THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Biophysical Interpretation of Evolutionary Consequences on the SARS-CoV2 Main Protease through Molecular Dynamics Simulations and Network Topology Analysis

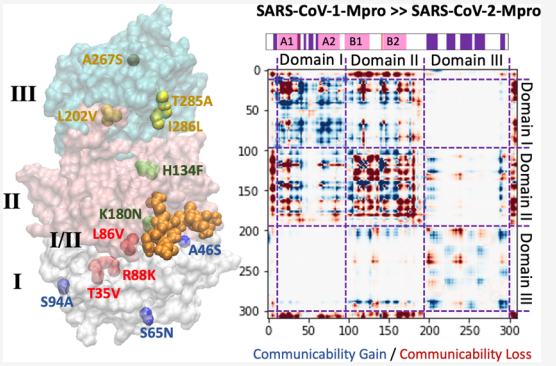
Nuttawat Sawang, Saree Phongphanphanee, Jirasak Wong-ekkabut, and Thana Sutthibutpong*

Cite This: *J. Phys. Chem. B* 2023, 127, 2331−2343

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In this study, we present a combined analysis procedure between atomistic molecular dynamics (MD) simulations and network topology to obtain more understanding on the evolutionary consequences on protein stability and substrate binding of the main protease enzyme of SARS-CoV2. Communicability matrices of the protein residue networks (PRNs) were extracted from MD trajectories of both Mpro enzymes in complex with the nsp8/9 peptide substrate to compare the local communicability within both proteases that would affect the enzyme function, along with biophysical details on global protein conformation, flexibility, and contribution of amino acid side chains to both intramolecular and intermolecular interactions. The analysis displayed the significance of the mutated residue 46 with the highest communicability gain to the binding pocket closure. Interestingly, the mutated residue 134 with the highest communicability loss corresponded to a local structural disruption of the adjacent peptide loop. The enhanced flexibility of the disrupted loop connecting to the catalytic residue Cys145 introduced an extra binding mode that brought the substrate in proximity and could facilitate the reaction. This understanding might provide further help in the drug development strategy against SARS-CoV2 and prove the capability of the combined techniques of MD simulations and network topology analysis as a "reverse" protein engineering tool.



1. INTRODUCTION

The global outbreak of COVID-19 through the infection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV2) has caused over 600 million reported cases and over 6 million deaths around the world within 33 months (as of September 2022) and caused significant public health and economic impacts.[1,2] Although the severity of infections has reduced so that COVID-19 has been announced as an "endemic" in many countries,[3] attempts toward drug and immunity booster development have been continuing to lead to the "herd immunity" of the global population.[4,5] SARS-CoV2, classified in the subgenus *sarbecovirus* of the genus *betacoronavirus*, is one of the enveloped, positive-sense, single-stranded RNA viruses.[6] A SARS-CoV2 virion, as well as other types of coronaviruses, consists of structural proteins to form spherical shells to contain the viral RNA and spike glycoproteins to attach and fuse with the target cell.[7] Meanwhile, the nonstructural proteins (NSPs) play their roles of transcription factors and enzymes with other biological functions.[8] The genome of SARS-CoV2 contains around 32 kb, in which two-third of the genome length is translated into polyproteins pp1a and pp1ab.[9,10] The nonfunctional polyproteins are processed by the 3C-like protease or the main protease (Mpro),[11] along with one or two papain-like proteases,[12] to provide 16 fully functional nonstructural proteins. Therefore, Mpro becomes one of the important targets for drug development against SARS-CoV2,[13−15] inspiring many X-ray crystallographic[11,16,17] and computational studies to search for either possible drug repurposing[18−22] or alternative immunity boosts from herbal extracts.[23]

The structure of an Mpro as shown in Figure 1a consists of three domains: domains I (residues 15−99) and II (residues 1−14, 100−197) consist of antiparallel β-barrel structures, while domain III (residues 198−301) forms a compact α-helical domain connected to domain II by a long linker loop. The nomenclature for each α helix and β strand within each domain is also provided in Figure 1a along with the amino acid sequence. The active site is in a cleft between domains I and II, and it holds a histidine/cysteine catalytic dyad. In the SARS-CoV2 main protease, Cys145 acts as a nucleophile during the first step of the hydrolysis reaction, assisted by His41 as a base

Received: November 27, 2022
Revised: February 25, 2023
Published: March 13, 2023

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

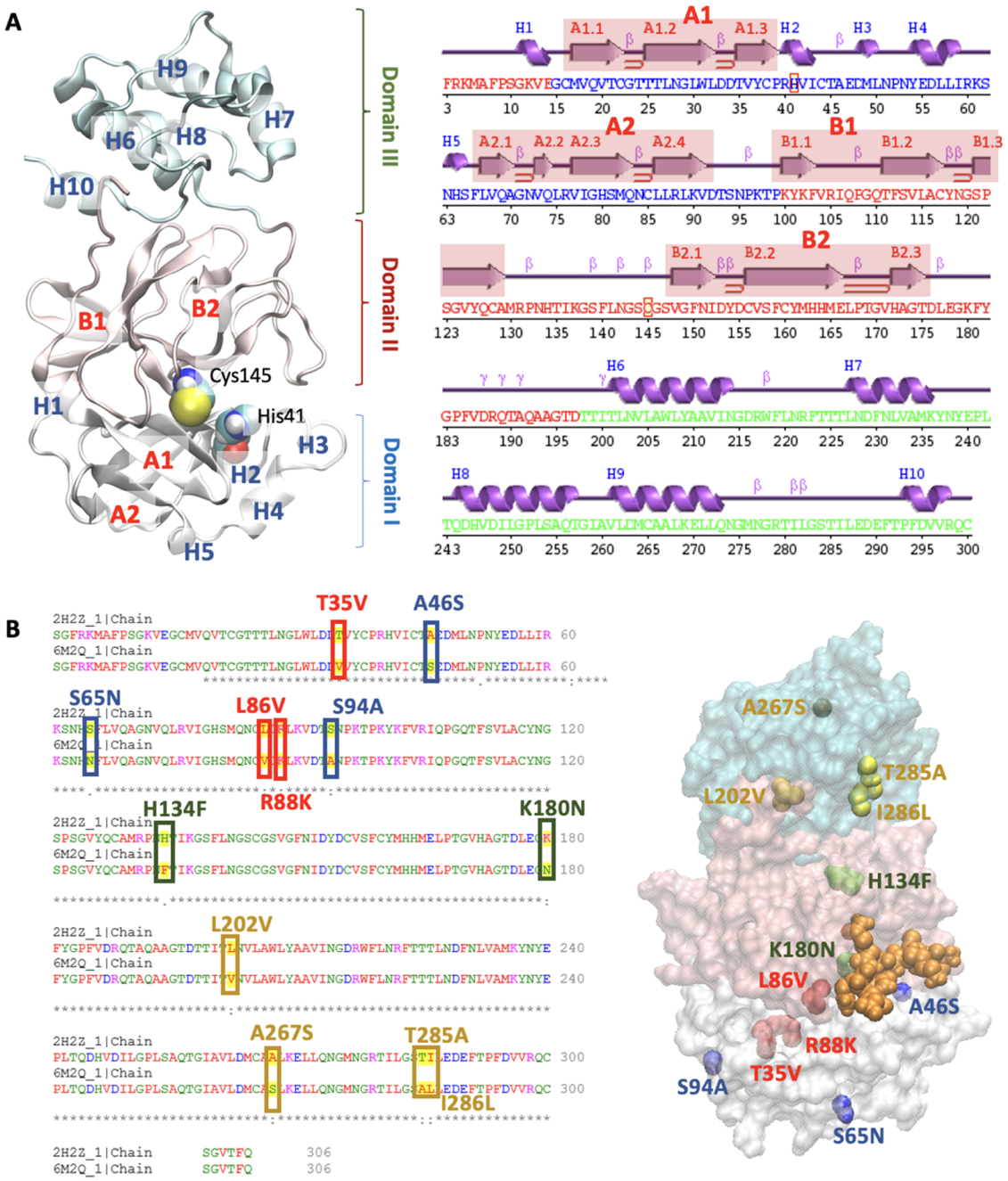


Figure 1. (a) Amino acid sequence and three-dimensional model of a SARS-CoV1 main protease (CoV1-MPro) enzyme showing three different domains. Catalytic residues H41 and C145 were highlighted, and nomenclatures are also given for each α helix and β sheet. (b) Amino acid sequence alignment between a SARS-CoV1 main protease (CoV1-MPro) and a SARS-CoV2 main protease (CoV2-MPro). All 12 mutated sites are highlighted on the three-dimensional model.

catalyst.[24,25] Meanwhile, X-ray crystallographic studies revealed that domain III facilitated a dimerization of Mpro protomers,[11] which could increase the catalysis efficiency. One of the strategies for the drug design and development of SARS-CoV2 is utilizing the comparative view with SARS-CoV or SARS-CoV1 of the previous outbreak. It is found that the main protease enzyme of SARS-CoV-2 (CoV2-Mpro) is highly conserved relative to the main protease of SARS-CoV-1 (CoV1-Mpro) so that it has become an interesting drug target, as some inhibitors targeting CoV1-Mpro could also be active against CoV2-Mpro.[11] Figure 1b displays the aligned amino acid sequences of the main protease enzymes of SARS-CoV1 (CoV1-Mpro; referred from PDB ID: 2H2Z) and SARS-CoV2 (CoV2-Mpro; referred from PDB ID: 6M2Q), in which 12 mutated amino acid residues ca n be seen and more than 96% of the amino acids are conserved. Six mutations in domain I consisting of A46S, S65N, and S94A are found near the outer surface, while T35V, L86V, and R88K mutations are found near the interface between domains I and II. For domain II mutations, K180N appears near the substrate binding site, while H134F appears near the outer surface. The L202V mutation is found near the interface between domains II and III, while A267S is found buried inside domain III. Mutations T285A and I286L are found at the outer surface of domain III,

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

and the sites have been proposed to contribute to an Mpro dimerization.[26]

An extensive review by B.Goyal and D.Goyal[27] provided some information about how Mpro dimerization could be an important drug target. Dimerization was proposed as a stabilizer for both C- and N-termini, which also corrected the orientation of the Mpro binding pockets. Moreover, Mpro was among the most conserved genes in the genomes of SARS-CoV1 and SARS-CoV2, so that any drugs that target CoV2-Mpro dimerization should also inhibit the dimerization of other variants. The same study also suggested that 12 mutated residues found in CoV2-Mpro rarely affected the protein structure and catalytic activity, as the 96% conserved protein structures were almost perfectly aligned. However, a previous study by Li et al.[28] on the Ser139-Leu141 loop of CoV1-Mpro disrupting the catalytic activity of an inactive Mpro protomer demonstrated that a conformational transition caused by the L141T mutation destabilized the Ser139-Leu141 loop but surprisingly maintained the catalytic activity of the mutated Mpro in a monomeric form, while another study by Rocha et al.[29] revealed the enhanced flexibility of the CoV2-Mpro active site that facilitated the entrance of both substrates and inhibitors. Therefore, indirect changes induced by naturally occurring mutations on an Mpro during its evolution from SARS-CoV1 to SARS-CoV2 have attracted our attention to "reverse-engineer" the molecular evolution of an important protein converting engine.

To investigate these subtle changes caused by altering chemical details of the 12 mutated side chains, and thus altering the interaction network created by amino acids contained in a protomer, network topology analysis was chosen as a tool.[30] Protein residue network (PRN) or residue interaction network (RIN) simplifies the definition of a folded protein structure by picking up a representative point from each amino acid residue as a "node," followed by defining a criterion for the connectedness or the "edge" between each pair of nodes. Therefore, protein folding can be characterized through the patterns of connecting edges and clustering parameters.[31] From some previous studies, PRN has been used for investigating molecular interactions,[32] active site identification,[33] and the identification of local networks with high betweenness,[34] where the nodes lie within paths between distal node pairs and act as a "mediator" for interactions or mechanical signals between nodes. Another study by Estrada[35] investigated the PRNs of CoV1-Mpro and CoV2-Mpro from the X-ray crystallographic database, employing the second and third kinds of network topological measurements on the communicability and the long-range communicability between residue pairs, signifying the transfer of mechanical or allosteric signals through different parts of residue networks, which could not be captured by only considering the nearest neighbors within adjacency matrices. The analysis of communicability matrices found that CoV2-Mpro possessed significantly higher communicability than CoV1-Mpro. Also, it was found that regions with high communicability were located near the binding pocket of the Mpro and represent their capability of mechanical signal transfer through the PRN, which was the hidden feature of the protein that could only be unveiled by network topology analysis.

In this work, we extended the investigation of Estrada (2020)[35] to investigate the enhanced communicability of CoV2-Mpro through the naturally occurring mutations from CoV1-Mpro by adding the protein dynamics within a simulated explicit solvent environment. Atomistic molecular dynamics (MD) simulations on the main protease (Mpro) enzymes of CoV1-Mpro and CoV2-Mpro in complex with an nsp8/9 peptide substrate were performed to generate an ensemble of folded protein configurations before performing the topological network analysis. The results obtained in terms of local communicability were then discussed along with protein conformational analysis from MD simulations to provide a biophysical interpretation of changes in the network topology and a more complete picture of the evolutionary consequence to the function of CoV2-Mpro for an even better drug development strategy.

2. METHODOLOGY

2.1. Molecular Dynamics Simulations and Analysis of Main Proteases from SARS-CoV1 and SARS-CoV2.

Atomistic structures of the main protease enzyme from SARS-CoV2 in complex with an nsp8/9 substrate were obtained from the crystallographic data (PDB ID: 7MGR).[36] The catalytic residue Cys145 of the crystallography sample underwent an alanine mutation, which prevented hydrolysis and preserved the nsp8/9 substrate to form a complex with a favorable binding mode. To prepare the starting structure for an atomistic simulation, a cysteine was added back to the catalytic residue 145 of the SARS-CoV2 main protease. Then, to obtain the atomistic structure of the main protease enzyme from SARS-CoV1 in complex with an nsp8/9 substrate with a similar binding mode to SARS-CoV2, homology modeling was performed through the SWISSMODEL web server[37] using the 7MGR structure as the template. After that, both structures were parameterized through the GROMOS54A7 forcefield before being explicitly solvated with the SPC water model before adding Na+ counterions to neutralize the charge within the box. A 5000-step energy minimization was then performed by the steepest descent algorithm, followed by an equilibration stage that gradually increased the temperature from 100 to 300 K within 1 ns by simulated annealing molecular dynamics (MD). After that, three replicas of 100 ns productive MD runs were performed for both CoV1-Mpro and CoV2-Mpro in complex with the nsp8/9 substrate under a constant temperature of 300 K and a constant pressure of 1 atm using the GROMACS 5.1.2 suite.[38] The temperature and pressure in all of these NPT simulations were regulated by velocity rescaling[39] and Nose−Hoover[40] algorithms. The six simulations of Mpro−nsp8/9 complexes were named according to their sources as CoV1-r0, CoV1-r1, CoV1-r2, CoV2-r0, CoV2-r1, and CoV2-r2.

After each simulation was complete, root-mean-square deviation (RMSD) was calculated along the whole 100 ns simulations using the starting structures as the references as a measure for global conformational changes, and per-residue root-mean-square fluctuation (RMSF) was performed for the last 50 ns of both trajectories to address the local flexibility of each amino acid residue. The radius of gyration (Rg) was calculated to measure the compactness of each domain within both proteases along the whole 100 ns MD trajectories. Additionally, the time series of Rg for the system containing domains I and II forming a binding cleft were also calculated for all trajectories. Numbers of hydrogen bonds between pairs of atom groups of interest along the MD trajectories were analyzed through the gmx hbond module within the GROMACS 5.1.2 suite, as the <3.5 Å donor−acceptor distance and the <45° donor−hydrogen-acceptor angle were

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

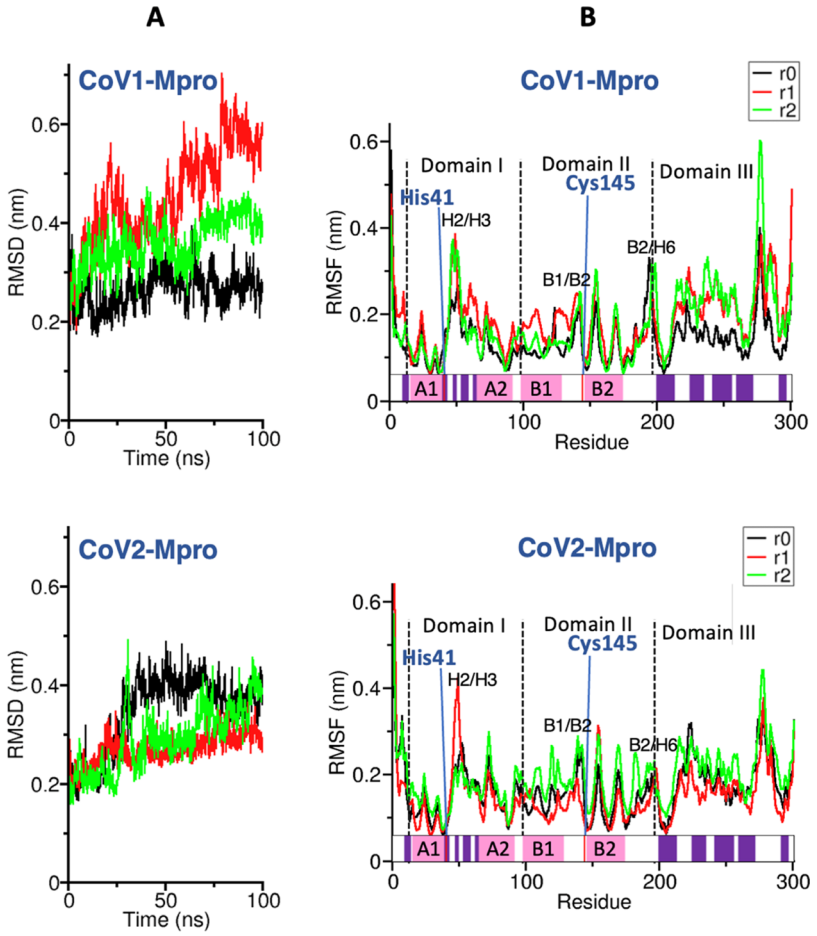


Figure 2. (a) Root-mean-square deviation (RMSD) calculated from all 100 ns MD trajectories of C-α atoms within domain I and domain II of CoV1-MPro (top) and CoV2-MPro (bottom), and (b) per-residue root-mean-square fluctuation (RMSF) of both structures calculated from the last 50 ns of MD trajectories. The colorbar at the bottom of the graph indicates the regions containing α helices (purple) and β strands/sheets (pink).

set as the criteria. Visual molecular dynamics (VMD) software[41] was used for all of the visualizations.

2.2. Network Topology Analysis. An in-house python script was created based on the commands from the "MDAnalysis" open library[42] to extract the coordinates of C-α atoms from all amino acid residues of both Mpro and nsp8/9 substrates and define the extracted coordinates as "nodes" of the protein residue network (PRN). "Edges" of the network were defined through the distance r_{ij} between C-α atoms of nodes i and j, as r_{ij} needs to be smaller than r_0 = 7.4 Å. The cutoff distance of 7.4 Å was specified from the first RDF shell of C-α atoms of all residue pairs, excluding the pairs connected by peptide bonds. This cutoff distance estimated the range of pairwise interactions between amino acids. As one of the amino acid residues was perturbed, the motion of the interacting amino acid pairs within the cutoff distance would also be detected. Then, an element $A_{ij}(t)$ of an adjacency matrix $A(t)$ for a protein residue network (PRN) at a time step t was defined by

$$A_{ij}(t) = H(r_0 - r_{ij}) = \begin{cases} 1, & r_{ij} < r_0 \\ 0, & r_{ij} \geq r_0 \end{cases} \quad (1)$$

The adjacency matrix $A(t)$ not only described the connectivity between amino acid residues at any time step but also represented the ability of transmitting some mechanical signal between any node pair (i,j) with $A_{ij}(t) = 1$. Therefore, $A(t)$ represented a signal hopping between each "adjacent" residue pair, and its higher order $A^n(t)$ represented the number of possible paths of signal transmission between residue pairs requiring n hopping steps. The transmission of the mechanical signal between any node pairs, including nonadjacent amino acids, could then be described through the communicability matrix

$$G(t) = \exp(A(t)) = \sum_{n=0}^{\infty} \frac{A^n(t)}{n!} = \hat{I} + A(t) + \frac{A^2(t)}{2!} + \frac{A^3(t)}{3!} + \cdots \quad (2)$$

From eq 2, the number of transmission paths requiring n hopping steps was weighted out by $n!$, which could roughly estimate the lossy mechanical signal through complex biomolecules.

From both of the simulated MD trajectories of the main proteases from SARS-CoV1 and SARS-CoV2, coordinates of all C-α atoms were extracted from every 0.1 ns of the last 50 ns, as they contained an ensemble of equilibrated molecular configurations. Adjacency and communicability matrices were then calculated from all of the time steps before the

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

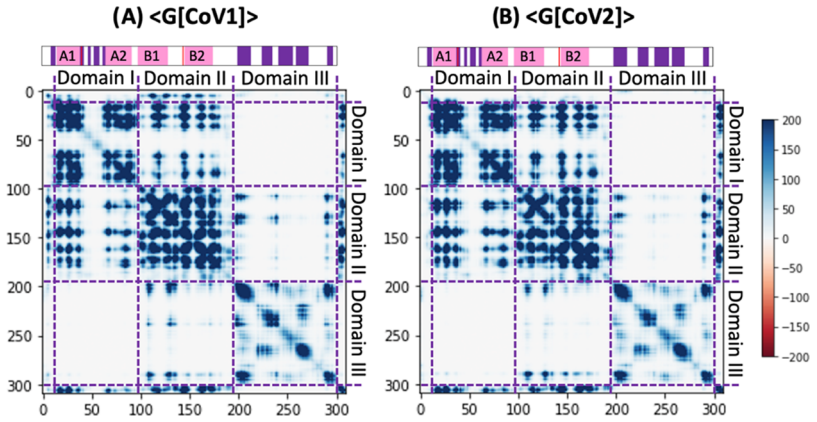


Figure 3. (a, b) Communicability matrices of the C-α atoms within (a) CoV1-Mpro and (b) CoV2-Mpro. Each matrix was the average over the last 50 ns of all three replicas.

calculations of time averages $A \equiv \langle A(t) \rangle$ and $G \equiv \langle G(t) \rangle$ were done along with the standard deviations.

The information contained within the time-averaged communicability matrix G could also be presented as submatrices, in which the average of all elements within a submatrix of interest could be interpreted as the communicability between domains or communicability between secondary structures. As Mpro proteins in this study consisted of three domains and the cleft between domains I and II was the catalytic site, the average communicability between each pair of domains was calculated. Furthermore, the communicability degree of a residue or node i or $G_i \equiv \sum_j G_{ij}$ was defined as the row sum of communicability that involved the residue or node i. Either the row sum or the average communicability within each row of G could be interpreted as the ability of each individual node to transmit mechanical signals, along with the subgraph centrality ($SC_i \equiv G_{ii}$) previously defined from the diagonal elements of G.[43]

Moreover, the difference in the matrix element values between those of communicability matrices calculated from the simulations of main proteases from SARS-CoV1 and SARS-CoV2 or $\Delta G_{ij} = G_{ij}[CoV2 - MPro] - G_{ij}[CoV1\text{-}MPro]$ was defined as the "communicability gain" through mutations. In the case of obtaining a negative value, ΔG_{ij} became "communicability loss," instead. Indices i and j in ΔG_{ij} could represent residues, secondary structures, or domains. For the calculations of communicability gain/loss between secondary structures and domains, communicability gain/loss will be averaged over all residue pairs within the submatrix or the subnetwork of interest. The reference communicability matrix prior to mutations, which contained elements $G_{ij}[CoV1 - MPro]$, was defined through the average of the communicability matrix from the last 50 ns of all three replicas of CoV1-Mpro simulations. In this study, the matrices representing communicability gain/loss for all replicas of CoV2-Mpro simulations were visualized, along with the communicability gain/loss at all mutated residues determined from the average of ΔG_{ij} over the rows of mutated residues to address the effects of mutation.

3. RESULTS

3.1. Decreased Flexibility of the SARS-CoV2 Main Protease near the Active Sites. Atomistic molecular dynamics (MD) simulations were performed under an explicit solvent for two protease/peptide complexes consisting of the main proteases of SARS-CoV1 (CoV1-Mpro) and SARS-CoV2 (CoV2-Mpro) interacting with the nsp8/9 substrate. After 100 ns production run of each replica, the root-mean-square deviation (RMSD) was calculated. From Figure 2a, the RMSDs of peptide-bound Mpro proteins were found to converge after 50 ns as the fluctuation of RMSD was significantly smaller than the RMSD values, suggesting that the trajectories were well equilibrated. Therefore, the last 50 ns of all simulations were considered for further analysis. To visualize the consequence of molecular evolution from CoV1-Mpro to CoV2-Mpro on the flexibility of protein regions, per-residue root-mean-square fluctuation (RMSF) was calculated from the last 50 ns of both simulations (see Figure 2b). For both CoV1-Mpro and CoV2-Mpro proteins, peaks of RMSFs represented the flexible loops connecting α helices, while the local minima of RMSFs represented either well-defined α helices or β strands. According to the profiles, the α helices H2 containing the His41 catalytic residue had very low RMSFs for all replicas of CoV2-Mpro simulations. The nearby H2/H3 flexible loops of CoV2-r0 and CoV2-r2 simulations had relatively low RMSFs compared with those of the CoV2-r1 simulation and all of the CoV1-Mpro simulations. However, the B1/B2 loop connecting two β sheets within domain II and containing another catalytic residue Cys145 for CoV2-r1 had the lowest RMSF when compared with all other simulations. Remarkably, RMSF calculated from the B2/H6 loop region connecting domains II and III of all CoV2-Mpro replicas was lower than those of all CoV1-Mpro replicas. Further elucidation of these conventional analyses of proteins from MD simulations was carried out later in this study through network topology analysis for the interrelated dynamical behaviors of amino acid units.

3.2. Characterization of Protein Residue Networks of CoV1-Mpro and CoV2-Mpro through Communicability Matrices: Some Common Features. Protein residue networks (PRNs) of CoV1-Mpro and CoV2-Mpro, along with their binding nsp8/9 peptides, were created from the C-α atoms of all amino acid residues, defined as nodes. The edges of the network were defined from pairs of nodes with geometrical distance lower than the 7.4 Å cutoff, defined by the first peak of the radial distribution function (RDF) between all C-α atoms. The information of the defined edges was contained within elements of the adjacency matrix (A) for each protein structure. Then, the communicability matrix (G) was defined by the exponent of the adjacency matrix so that

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

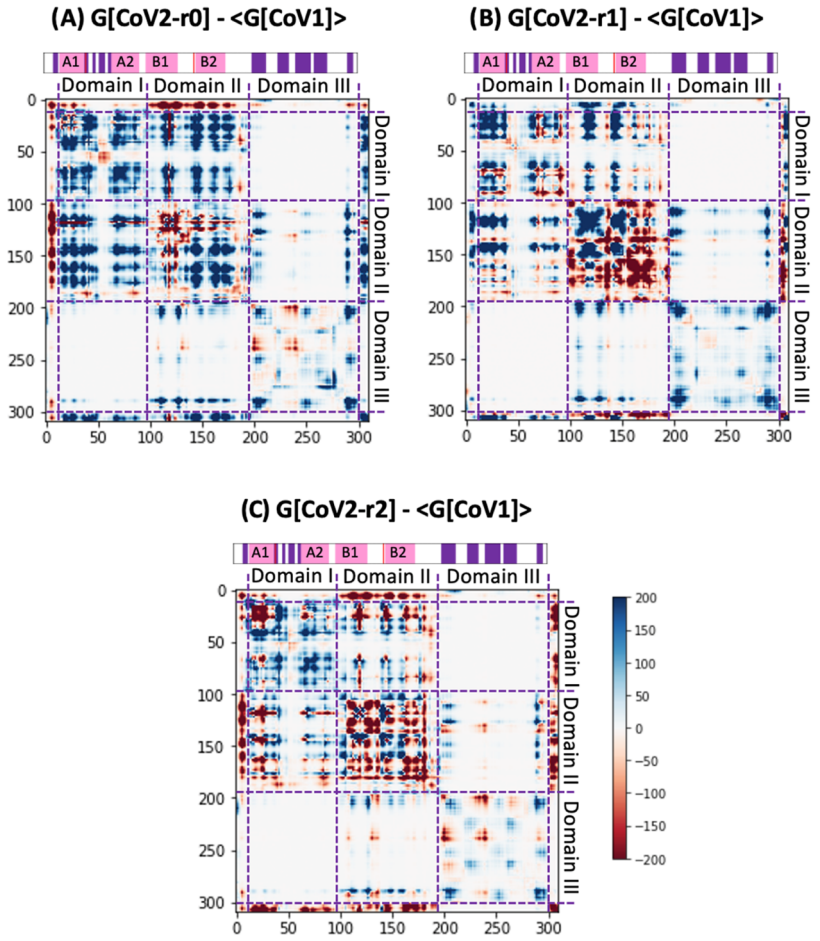


Figure 4. (a−c) Difference between the communicability matrix calculated from the last 50 ns of each replica of CoV2-Mpro (G[CoV2-r0], G[CoV2-r1], and G[CoV2-r2]) and the averaged communicability matrix over the three replicas of CoV1-Mpro simulations (⟨G[CoV1]⟩). Red and blue colors represent the regions where the CoV1-Mpro and the CoV2-Mpro replica possessed higher communicability, respectively.

each element of G became the weighted sum of the number of paths connecting each pair of nodes. Therefore, an element of G for a PRN could be interpreted as the amount of mechanical signal between a pair of amino acids. Extracting A and G matrices from each conformational snapshot of an MD trajectory resulted in the time dependence of matrix element values within the dynamic networks. Therefore, an ensemble of possible network configurations for a PRN could be generated from an equilibrated MD trajectory.

Figure 3a,b displays the time-averaged communicability matrix (G) for the protein/peptide complexes containing CoV1-Mpro and CoV2-Mpro, respectively. Both complexes displayed common traits on several regions:

(i) high communicability of amino acids within and between the β sheets A1 and A2 from the protease domain I;

(ii) low communicability of the regions involving two flexible loops with short α helices H2−H5;

(iii) high communicability of amino acids within and between the β sheets B1 and B2 from protease domain II;

(iv) high communicability between the amino acids within the β sheets A1 and A2 from protease domain I and the amino acids within β sheets B1 and B2 from protease domain II;

(v) high communicability of amino acids between the α helices from protease domain III;

(vi) low communicability between the amino acids within domain III and the amino acids within other domains; and

(vii) high communicability between the amino acids of the nsp8/9 peptide substrates and the β sheets A1, A2, B1, and B2 from domains I and II, along with the long B1/B2 loop that contained the catalytic residue Cys145.

The common characteristics of communicability for both CoV1-Mpro and CoV2-Mpro mentioned in (iv) and (vii) depicted the formation of a binding cleft through the interaction network between domain I and domain II and the interaction network between the cleft and the substrate.

3.3. Local Difference in the Communicability and Interaction Network Induced by Naturally Occurring Mutations. Difference in the time-averaged communicability matrices (G) of the PRN from CoV1-Mpro/peptide and CoV2-Mpro/peptide complexes is shown in Figure 4. Time-averaged communicability matrix from the last 50 ns of each CoV2-Mpro replica was compared to the time-averaged communicability matrix from the last 50 ns of all CoV1-Mpro replicas. Common features observed from all MD replicas included significant communicability gain for the network of amino acids within two β sheet regions of domain I, which contained six mutated amino acid residues. In domain

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

II, communicability loss was observed between the B1 and B2 sheets for all CoV2-Mpro replicas. However, the communicability loss within each β sheet region of domain II became unique for each CoV2-Mpro replica as the CoV2-r0 simulation contained only the communicability loss within B1, while the CoV2-r1 simulation contained the communicability loss within a half of the B2 region and the communicability loss was found within different regions of both B1 and B2. For domain III, about an equal amount of communicability gain and loss was found for the CoV2-r0 and CoV2-r2 simulations, while CoV2-r1 possessed a communicability gain within domain III.

To further understand the molecular mechanisms underlying the communicability gain/loss in Figure 4 and to argue about the consequence of naturally occurring mutations of Mpro from SARS-CoV1 to SARS-CoV2, all 12 mutations are characterized in Table 1 by their location and the alteration

Table 1. Communicability Gain/Loss (dG_i) and Subgraph Centrality (SC_i) of All 12 Mutated Residues within Different Domains and Regions of CoV2-Mpro Relative to Those in CoV1-Mpro

residues	domain	region	mutation	SC_i	dG_i
L86V	I	A2	reduce size	649.9	15.35
R88K	I	A2	reduce size	810.4	9.46
A46S	I	H2/H3	adding polar group	85.2	8.41
L202V	III	H6	reduce size	314.4	6.94
S65N	I	A2	adding polar group	144.7	6.26
T35V	I	A1	removing polar group	646.5	6.09
H134F	II	B1/B2	removing polar group	260.1	2.81
T285A	III	H9/H10	removing polar group	75.0	2.59
I286L	III	H9/H10	isomer	41.8	2.09
S94A	I	A2/B1	removing polar group	117.8	2.04
A267S	III	H9	adding polar group	382.4	1.94
K180N	II	B2/H6	removing pos. charge	249.3	−4.44

of side-chain properties, along with the communicability gain/loss (dG_i) or the difference in average communicability of each residue. The dG_i values in Table 1 are the averaged values over three replicas of CoV2-Mpro simulations. The subgraph centrality (SC_i) of each mutated residue averaged between those of CoV1-Mpro and CoV2-Mpro is also presented in Table 1. As the subgraph centrality represents the number of possible paths for transferring the mechanical signal from a given residue, it could also be inferred as the potential of a mutated residue to affect nearby residues.

From Table 1, the L86V and R88K mutations had the highest communicability gain through reducing the sizes of their amino acid side chains while maintaining the hydrophobicity of residue 86 and the electrostatic property of residue 88. Their highest subgraph centrality was in concurrence with the location of both residues at the interfacial region between domains I and II. The next highest communicability gain was found for the A46S mutation through an additional polar group, for which the low subgraph centrality was in concurrence with the location of residue 46 at the outer surface of domain I. The L202V mutation at the interfacial region between domains II and III with the reduced size of the hydrophobic side chain was also associated with a substantial communicability gain, as well as an addition of the polar group within the S65N mutation at the surface of domain I. Interestingly, the T35V mutation with a polar group removal, which was located near the L86V and R88K mutated sites, also caused a high communicability gain, while the other three mutations with polar group removal were associated with less significant communicability gain. Finally, the only mutation fully associated with the communicability loss in domain II was the K180N mutation, in which a positively charged side chain was turned into a polar but uncharged side chain.

Figure 5 displays the effects of the additional hydroxyl group of a serine side chain associated with A46S mutation. Mutation at residue 46 became of interest due to the high communicability gain and its location near the binding cleft and the nsp8/9 substrate. The superimposed snapshots showing the positions of Ala46 and the nsp8/9 substrate from all MD replicas of CoV1-Mpro complexes in Figure 5a showed that those two atom groups were rarely in contact and hence the less probability of hydrogen-bond formation between the nsp8/9 substrate and the H2/H3 loop. The average number of hydrogen bonds between the nsp8/9 substrate and the H2/H3 loop from the last 50 ns of all three simulations was only 0.04.

In the case of the A46S mutation within the CoV2-Mpro complexes, superimposed snapshots from Figure 5b displayed different modes of contacts between Ser46 and nsp8/9 among the three MD replicas of CoV2-Mpro complexes. Snapshots from the last 50 ns of the CoV2-r0 simulation displayed contacts between the mutated Ser46 residue and the C-terminus of the nsp8/9 substrate, for which the average number of hydrogen bonds between the H2/H3 loop and the peptide substrate over the last 50 ns was around 2.28. Snapshots from the CoV2-r1 and CoV2-r2 simulations displayed the contacts between the H2/H3 loop and the N-terminal of nsp8/9. Time-averaged number of hydrogen bonds calculated from the last 50 ns of the CoV2-r1 and CoV2-r2 simulations were found to be around 0.55 and 0.39, respectively. The increased number of hydrogen bonds due to the A46S mutation corresponded to the addition of a hydroxyl (-O-H) group, which could be either a hydrogen-bond donor or acceptor. Additional hydrogen bonds between residue 46 from domain I and the nsp8/9 substrate residing between domains I and II could contribute to the average communicability gain between domains I and II mentioned earlier.

Figure 6 displays the superimposed snapshots of the regions associated with the communicability loss from the mutation K180N. Superimposed snapshots from the last 50 ns of all three CoV1-Mpro simulations in Figure 6a display the loop connecting B1.1 and B1.2 strands within the β sheet B, which also contained the positively charged Arg105 and the hydrophobic Ile106 residues. Next to residue Ile106, the B1/B2 loop connecting the B1 and B2 sheets also contained the mutated residue 134. Both B1.1/B1.2 and B1/B2 loops formed interaction networks with the B2/H6 loop connecting domains II and III. Time-averaged number of hydrogen bonds between B1.1/B1.2 and B2/H6 loops over the last 50 ns was found at 2.5 ± 1.1 for all MD replicas of CoV1-Mpro. Meanwhile, the average numbers of hydrogen bonds between the B1/B2 and B2/H6 loops were around 2.0 ± 1.3, 3.7 ± 1.1, and 3.6 ± 1.1 for the CoV1-r0, CoV1-r1, and CoV1-r2 simulations, respectively.

However, for the three MD replicas of CoV2-Mpro complexes in Figure 6b, the number of hydrogen bonds between loop regions associated with the mutations H134F and K180N decreased from those in CoV1-Mpro complexes; 2.4 ± 1.3, 1.1 ± 1.2, and 0.9 ± 0.7 hydrogen bonds were found

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

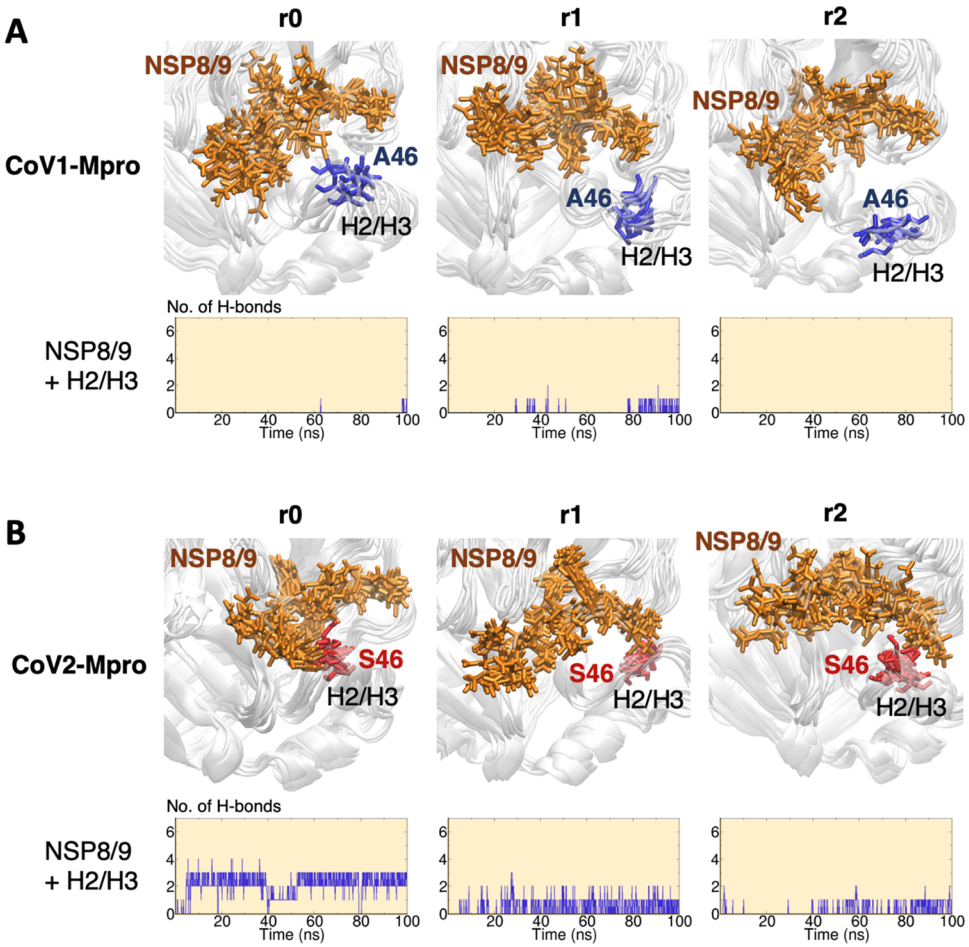


Figure 5. (a, b) Superimposed snapshots taken from the last 50 ns of all MD trajectories of (a) CoV1-Mpro (r0, r1, and r2 replicas) and (b) CoV2-Mpro (r0, r1, and r2 replicas). The nsp8/9 peptide substrate and the mutated residue 46 located within the H2/H3 loop are highlighted (top), with the number of hydrogen bonds between the nsp8/9 substrates and the H2/H3 loop counted along all MD replicas (bottom).

to contribute to the interaction network between B1.1/B1.2 and B2/H6 loops of the CoV2-r0, CoV2-r1, and CoV2-r2 simulations, respectively. Meanwhile, 1.2 ± 1.0, 1.2 ± 1.0, and 2.1 ± 1.0 hydrogen bonds were found between B1/B1 and B2/H6 loops of the same simulations, which were significantly lower than those of CoV1-Mpro simulations. Conformational snapshots showing the positions of important residues in Figure 6a,b showed that the mutated Phe134 tended to form additional hydrophobic contacts with Ile106, while the positive charge removal of the residue 180 resulted in a possible ionic bond between Arg105 and Glu178. These additional interactions caused by the H134F and K180N mutations occurred within the β sheet B and the B2/H6 loop themselves and thus competed with the interactions between different loops, which resulted in hydrogen-bonding loss. The loss of hydrogen bonding between different regions in domain II was in concurrence with the communicability loss observed from the network topology analysis. However, the low RMSF at B2/H6 loops for CoV2-Mpro simulations despite losing a number of hydrogen bonds suggested that the B2/H6 loops could be stabilized through the additional contacts between domains I and II, which in turn were reflected by the communicability gain between domains I and II.

3.4. Binding Cleft Closure and Substrate Locking through Local Communicability and Conformational Analysis. The consequences of amino acid alteration of CoV1-Mpro into CoV2-Mpro on a protein conformation were further investigated (Figure 7a). The superimposed snapshots of CoV1-Mpro displayed the connectivity between the nsp8/9 peptide substrate (orange) and domain II. However, superimposed snapshots of CoV2-Mpro displayed additional contacts between the substrate and the mutated residue 46 (highlighted in blue) in domain I, as shown in the hydrogen-bond analysis from an earlier section. The effects of all mutations mentioned above were further quantified by the pairwise averaged communicability between different protein domains and between each protein domain and the substrate, as shown in Table 2. Interestingly, the interaction network between domain I and the nsp8/9 substrate displayed an 8% communicability gain, which could be induced by the additional contact between nsp8/9 and mutated residue 46, while a 6% communicability loss was observed between domain II and nsp8/9. Moreover, the interaction network between domains I and II displayed a 22% communicability gain, suggesting that mutations from SARS-CoV1-Mpro to SARS-CoV2-Mpro induced a closure of the binding cleft formed between domains I and II.

To obtain further comparison between the communicability gain and other physical quantities, the radius of gyration (Rg) was calculated for domain I and domain II (Figure 7b) for all simulations. The CoV2-r0 and CoV2-r2 simulations possessed smaller Rg values for domain I and larger Rg values for domain

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

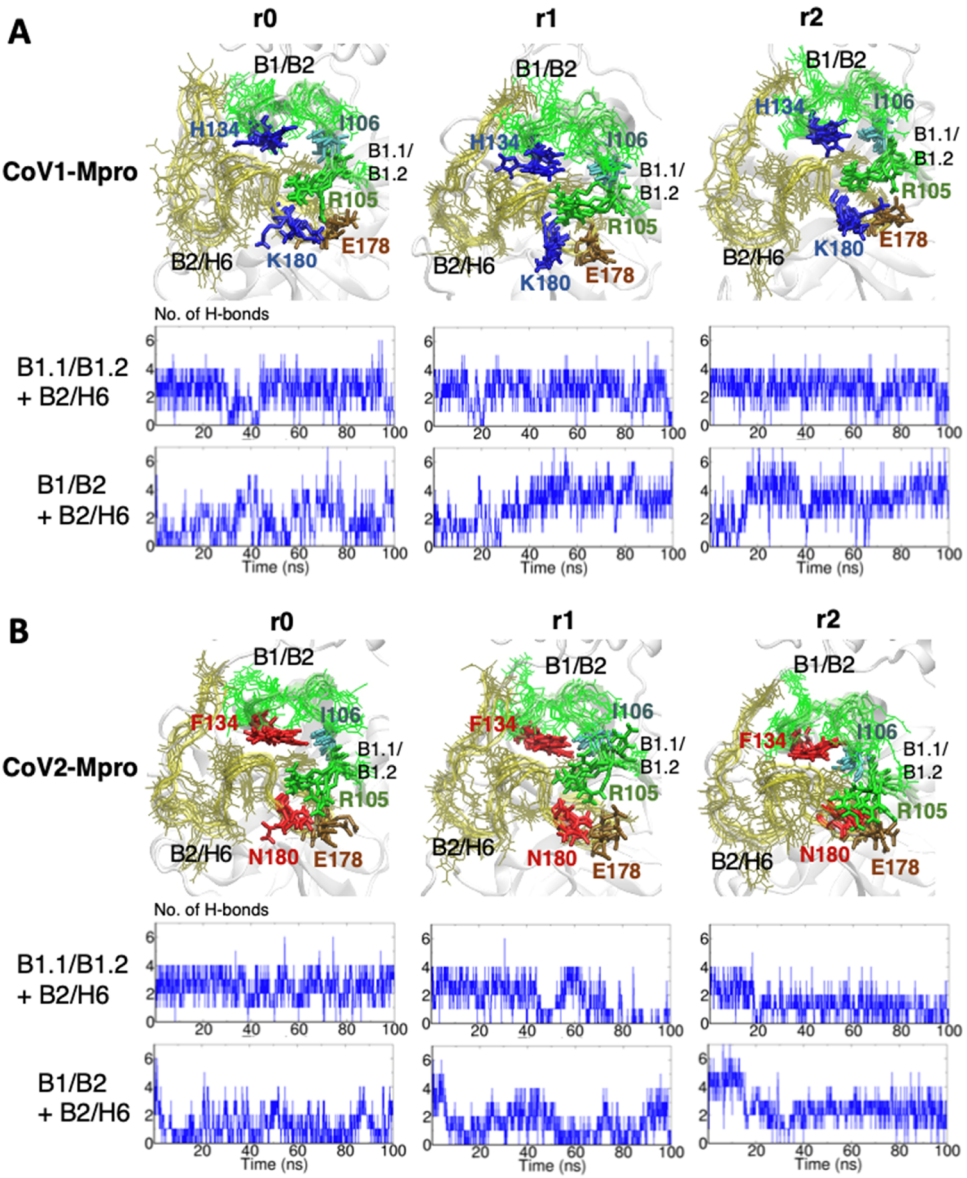


Figure 6. (a, b) Superimposed snapshots taken from the last 50 ns of all MD trajectories of (a) CoV1-Mpro (r0, r1, and r2 replicas) and (b) CoV2-Mpro (r0, r1, and r2 replicas). The loop connecting the B2 β sheet and H6 α helix regions (B2/H6; yellow), the loop connecting B1 and B2 β sheet regions (B1/B2; green), and the loop connecting B1.1 and B1.2 β strands (B1.1/B1.2; green) are highlighted along with the mutated residue 134 (H134F), the mutated residue 180 (K180N), and other residues involved in the subnetwork of interactions (top). The number of hydrogen bonds between B1 and B2/H6 regions and between B1/B2 and B2/H6 regions were counted along all MD replicas (bottom).

II than their CoV2 counterparts, which justified the local communicability gain within domain I and the local communicability loss within domain II. CoV2-r1 had a locally unfolded structure of the H2/H3 loop trapped by the B2.2/B2.3 loop, so Rg values were unusually high. Figure 7c displays the radius of gyration values Rg_{I+II} calculated along all of the 100 ns trajectories for the combined system of domains I and II. CoV2-r1 also showed an unusually high Rg_{I+II} contributed by the unfolded and trapped H2/H3 loop. Meanwhile, CoV2-r0 and CoV2-r2 displayed better convergence to 1.64 nm after 50 ns. The smallest Rg_{I+II} values after 50 ns of CoV2-r0 and CoV2-r2 could represent the binding cleft closure, and this was in concurrence with the communicability gain between domains I and II.

Domain I of both CoV1-Mpro and CoV2-Mpro consisted of two β sheets (A1 and A2) and four short α helices H2−H5 alternating with flexible loop regions, in which the mutated residue 46 was located within the H2/H3 loop between the H2 and H3 helices. While the Ala46 of CoV1-Mpro contained only a short hydrophobic side chain, the Ser46 of CoV2-Mpro contained an additional polar hydroxyl group. Two hydrogen bonds formed between Ser46 and a glutamic acid at the C-terminal of the nsp8/9 peptide substrate also brought H2 and H3 α helices into close contact with the surface of domain II. The closure of domains I/II corresponded to the significantly increased communicability between domain II and H2−H3 helices (Table 3). Interestingly, the increased communicability between domain II and H4−H5 helices also signified better mechanical signal transmission to the second nearest neighbors, which was caused by the A46S mutation.

The extended interaction networks and communicability between parts of domains I and II could be viewed as the

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

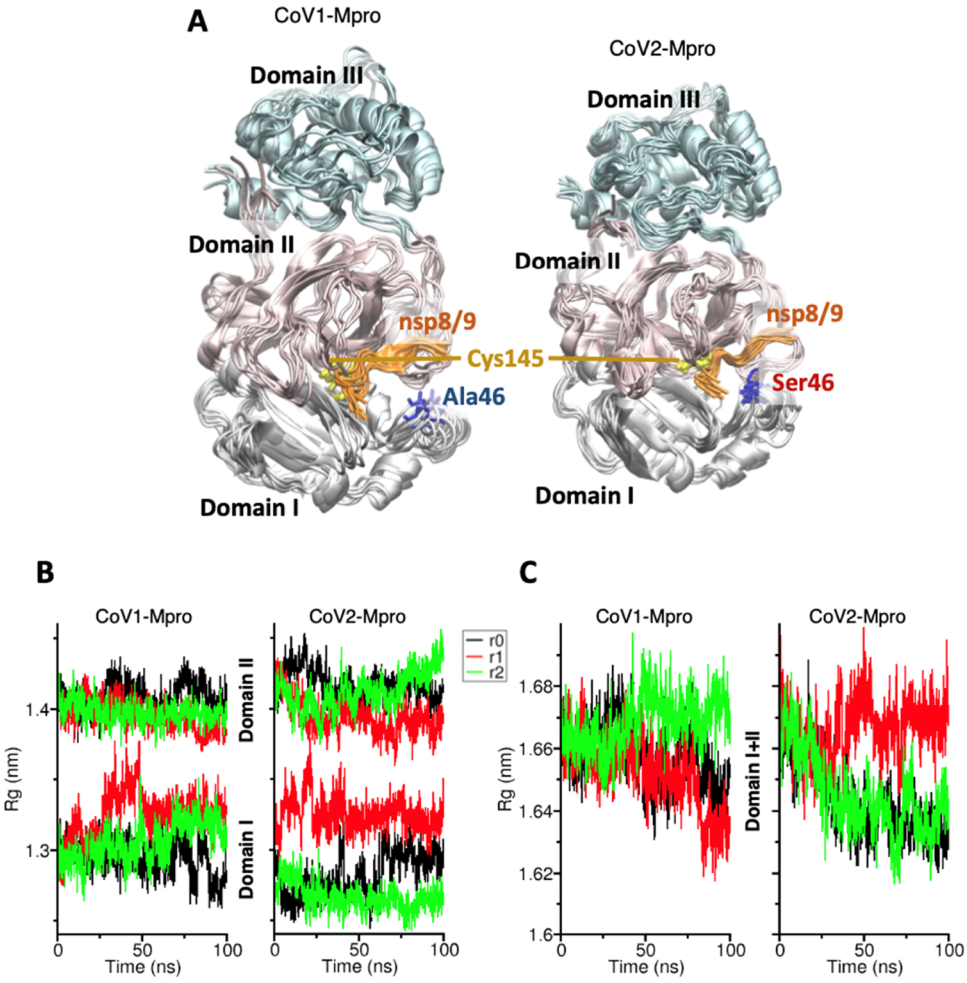


Figure 7. (a) Sample conformational snapshots taken from every 5 ns of equilibrated trajectories of the CoV1-Mpro-r0 and CoV2-Mpro-r0 simulations, highlighting all domains, nsp8/9 peptide substrates (orange), A46S mutated sites (blue), and Cys145 catalytic residues (yellow). (b) Radius of gyration calculated from domain I and domain II of CoV1-Mpro (left) and CoV2-Mpro (right) along 100 ns MD trajectories of the r0 (black), r1 (red), and r2 (green) replicas, and (c) radius of gyration calculated from the system containing both domains I and II for all MD replicas of CoV1-Mpro (left) and CoV2-Mpro (right).

Table 2. Averaged Communicability Calculated between All Pairs of Protein Domains and Substrates (i,j) = I, II, III, or nsp8/9, Communicability within Each Domain

$\langle G[\text{CoV1}]\rangle_{i,j}$	I	II	III	nsp8/9
I	161.88 ± 1.58	57.74 ± 1.04	0.16 ± 0.01	85.33 ± 2.98
II	57.74 ± 1.04	157.06 ± 3.28	8.00 ± 0.31	130.67 ± 1.91
III	0.16 ± 0.01	8.00 ± 0.31	49.23 ± 0.63	1.10 ± 0.07
nsp8/9	85.33 ± 2.98	130.67 ± 1.91	1.10 ± 0.07	237.10 ± 4.20
$\langle G[\text{CoV2}]\rangle_{i,j}$	I	II	III	nsp8/9
I	184.75 ± 1.44	70.32 ± 2.27	0.33 ± 0.01	92.30 ± 1.96
II	80.55 ± 2.02	159.25 ± 2.46	10.28 ± 0.43	123.16 ± 2.44
III	0.53 ± 0.02	10.28 ± 0.43	54.58 ± 0.48	1.18 ± 0.07
nsp8/9	127.53 ± 2.94	123.16 ± 2.44	1.18 ± 0.07	224.96 ± 4.60

Table 3. Averaged Communicability Calculated between the Whole Domain II and α Helices H2−H5 Containing Mutated Residue 46

$\langle G[\text{CoV1}]\rangle_{i,j}$	H2	H3	H4	H5
domain II	29.69 ± 1.32	5.67 ± 0.62	2.48 ± 0.11	19.81 ± 0.78
$\langle G[\text{CoV1}]\rangle_{i,j}$	H2	H3	H4	H5
domain II	42.67 ± 1.33	13.64 ± 1.04	2.92 ± 0.11	24.42 ± 0.90

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

closure of the binding pocket for CoV2-Mpro. Additional hydrogen bonds between the Ser46 of CoV2-Mpro and the nsp8/9 peptide also corresponded to the shorter minimum distance between the Cys145 active residue and the peptide substrate. From Table 4, enhancement of the communicability between the nsp8/9 peptide substrate and the Cys145 active residue found in CoV2-Mpro corresponded to the improved communicability between Cys145 and H2/H3 loops containing the mutated Ser46, which also confirmed the importance of A46S mutation that helped bring the peptide substrate closer to the active site, signifying the higher propensity of the hydrolysis reaction to occur when CoV2-Mpro was bound by the substrate.

Table 4. Averaged Communicability Calculated between the Catalytic Residue 145 and the nsp8/9 Substrate, and between the Catalytic Residue 145 and the H2/H3 Loop

$G[\text{CoV1-Mpro}]_{i,j}$	nsp8/9	H2/H3
Cys145	397.98 ± 7.35	30.69 ± 4.34
$G[\text{CoV2-Mpro}]_{i,j}$	nsp8/9	H2/H3
Cys145	409.78 ± 9.03	61.80 ± 2.38

Finally, the communicability gain of interaction subnetworks involving the nsp8/9 substrate was also reflected by the tendency of the substrate and catalytic residue Cys145 to come in proximity. Figure 8 displays the distance measured between the sulfur atom of catalytic residue Cys145 and the carbonyl group of the conserved Gln5 residue of the nsp8/9 substrate, signifying the tendency for a nucleophilic attack in the early stage of protein catalysis. The sulfur−carbonyl distance measured around 0.45 ± 0.05 nm during the last 50 ns of all CoV1-MPro simulations. Interestingly, bimodality was observed for the last 50 ns of all CoV2-MPro simulations, in which the sulfur−carbonyl distance could be either around 0.50 ± 0.05 nm or came in proximity for a nucleophilic attack around 0.40 ± 0.05 nm and could facilitate the catalytic reaction of CoV2-Mpro.

4. DISCUSSIONS AND CONCLUSIONS

In this study, we present an additional protein structure analysis procedure from atomistic MD trajectories based on network topology to gain more understanding of the evolutionary consequences on both protein stability and substrate binding at the molecular level. Communicability matrices were extracted from the protein residue networks of the main protease enzymes from SARS-CoV1 and SARS-CoV2, picked as our case study to elucidate the importance of mutations within binding domains I and II. Ensembles of protein−substrate complexes for network topological analysis were generated by atomistic molecular dynamics simulations of both SARS-CoV1 and SARS-CoV2 systems, equilibrated under explicit solvent environments. Also, postsimulation analyses were performed to provide more biophysical details on global protein conformation, flexibility, and contribution of amino acid side chains to both intramolecular and intermolecular interactions.

Conventional postsimulation analyses, e.g., RMSD and RMSF calculations, ensured that all simulations were well equilibrated after 50 ns and showed that CoV2-Mpro tended to be more mechanically stable than CoV1-Mpro, especially at domains I and II containing active residues and forming a binding cleft for a peptide substrate. Communicability matrices calculated from the protein residue networks of both proteins displayed some common characteristics, e.g., high communicability between α and β structures within the same domains, along with high interdomain communicability between domains I and II. The effects of mutations from CoV1-Mpro to CoV2-Mpro that enhanced the stability of peptide binding domains were still unclear up to this point.

Therefore, communicability gain/loss between each pair of amino acids or between any region was defined through the difference in communicability of amino acid pairs within CoV2-Mpro in comparison with its predecessor from CoV1-Mpro. To address the effects of 12 naturally occurring mutations on the interaction network, total communicability gain/loss was considered for all mutated residues, and six mutated sites were involved in binding domains I and II with

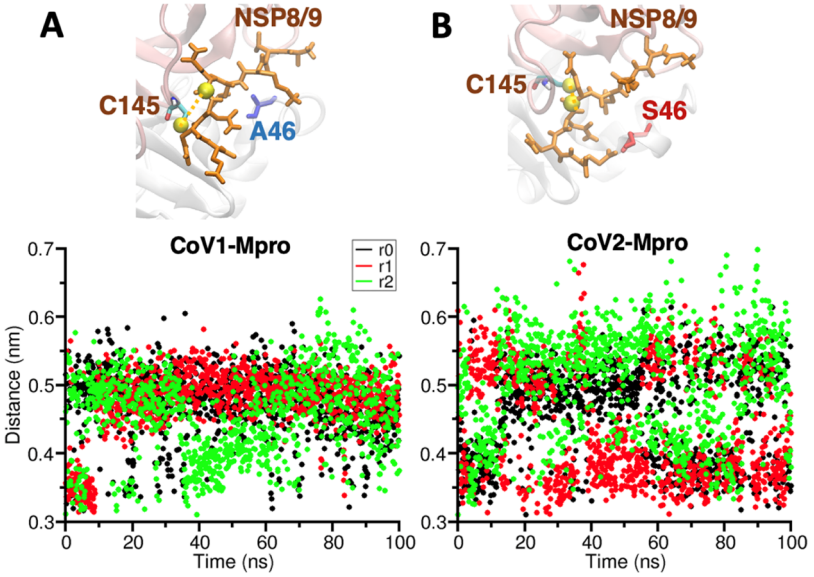


Figure 8. Distances measured between the C-α atoms (yellow) of the catalytic residue 145 and the carbonyl atom as the electron receptor for all MD replicas of (a) CoV1-Mpro and (b) CoV2-Mpro.

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343

significant communicability gain/loss. The A46S mutation was identified as the key contributor to the improved communicability of the SARS-CoV2 main protease due to the addition of a polar hydroxyl group that created extra hydrogen bonds with the substrate and was in concurrence with the higher tendency of electron transfer from the catalytic residue C145 during the reaction. In addition, mutations from CoV1-Mpro to CoV2-Mpro with significant communicability gain/loss could result in either hydrogen bonds or hydrophobic contact addition and deletion, which in turn resulted in communicability gain and loss, respectively. Interestingly, local communicability loss due to the local network disruption induced by the H134F and K180N mutations allowed the more flexible B1/B2 loop containing the active residue Cys145 to interact with the substrate and domain I, corresponding to a greater communicability gain and an improved stability of binding domains I and II. A higher amount of total communicability gain for the mutations from CoV1-Mpro to CoV2-Mpro was in concurrence with the improved compactness of the system containing binding domains I and II, suggesting a binding cleft closure that could facilitate catalytic reactions.

In conclusion, the analysis of the communicability matrix and communicability gain/loss served as a quick tool to visualize the fate of interaction networks and identify important mutated sites of the SARS-CoV2 main protease. Our analysis showed that, in addition to the dimerization interface, mutations occurring within the peptide binding domains could also promote enzymatic activity by either disrupting or repairing some regions of the interaction network within a protomer itself. This understanding of the evolutionary consequences of the SARS-CoV2 main protease enzyme through network topology analysis and atomistic MD simulations could benefit the design of new inhibitors as potential immunity boosters for COVID-19 based on the known SARS-CoV1 main protease inhibitors. Moreover, the analysis of network topology in this study could be viewed as a "reverse protein engineering" tool and could later be used in "protein engineering" of other enzymes to visualize the consequences of a proposed site-directed mutation to changes of network communicability affecting both stability and substrate binding of any enzyme.

AUTHOR INFORMATION

Corresponding Author

Thana Sutthibutpong − *Theoretical and Computational Physics Group, Department of Physics, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok 10140, Thailand; Center of Excellence in Theoretical and Computational Science (TaCS-CoE), Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok 10140, Thailand; Computational Biomodelling Laboratory for Agricultural Science and Technology (CBLAST), Faculty of Science, Kasetsart University, Bangkok 10900, Thailand;* orcid.org/0000-0002-4468-8885; Email: thana.sut@kmutt.ac.th

Authors

Nuttawat Sawang − *Theoretical and Computational Physics Group, Department of Physics, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok 10140, Thailand; Center of Excellence in Theoretical and Computational Science (TaCS-CoE), Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok 10140, Thailand*

Saree Phongphanphanee − *Computational Biomodelling Laboratory for Agricultural Science and Technology (CBLAST), Faculty of Science, Kasetsart University, Bangkok 10900, Thailand; Thailand Center of Excellence in Physics (ThEP Center), Ministry of Higher Education, Science, Research and Innovation, Bangkok 10400, Thailand; Department of Materials Science, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand*

Jirasak Wong-ekkabut − *Computational Biomodelling Laboratory for Agricultural Science and Technology (CBLAST), Faculty of Science, Kasetsart University, Bangkok 10900, Thailand; Thailand Center of Excellence in Physics (ThEP Center), Ministry of Higher Education, Science, Research and Innovation, Bangkok 10400, Thailand; Department of Physics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand;* orcid.org/0000-0002-3651-9870

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c08312

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was financially supported by the National Research Council of Thailand (NRCT) through the Research Grants for Talented Mid-Career Researchers grant no. N41A640080.

REFERENCES

(1) Zhang, Y.; Ma, Z. F. Impact of the COVID-19 Pandemic on Mental Health and Quality of Life among Local Residents in Liaoning Province, China: A Cross-Sectional Study. *Int. J. Environ. Res. Public Health* **2020**, *17*, No. 2381.

(2) Zhang, Q.; Bastard, P.; Liu, Z.; Le Pen, J.; Moncada-Velez, M.; Chen, J.; et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **2020**, *370*, No. eabd4570.

(3) Biancolella, M.; Colona, V. L.; Mehrian-Shai, R.; Watt, J. L.; Luzzatto, L.; Novelli, G.; Reichardt, J. K. V. COVID-19 2022 update: transition of the pandemic to the endemic phase. *Hum. Genomics* **2022**, *16*, No. 19.

(4) Hess, S.; Lancsar, E.; Mariel, P.; Meyerhoff, J.; Song, F.; van den Broek-Altenburg, E.; et al. The path towards herd immunity: Predicting COVID-19 vaccination uptake through results from a stated choice study across six continents. *Soc. Sci. Med.* **2022**, *298*, No. 114800.

(5) Raina, S.; Kumar, R. Herd immunity in COVID-19: Needs de-emphasizing. *J. Fam. Med. Primary Care* **2022**, *11*, No. 1595.

(6) Helmy, Y. A.; Fawzy, M.; Elaswad, A.; Sobieh, A.; Kenney, S. P.; Shehata, A. A. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *J. Clin. Med.* **2020**, *9*, 1225.

(7) Huang, Y.; Yang, C.; Xu, X.; Xu, W.; Liu, S. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* **2020**, *41*, 1141−1149.

(8) Low, Z. Y.; Zabidi, N. Z.; Yip, A. J. W.; Puniyamurti, A.; Chow, V. T. K.; Lal, S. K. SARS-CoV-2 Non-Structural Proteins and Their Roles in Host Immune Evasion. *Viruses* **2022**, *14*, No. 1991.

(9) Lee, J.; Kenward, C.; Worrall, L. J.; Vuckovic, M.; Gentile, F.; Ton, A-T.; et al. X-ray crystallographic characterization of the SARS-CoV-2 main protease polyprotein cleavage sites essential for viral processing and maturation. *Nat. Commun.* **2022**, *13*, No. 5196.

(10) Narwal, M.; Armache, J-P.; Edwards, T.; Murakami, K. S. SARS-CoV-2 polyprotein substrate regulates the stepwise Mpro cleavage reaction *Biophysics* 2022, Sep DOI: 10.1101/2022.09.09.507133.
(11) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289−293.
(12) Shen, Z.; Ratia, K.; Cooper, L.; Kong, D.; Lee, H.; Kwon, Y.et al. Potent, Novel SARS-CoV-2 PLpro Inhibitors Block Viral Replication in Monkey and Human Cell Cultures *Microbiology* 2021, Feb DOI: 10.1101/2021.02.13.431008.
(13) Mengist, H. M.; Dilnessa, T.; Jin, T. Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Front. Chem.* **2021**, *9*, No. 622898.
(14) Agost-Beltrán, L.; de la Hoz-Rodríguez, S.; Bou-Iserte, L.; Rodríguez, S.; Fernández-de-la-Pradilla, A.; González, F. V. Advances in the Development of SARS-CoV-2 Mpro Inhibitors. *Molecules* **2022**, *27*, 2523.
(15) Mohammad, T.; Shamsi, A.; Anwar, S.; Umair, M.; Hussain, A.; Rehman, MdT.; et al. Identification of high-affinity inhibitors of SARS-CoV-2 main protease: Towards the development of effective COVID-19 therapy. *Virus Res.* **2020**, *288*, No. 198102.
(16) Zhang, Y.; Gao, H.; Hu, X.; Wang, Q.; Zhong, F.; Zhou, X.; et al. Structure-Based Discovery and Structural Basis of a Novel Broad-Spectrum Natural Product against the Main Protease of Coronavirus. *J. Virol* **2022**, *96*, No. e01253-21.
(17) Li, J.; Lin, C.; Zhou, X.; Zhong, F.; Zeng, P.; Yang, Y.; et al. Structural Basis of the Main Proteases of Coronavirus Bound to Drug Candidate PF-07321332. *J. Virol.* **2022**, *96*, No. e02013-21.
(18) Rahman, MdM.; Saha, T.; Islam, K. J.; Suman, R. H.; Biswas, S.; Rahat, E. U.; et al. Virtual screening, molecular dynamics and structure−activity relationship studies to identify potent approved drugs for Covid-19 treatment. *J. Biomol. Struct. Dyn.* **2020**, 6231.
(19) Mohapatra, S.; Nath, P.; Chatterjee, M.; Das, N.; Kalita, D.; Roy, P.; et al. Repurposing therapeutics for COVID-19: Rapid prediction of commercially available drugs through machine learning and docking. *PLoS One* **2020**, *15*, No. e0241543.
(20) Lokhande, K. B.; Doiphode, S.; Vyas, R.; Swamy, K. V. Molecular docking and simulation studies on SARS-CoV-2 M^{pro} reveals Mitoxantrone, Leucovorin, Birinapant, and Dynasore as potent drugs against COVID-19. *J. Biomol. Struct. Dyn.* **2020**, *39*, 7294−7305.
(21) Dai, W.; Zhang, B.; Jiang, X-M.; Su, H.; Li, J.; Zhao, Y.; et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020**, *368*, 1331−1335.
(22) Lockbaum, G. J.; Henes, M.; Lee, J. M.; Timm, J.; Nalivaika, E. A.; Thompson, P. R.; et al. Pan-3C Protease Inhibitor Rupintrivir Binds SARS-CoV-2 Main Protease in a Unique Binding Mode. *Biochemistry* **2021**, *60*, 2925−2931.
(23) Issa, S. S.; Sokornova, S. V.; Zhidkin, R. R.; Matveeva, T. V. The Main Protease of SARS-CoV-2 as a Target for Phytochemicals against Coronavirus. *Plants* **2022**, *11*, No. 1862.
(24) Clark, A. C. Caspase Allostery and Conformational Selection. *Chem. Rev.* **2016**, *116*, 6666−6706.
(25) Verma, S.; Dixit, R.; Pandey, K. C. Cysteine Proteases: Modes of Activation and Future Prospects as Pharmacological Targets. *Front Pharmacol.* **2016**, *7*, No. 107.
(26) 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*, No. e101941.
(27) Goyal, B.; Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Comb. Sci.* **2020**, *22*, 297−305.
(28) Li, C.; Teng, X.; Qi, Y.; Tang, B.; Shi, H.; Ma, X.; Lai, L. Conformational Flexibility of a Short Loop near the Active Site of the SARS-3CLpro is Essential to Maintain Catalytic Activity. *Sci. Rep.* **2016**, *6*, No. 20918.
(29) Rocha, R. E. O.; Chaves, E. J. F.; Fischer, P. H. C.; Costa, L. S. C.; Grillo, I. B.; da Cruz, L. E. G.; et al. A higher flexibility at the SARS-CoV-2 main protease active site compared to SARS-CoV and its potentialities for new inhibitor virtual screening targeting multi-conformers. *J. Biomol. Struct. Dyn.* **2021**, 9214−9234.
(30) Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Giuliani, A. Protein Contact Networks: An Emerging Paradigm in Chemistry. *Chem. Rev.* **2013**, *113*, 1598−1613.
(31) Dokholyan, N. V.; Li, L.; Ding, F.; Shakhnovich, E. I. Topological determinants of protein folding. *Proc. Natl. Acad. Sci.* **2002**, *99*, 8637−8641.
(32) Liu, R.; Hu, J. Computational Prediction of Heme-Binding Residues by Exploiting Residue Interaction Network. *PLoS One* **2011**, *6*, No. e25560.
(33) Emerson, I. A.; Louis, P. T. Detection of Active Site Residues in Bovine Rhodopsin Using Network Analysis. *Trends Bioinf.* **2015**, *8*, 63−74.
(34) Hu, G.; Yan, W.; Zhou, J.; Shen, B. Residue interaction network analysis of Dronpa and a DNA clamp. *J. Theor. Biol.* **2014**, *348*, 55−64.
(35) Estrada, E. Topological analysis of SARS CoV-2 main protease. *Chaos: Interdiscip. J. Nonlinear Sci.* **2020**, *30*, No. 061102.
(36) MacDonald, E. A.; Frey, G.; Namchuk, M. N.; Harrison, S. C.; Hinshaw, S. M.; Windsor, I. W. Recognition of Divergent Viral Substrates by the SARS-CoV-2 Main Protease. *ACS Infect. Dis.* **2021**, *7*, 2591−2595.
(37) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **2014**, *42*, W252−W258.
(38) 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.
(39) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(40) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255−268.
(41) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(42) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.
(43) Benzi, M.; Klymko, C. Total communicability as a centrality measure. *J Complex Network* **2013**, *1*, 124−149.

https://doi.org/10.1021/acs.jpcb.2c08312
J. Phys. Chem. B 2023, 127, 2331−2343