Analysis of Ten Microsecond simulation data of SARS-CoV-2 dimeric main protease

Md. Rimon Parves[1,3], Yasir Mohamed Riza[1], Shafi Mahmud[2,3], Rajib Islam[3], Sinthyia Ahmed[3], Bibi Ashiana Evy[1], Md. Hasanuzzaman[4], Mohammad A Halim [3]

[1]Dept of Biochemistry and Biotechnology, University of Science and Technology Chittagong (USTC), Foy's Lake, Khulshi–4202, Chittagong, Bangladesh

[2]Dept of Genetic Engineering & Biotechnology, University of Rajshahi, Rajshahi, Bangladesh

[3]Division of Computer-Aided Drug Design, The Red-Green Research Centre, Dhaka, Bangladesh

[4]Dept of Animal Science and Nutrition, Chittagong Veterinary and Animal Sciences University, Bangladesh

*Corresponding author should be addressed to: Md. Rimon Parves

Email: rimonriju@gmail.com

Faculty of Basic Medical & Pharmaceutical Sciences

University of Science and Technology, Chittagong

BBMH, Foy's Lake, Chittagong-4202

Abstract

The dimeric main protease of SARS-CoV-2, has become a crucial target for inhibiting/modulating its catalytic activity. However, understanding of its conformational change, and atomistic flexibility, is very much lucrative for designing/developing small molecules. Fortunately, huge data has been revealed by a research group, performed about ten-microsecond molecular dynamics to paving the way for understanding the structural complexity of protease. Herein, we have done the basic structural analysis, advanced flexibility and conformational analysis like PCA, for revealing out the regions and residues, which are mostly flexible and likely to be responsible for different conformation of protease protein.

Introduction

The COVID-19 (2019-nCoV) outbreak was declared a pandemic by the World Health Organization (WHO) on March 11th, 2020. Coronaviruses (CoVs) are single-stranded positive-sense RNA viruses that can cause several pulmonary diseases in mammals (Masters, 2006; Weiss & Navas-Martin, 2005). This new virus was originally named SARS-CoV-2 (Gorbalenya et al., 2020), since its RNA genome was similar to the SARS coronavirus (about 82%). Furthermore, both viruses belonged to the clade b of the genus Betacoronavirus (Wu et al., 2020; Zhou et al., 2020). What initially started in Wuhan as a localized incident of the virus leaping across species, has led to an explosive growth in the number of cases due to its astonishing human-to-human mode of transmission, that has led to the virus spreading beyond its borders into the global community. As of April 4, there has been a confirmed number of 1,170,159 cumulative cases globally, of which 63,832 are deaths. It's not all bad news as there has been about 237,978 confirmed cases of patients who have recovered (WHO, n.d.).

The symptoms of COVID-19 bare striking similarities to the pneumonia-like clinical features of those reported from the MERS-CoV and SARS-CoV. Infected patients typically display shortness of breath (dyspnea), dry cough and fever, with a few patients exhibiting upper respiratory tract symptoms such as rhinorrhea and sore throats (Assiri et al., 2013; Lee et al., 2003; Zhou et al., 2020). However, COVID-19 differs from MERS-CoV and SARS-CoV in that, patients rarely showed any signs of diarrhea-like enteric symptoms (Assiri et al., 2013). The main mode of transmission of COVD-19 is by human-to-human contact through airborne droplets (from coughing or sneezing) or through contact with infected individuals (WHO, n.d.), which is why the WHO suggested 'Social Distancing' as the best preventative measure for limiting the spread of the disease and flatting down the curve to reduce the load on the healthcare system and to avoid preventable deaths. The mean incubation period of COVID-19 is estimated to be around 4-7 days; therefore, it is speculated that infected individuals who carry COVID-19 may be able to infect other individuals unbeknownst since they may have little to no symptoms (i.e. they are asymptomatic) (WHO, n.d.). To combat this deadly pandemic, attempts to streamline the development of new vaccines and antiviral therapeutics are underway all over the Globe.

COVID-19 is a 800 kDa polypeptide with a large genome that has a propensity to be cleaved by various proteases (such as chymotrypsin-like proteases and papain-like proteases). These proteases play a critical role in COVID-19's non-structural protein formation which in turn plays a vital role in the virus' replication(Haider et al., 2020). The main protease of coronaviruses, M^{pro} (also known as $3CL^{pro}$), is one of the best characterized drug targets (Anand et al., 2003). This enzyme in particular (along with papain-like proteases) is essential for the processing of polyproteins from the viral RNA (Hilgenfeld, 2014). Inhibition of this enzyme's activity would effectively block viral replication (Zhang et al., 2020).

The M^{pro} operates at 11 cleavage sites on the large polyprotein 1ab (replicase 1ab, ~790 kDa), with the recognition sequence at most sites being Leu-Gln↓(Ser, Ala, Gly) (↓ marks the cleavage site) (Zhang et al., 2020). What is furthermore promising is that the inhibitors developed against these proteases are most likely going to be non-toxic since human proteases do not share any similar cleavage specificity (Zhang et al., 2020). Therefore, these proteases are a promising target site for the development of antiviral drugs for the treatment of COVID-19.

Methodology

The total ten-microsecond simulated trajectories of SARS-CoV-2 dimeric main protease were retrieved from the dataset made available by the research group (KOMATSU, Teruhisa S.; KOYAMA, Yohei M.; OKIMOTO, Noriaki; MORIMOTO, Gentaro; OHNO, Yousuke; TAIJI, 2020). The Gromacs program (Abraham et al., 2015) was used for performing basic structural analysis. Subsequently, the trajectories were converted to dcd format by VMD (Humphrey et al., 1996), before being analyzed by Bio3d program. The PCA (principal component analysis), were analyzed through Bio3d program(Grant et al., 2006), using the method as described before (Dash et al., 2019, 2020).

Conformational analysis

The RMSD profile of the Apo protein was evaluated at the atomic scale to understand the overall stability and performance of the systems. The conformational variation and flexibility reflected by the root mean square deviation value of the protein structure where small fluctuation denotes fewer fluctuations(Hannan et al., 2019; Kumar et al., 2019). From **Figure 1a)** the RMSD of the protein deviate much more at the initial phase as the flexible nature of the protein was higher. However, after a greater degree of fluctuation in the first 2 microseconds, the protein structure gained its rigid nature and did not fluctuate much during the rest of the simulation periods.

However, the rising value of SASA (**Figure 1d)** signs the expanding nature of the protein system. From the beginning stage (0-0.5μs) the SASA value was stable and thereafter a sharp decrease of the SASA value was observed. However, it began to rise again after 1.5μs and declined again. This tendency was seen till 6.5μs and the complex finally was stable at (250 $Å^2$ - 260 $Å^2$) range.

Conversely, the tightness of the packaging system of protein complex assessed through the radius of gyration or Rg value (Dash et al., 2019). Like the SASA profile, Rg exhibited a similar trend at the beginning of simulation (**Figure 1c**). However, after 0.5μs the Rg decreases a little bit and has a similar range of Rg value till 2μs; after then, the protein system exhibits a somewhat similar profile. Interestingly, the Rg value deviates more in 5-6.5μs and began to slightly increasing trend till 10μs.

However, for understanding the regions of protein fluctuated during the simulation, the RMSF of A and B chain was evaluated (**Figure 1b**). Initially, in beta turn region (40-55) higher peak was observed and interestingly, both A and B chains showed a similar trend among these amino acid regions. Likewise, in the active site cavity (Arg188-Gln192) especially in in gamma turn region, there higher RMSF profile was observed. Moreover, another beta-turn and helix region (220-270) was seemed to show higher fluctuations compared to other regions and interestingly, in the helix region A chain had a higher level of RMSF profile than B chain.

Furthermore, the snapshots were taken from the trajectory, especially from the potential variations time frame before perform superimposition (**Figure 1e**). This figure, illustrates, how different variations were created at different times.

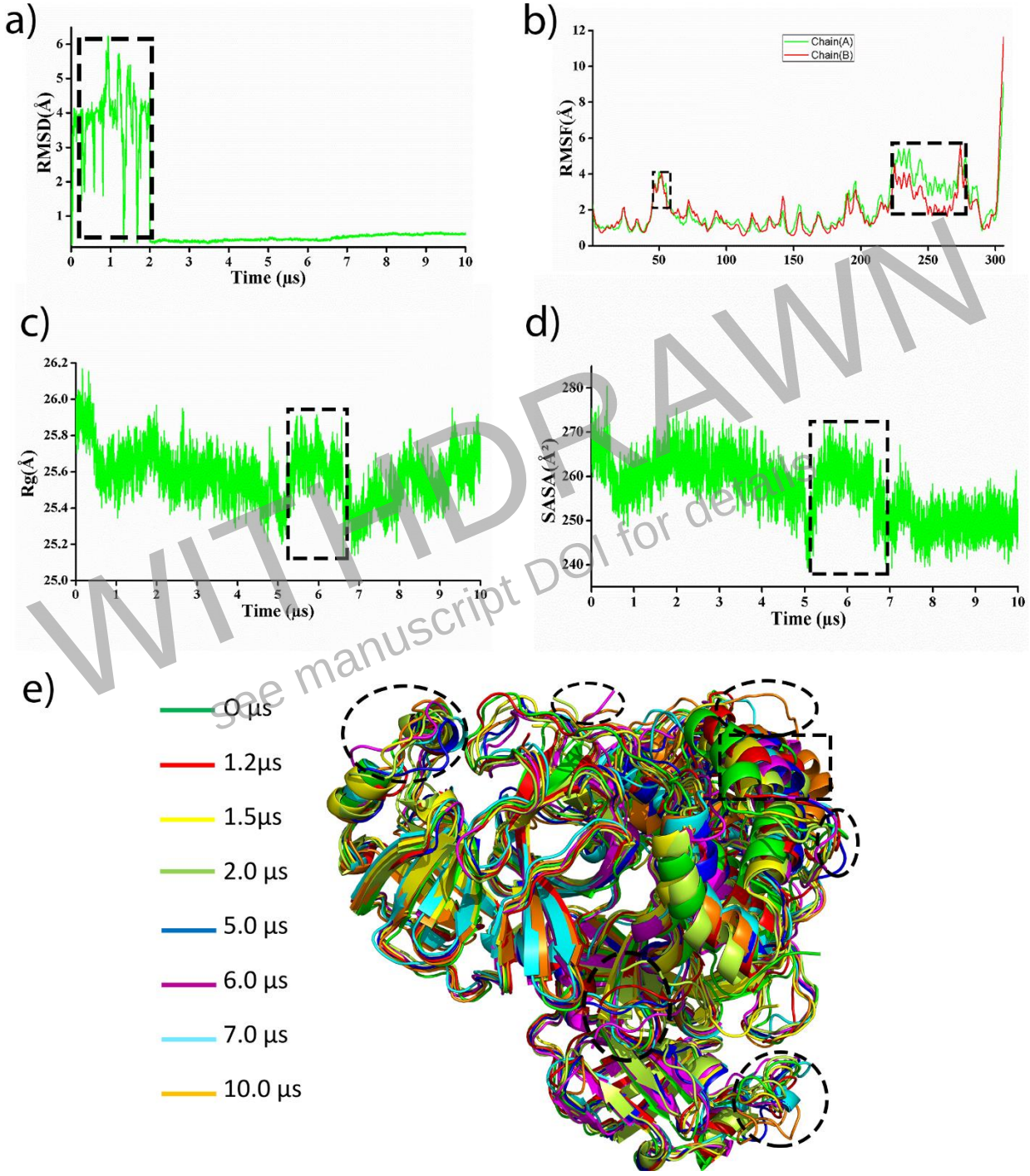

Figure 1. Timeseries analysis of RMSD, RMSF, Rg, and SASA. The marked regions onto the corresponding figure indicate the significant change that occurred during the simulation. a) The marked black box defines the highly fluctuated conformer within the time. b) The marked regions had greater flexibility. c) Marked region defines conformational states having the highest and lowest compactness and rigidity. d) Marked region defines conformational states having higher and lower protein expansion. e) Circular regions indicate the changes occurred in different timesteps as illustrated.

Residual flexibility analysis

Additionally, Principle component analysis (PCA) was performed. The PCA analysis can convert the high-dimensional data of protein dynamics into the low-dimensional space to obtain a series of eigenvectors and eigenvalues that reflect overall motions in the protein (Shukla & Singh, 2019; Yan et al., 2018). The PCA can be applied to any system and permits to study the influence of any varying parameters, by reducing the complexity of the collective motion (García, 1992; Kitao et al., 1991; Kitao & Go, 1999), which is associated with the phase space behavior related to protein functions and stability. Therefore, it is often used to characterize different conformational variances, which are involved in protein folding, open-close mechanism of ion channels, and conformational dynamics (Grottesi et al., 2005; Maisuradze et al., 2009, 2010; Spellmon et al., 2016). The first 34 PCs of the protein accounted for 47.49%, the total variations. The first PC1 pdbs were rendered (**Figure 2**) to getting insights into the regions of protein being fluctuated during the simulation and **Figure 2** depicts conformational variations on timestep. It can be seen from **Figure 2** that, residues ranging from 15-45, 58-85, 116-125, 131-145, 165-178, 183-201 (mostly active site), of B chain of dimeric protease have demonstrated highest flexibility than A chain. But notably higher fluctuations were observed for A chain, within the residues ranging from 47-49, 202-216, 219-273, 274-299. Thus, this compressed analysis supports earlier RMSF calculation.

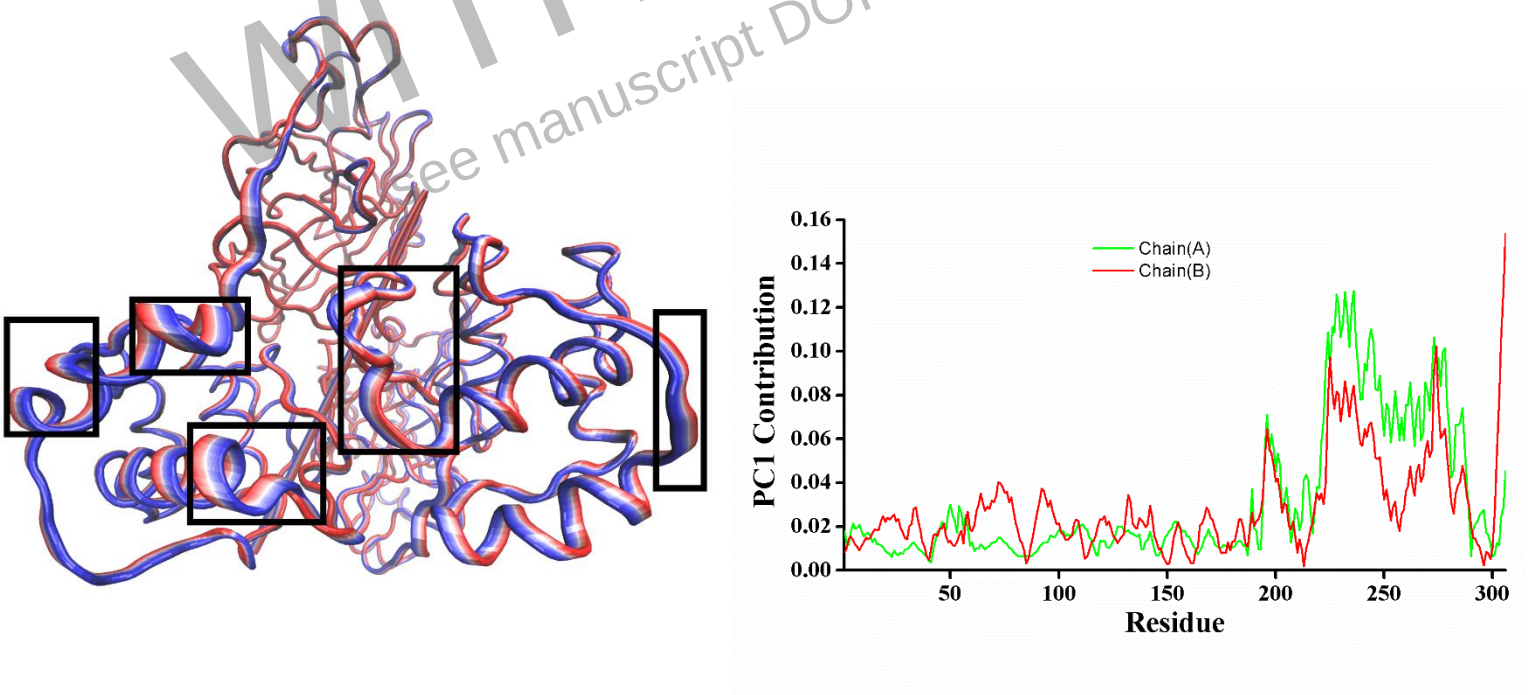

Figure 2. The left panel is the tube view of protease. The rectangular black shaped box marks the most flexible region. The right panel is the residual contribution to PC1.

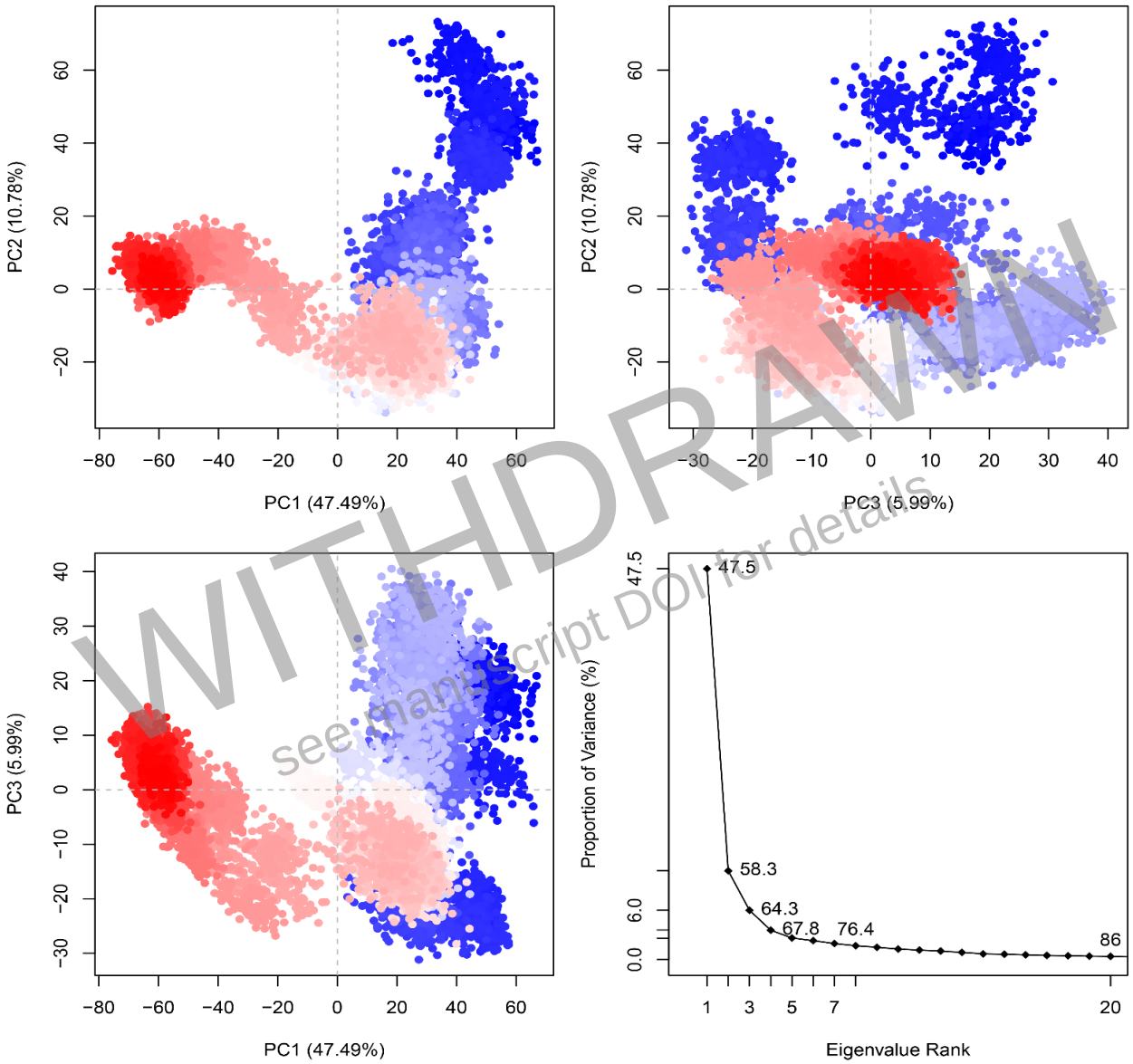

Figure 3. Principle component analysis of Protease. Each dot denotes its conformation of the protein throughout the X and Y axis. The spread of blue and red color dots described the degree of conformational changes in the simulation, where the color spectrum from blue to white to red is equivalent to simulation time. The blue specifies initial timestep, white specifies intermediate, and final timestep is represented by red color.

Salt-Bridge analysis

Since it been told that two residues; GLU290 of A chain and ARG4 of B chain, contribute to the dimerization of protease, we attempted to understand how was their interaction during the whole simulation. We found that several higher distance at 1.90 µs, 2.001-2.004 µs, 3.159 µs, 3.26 µs, 3.396 µs, 3.448 µs, 5.207 µs, and 8.851 µs. Thus, the conformers resembling the major distance might reveal their contribution to dimerization.

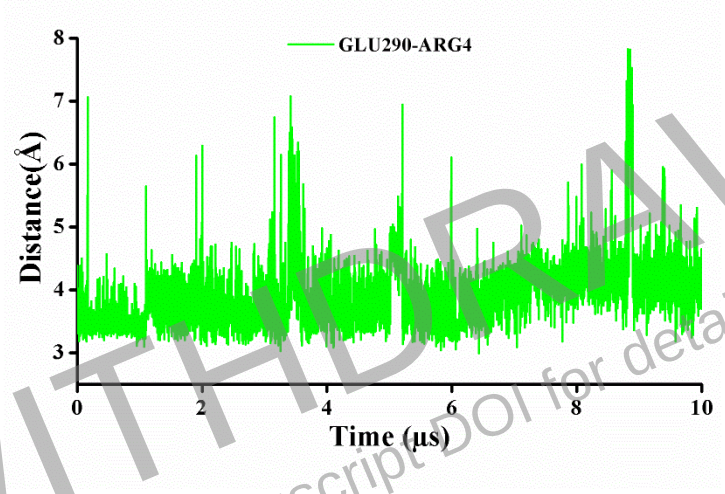

Figure 4. Salt bridge interaction distance between GLU290 of A chain and ARG4 of B chain.

Discussion

Molecular dynamics has become an attractive method for understanding possible or reliable biologically relevant protein conformation before revealing the structural complexity. By understanding atomistic flexibility, a structural biologist and medicinal chemist are continuously fixing their target to come with a better drug candidate or other therapeutic agents. However, with the retrieval of ten microsecond simulation data, we tried to simplify a few basic structural properties of the dimeric main protease of Covid-19. Thus, we have analyzed different properties, such as RMSD, RMSF, SASA, Rg, and PCA (**Figure 1 and Figure 2)**. The RMSD was analyzed to observe the stability of the protein during the simulation time. The RMSD revealed strong stability of the protein after 1.8 µs, and till this time, the convergence of protein conformation was not good. However, The Rg and SASA analysis exposed higher and lower fluctuations mostly by five to seven microseconds, suggesting that protein might have gone folding state, the higher degree of fluctuation in Rg indicates less compactness or rigidity due to protein folding or upon ligand binding, whereas the similar trends in SASA indicate protein expansion. A simple atomic fluctuation might change molecular shape, which may affect SASA. So, RMSF analysis could bring out some important information on residual or atomic flexibility, that might contribute to different conformational states. As shown in **Figure 2**, the residues, which exhibit a higher contribution to the PC1 score, have been plotted, which are likely to change protein conformer. However, the catalytic residues as resided the gap between domains I (residues 8–101), and II (residues 102–184) are not deemed to be most flexible as compared to others. Until now, the common residues that interact with antiviral compounds or drugs are mostly THR24, THR25, THR26, LEU27, HIS41, MET49, TYR54, PHE140, LEU141, ASN142, GLY143, SER144, CYS145, HIS163, HIS164, MET165, GLU166, LEU167, PRO168, ASP187, ARG188 GLN189, THR190, ALA191, and GLN192, among which, HIS41 and CYS145 are two key hydrophobic catalytic

residues(Zhang et al., 2020). However, the main key catalytic residues are found to be less flexible, which was approximately ~1.25Å for HIS41, and ~0.90 Å for CYS145 compared to other residues.

Conclusion

This study concludes with a very primary understanding of SARS-CoV-2 dimeric main protease. Total conformers generated by MD protocol, was very much related to each other. Although few major deviations were marked from RMSD, but those did not remain afterward. However, this study found a significant change in residual flexibility and its contribution to possible different conformers, as explained in RMSF and PCA analysis. The next version of the analysis will reveal the dimerization mechanism, dynamic cross-correlation matrix, and detailed salt bridge analysis.

Acknowledgments

We would like to express our deepest gratitude and thanks to our creator, the simulation group for their dataset, and all those who've helped and supported us along the way.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding statement

The authors received no funding from an external source.

References:

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/https://doi.org/10.1016/j.softx.2015.06.001

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) Structure: Basis for design of anti-SARS drugs. *Science*, *300*(5626), 1763–1767. https://doi.org/10.1126/science.1085658

Assiri, A., Al-Tawfiq, J. A., Al-Rabeeah, A. A., Al-Rabiah, F. A., Al-Hajjar, S., Al-Barrak, A., Flemban, H., Al-Nassir, W. N., Balkhy, H. H., Al-Hakeem, R. F., Makhdoom, H. Q., Zumla, A. I., & Memish, Z. A. (2013). Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: A descriptive study. *The Lancet Infectious Diseases*, *13*(9), 752–761. https://doi.org/10.1016/S1473-3099(13)70204-4

Dash, R., Ali, M., Dash, N., Azad, M., Hosen, S. M., Hannan, A., & Moon, I.-S. (2019). Structural and Dynamic Characterizations Highlight the Deleterious Role of SULT1A1 R213H Polymorphism in Substrate Binding. *International Journal of Molecular Sciences*, *20*, 6256. https://doi.org/10.3390/ijms20246256

Dash, R., Choi, H. J., & Moon, I.-S. (2020). Mechanistic insights into the deleterious roles of Nasu-Hakola disease associated TREM2 variants. *Scientific Reports*, *10*. https://doi.org/10.1038/s41598-020-60561-x

García, A. E. (1992). Large-amplitude nonlinear motions in proteins. *Physical Review Letters*, *68*(17), 2696–2699. https://doi.org/10.1103/PhysRevLett.68.2696

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I., & Ziebuhr, J. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*. https://doi.org/10.1038/s41564-020-0695-z

Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., & Caves, L. S. D. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics (Oxford, England)*, *22*(21), 2695–2696. https://doi.org/10.1093/bioinformatics/btl461

Grottesi, A., Domene, C., Hall, B., & Sansom, M. S. P. (2005). Conformational Dynamics of M2 Helices in KirBac Channels: Helix Flexibility in Relation to Gating via Molecular Dynamics Simulations. *Biochemistry*, *44*(44), 14586–14594. https://doi.org/10.1021/bi0510429

Haider, Z., Muneeb Subhani, M., Ansar Farooq, M., Ishaq, M., Khalid, M., Sohail Ahmad Khan, R., & Khan Niazi, A. (2020). *In Silico discovery of novel inhibitors against main protease (M pro) of SARS-CoV-2 using pharmacophore and molecular docking based virtual screening from ZINC database*. *March*. https://doi.org/10.20944/preprints202002.0431.v2

Hannan, M., Dash, R., Sohag, A. A. M., & Moon, I. S. (2019). Deciphering Molecular Mechanism of the Neuropharmacological Action of Fucosterol through Integrated System Pharmacology and In Silico Analysis. *Marine Drugs*, *17*(11), 639.

Hilgenfeld, R. (2014). From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS Journal*, *281*(18), 4085–4096. https://doi.org/10.1111/febs.12936

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of*

Molecular Graphics, *14*(1), 33–38. https://doi.org/https://doi.org/10.1016/0263-7855(96)00018-5

Kitao, A., & Go, N. (1999). Investigating protein dynamics in collective coordinate space. *Current Opinion in Structural Biology*, *9*(2), 164–169. https://doi.org/https://doi.org/10.1016/S0959-440X(99)80023-2

Kitao, A., Hirata, F., & Gō, N. (1991). The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chemical Physics*, *158*(2), 447–472. https://doi.org/https://doi.org/10.1016/0301-0104(91)87082-7

KOMATSU, Teruhisa S.; KOYAMA, Yohei M.; OKIMOTO, Noriaki; MORIMOTO, Gentaro; OHNO, Yousuke; TAIJI, M. (2020). *COVID-19 related trajectory data of 10 microseconds all atom molecular dynamics simulation of SARS-CoV-2 dimeric main protease. (2020). 1. https://doi.org/10.17632/VPPS4VHRYG.1COVID-19 related trajectory data of 10 microseconds all atom molecular dynamics*. *1*. https://doi.org/10.17632/VPPS4VHRYG.1

Kumar, D. T., Iyer, S., Christy, J. P., Siva, R., Tayubi, I. A., Doss, C. G. P., & Zayed, H. (2019). A comparative computational approach toward pharmacological chaperones (NN-DNJ and ambroxol) on N370S and L444P mutations causing Gaucher's disease. In *Advances in protein chemistry and structural biology* (Vol. 114, pp. 315–339). Elsevier.

Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G. M., Ahuja, A., Yung, M. Y., Leung, C. B., To, K. F., Lui, S. F., Szeto, C. C., Chung, S., & Sung, J. J. Y. (2003). A major outbreak of severe acute respiratory syndrome in Hong Kong. *New England Journal of Medicine*, *348*(20), 1986–1994. https://doi.org/10.1056/NEJMoa030685

Maisuradze, G. G., Liwo, A., & Scheraga, H. A. (2009). Principal component analysis for protein folding dynamics. *Journal of Molecular Biology*, *385*(1), 312–329. https://doi.org/10.1016/j.jmb.2008.10.018

Maisuradze, G. G., Liwo, A., & Scheraga, H. A. (2010). Relation between free energy landscapes of proteins and dynamics. *Journal of Chemical Theory and Computation*, *6*(2), 583–595. https://doi.org/10.1021/ct9005745

Masters, P. S. (2006). The Molecular Biology of Coronaviruses. In *Advances in Virus Research* (Vol. 65, pp. 193–292). https://doi.org/10.1016/S0065-3527(06)66005-3

Shukla, R., & Singh, T. R. (2019). Virtual screening, pharmacokinetics, molecular dynamics and binding free energy analysis for small natural molecules against cyclin-dependent kinase 5 for Alzheimer's disease. *Journal of Biomolecular Structure and Dynamics*, *0*(0), 1–15. https://doi.org/10.1080/07391102.2019.1571947

Spellmon, N., Sun, X., Sirinupong, N., Edwards, B., Li, C., & Yang, Z. (2016). Molecular Dynamics Simulation Reveals Correlated Inter-Lobe Motion in Protein Lysine Methyltransferase SMYD2. *PLOS ONE*, *10*(12), e0145758. https://doi.org/10.1371/journal.pone.0145758

Weiss, S. R., & Navas-Martin, S. (2005). Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus. *Microbiology and Molecular Biology Reviews*, *69*(4), 635–664. https://doi.org/10.1128/mmbr.69.4.635-664.2005

WHO. (n.d.). *Situation reports*.

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China.

Nature, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3

Yan, F., Liu, X., Zhang, S., Su, J., Zhang, Q., & Chen, J. (2018). Electrostatic interaction-mediated conformational changes of adipocyte fatty acid binding protein probed by molecular dynamics simulation. *Journal of Biomolecular Structure and Dynamics*, *0*(0), 000. https://doi.org/10.1080/07391102.2018.1520648

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science (New York, N.Y.)*, *3405*(March), 1–9. https://doi.org/10.1126/science.abb3405

Zhou, P., Yang, X. Lou, Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. Di, Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., … Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7