



# Structural differences in 3C-like protease (Mpro) from SARS-CoV and SARS-CoV-2: molecular insights revealed by Molecular Dynamics Simulations

Meet Parmar<sup>1</sup> · Ritik Thumar<sup>1</sup> · Bhumi Patel<sup>1</sup> · Mohd Athar<sup>2,5</sup> · Prakash C. Jha<sup>3</sup> · Dhaval Patel<sup>1,4</sup> 

Received: 14 April 2022 / Accepted: 29 October 2022 / Published online: 23 November 2022  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Novel coronavirus SARS-CoV-2 has infected millions of people with thousands of mortalities globally. The main protease (Mpro) is vital in processing replicase polyproteins. Both the CoV's Mpro shares 97% identity, with 12 mutations, but none are present in the active site. Although many therapeutics and vaccines are available to combat SARS-CoV-2, these treatments may not be practical due to their high mutational rate. On the other hand, Mpro has a high degree of conservation throughout variants, making Mpro a stout drug target. Here, we report a detailed comparison of both the monomeric Mpro and the biologically active dimeric Mpro using MD simulation to understand the impact of the 12 divergent residues (T35V, A46S, S65N, L86V, R88K, S94A, H134F, K180N, L202V, A267S, T285A and I286L) on the molecular microenvironment and the interaction between crucial residues. The present study concluded that the change in the microenvironment of residues at the entrance (T25, T26, M49 and Q189), near the catalytic site (F140, H163, H164, M165 and H172) and in the substrate-binding site (V35, N65, K88 and N180) is due to 12 mutations in the SARS-CoV-2 Mpro. Furthermore, the involvement of F140, E166 and H172 residues in dimerization stabilizes the Mpro dimer, which should be considered. We anticipate that networks and microenvironment changes identified here might guide repurposing attempts and optimization of new Mpro inhibitors.

**Keywords** COVID-19 · SARS-CoV-2 · Mpro · Molecular dynamics · 3C-like proteases · Drug repurposing

## Introduction

The year 2019 ended with a fatal outbreak of a novel coronavirus (SARS-CoV-2) identified as a causative agent for a series of unusual pneumonia cases in Wuhan City, Hubei Province of China [1, 2]. As the cases were reported sporadically, the WHO announced a Public Health Emergency of

International Concern (PHEIC) for the 2019-nCoV outbreak in January 2020 (WHO.int). Furthermore, since the infection crossed geographical barriers, the WHO permanently named the 2019-nCoV pathogen SARS-CoV-2 and the causing coronavirus disease 2019 (COVID-2019) by declaring it a pandemic situation in March 2020 (WHO.int). The disease caused by SARS-CoV-2 presents vast pathophysiological symptoms, including fever, coughing and shortness of breath in typical cases, whereas pneumonia, kidney failure and severe acute respiratory failure in severe cases [2, 3].

SARS-CoV-2 belongs to the subgenus *Sarbecovirus*, genus *Betacoronavirus*, Coronaviridae family and the Orthocoronavirinae subfamily. It is a  $\beta$ -coronavirus of 2B ( $\beta$ -coronavirus) group and to date not reported zoonotic unlike earlier coronaviruses [4, 5]. SARS-CoV-2 is an enveloped single-stranded RNA virus (+ve ssRNA) with a genome size of 29.9 kb genome that spreads widely among humans and other mammals, causing a wide range of symptoms from the common cold to fatal diseases, such as severe acute respiratory syndrome (SARS) [6, 7]. Genome sequencing of SARS-CoV-2 has revealed 96% identical to the bat

✉ Dhaval Patel  
pateldhaval.in@gmail.com

<sup>1</sup> Department of Biotechnology and Bioengineering, Institute of Advanced Research, Gandhinagar, Gujarat 382426, India

<sup>2</sup> Center for Chemical Biology and Therapeutics, InStem, Bangalore 560065, Karnataka, India

<sup>3</sup> School of Applied Material Sciences, Central University of Gujarat, Gandhinagar 382030, Gujarat, India

<sup>4</sup> Gujarat Biotechnology University, Gujarat International Finance Tec-City, Gandhinagar 382355, Gujarat, India

<sup>5</sup> Physics Department, University of Cagliari, Monserrato (CA), Italy

coronavirus and 79.6% sequence identical to SARS-CoV [3]. Other known members of *Betacoronavirus* pathogenic to humans are SARS-CoV, the causative agent of the 2002–2003 outbreak [8], and Middle East respiratory syndrome (MERS-CoV), which broke out in 2012 [9]. Both have high mortality rates, 9.6% for SARS-CoV and 34.4% for MERS-CoV. Despite SARS-CoV-2 being genetically distinct from SARS-CoV, the proteome is quite similar. The two major types of protein are being identified for pathogenicity: nonstructural proteins (nsps) and structural proteins. The SARS-CoV-2 virus genome has 15 open reading frames (ORFs): ORF 1a (encodes nsp1- nsp11); ORF 1b (encodes nsp12 - nsp16); ORF S (encodes a spike [S] protein); ORFs 3a, 3b, 6, 7a, 7b, 8, 9a, 9b and 10 (encode accessory proteins); ORF E (encodes envelope protein, virus assembly and morphogenesis proteins); ORF M (encodes membrane protein and virus assembly proteins); and ORF N (encodes nucleocapsid protein) [10, 11]. With ribosomal frameshifting during translation, the replicase gene of SARS-CoV-2 encodes two overlapping translation products, polyproteins 1a and 1ab (pp1a and pp1ab) [12]. Each of these polyproteins is then cleaved by the main protease (Mpro) and papain-like protease (PLpro) of SARS-CoV-2 to release 11 (pp1a) and 5 (pp1ab) functional proteins necessary for viral replication.

As viral proteases play a vital role in processing the polyproteins translated from viral RNA, inhibitors designed against these proteases could effectively block coronavirus replication, thus making the proteases attractive targets for drug discovery [6]. Viral proteases are suggested to be involved in SARS-CoV-2's mechanisms of infection and pathogenicity [13]. The Nsp5 is the central protease (Mpro) of SARS-CoV-2, a chymotrypsin-like cysteine protease also known as “3C-like protease” (3CLpro); it is necessary for viral replication, structural assembly and pathogenicity [14]. The Mpro cleaves at least 11 sites during the proteolytic processing of pp1a and pp1ab. The recognition sequence at most sites is Leu-Gln↓(Ser, Ala, Gly) (↓ marks the cleavage site) [15, 16]. SARS-CoV-2 Mpro is classified as a chymotrypsin-like cysteine protease with a molecular mass of ~33.8 kDa [17]. Its recently discovered structure contains three domains: domains I (residues 1–101) and II (residues 102–184), which are made up of antiparallel  $\beta$ -barrel structures (13  $\beta$ -strands) in a chymotrypsin-like fold responsible for catalysis; and the  $\alpha$ -helical domain III (residues 201–306; 5  $\alpha$ -helices), which is required for enzymatic activity [16, 17]. A long loop (residues 185–200) connects domains I and II with III. The substrate-binding site is located in a cleft between domains I and II, composed of a Cys-His dyad. The overall structure is similar to SARS-CoV Mpro and with other  $\beta$ -coronaviruses Mpro (MERS-Mpro, HKU5-Mpro and HKU4-Mpro) [17]. The structures of SARS-CoV-2 Mpro and SARS-CoV Mpro are similar,

the main difference being the surfaces of the proteins, where 12 different amino acids are located. Since the structure of SARS-CoV was first explained and, more recently, that of SARS-CoV-2 was available, numerous efforts have been made to inhibit viral-polyprotein cleavage by blocking active sites of the protein in order to prevent the spread of infection [15, 16]. Our present study attempted to identify the sequence and structural differences between SARS-CoV Mpro and SARS-CoV-2 Mpro and their dynamics differences using the Molecular Dynamics (MD) approach. Our findings could help researchers identify new compounds/inhibitors against SARS-CoV-2 Mpro and add to the collective knowledge of how these two Mpro differ in their structural dynamics, which will aid in repurposing their existing antiviral protease inhibitors.

## Materials and methods

### Sequence analysis

A comparison of the SARS-CoV-2 Mpro protein sequence with other Mpro sequences from coronaviruses known to cause SARS in humans, i.e. SARS-CoVs, MERS-CoV and HCoV, was performed. Additionally, a sequence from intermediate host Bat-CoV-RaTG13 was also considered for the analysis. The basis for selecting these sequences was based on the historic outbreak of coronavirus infecting humans to assess the sequence-level similarity and divergence from an evolutionary point of view. The Mpro sequences in FASTA format were downloaded from NCBI for the SARS-CoV-2 (GI: 1820435677), SARS-CoV (GI: 157834798), MERS-CoV (GI: 70842105), Bat-CoV-RaTG13 (NCBI: QHR63299.1) and four other human coronaviruses HCoV-HKU (GI: 203282419), HCoV-OC43 (GI: 203282418), HCoV-NL63 (GI: 403071823) and HCoV-229E (GI: 403071823), which have been known to cause SARS in human. The SARS-CoV-2 was also checked individually with other Mpro sequences using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) [18] to identify pair-wise sequence identity. Multiple sequence alignment was performed using ClustalW (<http://www.clustal.org/>) [19] to identify conserved residues, followed by phylogenetic tree construction using MEGA6 [20]. ESPript3 (<https://espript.ibcp.fr/ESPript/ESPript/>) was used for structure-based sequence alignment using the SARS-CoV-2 structure as a reference [21]. The Maximum Parsimony method was used for tree construction, followed by 1000 bootstrap iterations for statistical significance. Physicochemical parameters of all CoV Mpro, including isoelectric point, instability index, grand average of hydropathicity (GRAVY) and amino acid composition, were computed using the ProtParam (<https://web.expasy.org/protparam/>) tool of ExPASy.

## Structure comparison of Mpro from SARS-CoV and SARS-CoV-2

Numerous studies have been going on for Mpro from SARS-CoV-2, which has recently led to rapid identification and deposition of 3D X-ray structures coordinates in PDB. The available structures of SARS-CoV-2 Mpro were downloaded from PDB for structure comparison. The SARS-CoV-2 Mpro structures have been solved in apo- and ligand-bound forms. For our comparative study, we used PDB ID: 6M03 (the crystal structure of COVID-19 main protease in an apo form) for SARS-CoV-2 and PDB ID: 2DUC (crystal structure of SARS coronavirus main proteinase (Mpro)) for SARS-CoV. For structure-based comparison, Mpro from CoVs infecting humans as well as other similar viruses infecting other hosts (i.e. infectious bronchitis virus (IBV), porcine transmissible gastroenteritis coronavirus, feline infectious peritonitis virus, porcine epidemic diarrhoea virus and murine hepatitis virus) was undertaken to access the fold similarity in Mpro known structures from different coronavirus infecting hosts other than humans. In the case of multiple structures available in PDB for the same protein, the one representing the structure for the complete protein with the highest structure resolution was chosen for structural comparison analysis. Mpro structures from other CoVs: SARS-CoV (PDB ID: 5B6O), MERS-CoV (PDB ID: 5WKJ), BtCoV-HKU4 (PDB ID: 2YNA), HCoV-HKU1 (PDB ID: 3D23), MHV-A59-CoV (PDB ID: 6JIJ), PEDV-CoV (PDB ID: 5ZQG), FIPV-CoV (PDB ID: 5EU8), TGEV-CoV (PDB ID: 4F49), HCoV-NL63 (PDB ID: 5GWY), HCoV-229E (PDB ID: 2ZU2) and IBV-CoV (PDB ID: 2Q6D) were downloaded from PDB (<https://www.rcsb.org/>) for structural superposition. PyMOL [22], Chimera [23] and other tools were used for structural superposition and analysis of structural features in SARS-CoV-2 Mpro and differences with other CoV Mpro.

## Molecular Dynamics Simulations of Mpro from SARS-CoV-2 and SARS-CoV

For comparison of dynamics of Mpro from SARS-CoV and SARS-CoV-2, PDB ID: 2DUC and PDB ID: 6M03 were used as the starting conformation in Molecular Dynamics Simulations (MDS). To assess the comparison between monomer and dimer, four MDS were set up: SARS-CoV\_monomer, SARS-CoV\_dimer, SARS-CoV-2\_monomer and SARS-CoV-2\_dimer. Each MDS setup was performed for a 50-ns production run, as reported in our earlier studies [24–26] using GROMACS ver.2016.4 [27] with Amber99SB force-field [28]. All the MDS systems were solvated with a simple point charge (SPC) water model in a dodecahedron box configuration with a distance of 1 nm from all the protein directions and periodic boundaries. The simulation setup was neutralized by adding an equal number of counter ions ( $\text{Na}^+$

$\text{Cl}^-$ ) and subjected to energy minimization using the steepest descent algorithm to remove any steric clashes and wrong contacts before the actual MD run. Following minimization, equilibration with position restraint was carried out under NVT (constant number [N], constant volume [V] and constant temperature [T]) and NPT (constant number [N], constant pressure [P] and constant temperature [T]) ensemble for 100 ps each. For NVT equilibration, a modified Berendsen thermostat algorithm [29] was used to maintain the system at a constant volume for 100 ps and a constant temperature of 300 K. NPT equilibration was performed at a constant pressure of 1 bar for 100 ps maintained by Parrinello-Rahman barostat [30]. For calculations of long-range electrostatic interactions, Particle Mesh Ewald approximation was applied with a 1-nm cut-off [31] and computing coulomb & the van der Waals interactions, and the bond length was constraint using the LINCS algorithm [32]. Final production MD was simulated for a 50-ns run with default parameters. The trajectories were visualized using VMD [33] and Chimera [23]. For the calculation of root mean square deviation (RMSD), root mean square fluctuation (RMSF), hydrogen bonds (H-bonds) and other analysis, in-built gmx commands were used in GROMACS, and the plotting tool GRACE was used for the generation and visualization of the plots (<http://plasma-gate.weizmann.ac.il/Grace>) as reported in our previous study [34].

## Clustering of conformations for ensemble generation and Essential Dynamics

To assess the dominant conformation acquired by CoV Mpro through the entire simulations was studied by clustering analysis. The entire MDS trajectories were subjected to RMSD-based clustering via the ‘gmx cluster’ that explores the conformational landscape among the ensemble of protein structures. The GROMOS algorithm, as described by Daura et al. [35], was used to determine the dominant conformation with a 0.15-nm  $\text{C}\alpha$  RMSD cut-off. For investigation of the global motion of SARS-CoV and SARS-CoV-2 Mpro in dimer and monomer form, Principal Component Analysis (PCA) or Essential Dynamics (ED) analysis was carried out. Congruent to our earlier studies, the collective motion and essential dynamics of  $\text{C}\alpha$  backbone atoms were also examined for the entire simulations trajectories, as computed using gmx covar and gmx ana eig tools [36–39]. PCA is a method that reduces the complexity of the data and results in the concerted motion in the MD simulations. These motions are essentially correlated and significant for biological functions. The set of eigenvectors and eigenvalues was computed by diagonalizing the covariance matrix. The eigenvalues represent the amplitude of the eigenvector along with the multidimensional space, while the  $\text{C}\alpha$  displacement along each eigenvector shows the concerted motions of the protein in each direction. FES (free energy surface) (kcal/

mol) was also computed for all four MD systems considering the conformational variability in terms of ROG and RMSD taken together and represented by Gibbs free energy. It represents a mapping of all possible protein conformations adopted during the entire simulation trajectory. RMSD and ROG calculated to the average structure for computing FES were used to ensure adequate sampling for FES calculations.

## Results and discussion

### Sequence analysis

We performed multiple-sequence alignment of SARS-CoV-2 Mpro with its closest known homologs SARS-CoV, MERS-CoV, Bat-CoV-RaTG13, HCoV-HKU, HCoV-OC43, HCoV-NL63 and HCoV-229E to identify conserved segments in Mpro (Fig. S1). Pair-wise alignment revealed that the Mpro of SARS-CoV-2 resembled Bat-CoV-RaTG13 (99.35%) and SARS-CoV (96.08%). It shared 50.65%, 49.02%, 48.37%, 44.30% and 41.04% sequence identity with other human coronaviruses HCoV-HKU, HCoV-OC43, HCoV-NL63 and HCoV-229E, respectively (Table 1). The catalytic dyad of Cys-His was also conserved throughout all sequences in the Mpro family as it is essential for enzyme activity. In addition, the physicochemical properties of SARS-CoV-2 Mpro were in the same range as those of other Mpro, as shown in Table 1. Mpro from SARS-CoV-2, SARS-CoV, MERS-CoV and Bat-CoV-RaTG13 were 306 amino acids long, and Mpro from the other four human coronaviruses were 303 amino acids long. Sequence comparison of Mpro of SARS-CoV-2 and SARS-CoV revealed that SARS-CoV-2 Mpro had 12 divergent amino acids, of which 5 positions had residues with structurally/chemically similar amino acid substitution.

The majority (8/12) of variable residues were found in the Mpro  $\beta$ -strand-rich domains I and II, where the inhibitor/catalytic site is located; the remaining (4/12) residues were found in domain III. The connecting loop residues (185–200)

had no variable residues (Fig. S1). Also, SARS-CoV-2 Mpro had only two changes compared with Bat-CoV Mpro, one of which was inclusive of the 12 changes in Mpro between SARS-CoV and SARS-CoV-2; this indicated that Bat-CoV was intermediate between SARS-CoV and SARS-CoV-2 (Fig. S2). This finding was consistent with an initial report that SARS-CoV-2 is more comparable to SARS-CoV than MERS-CoV and shares a common ancestor with bat coronaviruses, suggesting bats as an intermediate host between SARS-CoV and SARS-CoV-2 [6].

The 12 residues in SARS-CoV-2 Mpro that differed from those in SARS-CoV Mpro were T35V, A46S, S65N, L86V, R88K, S94A, H134F, K180N, L202V, A267S, T285A and I286L. The substrate-binding pockets, including the active site residues T25, T26, H41, M49, F140, N142, G143, S144, C145, H163, H164, M165, E166, P168, H172, Q189, T190, A191 and Q192, were conserved in Mpro from both CoVs. Therefore, our analysis will collectively refer to these residues as active-site residues elsewhere.

### Structure comparison of Mpro from SARS-CoV and SARS-CoV-2

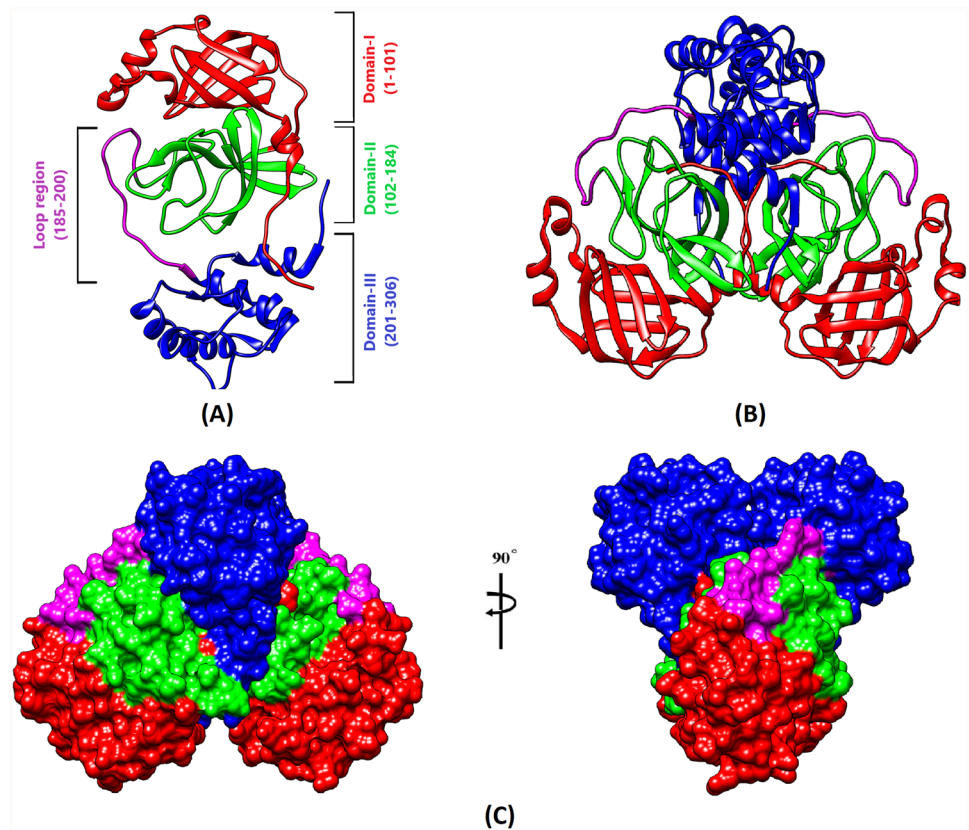
We downloaded the available three-dimensional (3D) X-ray structure coordinates for SARS-CoV and SARS-CoV-2 Mpro PDB. In this study, we used PDB ID: 6M03; for SARS-CoV-2, the structure was in the apo form, i.e. without any other molecule except the protein. While for SARS-CoV Mpro, we used PDB ID: 2DUC (crystal structure of SARS coronavirus Mpro [Mpro]), also in apo form. PDB ID: 2DUC was in dimer form, and PDB ID: 6M03 was in monomer form. To compare the dimer forms, we transformed the coordinates of 6M03 accordingly. SARS-CoV-2 Mpro contained three distinct domains: domains I (residues 1–101) and II (residues 102–184), made up of antiparallel  $\beta$ -barrel structures (13  $\beta$  strands), connected to the  $\alpha$ -helical domain III (residues 201–306) (5  $\alpha$ -helices) by a long loop (residues 185–200) (Fig. 1). The dimer interface

**Table 1** Sequence identities and physicochemical properties of SARS-CoV-2 Mpro compared with Mpro of other human coronaviruses

Mpro	(%) similarity with SARS-CoV-2	Length (a.a.)	Mol. Wt. (kDa)	pI	Grand average of hydropathicity (GRAVY)	Instability index
SARS-CoV-2	100%	306	33.8	5.95	−0.019	27.65
Bat-CoV-RaTG13	99.35%	306	33.8	5.95	−0.007	28.36
SARS-CoV	96.08%	306	33.8	6.22	−0.049	29.67
MERS-CoV	50.65%	306	33.3	5.86	0.129	27.33
HCoV-HKU1	49.02%	303	33.2	5.78	0.16	34.51
HCoV-OC43	48.37%	303	33.4	6.06	0.105	22.61
HCoV-NL63	44.30%	303	32.8	6.39	0.145	31.67
HCoV-229E	41.04%	303	33.1	6.38	0.059	30.42



**Fig. 1** SARS-CoV-2 Mpro in monomer and dimer state: **A** monomer structure of SARS-CoV-2 Mpro with domain assignment marked with residue ranges. **B** Dimer structure of SARS-CoV-2 Mpro coloured by domain assignment. **C** Dimer structure of SARS-CoV-2 Mpro in surface representation coloured by domain assignment at two different angles. In all panels, domain I is marked in red, domain II in green, domain III in blue and the loop region with magenta



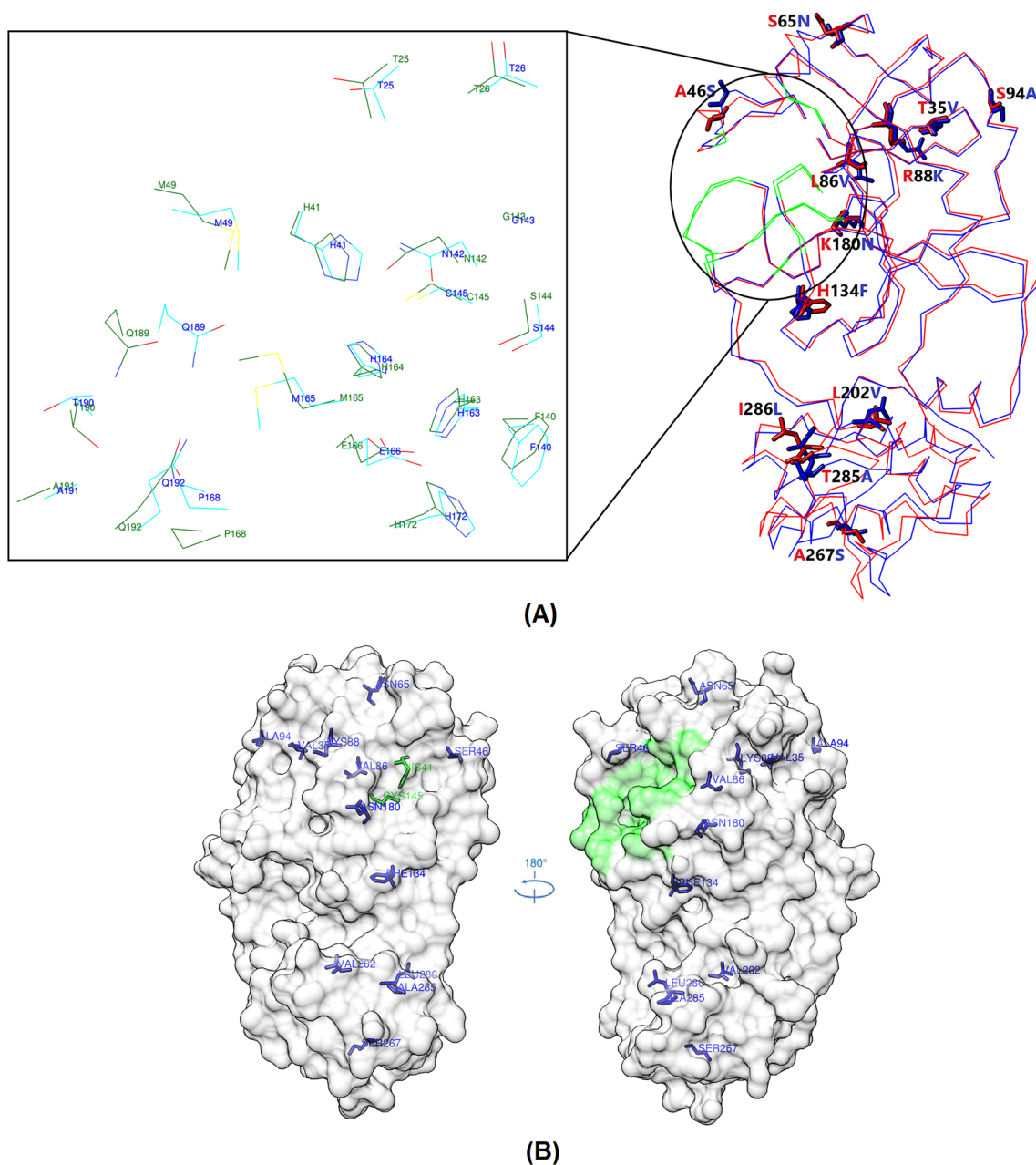
residues were not involved in active site formation except F140, E166 and H172, which formed fundamental interactions that served to open and close active sites, making the dimeric form biologically active [40, 41].

The 3D structure of SARS-CoV-2 Mpro was highly conserved compared with that of SARS-CoV Mpro, as expected from the 96% sequence identity; the root-mean-square deviation (RMSD) was 0.67 Å for all C $\alpha$  positions (comparison between the two apo-enzyme structures SARS-CoV-2 Mpro; PDB ID: 6M03 and SARS-CoV Mpro; PDB ID: 2DUC). Mpro formed an active catalytic dimer in which each monomer consisted of an N-terminal catalytic region and a C-terminal region. Dimerization was regulated by domain III (residues 201–306), a globular cluster of five helices [15, 40]. When superimposed on SARS-CoV Mpro, homo-dimer SARS-CoV-2 Mpro revealed high structural identity in domain orientations and active sites (Fig. S3). The tight dimer formed by SARS-CoV-2 Mpro had a contact interface of ~1558 Å and ~1306 Å for SARS-CoV Mpro, between domain II of chain A and the NH<sub>2</sub>-terminal residues (‘N-finger’) of chain B, with the two molecules, oriented perpendicular to one another (Fig. 1) as computed using the PDBSUM program (<http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/>) [42].

We also analysed the 12 residues between SARS-CoV-2 Mpro and SARS-CoV Mpro (T35V, A46S, S65N, L86V, R88K, S94A, H134F, K180N, L202V, A267S, T285A and

I286L) in terms of presence with the active/catalytic site. We found none in the active site, but just one of them, S46 in SARS-CoV-2 (A46 in SARS-CoV), was located near the entrance to that site (Fig. 2).

To identify the structural conservancies of SARS-CoV-2 Mpro from other Mpro of alpha, beta, gamma and delta coronaviruses, we identified structural homologs available in PDB using the Basic Local Alignment Tool (BLAST). The Mpro homolog structures identified from other CoVs were as follows: alphacoronavirus (PEDV-CoV (PDB ID: 5ZQG), HCoV-NL63 (PDB ID: 5GWY), HCoV-229E (PDB ID: 2ZU2), FIPV-CoV (PDB ID: 5EU8) and TGEV-CoV (PDB ID: 4F49)), betacoronavirus (SARS-CoV (PDB ID: 5B6O), MERS-CoV (PDB ID: 5WKJ), BtCoV-HKU4 (PDB ID: 2YNA), HCoV-HKU1 (PDB ID: 3D23) and MHV-A59-CoV (PDB ID: 6JIJ) and gammacoronavirus (IBV-CoV (PDB ID: 2Q6D)). Sequence-based structural alignment using SARS-CoV-2 Mpro as a reference structure revealed a striking resemblance at the sequence and the structural level (Fig. S4). All of the secondary structural elements for domains I and II, i.e. antiparallel  $\beta$ -barrel structures (13  $\beta$  strands), and domain III, i.e. 5  $\alpha$ -helices, were present in all of the CoV Mpro (Fig. S4). Phylogenetic analysis of these 12 sequences and the newly identified BatCoV-RaTG13 Mpro revealed clear, distinct clades of alpha, beta and gammacoronaviruses. The SARS-CoV and SARS-CoV-2 Mpros formed



**Fig. 2** Monomer structure alignment of SARS-CoV Mpro (red) and SARS-CoV-2 Mpro (blue) aligned with the C $\alpha$  backbone. **A** The overall structure of SARS-CoV and SARS-CoV-2 Mpro with 12 differing amino acids is marked and shown in the sticks (SARS-CoV Mpro) and blue (SARS-CoV-2 Mpro). The active site residues are coloured green. Close-up of active site residues represented in

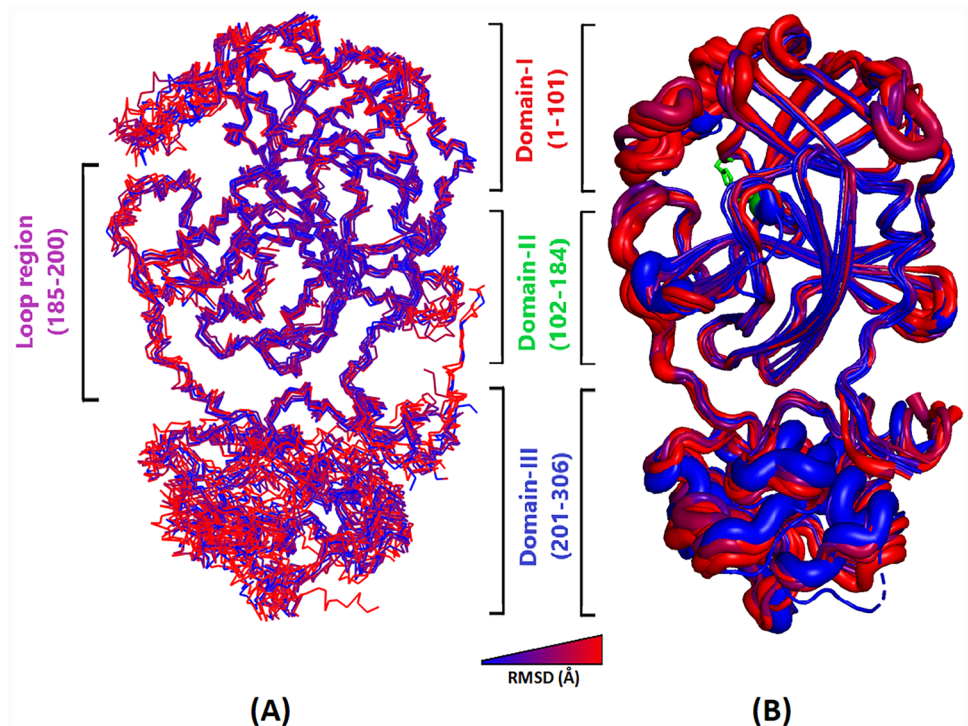
the sticks (side chains) as cyan (SARS-CoV-2 Mpro) and dark green (SARS-CoV Mpro). **B** The structure of SARS-CoV-2 Mpro in surface display in two angles (180° rotation) with 12 different residues from SARS-CoV Mpro marked and shown in blue sticks. The catalytic HIS41 and CYS145 residues are shown in green sticks, and other active site residues are shown on the green surface

a distinct clade from the rest of the sequences, with the latter very close to the recently identified BatCoV-RaTG13 Mpro [6].

Looking at the RMSD ( $\text{\AA}$ ) of the 12 superposed Mpro crystal structures, the most flexible regions seemed to be helical domain III (residues 201–306) and loops on the surface. However, substrate-binding pockets were located in a cleft between domains I and II, which were still highly

conserved among all CoV Mpro, suggesting the feasibility of designing antiviral inhibitors targeting this site (Fig. 3). It shows a variable tube representation of the C $\alpha$  trace of 12 Mpro homologous protein structures superimposed on SARS-CoV-2 Mpro; PDB ID: 6M03, and the tube size is proportional to the mean RMSD per residue between C $\alpha$  pairs. The blue-to-red colour gradient shows sequence conservation from high to low. This analysis identified weak

**Fig. 3** Structural superposition of 12 CoV Mpro from different strains. **A** Superposition of 12 Mpro structures ( $C\alpha$  backbone) from different CoV onto SARS-CoV-2 Mpro as a query. **B** Representation of 12 Mpro superposed structures using variable tube depiction, where the radius is proportional to the RMSD differences in  $C\alpha$  between SARS-CoV-2 Mpro and 12 other homologous Mpro structures. In both panels, blue-to-red colour ramping is used to visualize and correlate conservation from strong to weak conservation areas. Domain assignments are marked with residue ranges in both panels



and strong structural conservation areas correlating with sequence similarity in substrate-binding pockets located in the cleft between domains I and II (Fig. 3). Conversely, we found sequence level variation (8/12) residues Mpro  $\beta$  strand-rich domains I and II, where the inhibitor/catalytic site was located; the remaining (4/12) residues were found in domain III. Our analysis indicated that the variations found in SARS-CoV-2 Mpro compared with SARS-CoV did not confer flexibility on the region where the inhibitor/catalytic site was located based on the RMSD ( $\text{\AA}$ ) deviation of  $C\alpha$  atoms.

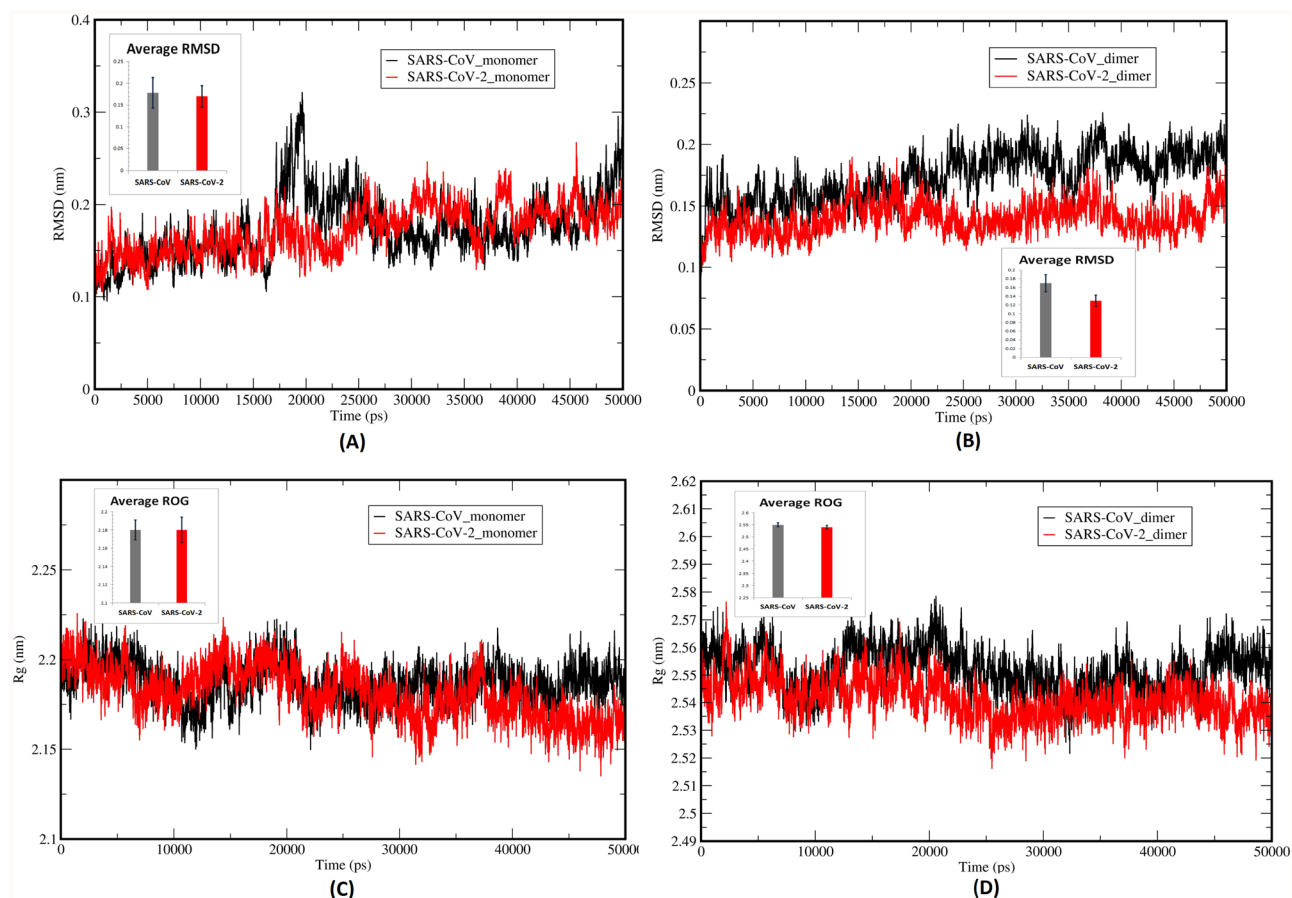
### Molecular Dynamics Simulations (MDS) of Mpro from SARS-CoV-2 and SARS-CoV

The simulations were conducted to study the conformational behaviour, structural dynamics and stability of the Mpro of SARS-CoV-2 and SARS-CoV. We analysed the dimeric and monomeric forms' dimeric and monomeric forms' RMSD, RMSF, H-bonds and PCA properties. A series of gmx commands and analysis was run for all four (SARS-CoV\_monomer, SARS-CoV\_dimer, SARS-CoV-2\_monomer and SARS-CoV-2\_dimer) MDS to interpret the difference between Mpro from SARS-CoV-2 and SARS-CoV for single configurations according to functions like RMSD, RMSF and ROG by obtaining value/s for each time point throughout the trajectory. Subsequently, the dynamics of both Mpro were also investigated in the time domain by averaging the fluctuations across entire simulations.

### Stability of SARS-CoV-2 and SARS-CoV Mpro

We investigated the overall stability of Mpro using an MDS run to assess configurational interpretation at each time point. Each 50-ns simulation RMSD value was evaluated to access the convergence of the simulation toward equilibrium in terms of the Euclidean distance from the average structure to a reference (crystal) structure. It is evident from the plots shown in Fig. 4 that most of the MDS reached equilibrium after 25 ns. Mean RMSD values of protein  $C\alpha$  backbones for SARS-CoV\_monomer, SARS-CoV-2\_monomer, SARS-CoV\_dimer and SARS-CoV-2\_dimer were  $0.17 \pm 0.035$ ,  $0.17 \pm 0.025$ ,  $0.17 \pm 0.020$  and  $0.13 \pm 0.013$  nm, respectively. Mean RMSD values of the Mpro monomer of both the CoVs and Mpro from the CoV dimer were similar. The lower RMSD values of SARS-CoV-2\_dimer compared with those of both Mpro monomer and SARS-CoV\_dimer indicated the formation of a stable molecule. Further, we subjected all four MDS to Radius of Gyration (ROG) calculation for the entire 50-ns simulation. ROG measures the protein's shape at each point by comparing it with the experimentally available hydrodynamic radius. Mean ROG values of protein  $C\alpha$  backbones for SARS-CoV\_monomer, SARS-CoV-2\_monomer, SARS-CoV\_dimer and SARS-CoV-2\_dimer were  $2.18 \pm 0.011$ ,  $2.18 \pm 0.014$ ,  $2.55 \pm 0.0081$  and  $2.54 \pm 0.0074$  nm, respectively (Fig. 4). ROG values of Mpro were similar between monomer and dimer forms in both CoVs, indicating that the overall shape was consistent irrespective of the changes at the residue level.





**Fig. 4** Molecular Dynamics Simulations of Mpro from SARS-CoV and SARS-CoV-2 in monomer and dimer structures, computing the deviation (nm) vs function of time (50 ns): RMSD of the protein C $\alpha$  backbone atoms of SARS-CoV and SARS-CoV-2 Mpro in monomer (A) and dimer (B) forms. ROG of the protein C $\alpha$  backbone atoms

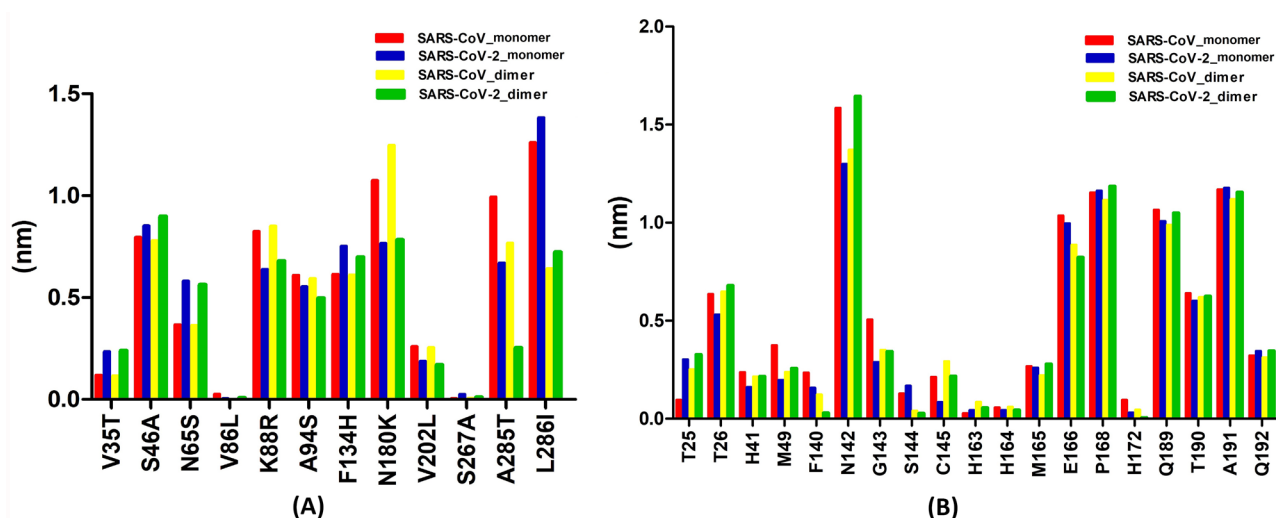
of SARS-CoV Mpro and SARS-CoV-2 Mpro in monomer (C) and dimer (D) forms. The inset graph represents average values with standard deviations, and SARS-CoV and SARS-CoV-2 Mpro are plotted in black and red, respectively

We also observed the formation of H-bonds for entire simulation trajectories of MD systems. The secondary structures that form the cores of protein structures underpinning the protein folding are stabilized by H-bonds and thus indicate the rigidity of the protein structure and specificity of intermolecular interactions between the secondary structures. Mean H-bond values of intramolecular interactions of SARS-CoV\_monomer, SARS-CoV-2\_monomer, SARS-CoV\_dimer and SARS-CoV-2\_dimer were  $211.73 \pm 6.96$ ,  $217.415 \pm 7.25$ ,  $441.363 \pm 9.83$  and  $451.954 \pm 9.92$  (Fig. S5). H-bond analysis indicated no substantial difference between Mpro or CoV in dimer or monomer form. To access the surface area of protein accessible to the solvent in which it was simulated, we calculated the surface accessible solvent area (SASA) variable for the entire trajectory. Total SASAs calculated for SARS-CoV\_monomer, SARS-CoV-2\_monomer, SARS-CoV\_dimer and SARS-CoV-2\_dimer were  $150.48 \pm 2.28$ ,  $149.38 \pm 2.18$ ,  $272.744 \pm 3.18$  and  $265.811 \pm 2.65$  (Fig. S5).

For the Mpro monomer, there is no significant difference between SARS-CoV and SARS-CoV-2, but for the dimer, the SARS-CoV-2 Mpro had little SASA values compared to SARS-CoV indicating a lesser magnitude of flexibility and instability.

SASA values of the 12 different residues and the active site residues are plotted in Fig. 5 for all four MD systems. For residues T35V, S65N and H134F, the SASA area was higher in SARS-CoV-2 than in SARS-CoV. This is because variable residue S65N was close to the binding site (T25, T26, M49 and Q189). Variable residue H134F was also critical, given that many functionally crucial residues (H172, E166, F140 (which are also involved in dimerization) and the oxyanion loop (140–144)) were found in its vicinity. In addition, H134F was also in the loop leading to the catalytic residue C145. Conversely, for residues R88K, K180N and L202V, SASA values were lower in SARS-CoV-2 than in SARS-CoV. Again, this is because the variable residues R88K and K180N were close to the catalytic site.





**Fig. 5** The plot of SASA values for individual residues in all four MD systems: **A** the plot of 12 divergent residues in SARS-CoV-2 Mpro. The X-axis represents the residue present in SARS-CoV-2, followed

by the residue number and ending with the residue found in SARS-CoV. **B** The plot of residues in the active site conserved in SARS-CoV and SARS-CoV-2 Mpro

SASA values of T285A and I286L were very different and lower for dimer than monomer. Specifically, T285A values were significantly lesser in SARS-CoV-2\_dimer compared with SARS-CoV\_dimer. This was correlated with the reduction in the distance between C $\alpha$  of A285 in SARS-CoV-2\_dimer, which allowed the two domains III to approach each other closely, a finding supported by an earlier study [16, 41]. Our findings suggested that despite having overall structural similarity with those in SARS-CoV Mpro, the 12 divergent residues in SARS-CoV-2 Mpro might affect the catalytic activity by modifying the microenvironment of the catalytic site, conferring a potential change in SARS-CoV-2.

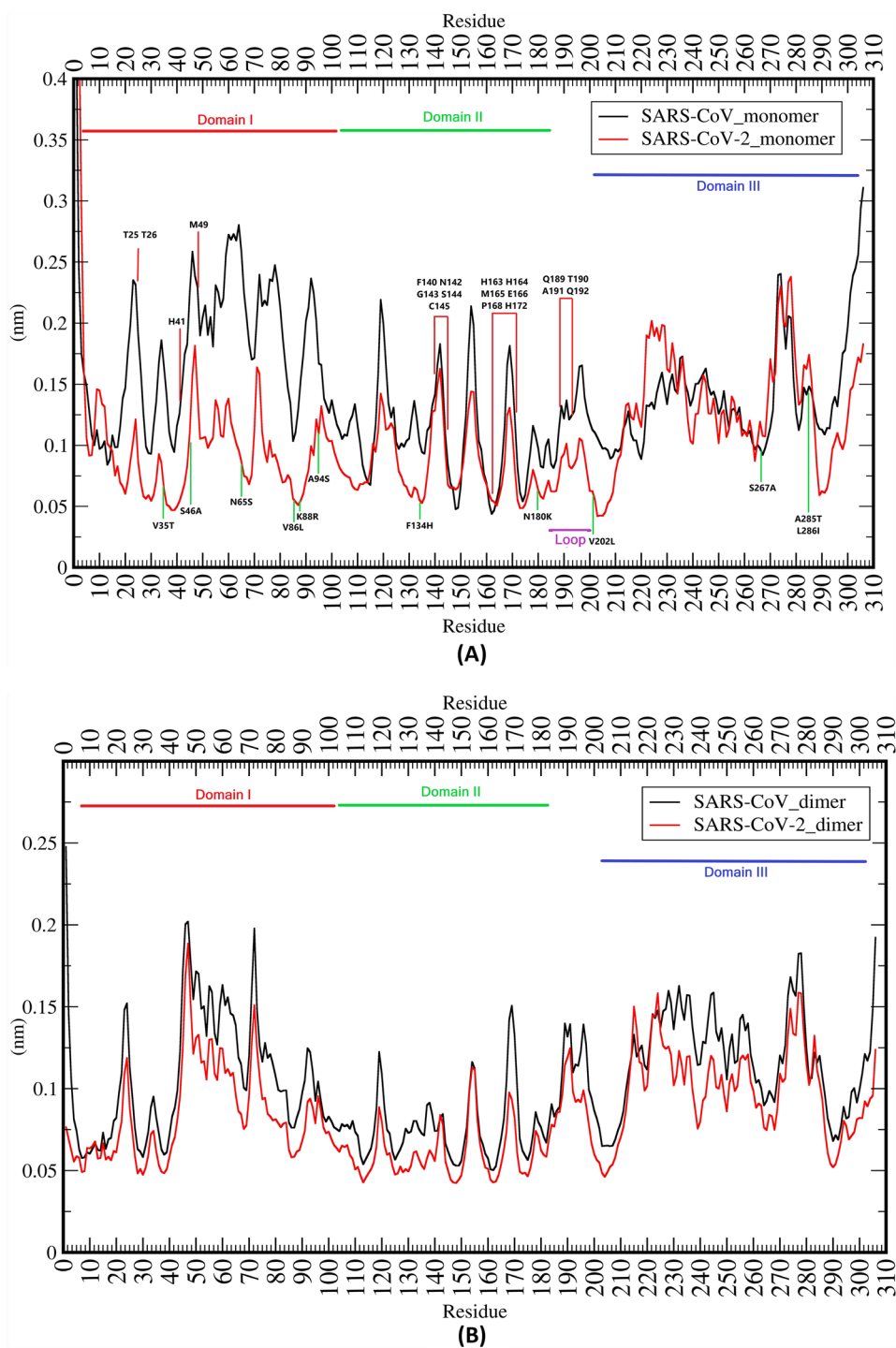
The SASA values for active site residues were also compared to compute the change in their exposure to solvent vis a vis 12 different residues. The dimer values of T25 were more significant than the corresponding monomer values in both the Mpro. For M49 and G143, SARS-CoV-2 monomer and dimer values were lower than those for SARS-CoV Mpro. T25 and M49 were very close to the variable positions 46 and 65 and were possibly affected by them. Values of the residues at F140, H172 and E166 in the case of the SARS-CoV-2 dimer were low compared to the SARS-CoV dimer, and these key residues are also involved in dimerization and vicinity to a variable residue H134F in SARS-CoV-2 Mpro. The catalytic residue at C145 had a lower SASA value in SARS-CoV-2 than in SARS-CoV in monomer and dimer forms. Notably, variable positions A46S and H134F were found in a loop leading to this catalytic H41 and C145.

### Residual fluctuations and molecular interactions in Mpro from SARS-CoV and SARS-CoV-2

The root mean square fluctuations (RMSF) value is an important measure of each atom's fluctuation from its average position. RMSF analysis reveals important insights into the flexibility of molecular regions. RMSF plots for each MD system are shown in Fig. 6. The calculated average RMSF values for SARS-CoV\_monomer, SARS-CoV-2\_monomer, SARS-CoV\_dimer and SARS-CoV-2\_dimer were  $0.14 \pm 0.054$ ,  $0.10 \pm 0.056$ ,  $0.1015 \pm 0.037$  and  $0.0825 \pm 0.029$  nm, respectively. SARS-CoV-2 Mpro had lower RMSF values overall, with SARS-CoV-2\_dimer having the lowest indicating higher stability and lower fluctuations. RMSF analysis of the protein backbone in SARS-CoV\_monomer revealed the domain I region followed by domain II region as two of the most highly flexible regions in the case of SARS-CoV\_monomer. As shown in Fig. 6, we observed fewer fluctuations in domains I and II in the SARS-CoV-2\_monomer, indicating the formation of a more stable molecule. In SARS-CoV-2\_monomer, we found a continuous stretch of residues 20–110 to have many fewer fluctuations than its counterpart in SARS-CoV\_monomer; overall RMSF values of each domain in SARS-CoV\_dimer were lower than those in SARS-CoV\_dimer.

Further, we analysed the individual RMSF values of the 12 divergent residues and the residues forming active sites in Mpro and plotted them, as shown in Fig. 7. In the case of SARS-CoV-2\_monomer, all the 12 divergent residues showed a significant decrease in fluctuations (RMSF) than SARS-CoV\_monomer Mpro, except for the three variant

**Fig. 6** Residue-wise RMSF deviations (nm) of Mpro from SARS-CoV and SARS-CoV-2 in monomer and dimer form: **A** RMSF plot of both CoV Mpro in monomer form. The 12 divergent residues in SARS-CoV-2 are marked with a green line in the lower plot. The active site residues are marked with a red line in the upper plot. **B** RMSF plot of both CoV Mpro in dimer form. In both panels, the domain I, II and III residues are marked in red, green and blue, respectively

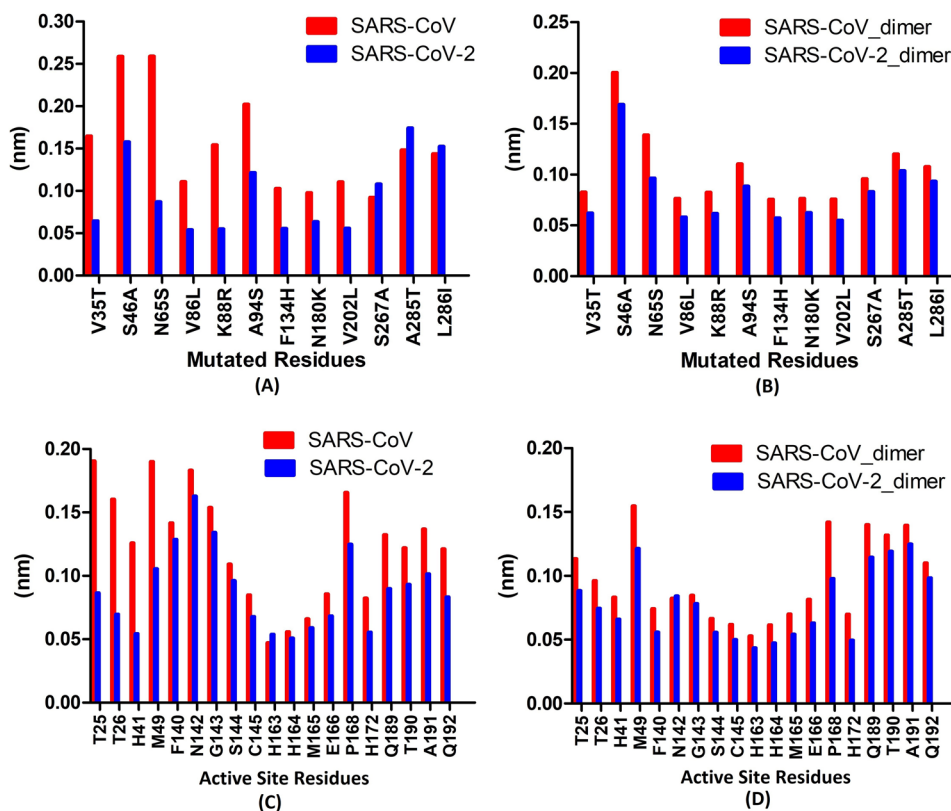


A267S, T285A and I286L, which are found in domain III. Alternatively, in both CoV dimers, all the variant residues in SARS-CoV-2 showed fewer fluctuations than SARS-CoV. The above three variants present in domain III also showed lesser values, maybe due to dimerization implications. Notably, in the case of SARS-CoV-2\_monomer Mpro, the T35V, S65N, R88K, H134F and L202V residues had at least 50% lesser RMSF values compared to SARS-CoV\_monomer.

The variable residue S65N is close to the binding site (T25, T26, M49 and Q189); R88K is close to the catalytic site; H134F is present in the loop that leads to catalytic C145, as well as H172, E166 and F140 (residues involved in dimerization) are present in its vicinity.

The individual RMSF values of residues from the active site showed a trend similar to that observed in the 12 divergent residues (Fig. 7). The residues of domain I had

**Fig. 7** RMSF plots showing deviations (nm) of selected residues from SARS-CoV and SARS-CoV-2 Mpro in monomer and dimer form: RMSF plots of 12 divergent residues from both the CoV Mpro in monomer (A) and dimer (B). The X-axis represents the residue present in SARS-CoV-2, followed by the residue number and ending with the residue found in SARS-CoV. RMSF plots of active site residues from the CoV Mpro in monomer (C) and dimer (D)



significantly lower values in SARS-CoV-2\_monomer than in SARS-CoV\_monomer. In SARS-CoV-2\_monomer Mpro, the T25, T26, H41 and M49 residues had significantly lower RMSF values than in SARS-CoV\_monomer Mpro. Notably, the environments of the above residues were directly affected by variable residues like R88K, S65N and A46S. Unlike the monomer CoV, in the case of both the CoV dimer, all the active site residues in SARS-CoV-2\_dimer showed fewer fluctuations than SARS-CoV\_dimer. We observed a similar trend in the RMSF values of the 12 variant residues. Surprisingly, the residues involved in dimerization (F140, H172 and E166) showed no significant RMSF differences between monomer and dimer CoV. Overall, the results of our RMSF analyses correlated with those of our SASA analyses.

We analysed molecular interactions of critical residues in CoV Mpro (monomer and dimer) using the *foldx* program to assess the final structure obtained after MD simulations. We assessed the change in molecular interactions or networks in key residues that might have been introduced due to mutations in SARS-COV-2. The variable position at A46S, located near the entrance of the binding site (Fig. 2), lost its interaction with M49 (active site residue) in both the monomer and dimer forms of SARS-CoV-2 Mpro, which was found in a previous study [43]. The variable residue at position H134F resulted in a substitution of +ve charge H134 to a hydrophobic F134 that altered the environment, and its location at the loop leading to the oxyanion loop

means it might serve to modulate the active site (interaction unique to SARS-CoV-2: A105 and SARS-CoV: G183). Variable residue at L86V, located near the catalytic site, resulted in a loss of interaction in SARS-CoV-2 (SARS-CoV: G179). Another variable position, R88K, located near the catalytic site, altered the electrostatic interaction profile of SARS-CoV-2 (unique to SARS-CoV-2: E55 & H164 and SARS-CoV: F103 & K180). A94's variable position resulted in the loss of its interaction with P96 in SARS-CoV-2.

Interestingly, the variable position at K180N resulted in a loss of all electrostatic interactions in SARS-CoV-2 due to the mutation of +ve charged residue (K) to a polar uncharged residue (N). Otherwise, K180 in SARS-CoV had electrostatic interactions with R40, R88, R105, D176, E178 and D187. The variable residues T285A and I286L in SARS-CoV\_2 resulted in changes in interaction compared with SARS-CoV and new connections with T280 and G283 (both chain B) at the interface of the dimer.

The catalytic residues H41 and C145 did not have a much-altered environment due to the mutations found in SARS-CoV-2. However, a new H-bond network was formed in SARS-CoV-2: H41 with H164, which is also one of the residues involved in active site formation. Such N-H...N hydrogen bonds formed by imidazole groups of two histidine residues (H41: H164) are rare in proteins [44]. Our SASA analysis indicated that H41 & H164 are relatively buried, and such N-H...N H-bonds formed by

a pair of buried histidine may significantly contribute to structural stability [44] of the SARS-CoV-2 Mpro, which is evident by our above RMSD and RMSF analysis. There is no change in the network of C145 catalytic residues, which indicates that the integrity of C145 may be essential for its conserved protease activity. The M49 and Q189 at the active site entrance are essential gatekeepers for substrate binding [45]. The network of M49 has been changed slightly in SARS-CoV-2 Mpro (gained in SARS-CoV-2: T45 & Q189 and lost: P52 & D187).

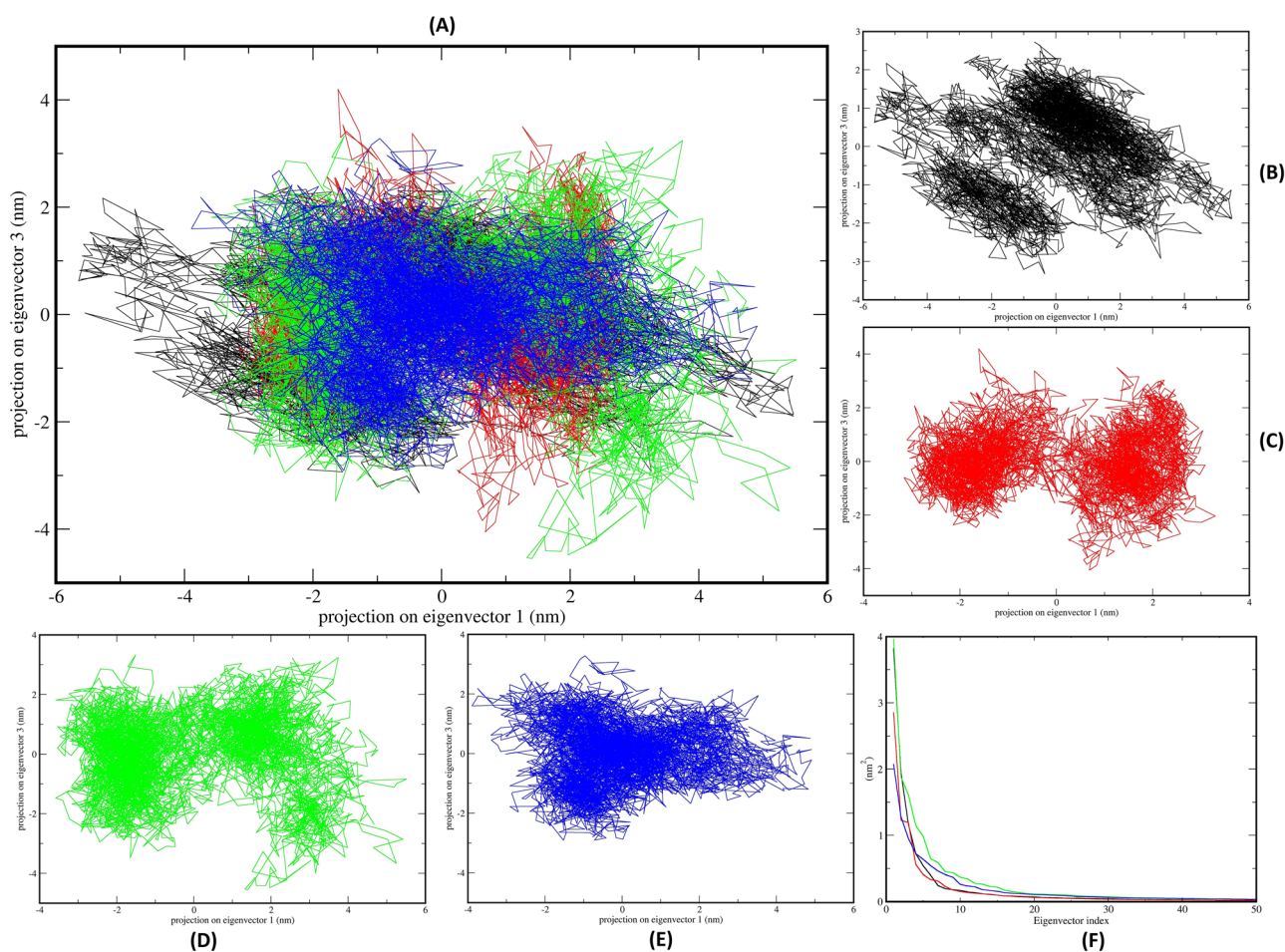
Similarly, in the other gatekeeper residue, Q192, its interaction with T190 was lost in SARS-CoV-2, while interaction with M165 was unique to SARS-CoV-2. G143 gained a new interaction with N28 in the case of SARS-CoV-2, which is absent in SARS-CoV. Similarly, a new interaction was gained by H163 to M165 in the case of SARS-CoV-2, which is absent in SARS-CoV. A critical residue, M165, found adjacent to the critical residue E166, which is required to open the substrate-binding site in CoV Mpro, has gained a unique connection in SARS-CoV-2 (HISA163, VALA186, ASPA187 and GLNA192) which are absent in SARS-CoV Mpro. P168 gained a new interaction with T169 in the case of SARS-CoV-2, which is absent in SARS-CoV. For position T190, a potential network with Q192, another active site residue was lost in the case of SARS-CoV-2 Mpro.

The Mpro dimer form is functional due to crucial interactions of residues F140, E166, H172 and H163 within and between the residues of another chain, specifically the N-finger that serves to open and close the active site as well as play an active role in dimerization [15, 40]. Our analysis of the interaction network of this essential residue revealed that the vital interaction of F140, E166 and E290 with N-finger residues of the other chain was maintained, in particular with residues S1 and R4 of the other chain. Apart from the N-finger residues, E166 also interacted with N214, D216 and C300 of another chain in SARS-CoV and SARS-CoV-2 Mpro, but we also observed a new interaction with SARS-CoV-2: R217. Interestingly, S1 from the N-finger of the other chain formed a unique H-bond with H172 in SARS-CoV-2 but was absent in SARS-CoV. H172 was one of the critical residues in the active site of CoV Mpro that formed a typical H-bond with S1 of another chain, which might also contribute to the restructuring process of SARS-CoV-2 Mpro. E290 was another crucial residue that interacted with R4 of another chain, forming a salt bridge interaction, which we observed in both CoV. Our collective analysis of RMSF, SASA and molecular interactions revealed that the 12 divergent residues in SARS-CoV-2 Mpro altered the microenvironments of neighbouring residues. These modified interaction networks ultimately restructured the molecular environment of the Mpro active-site residues at the entrance (T26, M49 and Q192) and near the catalytic region (F140, H163, H164, M165 and H172).

## Clustering of conformations for ensemble generation and Essential Dynamics

The conformational space and transitions in the SARS-CoV and SARS-CoV-2 Mpro for monomer and dimer were inspected by Principle Component Analysis (PCA) analysis. The PCA is a statistical computation that decreases the complexity of the MDS trajectories by extracting only the collective motion of C $\alpha$  atoms while preserving most of the other variations. It calculates the covariance matrix of positional fluctuations for backbone atoms which may decipher the dynamics and coherent motions of Mpro from both SARS-CoV and SARS-CoV-2. Figure 8 shows a plot of eigenvalues calculated from the covariance matrix of backbone fluctuations, plotted in decreasing order vs the respective eigenvector indices for all MD systems. Top 15 eigenvectors accounted for 77.50 %, 75.36 %, 68.08 % and 60.83 % of motions observed for 50-ns trajectory for SARS-CoV\_monomer, SARS-CoV-2\_monomer, SARS-CoV\_dimer and SARS-CoV-2\_dimer, respectively. The plot, shown in Fig. 8, is the 2D projection of the trajectories for two major principal components, PC1 and PC3, for SARS-CoV\_monomer, SARS-CoV-2\_monomer, SARS-CoV\_dimer and SARS-CoV-2\_dimer which represents different conformations in 2D space. The PCA analysis revealed the following observations. First, the 2D projection of SARS-CoV\_monomer (Fig. 8B) has more variation compared to the other three: SARS-CoV-2\_monomer (Fig. 8C), SARS-CoV\_dimer (Fig. 8D) and SARS-CoV-2\_dimer (Fig. 8E). Second, it is evident from the 2D plot that the SARS-CoV-2\_monomer (Fig. 8C) and SARS-CoV-2\_dimer (Fig. 8E) showed higher stability and occupied lesser phase space compared to SARS-CoV\_monomer (Fig. 8B) and SARS-CoV\_dimer (Fig. 8D). This indicates that the Mpro of SARS-CoV-2 is stable in both forms and less flexible than SARS-CoV Mpro. The covariance and 2D plot analysis also indicated the presence of two well-defined clusters in SARS-CoV\_monomer, SARS-CoV-2\_monomer & SARS-CoV\_dimer and only one defined cluster in the case of SARS-CoV-2\_dimer. The covariance plots depict the positive and negative limits are depicted by the covariance plots where positive values are related to the motion of the atoms occurring along the same direction (correlated), whereas the negative values indicate the motion of the atoms in the opposite direction (anti-correlated). Our PCA analysis from MD simulations (50 ns) revealed that the SARS-CoV\_monomer (Fig. 8B) and SARS-CoV\_dimer (Fig. 8D) had large anti-correlated motion and the SARS-CoV-2\_monomer (Fig. 8C) and SARS-CoV-2\_dimer (Fig. 8E) had a balance of correlated as well as anti-correlated motion. Thus, the above results concluded that SARS-CoV\_monomer and SARS-CoV\_dimer had increased flexibility and conformational space compared to the SARS-CoV-2\_monomer and SARS-CoV-2\_dimer





**Fig. 8** PCA 2D projection scatters plot of SARS-CoV and SARS-CoV-2 Mpro: **A** Overlay of 2D scatter plot projection of the motion of the proteins in phase space for the two principal components, PC1 and PC3, derived from four MD simulation setups. Panel **B**, **C**, **D** and **E** represent individual 2D plots of SARS-CoV\_monomer, SARS-CoV-2\_monomer, SARS-CoV\_dimer and SARS-CoV-2\_dimer, respectively.

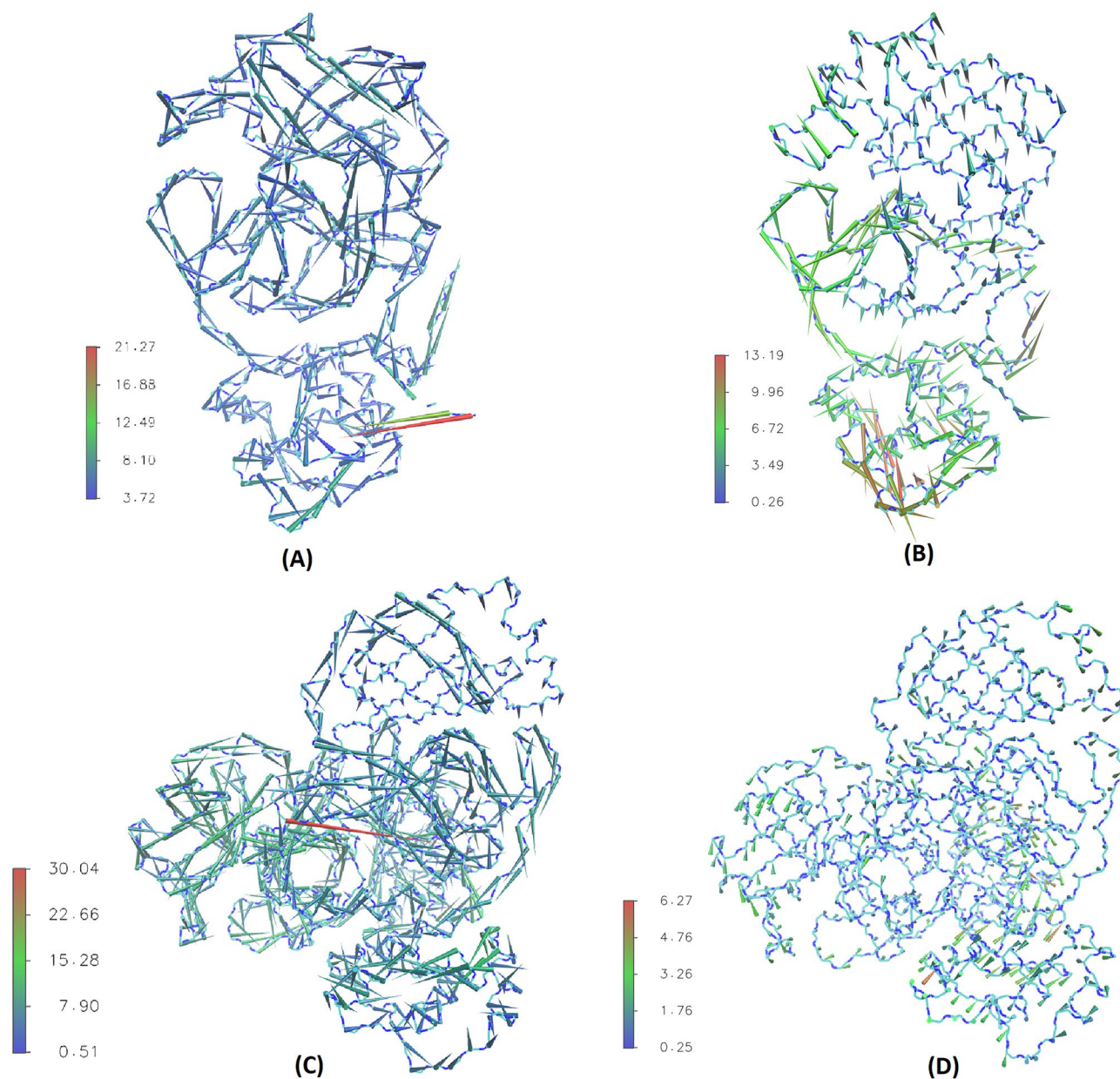
and notably, the most stable and less flexible was of SARS-CoV-2\_dimer. Fifty structures were extracted from each MD simulation for essential dynamics visualization, projecting the extremely selected eigenvectors (Fig. S1). The extreme motion of SARS-CoV-2\_monomer and SARS-CoV-2\_dimer was less deviating compared to SARS-CoV\_monomer and SARS-CoV\_dimer, indicating a stable conformational space. The width of the main-chain trace represents fluctuations throughout the timescale of MD simulations.

For visualizing the direction and extent of the principal motions, the first and last eigenvector was plotted in a porcupine plot representation in which the arrows indicate the direction of the eigenvector and the magnitude of the corresponding value (Fig. 9). The plot shows that SARS-CoV\_monomer had cone projection throughout the three domains, while the SARS-CoV-2\_monomer had less coherent motion in domains I and II. At the same time, the most stable and

dimer, respectively. **F** Plot representing Eigenvalues calculated from the covariance matrix of C $\alpha$  backbone fluctuations vs the respective eigenvector indices for the first 50 eigenvectors from 1000 eigenvectors. For all the panels, the colour representation is SARS-CoV\_monomer (black), SARS-CoV-2\_monomer (red), SARS-CoV\_dimer (green) and SARS-CoV-2\_dimer (blue)

with the tiniest cone projection was the SARS-CoV-2\_dimer indicating the most stable system formation. This result was in correlation with PCA analysis.

GROMOS 36-based clustering algorithm was applied to all four MDS setups to generate representative structures for the conformational space traversed by the MDS. The method creates representative RMSD-based clusters from the trajectory frames. It counts the number of the neighbouring structure using a 0.15-nm cut-off and then forms a cluster set with the largest number of neighbour structures, followed by its elimination from the pool of clusters. The process is repeated for the remaining frames to identify other clusters with decreasing number of neighbour structures, and each cluster's centroid is used as a representative structure. These centroid structure members from each cluster are representative structures of different frames. The RMSD values in the clusters range from 0.0519 to 0.41 nm (average 0.158004), 0.0501

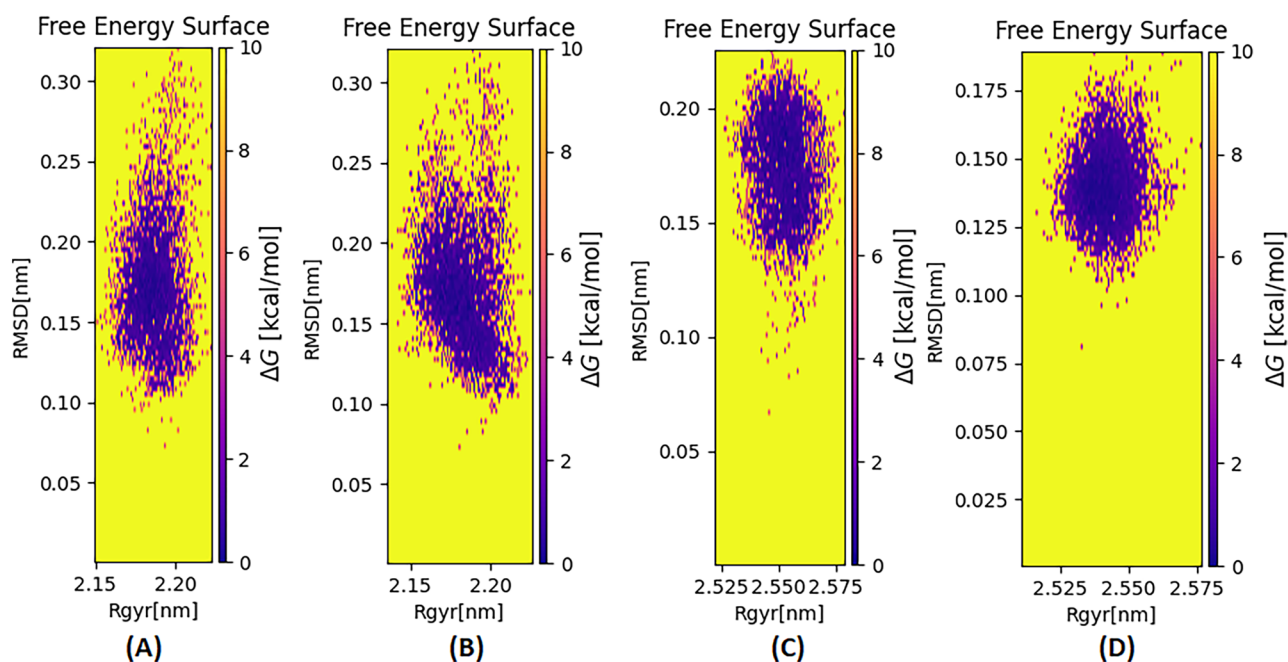


**Fig. 9** The porcupine plot for conformational variability is computed from the crystal structure and average MD simulations ensemble. Porcupine plots of **A** SARS-CoV\_monomer, **B** SARS-CoV-2\_monomer, **C** SARS-CoV\_dimer and **D** SARS-CoV-2\_dimer. The length of the

cone is proportional to the conformational variability, while the colour of the cone is represented by deviation in RMSD as indicated in the respective colour scale in each plot

to 0.304 nm (average 0.146737), 0.0517 to 0.273 nm (average 0.139786) and 0.0532 to 0.247 nm (average 0.128409) for SARS-CoV\_monomer, SARS-CoV-2\_monomer, SARS-CoV\_dimer and SARS-CoV-2\_dimer, respectively. Nineteen, 16, 12 and 5 clusters with 798, 1273, 637 and 158 transitions were observed for SARS-CoV\_monomer, SARS-CoV-2\_monomer, SARS-CoV\_dimer and SARS-CoV-2\_dimer, respectively. The representative structures from each cluster from SARS-CoV\_dimer and SARS-CoV-2\_dimer simulations are shown in Supplementary Figs. S2 and S3. Our

cluster analysis indicated that the Mpro of SARS-CoV-2 in both forms is stable and less flexible than SARS-CoV Mpro, while the SARS-CoV-2\_dimer had a stable and least number of conformations. The free energy surface analysis was performed for all the MD systems, representing the conformational variability in ROG and RMSD taken together and represented by Gibbs free energy (Fig. 10). The same trend as PCA and cluster analysis was also observed in FES. CoV\_dimer Mpro had converged free energy, and specifically, the SARS-CoV-2\_dimer had the most converged free energy



**Fig. 10** Free energy surface of Mpro computed over entire simulations (50ns): FES (in kcal/mol) for Mpro from **A** SARS-CoV\_monomer, **B** SARS-CoV-2\_monomer, **C** SARS-CoV\_dimer and **D** SARS-CoV-2\_

dimer considering the conformational variability in terms of ROG and RMSD took together and represented by Gibbs free energy

representing clustered RMSD and ROG values observed in its entire simulations.

## Conclusion

A fatal outbreak of a novel coronavirus (SARS-CoV-2) was identified as a causative agent for SARS disease in Wuhan; China infected millions of lives, affecting ~213 countries globally. WHO announced a Public Health Emergency of International Concern (PHEIC), referring to the pathogen as SARS-CoV-2 and declaring it a pandemic [46]. Mpro is the central protease involved in polyprotein processing, thus assisting the virus in infection and proliferation. Despite having a high sequence similarity level, these homologous proteins' active sites differed significantly in shape and size, implying that repurposing drugs for COVID-19 may be futile. The Mpro cleaves the pp1ab (polyprotein) into mature nsp(s), which are essential for viral replication and are rather evolutionary conserved among coronaviruses [47]. The human counterpart of viral proteases with the same specificity has not been discovered, making Mpro an ideal target for drug discovery against coronavirus infections [48]. The Mpro monomer is enzymatically inactive, unlike the dimeric Mpro. Generally, studies targeting Mpro have employed two strategies: (i) lead molecules targeting the substrate-binding pocket to block the catalytic activity and (ii) molecules inhibiting the Mpro dimerization [49].

However, no lead molecule targeting the catalytic binding pocket has reached clinical trials. The latter strategy targeting dimerization is not explored with full potential, and there are few reports on inhibitors targeting Mpro dimerization [50, 51]. The residues involved in Mpro dimerization are Arg4, Ser10, Gly11, Glu14, Asn28, Ser139, Phe140, Ser147, Glu290 and Arg298 [49]. An earlier study on SARS-CoV Mpro identified conserved serine residues (Ser139, Ser144 and Ser147) located in the proximity of the active site that can be targeted to inhibit the protease activity [52]. An earlier study indicated that the alanine substitution of Ser147 leads to dimer instability, and this Ser147 forms H-bonds with the backbone of Ser144 and His163 residues [53]. Our study revealed that Met165 had a unique connection in SARS-CoV-2, particularly with His163, absent in SARS-CoV Mpro.

The detailed analysis of the conformational changes in the binding pocket throughout the simulation shows the changes in flexibility and plasticity, which shows the possibility of reliable drug design. The effect of the variant residues on overall protein stability, active site residues and fluctuations were also analysed for both the CoVs Mpro in a monomer and a dimer. The insights revealed in our study agree with an earlier similar MD simulation study of the dimeric and monomeric form of a SARS-CoV Mpro by Zheng et al. [54]. The present MD simulation study attempted to test how the mutation accumulated in SARS-CoV-2 Mpro indirectly reconfigures the critical molecular networks around



the active site in SARS-CoV-2. Suarez and Diaz have performed a similar study investigating the structure and flexibility of the hydrated SARS-CoV-2 Mpro with 2- $\mu$ s MD simulation [55]. Our present findings are somewhat consistent with this report; particularly the inter-domain motions, interactions, inter-residue contacts and molecular networks are the catalytic site. A complete comparison of our study with the previously reported ones was not feasible. The primary reason is that most of the similar reports are focused on structure-activity relationships, whereas the present study aims to compare the SARS-CoV and SARS-CoV-2 Mpro monomer and dimer and study the effect of variable residues (mutations) on the structure.

Besides, it was also found that SARS-CoV-2 Mpro is more stable than SARS-CoV Mpro; the biological active SARS-CoV-2 dimer was more stable. A similar finding was reported by Goyal and Goyal, where emphasis on considering the dimer form of Mpro for drug designing was highlighted [49]. Although our results differ from a similar study by Tekpinar and Yildirim which concluded that SARS-CoV Mpro and SARS-CoV-2 Mpro display similar dynamics, the study included an N3 ligand, which could have impacted such behaviour [56]. Furthermore, another study attempted to identify the dimer-monomer equilibrium of SARS-CoV-2 Mpro by small molecule inhibitor using SAXS experiments [57]. Therefore, we suggest utilizing the present findings for any further drug identification and repurposing studies.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11224-022-02089-6>.

**Author contribution** Dhaval Patel: Conceptualization, methodology, investigation, analysis, writing—original draft preparation. Meet Parmar: Writing—original draft preparation, methodology, analysis. Ritik Thumar: Methodology, analysis. Bhumi Patel: Methodology, analysis. Mohd Athar: Writing—reviewing and editing, validation. Prakash Jha: Resources, validation.

**Funding** The authors acknowledge the Gujarat State Biotechnology Mission (GSBTM), Department of Science & Technology (DST), Gujarat, for financial support and resources in a project grant (GSBTM/JD(R&D)/610/20-21/351). The authors also acknowledge the Institute of Advanced Research, Gandhinagar, for resources and infrastructure facilities. The authors also thank the Gujarat Council on Science and Technology (GUJCOST), Gujarat, for the PARAM SHAVAK supercomputing facility. Finally, the author acknowledges the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), for a project grant, EMR/2016/003025.

**Data availability** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

**Code availability** Not applicable.

## Declarations

**Competing interests** The authors declare no competing interests.

## References

- Wang C, Horby PW, Hayden FG, Gao GF (2020) A novel coronavirus outbreak of global health concern. *Lancet* 395:470–473
- Chan JFW, Yuan S, Kok KH et al (2020) A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 395:514–523. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9)
- Zhou P, Yang X-L, Wang X-G et al (2020) Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *bioRxiv*. <https://doi.org/10.1101/2020.01.22.914952>
- Hui DS, Azhar EI, Madani TA et al (2020) The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis* 91:264–266
- Weiss SR, Leibowitz JL (2011) Coronavirus pathogenesis. In: *Advances in Virus Research*. Academic Press Inc., pp 85–164
- Zhu N, Zhang D, Wang W et al (2020) A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. <https://doi.org/10.1056/nejmoa2001017>
- Zhou P, Yang X-L, Wang X-G et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>
- Peiris JSM, Guan Y, Yuen KY (2004) Severe acute respiratory syndrome. *Nat Med* 10:S88–S97
- Zaki AM, Van Boheemen S, Bestebroer TM et al (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367:1814–1820. <https://doi.org/10.1056/NEJMoa1211721>
- Lu H, Stratton CW, Tang Y-W (2020) Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle. *J Med Virol* 92:401–402. <https://doi.org/10.1002/jmv.25678>
- Wu C, Liu Y, Yang Y et al (2020) Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm Sin B*. <https://doi.org/10.1016/j.apsb.2020.02.008>
- Chen Y, Liu Q, Guo D (2020) Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol* 92:418–423
- Benvenuto D, Giovannetti M, Ciccozzi A et al (2020) The 2019-new coronavirus epidemic: evidence for virus evolution. *bioRxiv*. <https://doi.org/10.1101/2020.01.24.915157>
- Anand K, Palm GJ, Mesters JR et al (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J* 21:3213–24. <https://doi.org/10.1093/emboj/cdf327>
- Anand K, Ziebuhr J, Wadhwani P et al (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* (80-) 300:1763–1767. <https://doi.org/10.1126/science.1085658>
- Zhang L, Lin D, Sun X et al (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science*. <https://doi.org/10.1126/science.abb3405>
- Jin Z, Du X, Xu Y et al (2020) Structure of Mpro from COVID-19 virus and discovery of its inhibitors. 2020.02.26.964882
- Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Larkin MA, Blackshields G, Brown NP et al (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725. <https://doi.org/10.1093/MOLBEV/MST197>



21. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31:3320–3323
22. Schrodinger LLC (2015) The PyMOL molecular graphics system. Version 1:8
23. Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera?A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. <https://doi.org/10.1002/jcc.20084>
24. Patel P, Parmar K, Vyas VK et al (2017) Combined in silico approaches for the identification of novel inhibitors of human islet amyloid polypeptide (hIAPP) fibrillation. *J Mol Graph Model* 77. <https://doi.org/10.1016/j.jmgm.2017.09.004>
25. Patel B, Patel D, Parmar K et al (2018) L. donovani XPRT: molecular characterization and evaluation of inhibitors. *Biochim Biophys Acta - Proteins Proteomics* 1866. <https://doi.org/10.1016/j.bbapap.2017.12.002>
26. Chaudhari A, Chaudhari M, Mahera S et al (2021) In-silico analysis reveals lower transcription efficiency of C241T variant of SARS-CoV-2 with host replication factors MADP1 and hnRNP-1. *Informatics Med Unlocked* 25:100670
27. Van Der Spoel D, Lindahl E, Hess B et al (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701–1718. <https://doi.org/10.1002/jcc.20291>
28. Lindorff-Larsen K, Piana S, Palmo K et al (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct Funct Bioinforma NA-NA*. <https://doi.org/10.1002/prot.22711>
29. Berendsen HJC, Postma JPM, Van Gunsteren WF et al (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690. <https://doi.org/10.1063/1.448118>
30. Parrinello M, Rahman A (1980) Crystal structure and pair potentials: a molecular-dynamics study. *Phys Rev Lett* 45:1196–1199. <https://doi.org/10.1103/PhysRevLett.45.1196>
31. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092. <https://doi.org/10.1063/1.464397>
32. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12%3c1463::AID-JCC4%3e3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3c1463::AID-JCC4%3e3.0.CO;2-H)
33. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
34. Patel P, Parmar K, Patel D et al (2018) Inhibition of amyloid fibril formation of lysozyme by ascorbic acid and a probable mechanism of action. *Int J Biol Macromol* 114:666–678. <https://doi.org/10.1016/j.ijbiomac.2018.03.152>
35. Daura X, Gademann K, Jaun B et al (1999) Peptide folding: when simulation meets experiment. *Angew Chemie Int Ed* 38:236–240. [https://doi.org/10.1002/\(SICI\)1521-3773\(19990115\)38:1%3c236::AID-ANIE236%3e3.0.CO;2-M](https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1%3c236::AID-ANIE236%3e3.0.CO;2-M)
36. Manhas A, Patel D, Lone MY, Jha PC (2019) Identification of natural compound inhibitors against *Pf*DXR: a hybrid structure-based molecular modeling approach and molecular dynamics simulation studies. *J Cell Biochem* 120:14531–14543. <https://doi.org/10.1002/jcb.28714>
37. Patel D, Athar M, Jha PC (2021) Exploring ruthenium-based organometallic inhibitors against plasmodium falciparum calcium dependent kinase 2 (PfCDPK2): a combined ensemble docking, QM/MM and molecular dynamics study. *ChemistrySelect* 6:8189–8199. <https://doi.org/10.1002/SLCT.202101801>
38. Patel D, Athar M, Jha PC (2022) Computational investigation of binding of chloroquinone and hydroxychloroquinone against PLPro of SARS-CoV-2. *J Biomol Struct Dyn* 40:1–11
39. Sharma A, Vora J, Patel D et al (2020) Identification of natural inhibitors against prime targets of SARS-CoV-2 using molecular docking, molecular dynamics simulation and MM-PBSA approaches. *J Biomol Struct Dyn* 40:1–16. <https://doi.org/10.1080/07391102.2020.1846624>
40. Shi J, Sivaraman J, Song J (2008) Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. *J Virol* 82:4620–4629. <https://doi.org/10.1128/jvi.02680-07>
41. Lim L, Shi J, Mu Y, Song J (2014) Dynamically-driven enhancement of the catalytic machinery of the SARS 3C-like protease by the S284–T285–I286/A mutations on the extra domain. *PLoS One* 9:101941. <https://doi.org/10.1371/journal.pone.0101941>
42. Laskowski RA, Jabłońska J, Pravda L et al (2018) PDBsum: structural summaries of PDB entries. *Protein Sci* 27:129–134. <https://doi.org/10.1002/PRO.3289>
43. Proteases SM, Krishnamoorthy N (2020) Variable structural networks at the active site of the SARS-CoV and PrePrint.org 1–17. <https://doi.org/10.20944/preprints202003.0423.v1>
44. Iyer AH, Krishna Deepak RNV, Sankararamkrishnan R (2018) Imidazole nitrogens of two histidine residues participating in N-H...N hydrogen bonds in protein structures: structural bioinformatics approach combined with quantum chemical calculations. *J Phys Chem B* 122:1205–1212. <https://doi.org/10.1021/acs.jpcc.7b11737>
45. Xue X, Yang H, Shen W et al (2007) Production of authentic SARS-CoV Mpro with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J Mol Biol* 366:965–975. <https://doi.org/10.1016/j.jmb.2006.11.073>
46. Cucinotta D, Vanelli M (2020) WHO declares COVID-19 a pandemic. *Acta Biomed* 91:157–160. <https://doi.org/10.23750/ABM.V91I11.9397>
47. Maffucci I, Contini A (2020) In silico drug repurposing for SARS-CoV-2 main proteinase and spike proteins. *Cite This J Proteome Res* 19:4648. <https://doi.org/10.1021/acs.jproteome.0c00383>
48. Ramos-Guzmán CA, Ruiz-Pernía JJ, Tuñón I (2020) Unraveling the SARS-CoV-2 main protease mechanism using multiscale methods. *ACS Catal* 10:12544–12554. [https://doi.org/10.1021/ACSCATAL.0C03420/SUPPL\\_FILE/CS0C03420\\_SI\\_003.MP4](https://doi.org/10.1021/ACSCATAL.0C03420/SUPPL_FILE/CS0C03420_SI_003.MP4)
49. Goyal B, Goyal D (2020) Targeting the dimerization of the main protease of coronaviruses: a potential broad-spectrum therapeutic strategy. *ACS Comb Sci* 22:297–305. <https://doi.org/10.1021/acscombsci.0c00058>
50. Wei P, Fan K, Chen H et al (2006) The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochem Biophys Res Commun* 339:865–872. <https://doi.org/10.1016/J.BBRC.2005.11.102>
51. Ding L, Zhang XX, Wei P et al (2005) The interaction between severe acute respiratory syndrome coronavirus 3C-like proteinase and a dimeric inhibitor by capillary electrophoresis. *Anal Biochem* 343:159. <https://doi.org/10.1016/J.AB.2005.04.027>
52. Bacha U, Barrila J, Velazquez-Campoy A et al (2004) Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. *Biochemistry* 43:4906–4912. <https://doi.org/10.1021/BIO361766>
53. Barrila J, Bacha U, Freire E (2006) Long range cooperative interactions modulate dimerization in SARS 3CLpro. *Biochemistry* 45:14908. <https://doi.org/10.1021/BIO616302>
54. Zheng K, Ma G, Zhou J et al (2007) Insight into the activity of SARS main protease: molecular dynamics study of dimeric and monomeric form of enzyme. *Proteins* 66:467–479. <https://doi.org/10.1002/PROT.21160>
55. Suárezsuárez D, Díaz N (2020) SARS-CoV-2 main protease: a molecular dynamics study. <https://doi.org/10.1021/acs.jcim.0c00575>

56. Tekpinar M, Yildirim A (2021) Impact of dimerization and N3 binding on molecular dynamics of SARS-CoV and SARS-CoV-2 main proteases. <https://doi.org/10.1080/07391102.2021.1880481>
57. Silvestrini L, Belhaj N, Comez L et al (123AD) The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors. *Sci Reports* 11:9283. <https://doi.org/10.1038/s41598-021-88630-9>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.