


RESEARCH ARTICLE

Comparative protein structure network analysis on 3CL^{pro} from SARS-CoV-1 and SARS-CoV-2

Surabhi Lata | Mohd. Akif 

Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad, India

Correspondence

Mohd. Akif, Department of Biochemistry, School of Life Sciences, University of Hyderabad, Prof. CR Rao Road, Gachibowli, Hyderabad, 500046 India.
Email: akif@uohyd.ac.in

Funding information

SERB; University of Hyderabad

Abstract

The main protease M^{pro}, 3CL^{pro} is an important target from coronaviruses. In spite of having 96% sequence identity among M^{pro}s from SARS-CoV-1 and SARS-CoV-2; the inhibitors used to block the activity of SARS-CoV-1 M^{pro} so far, were found to have differential inhibitory effect on M^{pro} of SARS-CoV-2. The possible reason could be due to the difference of few amino acids among the peptidases. Since, overall 3-D crystallographic structure of M^{pro} from SARS-CoV-1 and SARS-CoV-2 is quite similar and mapping a subtle structural variation is seemingly impossible. Hence, we have attempted to study a structural comparison of SARS-CoV-1 and SARS-CoV-2 M^{pro} in apo and inhibitor bound states using protein structure network (PSN) based approach at contacts level. The comparative PSNs analysis of apo M^{pro}s from SARS-CoV-1 and SARS-CoV-2 uncovers small but significant local changes occurring near the active site region and distributed throughout the structure. Additionally, we have shown how inhibitor binding perturbs the PSG and the communication pathways in M^{pro}s. Moreover, we have also investigated the network connectivity on the quaternary structure of M^{pro} and identified critical residue pairs for complex formation using three centrality measurement parameters along with the modularity analysis. Taken together, these results on the comparative PSN provide an insight into conformational changes that may be used as an additional guidance towards specific drug development.

KEYWORDS

Covid19, main protease (M^{pro}, 3CL^{pro}), protein structure graph, protein structure network, SARS coronavirus

1 | INTRODUCTION

Coronaviridae family of virus usually possesses enveloped, positive sense RNA virus that generally includes three highly pathogenic viruses such as Severe Acute Respiratory Syndrome Coronavirus 1 (SARS-CoV-1), Middle East Respiratory Syndrome Coronavirus (MERS-CoV) and SARS-CoV-2.¹ SARS-CoV-1 originated in China and caused a global pandemic in 2003 with about a 10% fatality rate.^{2,3} MERS-CoV was first reported from Saudi Arabia in 2012 and has infected the human population with limited human-to-human transmission.⁴ SARS-CoV-2, a new coronavirus reported for the first time

from Wuhan, China in December 2019, causes severe human respiratory disease.⁵ It has also been characterized as a very contagious pathogenic virus with rapid transmission capability among human-to-human that has caused an outbreak of the severe pulmonary diseases in almost 216 countries, resulting approximately 814 438 confirmed deaths globally till date (WHO report, 2020). WHO has coined SARS-CoV-2 causing disease as COVID-19 pandemic and that has now become a global health emergency and expected to have severe ramifications on the global economy.

Currently, there is no specific treatment available to control COVID-19 pandemic. Efforts are being made towards design of

vaccines as well as drugs against COVID-19. In the past, therapies have been developed against SARS-CoV-1 targets such as proteases, helicases and polymerases. Moreover, Immune modulators such as interferons and corticosteroids have also been used as therapeutics.⁶ Among the viral targets, main protease (M^{Pro}, 3CL^{Pro}) has been designated as an important drug target because of its essential role in the processing of polyprotein translated from the viral RNA.^{7,8} The M^{Pro} is a homodimer cysteine protease where each protomer consists of three domains, domain I (residue 8-101), domain II (residues 102-184) and domain III (residues 201-303) and catalytic residues and the substrate binding sites are situated between domains I and II of M^{Pro}.⁷ Recent crystal structure of SARS-CoV-2 M^{Pro}⁹ reveals its structural similarity with M^{Pro} of SARS-CoV-1 and has a high degree of sequence identity (96.1%) among the two.⁹ Previous studies showed that HIV-1 protease inhibitors block SARS-CoV-1 M^{Pro}.¹⁰ Hence, having a structural similarity to the M^{Pro} from SARS-CoV-1 and SARS-CoV-2, the known inhibitors could also impart a similar effect on the M^{Pro} of SARS-CoV-2. However, the HIV protease inhibitors show different binding effect on M^{Pro} of SARS-CoV-2.¹¹ One of the HIV protease inhibitors, Lopinavir, was shown to inhibit M^{Pro} of SARS-CoV-1, in-vitro.¹² While, none of the HIV inhibitors was able to significantly inhibit M^{Pro} of SARS-CoV-2, in-vitro.¹³ Other known potent inhibitors such as α -ketoamide and N3 are also reported to have differential inhibition on the activity of M^{Pro}s from SARS-CoV-1 and SARS-CoV-2.^{9,14,15}

A small rearrangement of protein at the structural level by the substitution of a few amino acids at the substrate binding pockets or allosteric sites, results in changes in internal interactions which may lead to differing patterns of inhibitors sensitivity.¹⁶ Similar might be the case with the M^{Pro} from SARS-CoV-2, where few changes in the amino acid sequence in comparison to M^{Pro} of SARS-CoV-1, may attribute towards the differential effect on M^{Pro}s. Since no significant structural changes are noticed at the active site, a subtle change of interactions at the allosteric sites of the proteins may have an effect of sensitivity of the inhibitors. Hence, a protein structure network (PSN) based approach can investigate the negligible conformational changes associated in the protein structure.

A PSN mainly depicted on a protein structure as a system of networks that comprises nodes and links. Nodes are represented by amino acids residues and links are represented as long and short range interactions among the nodes. Interestingly, this method identifies small changes in structures of protein which are otherwise not easily detectable.¹⁷⁻¹⁹ Moreover, similar methods have already been implicated in investigating various features of a protein such structural flexibility,²⁰ protein domain folding,²¹ key residue in folding,²² structural pattern,²³ cluster of residue flexibility²⁴ and identification of functional residues.²⁵ Similarly, a PSN-ENM based method has also been used to construct a PSN on a protein 3-D structure by integrating the information from systems dynamic supplied from the Elastic Network mode analysis (ENM-NMA).²⁶ A global (average) network parameters generated from these methods reveal diminutive structural changes among proteins.

In order to probe subtle conformational changes occurring due to differences of few amino acids in the M^{Pro} sequences, alteration of local contacts as well as residue specific network parameters were investigated on the structures of M^{Pro} (apo and inhibitor bound states) from both SARS-CoV-1 and SARS-CoV-2 by using the PSN and PSN-EMA methods. Recently, topological interaction properties of M^{Pro} were analyzed.²⁷ In another similar study, an analysis of changes in residue interactions of M^{Pro} when bound to N3 inhibitor was also investigated.²⁸ However, in both the cases, monomeric unit of M^{Pro} was considered for the analysis. While, it is well known that a biological active M^{Pro} molecule exists as a dimer. In fact, as reported earlier that the subunits interfacial region of the M^{Pro} can be a possible target for a rational drug design against the SARS-CoV.^{29,30} Hence, it is essential to understand the network connectivity in the biologically active state. Here, we analyzed a comparative PSN in both the proteases (in a biological active state). Our PSN study showed differences in contacts and communication patterns in M^{Pro} of SARS-CoV-2 as compared to M^{Pro} of SARS-CoV-1. Further, we elucidated the negligible changes throughout the protein structure by quantifying their residues connectivity pattern and mapped the network parameters on 3D structures of protein. We also applied a graph theory centrality concept such as betweenness, closeness, hubs and modularity to highlight critical residues for complex formation. This study will provide an understanding about the sensitivity and effectiveness of the existing inhibitors and this would further be helpful to design specific inhibitors.

2 | MATERIALS AND METHODS

2.1 | PSN construction

The 3-D coordinates of 3CL^{Pro} were downloaded from Protein Data Bank (1Z1I³¹ & 6M03 for apo M^{Pro} of SARS-CoV-1 & 2, respectively; 5N19 & 6Y2G⁹ for the inhibitor bound complex of M^{Pro} from SARS-CoV-1 & 2, respectively). Here, C_{alpha} atom of amino acid residues is considered as a node and it forms an edge with another C_{alpha} atom if the distance cutoff is 7 Å. Edge weighted C_{alpha} network that is based on Euclidean distance was constructed using NAPS³² and protein network global parameters such as Degree, Betweenness Centrality, and Clustering Coefficient were analyzed.

2.1.1 | Network parameters

Degree

Degree is total number of direct links between two connected nodes where C_{alpha} atom were considered as contact type. The average degree (D) of a network with N nodes can be computed as $D = 1/N \sum_{i=1}^N D_i$.

Hubs

Hubs are nodes with higher degrees.

Shortest path length

Shortest path length (SPL) is the minimum number of links required to span through one node to another in a protein topological network.

Clustering coefficient

Clustering coefficient (CC) computes the cliquishness for each node in a protein network graph. Cliquishness is defined with respect to total possible edges between them. CC varies between 0 (for no clustering) and 1 (for maximum clustering).

Betweenness centrality

Betweenness centrality (BC) is a centrality measurement in a network graph which is based on the total number of shortest path passing between connected nodes in such a way that edges passing for weighted graphs is minimized.

Closeness centrality

Closeness centrality (CCen) represents the closeness of a node to other nodes. It is centrality measurement which calculates the sum of the shortest path.

Community or modularity

Community or modularity is the region in a network where nodes are more connected to each other.

2.2 | Structural communication analysis

PSN and elastic network model-normal mode analysis (ENM-NMA) approaches were used for long-space communication and effect of allostery on network connectivity.³³⁻³⁵ Previously, the ENM-NMA approach for PSN was applied to characterize the topological and allosteric communication pathways in proteins.³⁶ Other network parameters such as hubs, community, and structural communication analysis were analyzed using a mixed PSN ENM-NMA approach implemented in WebPSN.²⁶ It constructs Protein Structure Graph (PSG) based on interaction strength of two connected nodes

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} 100,$$

where interaction percentage (I_{ij}) of nodes i and j represents the number of side chain atoms pairs with given cut off (4.5 Å), N_i and N_j are normalization factors.³⁷⁻³⁹ The interaction strength (represented as percent) between residues i and j (I_{ij}) is calculated for all node pairs. If I_{ij} is more than the minimum interaction strength cutoff (I_{min}) among the residue pairs, then is considered to be interacting and hence represented as a connection in the PSG. It builds PSG on atomic cross-correlation motions using ENM-NMA.³⁹ All network parameters were visualized using PYMOL.

3 | RESULTS AND DISCUSSION

3.1 | PSN analysis of the SARS-CoV-1 and SARS-CoV-2 M^{PRO}

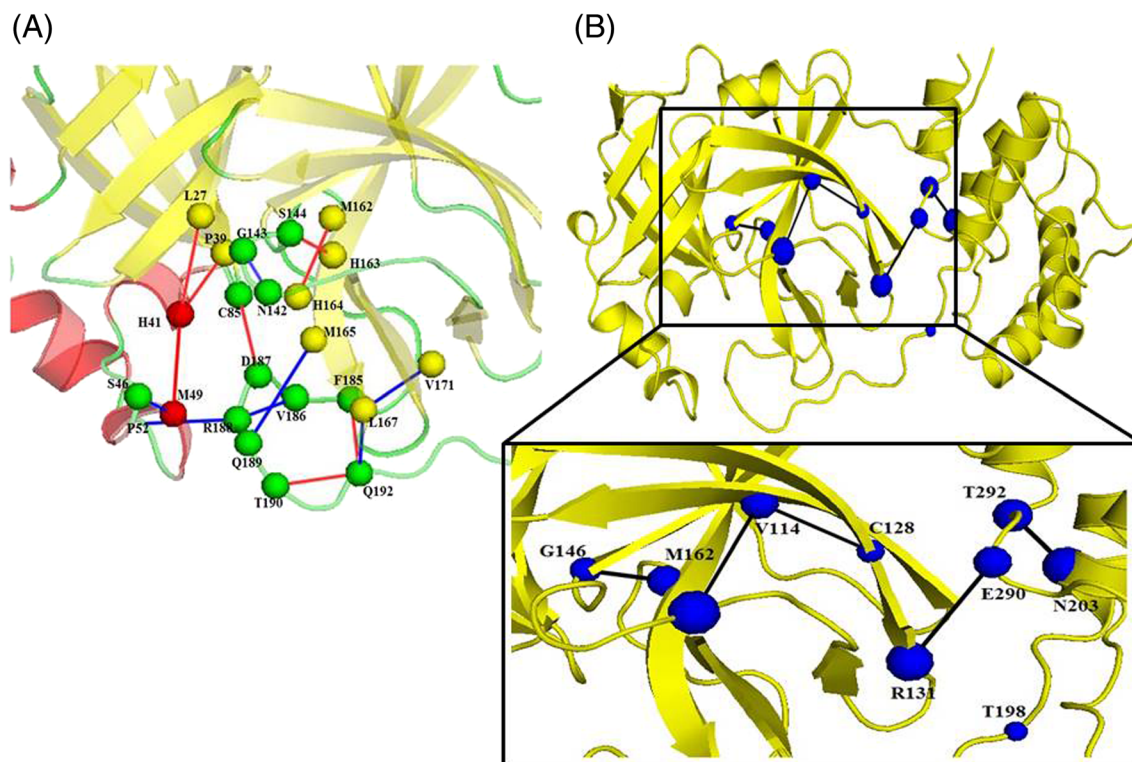
A PSN depicts a network of nodes and links. These nodes are represented by amino acids and links are represented as long and short range interactions among the nodes and that provide a useful information at the contact level in a protein structure. M^{PRO} has been designated as an attractive drug target. In spite of having 96% sequence identity and negligible variation in 3-D structure compared to SARS-CoV-1, the drugs/inhibitors developed so far against M^{PRO} of SARS-CoV-1 showed different inhibitory effect on the M^{PRO} of SARS-CoV-2. Since, structural changes among the two M^{PROs} are negligible, hence a network based approach has been utilized to map subtle conformational alteration arising in the protein structure. Network parameters such as Degree, BC, C Cen, CC, SP and Modularity were analyzed for both free and inhibitor bound forms of SARS-CoV M^{PRO} structures. A little difference was observed in the average network parameters of the M^{PRO} structures (Table 1) suggests a diminutive change in the overall structures.

Calculated degrees are compared among the two structures. It was found that the near active site residues (T26, I43, Q189 and Q192) of SARS-CoV-2 M^{PRO} showed an increase in degree by 2, compared to SARS-CoV-1 M^{PRO}, and while degree of D187 was observed to be decreased by 2 in SARS-CoV-2. It was also observed that the N and C-terminal residues (G2 and E290), crucial for dimerization, were associated with higher degree compared to the same residues of M^{PRO} in SARS-CoV-1. Additionally, few other residues of domain II & III of SARS-CoV-2 M^{PRO} also found to have changes in the calculated degree. List of residues showing the largest change in degrees among the two structures are listed in Table S1. Interestingly, replacement of A46 in SARS-CoV-1 M^{PRO} with S46 in SARS-CoV-2 M^{PRO} resulted in the rearrangement of contacts and observed to form new contacts with L27 and H41 from domain I of SARS-CoV-2 M^{PRO}. A recent report states that SARS-CoV-2 M^{PRO} possesses an active site with a solvent surface accessible area of 356 Å² and the solvent accessible surface area in case of SARS-CoV-1 M^{PRO} was observed to be only 256 Å².²⁸ These changes may be attributed due to the variation of amino acid, S46 which resulted in the rearrangement of contacts in SARS-CoV-2 M^{PRO}. Additionally, few residues of domain II of SARS-CoV-2 M^{PRO} are observed to form five new contacts (Figure 1A). Changes in the contact patterns surrounding the active site of SARS-CoV-2 M^{PRO} is due to change in amino acid, suggesting a subtle conformational change in the SARS-CoV-2 M^{PRO}, which may contribute towards the efficiency of inhibitors on M^{PROs}.

Hubs were also analyzed for the SARS-CoV-1 and SARS-CoV-2 M^{PROs}. Interestingly, a significant difference in the total number of hubs was observed among the two main proteases. The M^{PRO} from SARS-CoV-1 possesses 42 hubs, whereas the same from SARS-CoV-2 consists of 47 hubs, in total (Table 2). Many hubs were found to be similar among the two structures. Few hubs were distinctive to each

TABLE 1 Average network parameters of M^{pro} monomer and dimer units, generated through PSNs

	M^{pro}			
	Monomer		Dimer	
	SARS-CoV-1	SARS-CoV-2	SARS-CoV-1	SARS-CoV-2
No. of edges	1152	1185	2372	2436
Degree	7.65	7.75	7.87	7.96
Clustering coefficient	0.49	0.49	0.49	0.48
Shortest path	6.40	6.36	7.95	7.48

**FIGURE 1** Contact maps in M^{pro} . (A) New contacts in SARS-CoV-2 near the active site (new contacts made are in red line while contacts lost are in blue line). (B) Residues showing highest increase in BC along with the new contacts made mapped on the SARS-CoV-2 structure**TABLE 2** List of hubs in SARS-CoV-1 and SARS-CoV-2 M^{pro}

Hubs in M^{pro}	
SARS-CoV-1	SARS-CoV-2
3, 8, 19, 31, 32 , 38 , 39, 54, 88 , 95, 103, 106, 112, 113, 126, 131, 132 , 140, 159, 161, 164, 167 , 171 , 181, 182, 188 , 200, 207, 209, 213, 218 , 219, 229 , 239, 256, 268 , 273, 281, 282, 288 , 292, 293	3, 8, 19, 39, 40 , 41 , 54, 61 , 66 , 95, 101 , 103, 106, 112, 113, 118, 122 , 126, 131, 140, 150 , 156 , 157 , 159, 161, 163 , 164, 181, 182, 185 , 187 , 192 , 200, 207, 209, 213, 219, 228 , 239, 242 , 256, 269 , 273, 281, 282, 292

Note: Residues in bold are unique to corresponding PSN.

structure, suggesting their important role in interactions and stability. Hubs near the active site region such as H41, H163, D187 and Q192 from the SARS-CoV-2 are assumed to be crucial for the catalysis. The unique hubs are distributed in all the three domains of SARS-CoV-2

M^{pro} , and may suggest a subtle change in inter domain communication within the protease.

Betweenness Centrality (BC) have been reported to play an important role in the structural complexes. In our study, the residues from

the both M^{Pro} structures and their corresponding BC scores are plotted in Figure S1A,B. The trend of the plots is quite comparable in both the structures except few residues shows significant change in the BC scores. Residues with significantly high BC scores (z scores ≥ 0.4) from each M^{Pro} structures are listed in Table S2. Significantly high BC value of a residue signifies its involvement in the communication among different modules of the PCN. The residue V114 with a high BC value is observed to make a new contact with F140 in case of SARS-CoV-2. In addition, other residues such as C128, G146, and T292 found to have high BC values that are also involved in the formation of a new set of contacts in the SARS-CoV-2 (Figure 1B). The new contacts formed in the SARS-CoV-2 M^{Pro} , suggest their role in providing connectivity among residues of the network.

Residues wise CC were analyzed for both protease structures and values are depicted in Figure S1C,D. The residue at 46 positions in SARS-CoV-2 compared to SARS-CoV-1, resulted in an increase in CC of nearby N-finger active site residues such as G23, T24, and S46. Interestingly, T24 is also observed to form a direct contact with the active site residues in SARS-CoV-2. This suggests that changes in the interconnectedness among the residues at and near the active site region may play a role towards selectivity of the inhibitors.

Though the average parameters calculated from the PSN of M^{ProS} from both SARS-CoV-1 & 2, did not show significant changes. However, residues wise comparison of degree and BC values among the two exhibited noticeable change (Table S1 and Table S2). Moreover, these observations on the change in the network parameters suggested their effect on the local conformations of M^{ProS} , which is assumed to provide an insight into the sensitivity and selectivity of inhibitors.

3.2 | Community structure analysis of the PSNs

M^{Pro} structures from SARS-CoV-1 and SARS-CoV-2 do not show much difference, however, an analysis of the contact points either generated or lost due to change of few amino acids, may provide an insight into restructuring of modules within the M^{Pro} . Hence, the community structure for both SARS-CoV-1 & SARS-CoV-2 M^{ProS} was analyzed. The analysis resulted in 12 communities in case of SARS-CoV-1 M^{Pro} and 11 communities in SARS-CoV-2 M^{Pro} and residues of each community are shown in Figure 2. The residues at the active site region of SARS-CoV-2 M^{Pro} are observed to constitute the largest community shown as C1 red module in Figure 2A, that consists of 12 nodes, 18 links and seven hubs. Unlike SARS-CoV-2, the largest community (formed with eight nodes, 11 links, four hubs) in case of SARS-CoV-1 M^{Pro} is located at the interfacial residues of domain I and II, instead of active site region (Figure 2B). Moreover, the community formed at the active site region of SARS-CoV-1 M^{Pro} is found smaller than that of SARS-CoV-2 M^{Pro} . Rearrangement of modules was also observed throughout the structure which indicates the perturbation at global level in the 3-D structure of two proteins.

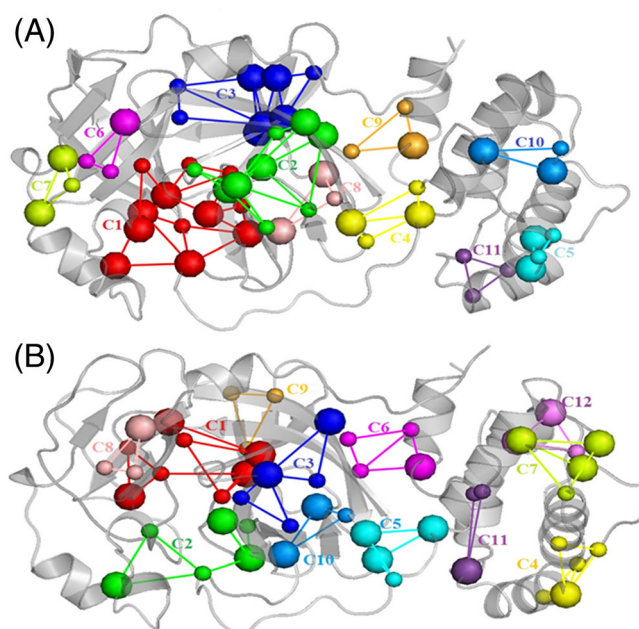


FIGURE 2 Community structure in the M^{Pro} . All communities are mapped on their tertiary structure of M^{ProS} and depicted with different color modules (A) SARS-CoV-2 and (B) SARS-CoV-1

3.3 | PSN analysis of inhibitor (ketoamide) bound complexes of M^{ProS}

3.3.1 | Inhibitor binding perturbs the PSG and the communication pathway in M^{Pro}

The PSG of SARS-CoV-1 M^{Pro} inhibitor bound complex is richer in nodes, links as well as hubs compared to its unbound state (Table 3). Binding of inhibitors to the M^{Pro} generated many hubs at the inhibitor binding site and these hubs are associated with residues such as H41, Y54, F140, S144, H163, H172, and Q192. Additionally, inhibitor complex specific hubs are also formed between the interface of domain I & II (residues C16, Y101, F150, and L115), while invariant hubs spanned throughout the structure. Interestingly, similar trends for nodes and links were not observed for SARS-CoV-2 M^{Pro} inhibitor bound complexes when compared with apo form of the same (Table 3). However, the total number of hub residues in the apo form of SARS-CoV-2 M^{Pro} was found to be 36, whereas the inhibitor complex of the same possesses 35 hub residues. Few hubs are found to be unique in each structure, suggesting their role towards the specificity of inhibitors. Unlike SARS-CoV-1, the residue H41 and Q192 from SARS-CoV-2 M^{Pro} inhibitor complex form do not participate in the active site hubs formation.

We mapped the perturbations on the 3-D structure which considers nodes and links unique to each structure (Figure 3A,B). In the case of the apo and inhibitor bound states of SARS-CoV-1 M^{Pro} , the bound inhibitor was observed to induce perturbations which are essentially consistent with a gain of intermolecular links and nodes. The perturbations associated with a gain of links are mostly located in

TABLE 3 Network components and its parameters for both SARS-CoV-1 and SARS-CoV-2 M^{Pro} monomer in apo and holo states and M^{Pro} dimer in apo state

	M ^{Pro} monomer				M ^{Pro} Dimer	
	SARS-CoV-1		SARS-CoV-2		SARS-CoV-1	SARS-CoV-2
	Apo	Holo	Apo	Holo	Apo	Apo
No. of linked nodes	282	290	288	291	577	578
No. of links	311	328	324	327	644	637
No. of hubs	30	37	36	35	58	53
No. of link mediated hubs	118	149	140	143	229	202

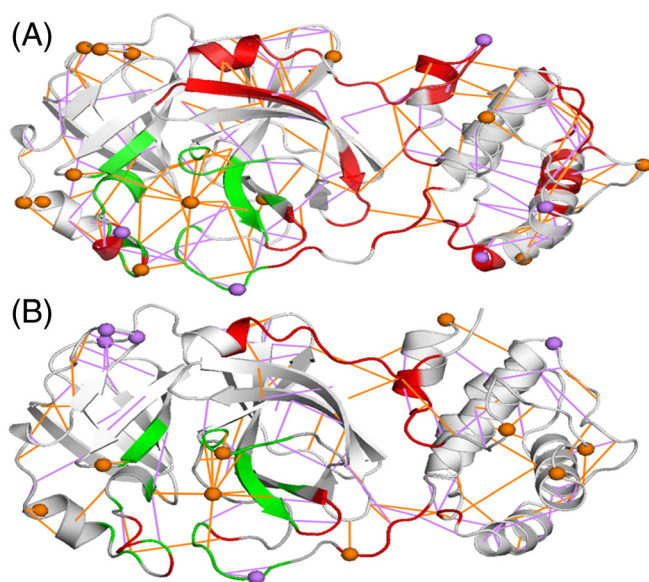


FIGURE 3 3-D PSG representations of the inhibitor bound complexes of M^{Pro}. (A) SARS-CoV-1 (B) SARS-CoV-2. Nodes and links peculiar to the unbound and bound states are in violet and orange, respectively. Major changes near the active site and other allosteric sites are in green and red respectively

the region of small helix near P2 group consists of residues S46- L50, β -hairpin loop near P3-P4 (Res E166-G170) and P5 loop (Res T190-A194). Additional gain of links is also observed in the interfacial region of domain I and II, along with N-finger residues making new links with C-terminal of domain III. However, in case of SARS-CoV-2 M^{Pro}, the comparison of the apo and inhibitor bound states was observed to have not very significant perturbations and the specific contacts show changes to a lesser extent than SARS-CoV-1 M^{Pro}.

Perturbations in inter and intra subunit communication due to binding of inhibitors were also analyzed (Table 4). To investigate more into the communication pattern within the whole structure, we analyzed meta-path and mapped residues participating in each path. The length of the shortest communication paths in SARS-CoV-1 M^{Pro} apo form was 62 345 and a total of 77 391 paths were observed in the inhibitor bound form of M^{Pro}. This indicates an increase in the pathways upon the inhibitor binding to SARS-CoV-1 M^{Pro}. In contrast, a decrease in possible pathways was observed in case of SARS-CoV-2 M^{Pro} inhibitor bound complex. However, the average path length increased in both inhibitors bound states of SARS-CoV-1 and SARS-

CoV-2 M^{Pro}. Changes in the most frequent nodes and links in the structural communication upon inhibitor binding was also observed (Figure 4). The significant inhibitor induced perturbations in the form of loss or frequency reduction of nodes were observed within the active site region of the complexes in either cases (for SARS-CoV-1: C44, P52, Y54, F140, S144, L167, R187, Q192 and for SARS-CoV-2: F140, S144, H163 and R187). Moreover, a redistribution of nodes was observed around the N-finger of M^{Pro} inhibitor complex from SARS-CoV-1, suggesting a role of intercommunication exchange between domain II & III, which may be crucial for the dimerization of M^{Pro}.

3.3.2 | Community analysis of apo and inhibitor bound complex

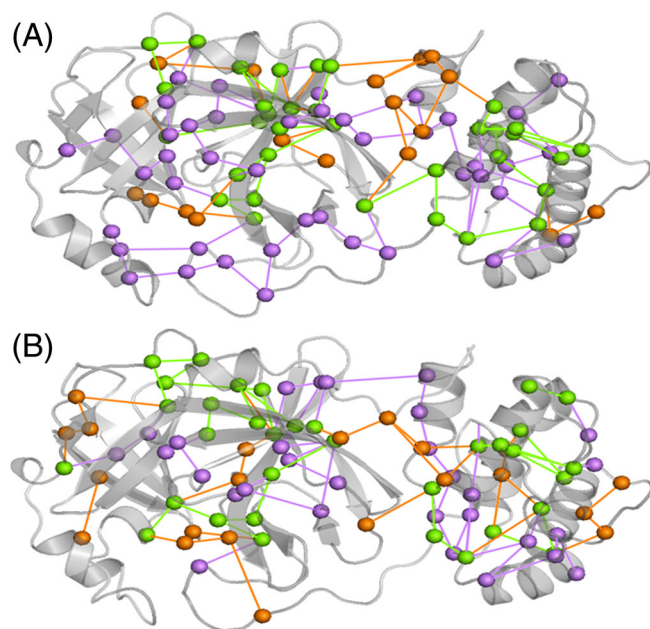
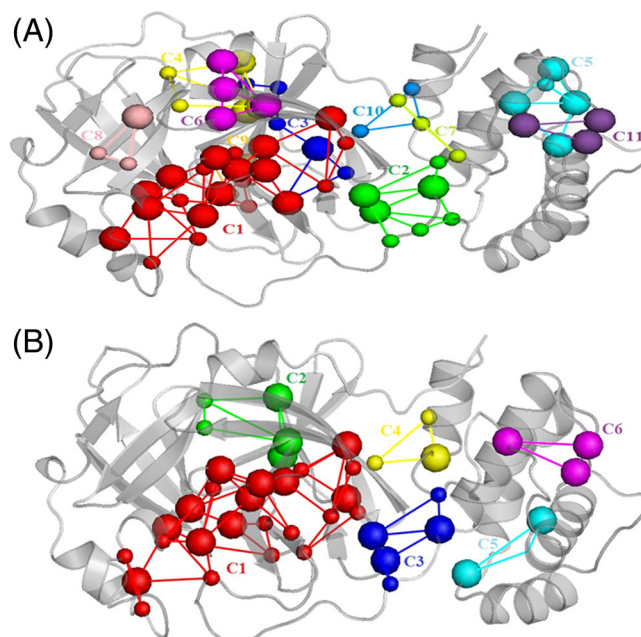
The community structure of PSNs of the M^{Pro} from SARS-CoV-1 and SAR-CoV-2, in apo and inhibitor bound states were analyzed. The SARS-CoV-1 M^{Pro} bound complex is found to have eleven communities, whereas the SARS-CoV-2 M^{Pro} inhibitor complex possesses six communities and these communities are mapped on PSNs. In both the complexes, inhibitor binding sites were part of a large community, C1 (Figure 5). The long loop connecting domain II & III was involved in the second largest community for SARS-CoV-1 complex which includes seven nodes, eleven links, and three hubs. This connecting loop is a part of community C3 in case of SARS-CoV-2 complex which includes five nodes, seven links and three hubs. The rearrangement of communities in the inhibitor complex forms in comparison to the apo forms, suggests that the inhibitors induce perturbations in the network connectivity. Our observations show a correlation with the previous report of RIN on M^{Pro}s of SARS-CoV-1 and SARS-CoV-2, with and without inhibitor N3.²⁸ Similarly, a recent study on topologies of M^{Pro}s by PCN methods highlights sensitive structural perturbations.

3.4 | PSN analysis on quaternary structure of SARS-CoV M^{Pro}

It has been known that biologically active SARS-CoV M^{Pro} exists as a dimer. Previously, mutagenesis of E290A in SARS-CoV-2 M^{Pro} reported loss of catalytic activity, indicating importance of domain III in dimerization.⁴⁰ Hence, In order to highlight structural differences

TABLE 4 Average path summary for both SARS-CoV-1 and SARS-CoV-2 M^{PRO} monomer in apo and holo states and M^{PRO} dimer in apo state

	M ^{PRO} monomer				M ^{PRO} dimer	
	SARS-CoV-1		SARS-CoV-2		SARS-CoV-1	SARS-CoV-2
	Apo	Holo	Apo	Holo	Apo	Apo
No. of nodes	82	54	65	55	153	78
No. of links	81	53	64	54	162	78
No. of shortest paths	62 345	77 391	108 495	83 636	290 236	468 290
Average path length	17.81	18.04	19.9	20.1	21.4	26.3
Average path force	6.87	7.90	7.47	6.87	7.57	6.74
Average path hub %	37.53	44.51	44.9	42.25	39.23	41.76

**FIGURE 4** Residues involved in global meta-path mapped on 3-D structure of M^{PRO}. (A) SARS-CoV-1 and (B) SARS-CoV-2. Residues involved in meta path for apo state is in purple nodes and links while for inhibitor bound state are in orange, shared nodes and links are in green**FIGURE 5** Community structure in the inhibitor bound states of M^{PRO}. All communities are mapped on their tertiary structure in different color module (A) SARS-CoV-1 and (B) SARS-CoV-2

and commonalities in the quaternary structures (homodimers) of M^{PRO}s, PSN parameters were computed and identified the crucial residues from the subunit interfacial region. Average network parameters for the two subunits of M^{PRO}s do not show any significant changes, hence no correlation was drawn (Table 1). So, we further evaluated the previously mentioned network components like links, hubs, and link mediated hubs. Interestingly, these parameters were observed to be slightly higher for SARS-CoV-1 dimeric form (Table 3). Moreover, significant changes were observed in the hubs and links mediated hubs. A total number of 58 hubs residues were observed in the quaternary SARS-CoV-1 M^{PRO}, whereas 53 hubs were noted in the SARS-CoV-2 M^{PRO} homodimer (Figure 6A). Interestingly, subtle rearrangements of hub residues in domain III in both homodimers

were also observed. In case of SARS-CoV-1 M^{PRO}, the residues at 206: A, 259: A, 289: A, 218: B, 230: B (the alphabets A & B represent sub-units) were observed to form a specific hub. Similarly, the residues at 288: A, 273: B, 288: B, 290: B from SARS-CoV-2 M^{PRO} were involved in a hub formation. In addition, specific hub residues were also observed in the domain II of SARS-CoV-2 (161: B, 181: B) and unfortunately these were not seen to form hubs in SARS-CoV-1 M^{PRO} homodimer. Other residues at 39: A, 141: B, 163: B, 172: B were also involved in hub formation in SARS-CoV-1 M^{PRO}. It was observed that residues at 185: A, 192: A, 192: B from SARS-CoV-2 M^{PRO} were engaged in the formation of active site hubs. These rearrangements in hub residues suggest some disruption in the inter-domain communication between both proteases in their quaternary structure.

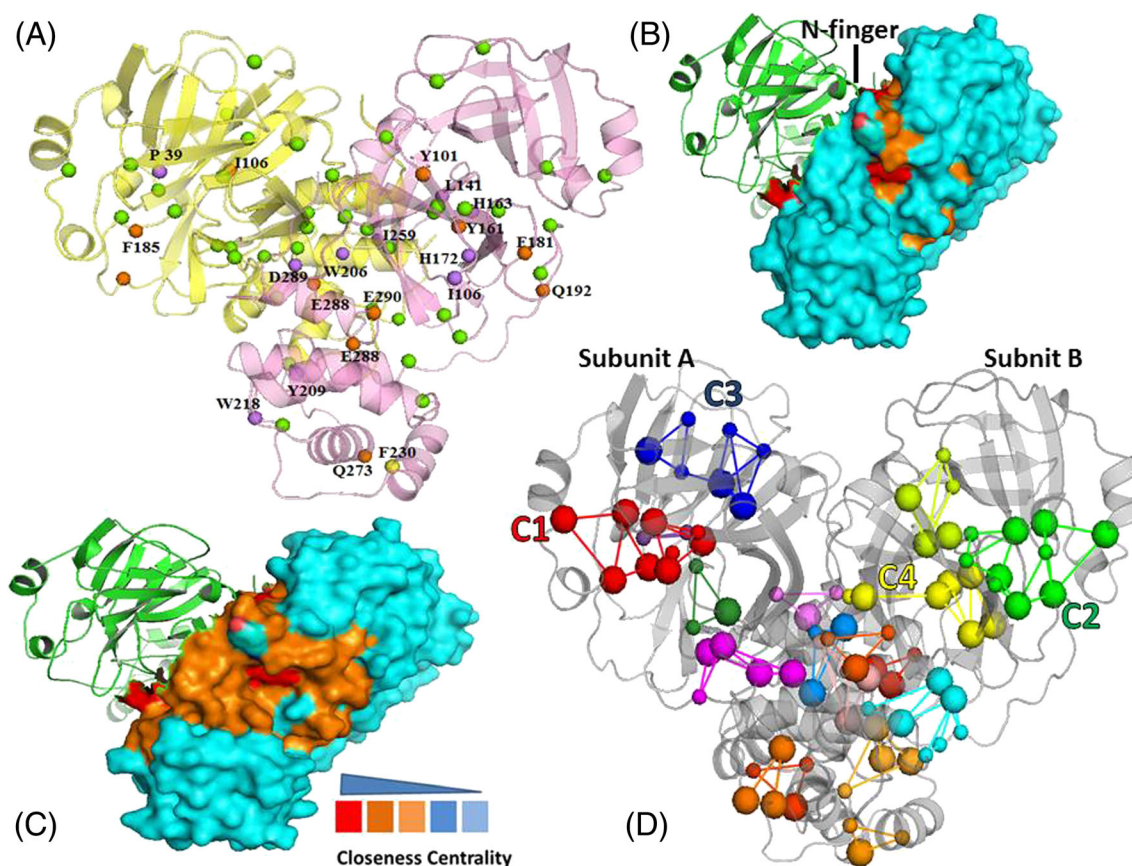


FIGURE 6 PSN analysis on the quaternary structure of SARS-CoV M^{Pro}. Chain A is in yellow, chain B is in light pink. (A) Residues involved in hub formation SARS-CoV-1 dimeric state is in purple nodes and links while for SARS-CoV-2 it is in orange color. Nodes and links shared in common are in green. Closeness Value for (B) SARS-CoV-1. (C) SARS-CoV-2. The residues are colored by their closeness values with red, orange and cyan corresponding to the high (top 20%), intermediate (20 ~ 60%) and low (below than 60%) respectively. (D) Consensus community structure analysis between both Proteases. Major communities in complex are numbered with respect to its specific module color

The BC for each residue of M^{Pro} from SARS-CoV-1 as well as SARS-CoV-2 dimeric complexes was analyzed and residues with high BC are listed in (Table S3). N-finger residues (Res 3 & 8) and C-terminal residues of domain III (Res 282 & 290) that are at the interface of both monomers were showing high betweenness values in both the structures, indicating the importance of these residues in the dimer formation.³⁹ Additionally, this interfacial residues showing high BC are also involved in hub formation which suggests their role in catalysis. The residues at positions 28 and 144 in SARS-CoV-2 M^{Pro} showed high BC value and their essentiality for enzyme activity and dimerization has been confirmed with experimental mutagenesis studies in homodimer formation.⁴¹

The closeness values of all residues were computed and classified them into three categories: i. High closeness value, ii. Intermediate closeness values, and iii. Small closeness values. Our results suggested that residues form N-finger (3–11), B2 (112–117), B3 (122–130) and B4 (149–151) of domain II might be considered as the most likely recognition sites (Figure 6B,C). Previously, it has been reported that residue C117 makes direct interaction with N28 and plays a major role in the dimer stability and enzymatic activity of SARS-CoV-1 M^{Pro}.⁴¹

Experimentally, it has been identified that N28A mutant plays a critical role in active site structural integrity and positions the important residues involved in dimer interface binding and catalysis of substrate.⁴² This suggests that residues showing high closeness might be responsible for long range interactions that are crucial for dimerization.

The search for shortest communication pathway led to a total of 290 236 and 468 290 paths for the M^{Pro}s of SARS-CoV-1 and SARS-CoV-2, respectively which indicates a significant increase in paths for SARS-CoV-2 M^{Pro} dimeric form (Table 4). The total number of nodes and links along with the specific nodes and links in global Metapath were observed to be 60% and 21.79% for SARS-CoV-1 and SARS-CoV-2, respectively. Additionally, we observed some of the interface residues are specific to SARS-CoV-1 and those are frequent nodes in communication pathways (Chain A: Res. 3, 6, 123, 126, 140, 290; Chain B: Res. 4, 6, 116, 141, 122, 126, 299). Few substrate binding residues (Chain A: Res. 41, 49, 144, 163, 165; Chain B: Res.163, 167) were also involved in the communication pathway of SARS-CoV-1 M^{Pro}, while these corresponding frequent nodes were absent in SARS-CoV-2 M^{Pro}. In addition, average hub percent involved in the communication pathway was also observed

to be decreased for SARS-CoV-2 (Table 4). These observations among the two suggest a change at structural communication level in the dimeric form of M^{pro}.

Common modules are shared among the two homo-dimeric M^{pro}s and depicted in Figure 6D. Two large communities CI and CII consisting of active site residues from both monomers possess 10 nodes, 15 links, and 7 hubs. Third large community CIII is distributed on the strands of domain II in both monomers. Additionally, a fourth community was observed at the interface residues of both monomers; N-term residue M6 of chain A and β -strands of domain II from the chain B. One of the residues F140 from this community has been previously reported to present on the dimer interface of SARS-CoV M^{pro} and mutation of this residue resulted in the conformational change of M^{pro}.⁴⁰

4 | CONCLUSIONS

Our study on comparative PSN analysis of M^{pro}s from SARS-CoV-1 and SARS-CoV-2, investigated the noticeable difference in the network parameters among the two proteases. Moreover, the study also highlights differential perturbation among the community structures in inhibitor bound form of proteins. Interestingly, the investigations helped us to probe subtle conformation changes associated throughout the structure of the two proteases, which otherwise are not evident from the crystals structures. Our observations gauge an insight into the diminutive structural changes which may provide an understanding towards selectivity of inhibitors towards M^{pro}s of SARS-CoV-2. In addition, the investigation of PSN on the quaternary structure of M^{pro}s suggests structural and network changes at the interface as well as long range interactions and highlights critical residue pairs for the complex formation using three centrality measurement parameters. This study is a thorough comparative investigation of subtle structural changes that may provide an insight into designing a specific inhibitor/drug.

ACKNOWLEDGEMENT

First author is grateful to the University of Hyderabad, Hyderabad for the student fellowship. This work is partially supported by the SERB, Govt. of India (EMR/2016/001183).

CONFLICT OF INTEREST

The authors declare no conflict of interest regarding this work.

AUTHOR CONTRIBUTIONS

Surabhi Lata and Mohd. Akif conceptualize the study, Surabhi Lata performed the study and analyzed the data. Mohd. Akif supervised the study, Surabhi Lata wrote the manuscript and Mohd. Akif reviewed and edited the manuscript.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/prot.26143>.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Mohd. Akif  <https://orcid.org/0000-0001-8462-7166>

REFERENCES

- Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol.* 2019;17(3):181-192.
- Zhong et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003 Elsevier connect, the company's public news and information. *Lancet.* 2020;362:1353-1358.
- Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med.* 2003; 348(20):1953-1966.
- Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med.* 2012;367(19):1814-1820.
- Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature.* 2020;579(7798):265-269.
- Zumla A, Chan JFW, Azhar EI, Hui DSC, Yuen KY. Coronaviruses-drug discovery and therapeutic options. *Nat Rev Drug Discov.* 2016;15(5): 327-347.
- Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain. *EMBO J.* 2002;21 (13):3213-3224.
- Yang H, Yang M, Ding Y, et al. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A.* 2003;100(23):13190-13195.
- Zhang L, Lin D, Sun X, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. *Science.* 2020;368(6489):409-412.
- Savarino A. Expanding the frontiers of existing antiviral drugs: possible effects of HIV-1 protease inhibitors against SARS and avian influenza. *J Clin Virol.* 2005;34(3):170-178.
- Ortega JT, Serrano ML, Pujol FH, Rangel HR. Unrevealing sequence and structural features of novel coronavirus using in silico approaches: the main protease as molecular target. *Excli J.* 2020;19: 400-409.
- Wu CY, Jan JT, Ma SH, et al. Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc Natl Acad Sci U S A.* 2004;101(27):10012-10017.
- Mahdi M, Mótóyán JA, Szojka ZI, Golda M, Miczi M, Tőzsér J. Analysis of the efficacy of HIV protease inhibitors against SARS-CoV-2's main protease. *Virol J.* 2020;17(1):1-8.
- Jin Z, Du X, Xu Y, et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature.* 2020;582(7811):289-293.
- Xue X, Yang H, Shen W, et al. Production of authentic SARS-CoV Mpro with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J Mol Biol.* 2007;366(3):965-975.
- Raugi DN, Smith RA, Gottlieb GS. Four amino acid changes in HIV-2 protease confer class-wide sensitivity to protease inhibitors. *J Virol.* 2016;90(2):1062-1069.
- Srivastava A, Sinha S. Uncoupling of an ammonia channel as a mechanism of allosteric inhibition in anthranilate synthase of *Serratia marcescens*: dynamic and graph theoretical analysis. *Mol Biosyst.* 2017;13 (1):142-155.
- Srivastava A, Sinha S. Thermostability of in vitro evolved bacillus subtilis lipase a: a network and dynamics perspective. *PLoS ONE.* 2014;9 (8):e102856. <https://doi.org/10.1371/journal.pone.0102856>

19. Kandhari N, Sinha S. complex network analysis of thermostable mutants of *Bacillus subtilis* lipase A. *Appl Netw Sci.* 2017;2(1).
20. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. *Proteins Struct Funct Genet.* 2001;44(2):150-165.
21. Dokholyan NV, Li L, Ding F, Shakhnovich EI. Topological determinants of protein folding. *Proc Natl Acad Sci U S A.* 2002;99(13):8637-8641.
22. Wangikar PP, Tendulkar AV, Ramya S, Mali DN, Sarawagi S. Functional sites in protein families uncovered via an objective and automated graph theoretic approach. *J Mol Biol.* 2003;326(3):955-978.
23. Vendruscolo M, Dokholyan NV, Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2002;65(6):4.
24. Atilgan AR, Akan P, Baysal C. Small-world communication of residues and significance for protein dynamics. *Biophys J.* 2004;86(1 l):85-91.
25. Amitai G, Shemesh A, Sitbon E, et al. Network analysis of protein structures identifies functional residues. *J Mol Biol.* 2004;344(4):1135-1146.
26. Seeber M, Fellingine A, Raimondi F, Mariani S, Fanelli F. WebPSN: a web server for high-throughput investigation of structural communication in biomacromolecules. *Bioinformatics.* 2015;31(5):779-781.
27. Estrada E. Topological analysis of SARS CoV-2 main protease. *Chaos.* 2020;30(6).
28. Griffin JWD. SARS-CoV and SARS-CoV-2 main protease residue interaction networks change when bound to inhibitor N3. *J Struct Biol.* 2020;211(3):107575.
29. Fan K, Wei P, Feng Q, et al. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J Biol Chem.* 2004;279(3):1637-1642.
30. Shi J, Wei Z, Song J. Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme. Defining the extra domain as a new target for design of highly specific protease inhibitors. *J Biol Chem.* 2004;279(23):24765-24773.
31. Hsu MF, Kuo CJ, Chang KT, et al. Mechanism of the maturation process of SARS-CoV 3CL protease. *J Biol Chem.* 2005;280(35):31257-31266.
32. Chakrabarty B, Parekh N. NAPS: network analysis of protein structures. *Nucleic Acids Res.* 2016;44(W1):W375-W382.
33. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: structure and dynamics. *Phys Rep.* 2006;424(4-5):175-308.
34. Tang QY, Kaneko K. Long-range correlation in protein dynamics: confirmation by structural data and normal mode analysis. *PLoS Comput Biol.* 2020;16(2):1-17.
35. Artymiuk PJ, Rice DW, Mitchell EM, Willett P. Structural resemblance between the families of bacterial signal-transduction proteins and of G proteins revealed by graph theoretical techniques. *Protein Eng des Sel.* 1990;4(1):39-43.
36. Raimondi F, Fellingine A, Seeber M, Mariani S, Fanelli F. A mixed protein structure network and elastic network model approach to predict the structural communication in biomolecular systems: the PDZ2 domain from tyrosine phosphatase 1E as a case study. *J Chem Theory Comput.* 2013;9(5):2504-2518.
37. Kannan N, Vishveshwara S. Identification of side-chain clusters in protein structures by a graph spectral method. *J Mol Biol.* 1999;292(2):441-464.
38. Brinda KV, Vishveshwara S. A network representation of protein structures: implications for protein stability. *Biophys J.* 2005;89(6):4159-4170.
39. Fellingine A, Seeber M, Fanelli F. webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules. *Nucleic Acids Res.* 2020;48(W1):W94-W103.
40. Chou CY, Chang HC, Hsu WC, Lin TZ, Lin CH, Chang GG. Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry.* 2004;43(47):14958-14970.
41. Barrila J, Gabelli SB, Bacha U, Amzel LM, Freire E. Mutation of Asn28 disrupts the dimerization and enzymatic activity of SARS 3CLpro. *Biochemistry.* 2010;49(20):4308-4317.
42. Hu T, Zhang Y, Li L, et al. Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure. *Virology.* 2009;388(2):324-334.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Lata S, Akif M. Comparative protein structure network analysis on 3CL^{pro} from SARS-CoV-1 and SARS-CoV-2. *Proteins.* 2021;89(9):1216-1225. <https://doi.org/10.1002/prot.26143>