Research Article

Insights into the structural properties of SARS-CoV-2 main protease

Ibrahim Yagiz Akbayrak a, Sule Irem Caglayan b, Lukasz Kurgan c,***, Vladimir N. Uversky d, e,***, Orkid Coskuner-Weber b,*

a Materials Sciences and Technologies, Turkish-German University, Sahinkaya Caddesi, No. 106, Beykoz, Istanbul, 34820, Turkey
b Molecular Biotechnology, Turkish-German University, Sahinkaya Caddesi, No. 106, Beykoz, Istanbul, 34820, Turkey
c Department of Computer Science, Virginia Commonwealth University, Richmond, VA, 23284, USA
d Molecular Medicine, USF Health Byrd Alzheimer’s Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA
e Laboratory of New Methods in Biology, Institute for Biological Instrumentation of the Russian Academy of Sciences, Federal Research Center “Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences”, Pushchino, Russia

ARTICLE INFO

Keywords:
SARS-CoV-2 main protease
Dynamics
Replica exchange MD simulations

ABSTRACT

SARS-CoV-2 is the infectious agent responsible for the coronavirus disease since 2019, which is the viral pneumonia pandemic worldwide. The structural knowledge on SARS-CoV-2 is rather limited. These limitations are also applicable to one of the most attractive drug targets of SARS-CoV-2 proteins – namely, main protease Mpro, also known as 3C-like protease (3CLpro). This protein is crucial for the processing of the viral polyproteins and plays crucial roles in interfering viral replication and transcription. In fact, although the crystal structure of this protein with an inhibitor was solved, Mpro conformational dynamics in aqueous solution is usually studied by molecular dynamics simulations without special sampling techniques. We conducted replica exchange molecular dynamics simulations on Mpro in water and report the dynamic structures of Mpro in an aqueous environment including root mean square fluctuations, secondary structure properties, radius of gyration, and end-to-end distances, chemical shift values, intrinsic disorder characteristics of Mpro and its active sites with a set of computational tools. The active sites we found coincide with the currently known sites and include a new interface for interaction with a protein partner.

Humans and animals are regularly infected by coronaviruses (CoVs). While typically this type of infection has rather mild respiratory symptoms, and most coronaviruses are not dangerous, there are also instances when a CoV infection can cause severe respiratory illness. Prior to the end of 2019, illustrative examples of such severe forms of CoV infection were given by the outbreaks of Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) caused by the SARS-CoV and MERS-CoV infections. Although at the time of their appearance (2003 for SARS and 2012 for MERS) these CoVs were considered as a major threat, their danger was not even close to that posed by the newest representative of the Coronaviridae family, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing the coronavirus disease 2019 (COVID-19). This is reflected in the values of the corresponding morbidity and mortality rates. In fact, by the time of its containment in October 4th, 2022, there were 624,083,852 COVID-19 cases in 230 countries and territories of the world, and infecting millions of people. Specifcally, according to the Worldometer, https://www.worldometers.info/coronavirus/, as of October 4th, 2022, there were 624,083,852 COVID-19 cases in 230 countries.

SARS-CoV-2 main protease (Mpro), also known as 3CLpro, is one of the most attractive drug targets of SARS-CoV-2, and its inhibition is expected to effectively block viral replication. SARS-CoV-2 Mpro is a dimeric aspartic protease, which inactivates the viral replication machinery by cleaving the viral polyprotein P0 into functional nonstructural proteins (NSPs) (Pang et al., 2020). The 3CLpro domain is a central element of the viral NSPs and contains residues that are highly conserved across different CoV species (Li et al., 2020). The proteolytic activity of 3CLpro is essential for the virus life cycle, and its inhibition is expected to provide a promising therapeutic strategy (Chen et al., 2020). The 3CLpro domain has been successfully targeted for the development of antiviral drugs (Chen et al., 2020; Liu et al., 2020). However, the crystal structure of this protein with an inhibitor was solved, and its conformational dynamics in aqueous solution is usually studied by molecular dynamics simulations without special sampling techniques. We conducted replica exchange molecular dynamics simulations on Mpro in water and report the dynamic structures of Mpro in an aqueous environment including root mean square fluctuations, secondary structure properties, radius of gyration, and end-to-end distances, chemical shift values, intrinsic disorder characteristics of Mpro and its active sites with a set of computational tools. The active sites we found coincide with the currently known sites and include a new interface for interaction with a protein partner.

**Corresponding author.
*** Corresponding author.
**** Corresponding author. Molecular Medicine, USF Health Byrd Alzheimer’s Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA.
E-mail addresses: lkurgan@vcu.edu (L. Kurgan), vuversky@health.usf.edu (V.N. Uversky), weber@tau.edu.tr (O. Coskuner-Weber).

https://doi.org/10.1016/j.crstbi.2022.11.001
Received 17 October 2022; Received in revised form 23 November 2022; Accepted 25 November 2022
2665-926X/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
countries.

SARS-CoV-2's RNA genome embraces 29,903 nucleotides and encodes three proteins (spike glycoprotein (S), membrane protein (M), and nucleocapsid protein (N)) and several non-structural proteins (NSPs) (Khaliyant et al., 2020). A single large replicate gene is responsible for the encoding of the proteins that are at the center of viral replication. The replicate gene encodes the overlapping polyproteins called pp1a and pp1ab that are necessary for viral transcription and replication. The larger replicate polyprotein 1 ab has 7,073 amino acid residues, which contains fifteen non-structural proteins. Nsp1, Nsp2, and Nsp3 are released from polyprotein through proteolytic processing using a viral papain-like proteinase (Nsp3/PL-Pro), while the rest are cleaved by viral 3C-like proteinase, Nsp3/3CL\textsuperscript{Pro} or main protease M\textsuperscript{Pro}, uses eleven or more conserved sites for digesting the protein. The digestion is initiated by an autocatalytic cleavage of this enzyme from pp1a and pp1ab. The functional importance of M\textsuperscript{Pro} in the viral life cycle is the main reason for it being an attractive target for antiviral treatment design.

Yang and co-workers used computer-aided drug design (CADD) to create a series of Michael acceptor inhibitors including an inhibitor named N3 that efficiently inhibits various CoV M\textsuperscript{Pro} species from SARS-CoV to SARS-CoV-2. This inhibitor forms a reversible complex with the protease, which in turn is prone to a chemical step that yields the formation of a stable covalent bond (Matthews et al., 1999). Recently, researchers solved a crystal structure of the M\textsuperscript{Pro} from SARS-CoV-2 in a complex with its inhibitor N3, reporting an active peptide fragment (Ala-Val-Leu) that binds specifically to the substrate-binding pocket of SARS-CoV-2 (Jin et al., 2020a). Immediately after its publication, this study raised significant interest from the scientific community, and since then guides some of the efforts towards the design of treatment for COVID-19 infection.

A crystal structure represents only static snapshots of a dynamic system, and therefore does not embrace the impacts of structural dynamics and the bulk water effects on M\textsuperscript{Pro} structure in water. This is a serious problem since protein dynamics can hold some important keys to the protein's structure and function. A solution to this problem can be provided by computational studies employing molecular dynamics (MD) simulations in bulk water environment. However, MD simulation scenarios require ergodic sampling of trajectories characterized by complex energy landscapes that possess minima and energy barriers between varying minima which can be challenging for crossing at ambient temperatures over currently available simulation time-scales. Despite this, replica-exchange molecular dynamics (REMD) simulations aim to enhance the conformational sampling via running independent temperature-dependent replicas and periodically exchanging the coordinates of the temperature-dependent replicas (Coskuner and Uversky, 2017; Coskuner and Wise-Scira, 2013; Coskuner-Weber and Uversky, 2019). Even though various MD simulations without special sampling techniques have been conducted on the SARS-CoV-2 main protease, these simulations did not utilize enhanced sampling algorithms (Suaires and Diaz, 2020; Kneller et al., 2020a; Diaz and Suarez, 2021; Komatsu et al., 2020). For instance, the structural flexibility of SARS-CoV-2 main protease was tested by means of the classical MD simulations without special sampling technique applications using f14SB version of AMBER parameters for the protein and the TIP3P model for the water (Suaires and Diaz, 2020). Based on these simulations, the native state configurations of the enzyme and those of its noncovalent complex with a model peptide for mimicking the polypeptide sequence were recognized at the active site configurations; the authors investigated the monomeric and dimeric states and showed that the domain III is not stable in the monomeric state, whereas in the presence of the peptide substrate, the monomeric protease exhibits a stable interdomain arrangement (Suaires and Diaz, 2020). Furthermore, they looked at the catalytic impact of the enzyme dimerization and concluded that the active site flexibility was induced by substrate binding (Suaires and Diaz, 2020). Furthermore, in a different study, the structural plasticity of SARS-CoV-2 3CL\textsuperscript{Pro} active site cavity was investigated at room temperature via X-ray crystallography and MD simulations (Kneller et al., 2020a). In this study, the conformational flexibility of the enzyme active site was detected by the comparisons between the room temperature ligand-free structure and the low temperature ligand-free and inhibitor-bound structures. This analysis indicated that the room-temperature structure of the 3CL\textsuperscript{Pro} ligand-free form may be a more physiologically relevant structure for performing docking studies (Kneller et al., 2020a). In addition, there have been various docking and drug binding studies on the SARS-CoV-2 main protease (Diaz and Suarez, 2021; Komatsu et al., 2020). For instance, Komatsu et al. performed MD simulations on dimeric SARS-CoV-2 main protease to examine the binding dynamics of small ligands (Komatsu et al., 2020). They used seven HIV inhibitors, darunavir, indinavir, lopinavir, nefilnavir, ritonavir, saquinavir, and tipranavir were utilized as potential lead drugs for studying access to the drug binding sites in the structures of M\textsuperscript{Pro} (Komatsu et al., 2020). The active sites were identified based on the probability calculations and contacts. Results showed differences in the shapes of the binding sites and binding poses of the ligands (Komatsu et al., 2020).

Different from the existing studies, we conduct extensive and computationally expensive REMD simulations and link these to several state-of-the-art structural bioinformatics tools for investigating the structural characteristics of the SARS-CoV-2 M\textsuperscript{Pro}. Specifically, we report herein the root mean square fluctuations, secondary structure element abundances per residue, K-means clustering results, chemical shift values as well as possible binding sites for peptides/proteins in general. Knowledge obtained from these structural analyses may help in designing more efficient COVID treatments including novel small molecule drugs.

1. Methods

All-atom REMD simulations of SARS-CoV-2 main protease were conducted in water with temperatures distributed between 280 K and 320 K using 40 replicas, which were exponentially distributed between these temperatures. For the SARS-CoV-2 main protease, we used the CHARMM36 parameters (Akbayrak et al., 2021). We selected the TIP3P parameters to model the solvent (PDB ID: 6LUT) due to the higher resolution in their experiments to isolate the initial M\textsuperscript{Pro} structure in our REMD simulations (Jin et al., 2020a). After solvating the structure in water, we initially conducted equilibration simulations using the NVT ensemble for 20 ns and next using the NPT ensemble for additional 20 ns per replica. Simulations were conducted for a total simulation time of 4.0 μs. Exchanges between replicas are attempted every 5 ps with a time step of 2 fs. Following our recent studies (Akbayrak et al., 2021), we use the Langevin dynamics for maintaining the temperature of each replica with a collision frequency of 2 ps\textsuperscript{-1}. Also, we applied the particle mesh Ewald (PME) method for treating the long-range interactions (Akbayrak et al., 2021). We utilized the SHAKE algorithm for constraining the bonds to hydrogen atoms and counterions were added for neutralizing the system (Akbayrak et al., 2021).

We calculated the structural properties of SARS-CoV-2 M\textsuperscript{Pro} from the conformations - obtained after convergence - from the replica closest to 310 K. We computed the contents of the secondary structure elements with the DSSP program linked to our own script (Akbayrak et al., 2021). Moreover, we calculated the end-to-end distance and radius of gyration of SARS-CoV-2 main protease in water. We applied the k-means clustering method to partition the structural observations into clusters, where each analysis belongs to the cluster with the nearest centroid. Therefore, we partitioned the data space onto Voronoi cells such that k-means clustering minimizes using squared Euclidean distances within cluster variances. To relate the ensemble generated in this study to available experimental data, we calculated the contents per residue, K-means clustering results, chemical shift values as well as possible binding sites for peptides/proteins in general. Knowledge obtained from these structural analyses may help in designing more efficient COVID treatments including novel small molecule drugs.
chemical shifts were compared by the shifts 4.3 program that uses a database of functional density shifts for more than 2000 peptides and empirical formulas to predict the chemical shift values in Mpro (Xu and Case, 2001).

We combined and aligned the above atomic-level analysis with residue-level studies that rely on a number of modern bioinformatics tools. We calculated and compared the REMD-derived root mean square fluctuations (RMSF) for each residue of the homodimer main protease in the vicinity of the inhibitor’s peptide fragment in water with findings from several popular residue-level intrinsic disorder predictors: PONDR® VLXT (Romero et al., 2001), PONDR® VSL2 (Obradovic et al., 2005; Peng et al., 2006), PONDR® FIT (Xue et al., 2010), and IUPred capable of predicting long and short disordered regions (Dosztaszos et al., 2005a, 2005b; Meszaros et al., 2018).

We evaluated predisposition of the SARS-CoV-2 Mpro for interaction with proteins and peptides with two complementary methods: HybridPBRpred (Zhang et al., 2020) and PepBCL (Wang et al., 2022). HybridPBRpred predicts protein-binding residues by combining outputs of the two tools: SCRIBER that targets predictions for structurized proteins (Zhang and Kurgan, 2019) and DisoRDPbind that focuses on the intrinsically disordered proteins (Peng and J’Kurgan, 2015; Peng et al., 2017). PepBCL is one of the most recent tools that predicts peptide binding residues (Wang et al., 2022). We also evaluate the nucleic acid binding potential of the SARS-CoV-2 Mpro with the DRNApred predictor (Yan and Kurgan, 2017). Combining HybridPBRpred, PepBCL and DRNApred allows us to annotate putative interactions with proteins, peptides, DNA and RNA along the SARS-CoV-2 Mpro sequence. Importantly, these methods generate predictions directly from the protein sequence, without the use of the homology modelling, and were shown to produce accurate results even in the absence of similarity to training proteins (Zhang et al., 2020; Wang et al., 2022; Yan and Kurgan, 2017). This means that they do not merely identify binding residues based on a similar protein complex, but are capable of finding new binding regions.

2. Results and discussion

Fig. 1 presents the selected structures for SARS-CoV-2 Mpro in water that were retrieved from our REMD simulations. The figure shows that the protein possesses noticeable structural flexibility, as evidenced by noticeable difference in spatial organization of these selected structures.

Fig. 2 represents the calculated root mean square fluctuations (RMSF) values that represent a range of fluctuations of a dynamical system about an average position for each residue of SARS-CoV-2 Mpro in water measured at several temperatures (280, 290, 300, 310, and 320 K). The figure shows that the degree of this protein per-residue flexibility was minimally affected by changes in temperature within the studied temperature interval. The mean RMSF value for the overall dynamics of Mpro in water at 310 K is 1.52 ± 1.31 Å. The maximum RMSF value of Mpro is 10.86 Å, while the minimum value is 0.55 Å.

Figs. 2 and 3 summarize the residue-level analyses. Fig. 2 shows that the SARS-CoV-2 Mpro is predicted mostly as an ordered protein by the majority of disorder predictors, since only short fragments of the N- and C-terminal regions and a short region in the vicinity of residue 95 have disorder scores exceeding the threshold of 0.5. However, we find several flexible regions in this protein (i.e., regions with disorder scores from 0.2 to 0.25). Fig. 2 shows that the majority of the dynamics features observed in our REMD simulations are correlated with the intrinsic disorder predisposition of the SARS-CoV-2 Mpro. The relation between these two independent results indicates that several peaks in disorder profile serve as envelopes which enclose the local RMSF peaks. Despite, in several regions, the heights of the RMSF peaks noticeably exceed the heights of the corresponding disorder peaks (e.g. residues 60–80 and 210–240). There are also several regions (e.g., residues 25–40 and 150–160), which are classified as ordered, however, still present significant structural variations/fluctuations. This may mean that the structural fluctuations of some regions of SARS-CoV-2 Mpro in water may be related to their intrinsic disorder predispositions, whereas other structural fluctuations of other regions are independent of their intrinsic disorder predisposition. Furthermore, the levels of local intrinsic disorder predisposition do not always scale up with the corresponding levels of structural flexibility.

Next, we looked at the predicted protein/peptide and nucleic acid binding sites in the SARS-CoV-2 Mpro. Results of this analysis are summarized in Fig. 3A, showing that the SARS-CoV-2 Mpro contains over 60 protein/peptide binding residues (residues Ser1, Gly2, Arg4-Gly11, Glu14, Gln19, Thr21, Thr24-Leu27, His41, Thr45, Leu46, Met 49, Leu50, Ty54, Leu67, Gln 69, Tyr118, Asn119, Ser123-Tyr126, Lys137-Cys145, His163-Pro168, His172, Asp187-Gln192, Ala285, Leu286, Gln290, Arg298, Gin299, Ser301, Val303, and Thr304). We extract these data from a comprehensive collection of SARS-CoV-2 Mpro structures in complex with various ligands that we collect from PDB and annotate using the BioLip tool (Yang et al., 2013). These PDB structures include 3aw (Akaji et al., 2011), 3av (Akaji et al., 2011), 5aw0,60, 6bu7,31 7byd (Jin et al., 2020b), 6xa436, 6xbg (Sacco et al., 2020), 6xbh (Sacco et al., 2020), 6xbi (Sacco et al., 2020), 6xch (Kneiller et al., 2020b), 6xfn (Sacco et al., 2020), 7c8b, 7kvg (Noske et al., 2021), 7cut (Wang et al., 2021), 7mg (MacDonald et al., 2021), 7ms (MacDonald et al., 2021), 7fn (Noske et al., 2021), 7mb (Li et al., 2021), 7em (Fu et al., 2021), 7s82, 7rnw (Johansen-Leete et al., 2022), 7dvp (Zhao et al., 2022), 7dxz (Zhao et al., 2022), 7dvw (Zhao et al., 2022), 7dmo,41 and 7d6w (Zhao et al., 2022). Using HybridPBRpred and PepBCL we identify 55 putative protein and peptide binding residues (Ser1, Phe2-Lys12, Gln19, Thr24, Asp48, Arg60, His80, Cys85, Pro122, Tyr126, Gln127, Arg131, Phe140-Cys145, Cys160-Met165, Pro122, Thr169, His172, Asp176, Thr198, Arg217, Arg222-Thr224, Tyr239,
Pro241-Gln244, His246, Pro252, Phe294, Arg298, Ser301, and Thr304-Gln306). Moreover, our analysis with the DRNApred tool suggests that there are no DNA- or RNA-binding residues/regions in this protein. The predictions are in a good agreement with the experimental data, with sensitivity (rate of correct prediction among native binding residues) of 43%, specificity (rate of correct predictions among the non-binding residues) of 88%, precision (rate of correct predictions among predicted binding residues) of 47%, and F1 score (harmonic mean of precision and sensitivity) of 0.45. We note that the current annotations of native binding residues are likely incomplete, which is why we consider the above metrics as relatively good. This is supported by an observation that the native annotations have grown from 45 binding residues (using data from 2011), to 51 residues (data from 2020 and earlier), and finally to 61 residues (using current data); Fig. 3A shows progression of these annotations using the color-coded horizontal lines at the bottom. Fig. 3B annotates these results onto the structure of the dimer complexed with the N3 inhibitor (PDB id: 6lu7), which is the structure used in this analysis. The predictions, which we show in green and
yellow, accurately identify binding pocket for the N3 inhibitor and the interface of the dimer. We also identify a novel putative binding region, shown in yellow in Fig. 3B (Arg131, Asp197, Thr198, Tyr239, Pro241-Gln244, His246, and Phe294).

We expect that this newly discovered putative peptide/protein binding site can be used by SARS-CoV-2 Mpro for interaction with yet to be discovered partners. The importance of the discovery of this putative site is difficult to overestimate, as it provides a novel target for small molecules that can affect functionality of SARS-CoV-2 Mpro site. This hypothesis is in line with the aforementioned notion that the current annotations of native binding residues of the SARS-CoV-2 Mpro are likely incomplete, as evidenced by the steady increase in the number of the experimentally validated binding residues of this protein.

**Panel B.** Visualization of the putative protein/peptide-binding residues predicted by HybridPBRpred and PepBCL (shown in green and yellow) in the tertiary structure of SARS-CoV-2 Mpro dimer complexed with N3 inhibitor (shown in blue). One of the Mpro chains is shown in black while the other is shown using a semi-transparent gray. The putative binding residues in green identify binding pocket for the N3 inhibitor and the interface of the dimer. The putative binding residues in yellow correspond to the new putative binding pocket. The right side of panel B zooms in on the predicted pocket. The images were produced with Mol* (Sehnal et al., 2021) using the 6lu7 PDB structure.

Next, we computed the per residue secondary structure propensities of SARS-CoV-2 Mpro in water with dynamics. **Fig. 4** illustrates the results of this analysis and shows the calculated α-helix, 3_{10}-helix, β-sheet, and turn structure abundances per residue with dynamics. Based on these calculations, we find that seven regions of Mpro adopt an α-helix conformation. These regions are: Ser10-Gly15, Tyr54-Arg60, Thr201-Asn214, Leu227-Tyr237, Gln244-Tyr257, Val261-Asn274, and Pro293-Cys300. On the other hand, two regions (Ser46-Asp48 and Asn63-Asn65) in the N-terminal region of Mpro adopt a 3_{10}-helix formation. Also, 13 regions located in the N-terminal and mid-domain regions of Mpro form β-sheet structures. These regions are: Met17-Cys22, Thr25-

**Fig. 5.** K-means clustering along with $R_g$ and $REE$ values of the SARS-CoV-2 Mpro in water. five k values were utilized and centroids were found to be located at $R_g = 22.54$ Å, $REE = 20.83$ Å (Centroid 1), $R_g = 22.47$ Å, $REE = 30.49$ Å (Centroid 2), $R_g = 22.52$ Å, $REE = 11.29$ Å (Centroid 3), $R_g = 22.54$ Å, $REE = 15.89$ Å (Centroid 4), $R_g = 22.51$ Å, $REE = 5.89$ Å (Centroid 5).

Leu32, Val35-Pro39, Phe66-Ala70, Val73-Leu75, Val77-Gln83, Val86-Val91, Tyr101-Phe103, Gln110-Tyr118, Ser121-Met130, Ser147-Ile152, Cys156-Leu167, and Val171-Thr175. There are 18 regions with turn structure formed in Mpro: Gly23-Thr24, Asp33-Asp34, Arg40-Pro52, Gly71, Asn72, Asn84, Cys85, Asn95-Thr98, Gln107-Gly109, Asn119, Gly120, Pro132-Phe134, Leu141-Gly146, Asp153-Asp155, Pro168-Gly170, Asp176-Gly179, Tyr182-Pro184, Arg217-Leu220, Met276-Thr280, Leu282-Gly283, and Ser301-Val303. Note that there are currently no experimental data (which include nuclear magnetic resonance spectroscopy measurements) on structural dynamics of the SARS-CoV-2 Mpro. Therefore, we cannot compare our findings with the experiments.

We show the k-means clustering algorithm results of the structures of the SARS-CoV-2 Mpro in water in **Fig. 5**. The average $REE$ value for the SARS-CoV-2 Mpro is $16.01 \pm 7.07$ Å, with a maximum value of $35.46$ Å and a minimum value of $3.93$ Å, meaning that, based on this property, SARS-CoV-2 Mpro is extremely flexible. Furthermore, the SARS-CoV-2 Mpro average $R_g$ value is $22.54$ Å ± $0.18$ Å and varies only between 22 and 24 Å, indicating that the
fluctuations in the compactness of M\textsuperscript{pro} in water are of rather low scale. To gain further insights into the performance of REMD simulations, we compared NMR chemical shift values for C\textalpha and H\textbeta atoms of M\textsuperscript{pro} with the experimental data (Cantrelle et al., 2021) (Fig. 6) in water since co-solvents impact the predicted secondary structure and chemical shift values. This analysis revealed that, in general, there is a strong correlation between the predicted and experimental data, indicating that REMD simulations generate relatively realistic picture.

3. Conclusions

We conducted REMD simulations of SARS-CoV-2 M\textsuperscript{pro} in aqueous solution and present the results for body temperature replica. We also analyzed several structural characteristics, including RMSF values with varying temperature, secondary structure propensities, and k-means clustering, C\textalpha and H\textbeta chemical shift values and we performed residue-level analysis of several key structural and functional characteristics that include propensity for intrinsic disorder and for protein-protein and protein-nucleic acids binding.

Our findings show that the residues of the SARS-CoV-2 M\textsuperscript{pro} are flexible based on RMSF values, and these results are supported by R\textbeta values and deviations. We also find that the flexibility of M\textsuperscript{pro} is related to the given temperature. However, the overall degree of structural compactness does not change much and varies only by 2.0 Å in water. We detect seven α-helix, two 3\text{10}-helix, thirteen β-sheet, and eighteen turn structure regions in the structures of the SARS-CoV-2 M\textsuperscript{pro} homodimer in water with dynamics in the presence of the inhibitor’s active peptide fragment. We also detect intra-molecular residue interactions between the mid-domain and N- or C-terminal regions, as well as between the N- and C-terminal regions. The calculated C\textalpha and H\textbeta chemical shift values are in excellent agreement with available experiments. To the best of our knowledge, we present herein the first REMD simulations results for SARS-CoV-2 M\textsuperscript{pro} in water. The reported results can be useful in the long run for developing COVID-19 treatments including small drug molecules.

4. Availability of data

The data that supports the findings of this study are available within this article. Further data that supports the findings are available from the corresponding author upon request.

CRediT authorship contribution statement

Ibrahim Yagiz Akbayrak: Software, Data curation. Sule Irem Caglayan: Data curation, Visualization. Lukasz Kurgan: Data curation, Visualization. Vladimir N. Uversky: Data curation, writing. Orkid Coskuner-Weber: Supervision, Data curation, Visualization, writing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The numerical calculations reported in this paper were performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources).

References

Akbayrak, I.Y., Caglayan, S.I., Durdagi, S., Kurgan, L., Uversky, V.N., Ulver, B., Dosztanyi, Z., Csizmok, V., Tompa, P., Simon, I., 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. Oxf. Engl. 21 (16), 3433–3434. https://doi.org/10.1093/bioinformatics/bti541.

Diaz, N., Suarez, D., 2021. Influence of charge configuration on substrate binding to SARS-CoV-2 main protease. Chem 57, 5314–5317.

Dosztanyi, Z., Csizmok, V., Tompa, P., Simon, I., 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. Oxf. Engl. 21 (16), 3433–3434. https://doi.org/10.1093/bioinformatics/bti541.

Diaz, N., Suarez, D., 2021. Influence of charge configuration on substrate binding to SARS-CoV-2 main protease. Chem 57, 5314–5317.
I.Y. Akbayrak et al. Current Research in Structural Biology 4 (2022) 349–355

Jin, Z., Da, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Lin, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L.W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., Yang, H., 2020. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. Nature. https://doi.org/10.1038/s41586-020-2223-y

Jin, Z., et al., 2020. Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. Nature 582, 289–293.
Johansen-Leete, J., et al., 2022. Antiviral cyclic peptides targeting the main protease of SARS-CoV-2. Chem. Sci. 13, 3826–3836.
Khailany, R.A., Safdar, M., Ozslan, M., 2020. Genomic characterization of a novel SARS-CoV-2. Gene Rep. 10, 100882. https://doi.org/10.1016/j.genrep.2020.100882.
Kneller, D.W., Phillips, G., et al., 2020. Structural plasticity of SARS-CoV-2 3CL M(pro) active site revealed by room temperature X-ray crystallography. Nat. Commun. 11, 3202.
Kneller, D.W., et al., 2020. Malleability of the SARS-CoV-2 3CL M(pro) active-site cavity facilitates binding of clinical antivirals. Structure 28, 1313–1320.
Komatsu, T.S., Okimoto, N., et al., 2020. Drug binding dynamics of the dimeric SARS-CoV-2 main protease, determined by molecular dynamics simulation. Sci. Rep. 10, 16986.
Li, L., et al., 2021. Self-masked aldehyde inhibitors: a novel strategy for inhibiting cysteine proteases. J. Med. Chem. 64, 11267–11287.
MacDonald, E.A., et al., 2021. Recognition of divergent viral subtypes by the SARS-CoV-2 main protease. ACS Infect. Dis. 7, 2591–2595.
Matthews, D.A., Dragovich, P.S., Webber, S.E., Fuhrman, S.A., Patick, A.K., Zalman, L.S., Hendrickson, T.F., Love, R.A., Prím, T.J., Maraković, J.T., Zhou, R., Tikoje, J., Ford, C.E., Meador, J.W., Ferre, R.A., Brown, E.L., Binford, S.L., Brothers, M.A., DeLadé, D.M., Wormald, S.T., 1999. Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11000–11007. https://doi.org/10.1073/pnas.96.20.11000.
Mészáros, B., Erdos, G., Dostáňá, Z., 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 46 (W1), W329–W337. https://doi.org/10.1093/nar/gkx384.
Noske, G.D., et al., 2021. A crystallographic snapshot of SARS-CoV-2 main protease maturation process. J. Mol. Biol. 433, 167118.
Obradovic, Z., Peng, K., Vučetic, S., Radivojac, P., Dunker, A.K., 2005. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 61 (Suppl. 7), 176–182. https://doi.org/10.1002/pro.20735.
Peng, Z., Kurgan, L., 2015. High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder. Nucleic Acids Res. 43, e121.
Peng, K., Radivojac, P., Vučetic, S., Dunker, A.K., Obradovic, Z., 2006. Length-dependent prediction of protein intrinsic disorder. BMC Bioinf. 7, 208. https://doi.org/10.1186/1471-2105-7-208.