper and lower loop likely plays a functional role in allowing M^{pro} to bind diverse substrates — not only the 11 cleavage sites in the polyproteins of SARS-CoV-2, but also proteins within the human proteome.[65] As substrate binding in one protomer is found to be correlated with the behaviour of the other, inhibitors that target allosteric pathways may provide an alternate mode of inhibition.

4 Methods

4.1 Experimental Design

We carried out extensive all-atom simulations of M^{pro} in several conditions (see Table 1 for a list of simulations). To investigate the effects of dimer against monomer, apo against substrate bound, and between different coronaviruses, we simulated several systems as follows. The crystal structure PDB 6XHU[33] was used for the apo SARS-CoV-2 M^{pro} monomer and dimer simulations, while a modified version of a crystal structure containing a covalent inhibitor (PDB 6LU7[8]) obtained from Suárez and Diaz[23] was used for substrate-bound simulations. In all SARS-CoV-2 M^{pro} simulations, lysines, arginines, aspartates, and glutamates were all charged. His_{64}, His_{163}, His_{164}, His_{172}, and His_{246} were all protonated at $N\epsilon$, while His_{41} and His_{80} were protonated at $N\delta$, consistent with a pH of 7.0.[23] The bound

systems contained a pentapeptide substrate mimic in the active site (details of the substrate can be found in SI Section 1.1). SARS-CoV and MERS-CoV M^{pro} simulations used crystal structures PDB 1UJ1[66] and PDB 5C3N,[67] respectively, as the starting structure for the simulations. The structure 1UJ1 is among the highest resolution apo structures of SARS-CoV M^{pro}, while 5C3N was the only published crystal structure of MERS-CoV M^{pro} at the start of the study. The protonation states of all residues in the SARS-CoV simulations match those in the apo dimer simulations. In the MERS-CoV simulations, His_8, His_{64}, His_{71}, His_{175}, and His_{194} are all protonated at Nϵ while His_{41}, His_{83} and His_{166} are protonated at Nδ, consistent with a pH of 7.0.[23] All systems studied, as well as the number of independent trajectories and the length of each simulation, are provided in Table 1. Simulations on a µs timescale were chosen based on the simulation timescale of other M^{pro} simulation studies.[22–25,27,28]

4.2 MD Simulations

GROMACS 2019.1 was used for all MD simulations.[68] The system was composed of the M^{pro} monomer or dimer with charged termini in a rhombic dodecahedral box filled with CHARMM-modified TIP3P water molecules[69] at a concentration of 0.15 M NaCl for a total of about 87000 atoms for the dimer systems and about 63000 atoms for the monomer systems. The protein had a minimum distance of 10 Å to all box edges in the initial structure, and periodic boundary conditions were applied. The CHARMM36m force field[70] was used, and virtual sites for all hydrogens were employed to allow for an integration timestep of 4 fs. MkVsites[71] was used to generate bond constraints for termini capping groups. Energy minimization was done prior to production runs using the steepest descent algorithm. Hydrogen bond lengths were constrained using the LINCS algorithm.[72] Prior to the production run, a restrained simulation was performed for 5 ns with position restraints on all heavy atoms. Short-range electrostatic interactions and van der Waals interactions were calculated with a 0.95 nm cut-off. Long-range electrostatic interactions were calculated using fast smooth particle-mesh Ewald summation with a fourth-order interpolation and a grid spacing of 0.12

nm.[73] Neighbour searching was done using the Verlet cutoff scheme. Temperature coupling was done using velocity rescaling to maintain a temperature of 298 K for all simulations.[74] Equilibration simulations were run in the NPT ensemble for 10 ns using Berendsen pressure coupling,[75] followed by 10 ns using the Parrinello-Rahman barostat.[76] Production runs were performed in the NPT ensemble at 298 K and 1 bar using the Parrinello-Rahman barostat and a velocity rescaling thermostat.[74,76]

4.3 Calculation of Predicted Residual Dipolar Couplings from Simulations

RDCs were computed from the M^{pro} apo dimer simulation ensembles. A total of 17000 structures from the equilibrated portions of the trajectories were used. Alignment tensors were fit to the experimental data from Robertson et al.[34] using the DC Order Matrix Fitting function (singular value decomposition) from NMRPipe.[77] We followed the same procedure as Robertson et al. in their comparison between RDCs computed from crystal structures and experimental RDCs. They optimized the experimental alignment tensor accuracy by iteratively restricting the fit to residues 5-115 and 199-272.[34] We likewise restricted the fit to these two contiguous regions in M^{pro}.

Q-factors were computed via NMRPipe[77] using the following equation:

$$Q = \frac{RMS(D_i^{\text{obs}} - D_i^{\text{pred}})}{RMS(D_i^{\text{obs}})} \tag{1}$$

where D_i^{obs} and D_i^{pred} are the observed (from NMR) and predicted (from simulations) RDCs for interaction i.[78]

4.4 Dimensionality Reduction and Clustering

UMAP[36] was performed on the upper and lower loops for each system using heavy atom positions as features after alignment of the protein by backbone atoms, excluding the loops

and termini. The number of neighbours parameter was chosen to maximize global topological structure in high dimensions over local structure (higher values), while the minimum distance parameter was optimized for clear clusters (lower values). The specific values used were number of neighbours = 400 and minimum distance = 0.1. Clustering was then carried out with HDBSCAN,[39] where the minimum cluster size and the minimum samples parameters were chosen to maximize relative validity indices (specific values used varied between minimum samples = 20-80 and minimum cluster size = 700-1700). Other choices of features (dihedral angles and backbone interatomic distances) were also investigated, but heavy atom positions provided the clearest categorization of conformational states. Loop states were categorized based on the distance between the loop tip (the Cα atoms of Leu_{50} on the upper loop and Arg_{188} on the lower loop) and Pro_9 in the N-terminus, which is used as a reference point due to its stability in the structure (Tables S4-S9). In the upper loop, clusters were determined to be closed if this distance was < 38.5 Å and open if ≥ 38.5 Å. In the lower loop, clusters were determined to be closed if the distance was ≤ 32.5 Å, intermediate if between 32.5 and 34.5 Å, and open if ≥ 34.5 Å. These bins were determined using the same distances in six apo crystal structures (6YB7, 6Y2E, 6WTM, 6M03, 6WQF, and 6XHU) as a guideline, where the upper and lower loops are in closed and intermediate states, respectively. In the crystal structures, these distances are: Leu_{50}-Pro_9 distance = 37.14 ± 0.03 Å and Arg_{188}-Pro_9 distance = 32.73 ± 0.02 Å. Additionally, structures belonging to each cluster were visually inspected to confirm that the assigned loop state was consistent with the position of the loop in simulation frames.

4.5 Cross-Correlation and Dynamic Network Analysis

Cross-correlation coefficients were calculated using CARMA.[79] Pairwise correlations were calculated as follows:

$$C_{ij} = \frac{\langle \Delta\vec{r}_i(t) \cdot \Delta\vec{r}_j(t) \rangle}{\sqrt{\langle \Delta\vec{r}_i(t)^2 \rangle \langle \Delta\vec{r}_j(t)^2 \rangle}} \tag{2}$$

where $\Delta\vec{r}_i(t) = \vec{r}_i(t) - \langle\vec{r}_i(t)\rangle$, $\vec{r}_i(t)$ is the position vector of the ith residue's α-carbon at time t and $\langle\vec{r}_i(t)\rangle$ is the ith node's mean position.[46]

Dynamic network analysis was performed using the dynetan 1.0.1 python package.[45] The network was created by assigning a node to the α-carbon of each residue and connecting nodes with edges if two nodes were within 4.5 Å for over 75% of equilibrated frames. After removing edges between nearest neighbours, the remaining edges were weighted by their normalized correlation between nodes:

$$w_{ij} = -log(|C_{ij}|) \tag{3}$$

where w_{ij} is the weight of an edge between nodes i and j and C_{ij} is the normalized correlation. w_{ij} represents the probability of information transfer across an edge. Inclusion of nearest neighbours did not significantly affect assigned communities or optimal paths.

The shortest distance between all nodes was calculated using the Floyd-Warshall algorithm, as implemented in the dynetan package. Communities were clustered using Louvain heuristics until a modularity metric was maximized. The modularity measures the probability difference between intra- and inter-community edges. Modularity values of all systems were about 0.75, which is slightly higher than scores in typical real-world networks (0.4-0.7).[80]

Author Contributions

E.L.: conceptualization, methodology, investigation, visualization, writing - original draft. S.R.: conceptualization, methodology, supervision, writing - review & editing.

Declaration of Interests

The authors declare no competing interests.

Data Availability Statement

MD simulation data (initial coordinates and simulation trajectories) are provided for all systems as a Zenodo repository accessible here: https://zenodo.org/records/13730633. Data underlying figures are also available in this Zenodo repository.

Acknowledgement

The authors thank Dr. Natalia Díaz from the University of Oviedo for providing the structure of the SARS-CoV-2 M^{pro} with the pentapeptide substrate mimic. This research was supported by a Connaught New Researcher Award to S.R. and a Natural Science and Engineering Research Council of Canada (NSERC) Discovery Grant. This research was enabled in part by support provided by Calcul Québec and the Digital Research Alliance of Canada (alliance.can.ca). Computations were performed on the Niagara supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation; the Government of Ontario; Ontario Research Fund – Research Excellence; and the University of Toronto.

Supporting Information Available

Supporting Information: Additional simulation methods, including system setup and analysis; B-factors and crystal contacts; comparison of crystal structures; RMSF profiles; sequence alignment; dimensionality reduction and clustering analysis; dynamic network analysis; comparison of monomer and dimer ensembles; analysis of SARS-CoV and MERS-CoV M^{pro}; principal component analysis; analysis of loop states in all systems; descriptions of Supporting Videos; References.

References

(1) Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; Chen, H.-D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R.-D.; Liu, M.-Q.; Chen, Y.; Shen, X.-R.; Wang, X.; Zheng, X.-S.; Zhao, K.; Chen, Q.-J.; Deng, F.; Liu, L.-L.; Yan, B.; Zhan, F.-X.; Wang, Y.-Y.; Xiao, G.-F.; Shi, Z.-L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273.

(2) Chan, J. F.-W.; Kok, K.-H.; Zhu, Z.; Chu, H.; To, K. K.-W.; Yuan, S.; Yuen, K.-Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections* **2020**, *9*, 221–236.

(3) Lee, J.; Worrall, L. J.; Vuckovic, M.; Rosell, F. I.; Gentile, F.; Ton, A.-T.; Caveney, N. A.; Ban, F.; Cherkasov, A.; Paetzel, M.; Strynadka, N. C. J. Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nature Communications* **2020**, *11*, 5877.

(4) Lee, J.; Kenward, C.; Worrall, L. J.; Vuckovic, M.; Gentile, F.; Ton, A.-T.; Ng, M.; Cherkasov, A.; Strynadka, N. C. J.; Paetzel, M. X-ray crystallographic characterization of the SARS-CoV-2 main protease polyprotein cleavage sites essential for viral processing and maturation. *Nature Communications* **2022**, *13*, 5196.

(5) Nashed, N. T.; Aniana, A.; Ghirlando, R.; Chiliveri, S. C.; Louis, J. M. Modulation of the monomer-dimer equilibrium and catalytic activity of SARS-CoV-2 main protease by a transition-state analog inhibitor. *Communications Biology* **2022**, *5*, 160.

(6) El-Baba, T. J.; Lutomski, C. A.; Kantsadi, A. L.; Malla, T. R.; John, T.; Mikhailov, V.; Bolla, J. R.; Schofield, C. J.; Zitzmann, N.; Vakonakis, I.; Robinson, C. V. Allosteric

Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays. *Angewandte Chemie International Edition* **2020**, *59*, 23544–23548.

(7) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020**, *368*, 409–412.

(8) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293.

(9) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catalysis* **2020**, *10*, 12544–12554.

(10) Li, G.; Hilgenfeld, R.; Whitley, R.; De Clercq, E. Therapeutic strategies for COVID-19: progress and lessons learned. *Nature Reviews Drug Discovery* **2023**, *22*, 449–475.

(11) Hammond, J.; Leister-Tebbe, H.; Gardner, A.; Abreu, P.; Bao, W.; Wisemandle, W.; Baniecki, M.; Hendrick, V. M.; Damle, B.; Simón-Campos, A.; Pypstra, R.; Rusnak, J. M. Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *New England Journal of Medicine* **2022**, *386*, 1397–1408.

(12) Chen, S. A.; Arutyunova, E.; Lu, J.; Khan, M. B.; Rut, W.; Zmudzinski, M.; Shahbaz, S.; Iyyathurai, J.; Moussa, E. W.; Turner, Z.; Bai, B.; Lamer, T.; Nieman, J. A.; Vederas, J. C.; Julien, O.; Drag, M.; Elahi, S.; Young, H. S.; Lemieux, M. J. SARS-CoV-2 M^{pro} Protease Variants of Concern Display Altered Viral Substrate and Cell Host Target Galectin-8 Processing but Retain Sensitivity toward Antivirals. *ACS Central Science* **2023**, *9*, 696–708.

(13) Iketani, S.; Mohri, H.; Culbertson, B.; Hong, S. J.; Duan, Y.; Luck, M. I.; Annavajhala, M. K.; Guo, Y.; Sheng, Z.; Uhlemann, A.-C.; Goff, S. P.; Sabo, Y.; Yang, H.; Chavez, A.; Ho, D. D. Multiple pathways for SARS-CoV-2 resistance to nirmatrelvir. *Nature* **2022**, *613*, 558–564.

(14) Gossen, J.; Albani, S.; Hanke, A.; Joseph, B. P.; Bergh, C.; Kuzikov, M.; Costanzi, E.; Manelfi, C.; Storici, P.; Gribbon, P.; Beccari, A. R.; Talarico, C.; Spyrakis, F.; Lindahl, E.; Zaliani, A.; Carloni, P.; Wade, R. C.; Musiani, F.; Kokh, D. B.; Rossetti, G. A Blueprint for High Affinity SARS-CoV-2 Mpro Inhibitors from Activity-Based Compound Library Screening Guided by Analysis of Protein Dynamics. *ACS Pharmacology & Translational Science* **2021**, *4*, 1079–1095.

(15) Cho, E.; Rosa, M.; Anjum, R.; Mehmood, S.; Soban, M.; Mujtaba, M.; Bux, K.; Moin, S. T.; Tanweer, M.; Dantu, S.; Pandini, A.; Yin, J.; Ma, H.; Ramanathan, A.; Islam, B.; Mey, A. S. J. S.; Bhowmik, D.; Haider, S. Dynamic Profiling of β-Coronavirus 3CL M^{pro} Protease Ligand-Binding Sites. *Journal of Chemical Information and Modeling* **2021**, *61*, 3058–3073.

(16) Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chemical Reviews* **2016**, *116*, 6391–6423.

(17) Jupin, I.; Ayach, M.; Jomat, L.; Fieulaine, S.; Bressanelli, S. A mobile loop near the active site acts as a switch between the dual activities of a viral protease/deubiquitinase. *PLOS Pathogens* **2017**, *13*, e1006714.

(18) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **2006**, *103*, 915–920.

(19) Freedberg, D. I.; Ishima, R.; Jacob, J.; Wang, Y.; Kustanovich, I.; Louis, J. M.;

Torchia, D. A. Rapid structural fluctuations of the free HIV protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Science* **2002**, *11*, 221–232.

(20) Verma, N.; Henderson, J. A.; Shen, J. Proton-Coupled Conformational Activation of SARS Coronavirus Main Proteases and Opportunity for Designing Small-Molecule Broad-Spectrum Targeted Covalent Inhibitors. *Journal of the American Chemical Society* **2020**, *142*, 21883–21890.

(21) Tran, N.; Dasari, S.; Barwell, S. A. E.; McLeod, M. J.; Kalyaanamoorthy, S.; Holyoak, T.; Ganesan, A. The H163A mutation unravels an oxidized conformation of the SARS-CoV-2 main protease. *Nature Communications* **2023**, *14*, 5625.

(22) Boby, M. L.; Fearon, D.; Ferla, M.; Filep, M.; Koekemoer, L.; Robinson, M. C.; The COVID Moonshot Consortium; Chodera, J. D.; Lee, A. A.; London, N.; von Delft, A.; von Delft, F. Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors. *Science* **2023**, *382*, eabo7201.

(23) Suárez, D.; Díaz, N. SARS-CoV-2 Main Protease: A Molecular Dynamics Study. *Journal of Chemical Information and Modeling* **2020**, *60*, 5815–5831.

(24) Díaz, N.; Suárez, D. Influence of charge configuration on substrate binding to SARS-CoV-2 main protease. *Chemical Communications* **2021**, *57*, 5314–5317.

(25) Grottesi, A.; Bešker, N.; Emerson, A.; Manelfi, C.; Beccari, A. R.; Frigerio, F.; Lindahl, E.; Cerchia, C.; Talarico, C. Computational Studies of SARS-CoV-2 3CLpro: Insights from MD Simulations. *International Journal of Molecular Sciences* **2020**, *21*, 5346.

(26) Ferreira, G. M.; Kronenberger, T.; Tonduru, A. K.; Hirata, R. D. C.; Hirata, M. H.; Poso, A. SARS-COV-2 M^{pro} conformational changes induced by covalently bound ligands. *Journal of Biomolecular Structure and Dynamics* **2021**, *40*, 12347–12357.

(27) Tekpinar, M.; Yildirim, A. Impact of dimerization and N3 binding on molecular dynamics of SARS-CoV and SARS-CoV-2 main proteases. *Journal of Biomolecular Structure and Dynamics* **2021**, *40*, 6243–6254.

(28) Ansari, N.; Rizzi, V.; Carloni, P.; Parrinello, M. Water-Triggered, Irreversible Conformational Change of SARS-CoV-2 Main Protease on Passing from the Solid State to Aqueous Solution. *Journal of the American Chemical Society* **2021**, *143*, 12930–12934.

(29) Jaffrelot Inizan, T.; Célerse, F.; Adjoua, O.; El Ahdab, D.; Jolly, L.-H.; Liu, C.; Ren, P.; Montes, M.; Lagarde, N.; Lagardère, L.; Monmarché, P.; Piquemal, J.-P. High-resolution mining of the SARS-CoV-2 main protease conformational space: supercomputer-driven unsupervised adaptive sampling. *Chemical Science* **2021**, *12*, 4889–4907.

(30) El Ahdab, D.; Lagardère, L.; Inizan, T. J.; Célerse, F.; Liu, C.; Adjoua, O.; Jolly, L.-H.; Gresh, N.; Hobaika, Z.; Ren, P.; Maroun, R. G.; Piquemal, J.-P. Interfacial Water Many-Body Effects Drive Structural Dynamics and Allosteric Interactions in SARS-CoV-2 Main Protease Dimerization Interface. *The Journal of Physical Chemistry Letters* **2021**, *12*, 6218–6226.

(31) Bram, Y.; Duan, X.; Nilsson-Payant, B. E.; Chandar, V.; Wu, H.; Shore, D.; Fajardo, A.; Sinha, S.; Hassan, N.; Weinstein, H.; TenOever, B. R.; Chen, S.; Schwartz, R. E. Dual-Reporter System for Real-Time Monitoring of SARS-CoV-2 Main Protease Activity in Live Cells Enables Identification of an Allosteric Inhibition Path. *ACS Bio & Med Chem Au* **2022**, *2*, 627–641.

(32) Sztain, T.; Amaro, R.; McCammon, J. A. Elucidation of Cryptic and Allosteric Pockets within the SARS-CoV-2 Main Protease. *Journal of Chemical Information and Modeling* **2021**, *61*, 3495–3501.

(33) Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Tan, K.; Joachimiak, A.; Coates, L.; Ko-

valevsky, A. Room-temperature X-ray crystallography reveals the oxidation and reactivity of cysteine residues in SARS-CoV-2 3CL Mpro: insights into enzyme mechanism and drug design. *IUCrJ* **2020**, *7*, 1028–1035.

(34) Robertson, A. J.; Courtney, J. M.; Shen, Y.; Ying, J.; Bax, A. Concordance of X-ray and AlphaFold2 Models of SARS-CoV-2 Main Protease with Residual Dipolar Couplings Measured in Solution. *Journal of the American Chemical Society* **2021**, *143*, 19306–19310.

(35) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS ONE* **2012**, *7*, e32131.

(36) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **2018**, *3*, 861.

(37) Klyshko, E.; Kim, J. S.-H.; McGough, L.; Valeeva, V.; Lee, E.; Ranganathan, R.; Rauscher, S. Functional protein dynamics in a crystal. *Nature Communications* **2024**, *15*, 3244.

(38) Oide, M.; Sugita, Y. Protein folding intermediates on the dimensionality reduced landscape with UMAP and native contact likelihood. *Journal of Chemical Physics* **2022**, *157*, 075101.

(39) Campello, R. J. G. B.; Moulavi, D.; Sander, J. *Lecture Notes in Computer Science*; Springer Berlin Heidelberg, 2013; p 160–172.

(40) Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Jedrzejczak, R.; Stols, L.; Langan, P.; Joachimiak, A.; Coates, L.; Kovalevsky, A. Structural plasticity of SARS-CoV-2 3CL M^{pro} active site cavity revealed by room temperature X-ray crystallography. *Nature Communications* **2020**, *11*, 3202.

(41) Firouzi, R.; Ashouri, M.; Karimi-Jafari, M. H. Structural insights into the substrate-binding site of main protease for the structure-based COVID-19 drug discovery. *Proteins: Structure, Function, and Bioinformatics* **2022**, *90*, 1090–1101.

(42) Vögele, M.; Thomson, N. J.; Truong, S. T.; McAvity, J.; Zachariae, U.; Dror, R. O. Systematic Analysis of Biomolecular Conformational Ensembles with PENSA. *arXiv* **2022**, 10.48550/ARXIV.2212.02714.

(43) Lin, J. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* **1991**, *37*, 145–151.

(44) Robertson, A. J.; Ying, J.; Bax, A. NMR Observation of Sulfhydryl Signals in SARS-CoV-2 Main Protease Aids Structural Studies. *ChemBioChem* **2022**, *23*, e202200471.

(45) Melo, M. C. R.; Bernardi, R. C.; de la Fuente-Nunez, C.; Luthey-Schulten, Z. Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *Journal of Chemical Physics* **2020**, *153*, 134104.

(46) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. Dynamical networks in tRNA:protein complexes. *Proceedings of the National Academy of Sciences* **2009**, *106*, 6620–6625.

(47) Floyd, R. W. Algorithm 97: Shortest path. *Communications of the ACM* **1962**, *5*, 345–345.

(48) Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C. D.; Resnick, E.; Strain-Damerell, C.; Aimon, A.; Ábrányi Balogh, P.; Brandão-Neto, J.; Carbery, A.; Davison, G.; Dias, A.; Downes, T. D.; Dunnett, L.; Fairhead, M.; Firth, J. D.; Jones, S. P.; Keeley, A.; Keserü, G. M.; Klein, H. F.; Martin, M. P.; Noble, M. E. M.; O'Brien, P.; Powell, A.; Reddi, R. N.; Skyner, R.; Snee, M.; Waring, M. J.; Wild, C.; London, N.; von Delft, F.; Walsh, M. A. Crystallographic and

electrophilic fragment screening of the SARS-CoV-2 main protease. *Nature Communications* **2020**, *11*, 5047.

(49) Cantrelle, F.; Boll, E.; Brier, L.; Moschidi, D.; Belouzard, S.; Landry, V.; Leroux, F.; Dewitte, F.; Landrieu, I.; Dubuisson, J.; Deprez, B.; Charton, J.; Hanoulle, X. NMR Spectroscopy of the Main Protease of SARS-CoV-2 and Fragment-Based Screening Identify Three Protein Hotspots and an Antiviral Fragment. *Angewandte Chemie International Edition* **2021**, *60*, 25428–25435.

(50) Krishnamoorthy, N.; Fakhro, K. Identification of mutation resistance coldspots for targeting the SARS-CoV2 main protease. *IUBMB Life* **2021**, *73*, 670–675.

(51) Günther, S.; Reinke, P. Y. A.; Fernández-García, Y.; Lieske, J.; Lane, T. J.; Ginn, H. M.; Koua, F. H. M.; Ehrt, C.; Ewert, W.; Oberthuer, D.; Yefanov, O.; Meier, S.; Lorenzen, K.; Krichel, B.; Kopicki, J.-D.; Gelisio, L.; Brehm, W.; Dunkel, I.; Seychell, B.; Gieseler, H.; Norton-Baker, B.; Escudero-Pérez, B.; Domaracky, M.; Saouane, S.; Tolstikova, A.; White, T. A.; Hänle, A.; Groessler, M.; Fleckenstein, H.; Trost, F.; Galchenkova, M.; Gevorkov, Y.; Li, C.; Awel, S.; Peck, A.; Barthelmess, M.; Schlünzen, F.; Lourdu Xavier, P.; Werner, N.; Andaleeb, H.; Ullah, N.; Falke, S.; Srinivasan, V.; França, B. A.; Schwinzer, M.; Brognaro, H.; Rogers, C.; Melo, D.; Zaitseva-Doyle, J. J.; Knoska, J.; Peña-Murillo, G. E.; Mashhour, A. R.; Hennicke, V.; Fischer, P.; Hakanpää, J.; Meyer, J.; Gribbon, P.; Ellinger, B.; Kuzikov, M.; Wolf, M.; Beccari, A. R.; Bourenkov, G.; von Stetten, D.; Pompidor, G.; Bento, I.; Panneerselvam, S.; Karpics, I.; Schneider, T. R.; Garcia-Alai, M. M.; Niebling, S.; Günther, C.; Schmidt, C.; Schubert, R.; Han, H.; Boger, J.; Monteiro, D. C. F.; Zhang, L.; Sun, X.; Pletzer-Zelgert, J.; Wollenhaupt, J.; Feiler, C. G.; Weiss, M. S.; Schulz, E.-C.; Mehrabi, P.; Karničar, K.; Usenik, A.; Loboda, J.; Tidow, H.; Chari, A.; Hilgenfeld, R.; Uetrecht, C.; Cox, R.; Zaliani, A.; Beck, T.; Rarey, M.; Günther, S.; Turk, D.; Hinrichs, W.; Chapman, H. N.; Pearson, A. R.; Betzel, C.; Meents, A. X-ray screening

identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021**, *372*, 642–646.

(52) Dubanevics, I.; McLeish, T. C. B. Computational analysis of dynamic allostery and control in the SARS-CoV-2 main protease. *Journal of The Royal Society Interface* **2021**, *18*, 20200591.

(53) Campitelli, P.; Lu, J.; Ozkan, S. B. Dynamic allostery highlights the evolutionary differences between the CoV-1 and CoV-2 main proteases. *Biophysical Journal* **2022**, *121*, 1483–1492.

(54) Verma, S.; Pandey, A. K. Factual insights of the allosteric inhibition mechanism of SARS-CoV-2 main protease by quercetin: an in silico analysis. *3 Biotech* **2021**, *11*, 67.

(55) Sencanski, M.; Perovic, V.; Pajovic, S. B.; Adzic, M.; Paessler, S.; Glisic, S. Drug Repurposing for Candidate SARS-CoV-2 Main Protease Inhibitors by a Novel In Silico Method. *Molecules* **2020**, *25*, 3830.

(56) Flynn, J. M.; Samant, N.; Schneider-Nachum, G.; Barkan, D. T.; Yilmaz, N. K.; Schiffer, C. A.; Moquin, S. A.; Dovala, D.; Bolon, D. N. Comprehensive fitness landscape of SARS-CoV-2 Mpro reveals insights into viral resistance mechanisms. *eLife* **2022**, *11*, e77433.

(57) Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J.; de Wilde, A.; Snijder, E. J.; Liu, H.; Hilgenfeld, R. α-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *Journal of Medicinal Chemistry* **2020**, *63*, 4562–4578.

(58) Anand, K.; Palm, G. J.; Mesters, J. R.; Siddell, S. G.; Ziebuhr, J.; Hilgenfeld, R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *The EMBO Journal* **2002**, *21*, 3213–3224.

(59) Ziebuhr, J.; Heusipp, G.; Siddell, S. G. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *Journal of Virology* **1997**, *71*, 3992–3997.

(60) Nashed, N. T.; Kneller, D. W.; Coates, L.; Ghirlando, R.; Aniana, A.; Kovalevsky, A.; Louis, J. M. Autoprocessing and oxyanion loop reorganization upon GC373 and nirmatrelvir binding of monomeric SARS-CoV-2 main protease catalytic domain. *Communications Biology* **2022**, *5*.

(61) Aniana, A.; Nashed, N. T.; Ghirlando, R.; Coates, L.; Kneller, D. W.; Kovalevsky, A.; Louis, J. M. Insights into the mechanism of SARS-CoV-2 main protease autocatalytic maturation from model precursors. *Communications Biology* **2023**, *6*.

(62) Lim, L.; Shi, J.; Mu, Y.; Song, J. Dynamically-Driven Enhancement of the Catalytic Machinery of the SARS 3C-Like Protease by the S284-T285-I286/A Mutations on the Extra Domain. *PLoS ONE* **2014**, *9*, e101941.

(63) Barrila, J.; Bacha, U.; Freire, E. Long-Range Cooperative Interactions Modulate Dimerization in SARS 3CLpro. *Biochemistry* **2006**, *45*, 14908–14916.

(64) Chen, S.; Zhang, J.; Hu, T.; Chen, K.; Jiang, H.; Shen, X. Residues on the Dimer Interface of SARS Coronavirus 3C-like Protease: Dimer Stability Characterization and Enzyme Catalytic Activity Analysis. *Journal of Biochemistry* **2007**, *143*, 525–536.

(65) Koudelka, T.; Boger, J.; Henkel, A.; Schönherr, R.; Krantz, S.; Fuchs, S.; Rodríguez, E.; Redecke, L.; Tholey, A. N-Terminomics for the Identification of In Vitro Substrates and Cleavage Site Specificity of the SARS-CoV-2 Main Protease. *PROTEOMICS* **2020**, *21*, 2000246.

(66) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. The crystal struc-

tures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proceedings of the National Academy of Sciences* **2003**, *100*, 13190–13195.

(67) Ho, B.-L.; Cheng, S.-C.; Shi, L.; Wang, T.-Y.; Ho, K.-I.; Chou, C.-Y. Critical Assessment of the Important Residues Involved in the Dimerization and Catalysis of MERS Coronavirus Main Protease. *PLOS ONE* **2015**, *10*, e0144865.

(68) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(69) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **1998**, *102*, 3586–3616.

(70) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2017**, *14*, 71–73.

(71) Larsson, P.; Kneiszl, R. C.; Marklund, E. G. MkVsites: A tool for creating GROMACS virtual sites parameters to increase performance in all-atom molecular dynamics simulations. *Journal of Computational Chemistry* **2020**, *41*, 1564–1569.

(72) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation* **2007**, *4*, 116–122.

(73) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G.

A smooth particle mesh Ewald method. *Journal of Chemical Physics* **1995**, *103*, 8577–8593.

(74) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *Journal of Chemical Physics* **2007**, *126*, 014101.

(75) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **1984**, *81*, 3684–3690.

(76) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182–7190.

(77) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **1995**, *6*, 277–293.

(78) Bax, A.; Grishaev, A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Current Opinion in Structural Biology* **2005**, *15*, 563–570.

(79) Glykos, N. M. Carma: A molecular dynamics analysis program. *Journal of Computational Chemistry* **2006**, *27*, 1765–1768.

(80) Girvan, M.; Newman, M. E. J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **2002**, *99*, 7821–7826.