Molecular Dynamics Simulations Indicate the SARS-CoV-2 Mpro Is Not a Viable Target for Small-Molecule Inhibitors Design

Maria Bzówka[1#], Karolina Mitusińska[1#], Agata Raczyńska[1], Aleksandra Samol[1], Jack Tuszyński[2,3], Artur Góra[1*]

1) Tunneling Group, Biotechnology Center, ul. Krzywoustego 8, Silesian University of Technology, Gliwice, 44-100, Poland
2) Department of Physics, University of Alberta, Edmonton, AB, T6G 2E1, Canada
3) DIMEAS, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Turin, 10129, Italy

[#] These authors contributed equally to this work
* corresponding author: Artur Góra **Email:** a.gora@tunnelinggroup.pl

Abstract

The coronavirus outbreak took place in December 2019 and continues to spread worldwide. In the absence of an effective vaccine, inhibitor repurposing may seem a fruitful attempt. Here, we compared Mpros from SARS-CoV-2 and SARS-CoV. Despite a high level of sequence similarity, the binding sites of analysed proteins show major differences in both shape and size indicating that repurposing SARS drugs for COVID-19 may be futile. Furthermore, the analysis of the pockets' conformational changes during the simulation time indicates their flexibility, which dashes hopes for rapid and reliable drug design. Conversely, structural stability of the SARS-CoV-2 Mpro with respect to mutations of the binding cavity and adjacent flexible loops indicates that the protein's mutability will pose a further challenge to the rational design of small-molecule inhibitors. However, few residues contribute significantly to the protein stability and thus can be considered as key anchoring residues for Mpro inhibitor design.

Introduction

In December 2019, the first atypical pneumonia outbreak associated with the novel coronavirus of zoonotic origin (SARS-CoV-2) appeared in Wuhan, China[1]. In humans, coronaviruses usually cause mild to moderate upper-respiratory tract illnesses, however, the rarer forms of coronaviruses can be lethal. Based on the genomic data analysis of SARS-CoV-2, Zhang et al. indicated that it is closely related to two SARS-CoV sequences isolated from bats in 2015 and 2017[2]. One of the key enzymes in the life cycle of coronaviruses is the main protease (Mpro). Together with non-structural proteins and the spike glycoprotein, it is essential for interactions between the virus and host cell receptors during viral entry[3]. While 12 residues differ between both coronaviruses, only one, namely S46 in SARS-CoV-2 (A46 in SARS-CoV), is located in the proximity of the entrance to the active site (Fig. 1A). Such a small structural change would not be expected to substantially affect the binding of small molecules[4] and would suggest that repurposing SARS drugs against COVID-19 may be fruitful. To this day, several *in silico* attempts have been made, including a massive virtual screening (VS) for SARS-CoV-2 Mpro inhibitors using Deep Docking[5] and only the clinically approved drugs[6–8]. Regrettably, as we show, this strategy is not likely to succeed.

Results and Discussion

The structures of SARS-CoV and SARS-CoV-2 Mpros are highly similar. Both enzymes share the same structural composition; they consist of three domains: an antiparallel β-barrel comprising I and II domains, and the mostly helical III domain, which is required for the enzymatic activity[9]. Both enzymes resemble the structure of cysteine proteases, although their active site comprises

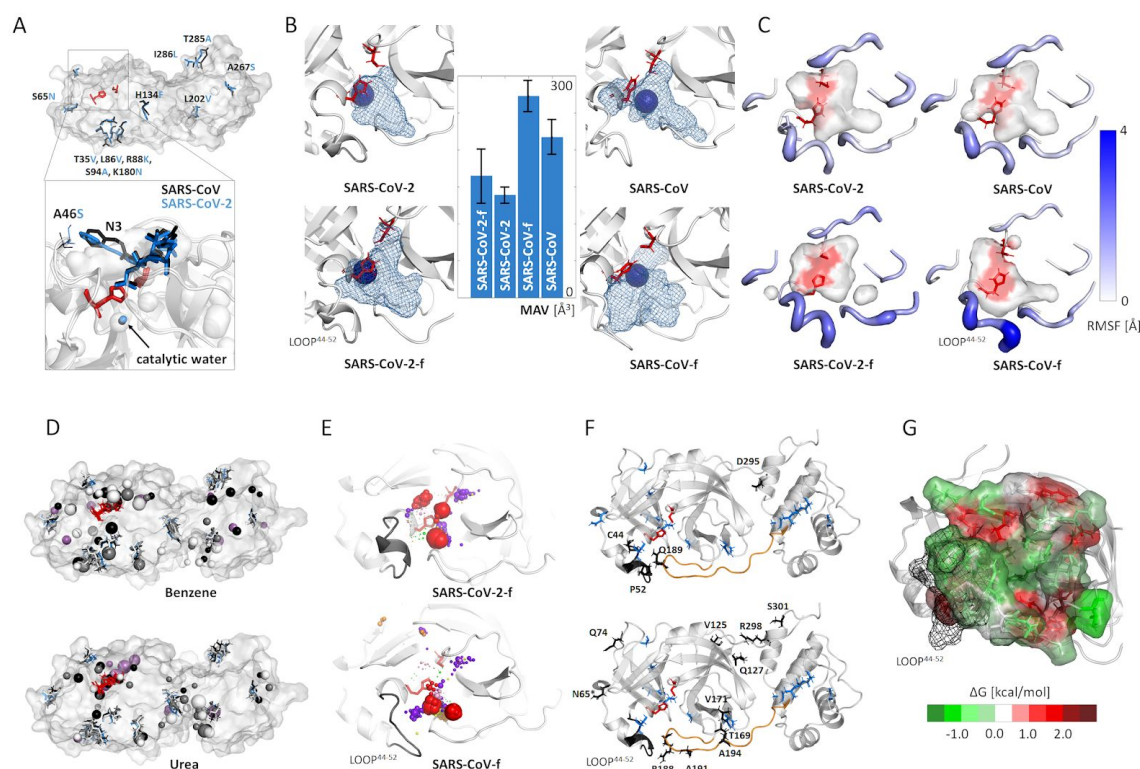a catalytic dyad, namely H41 and C145, and lacks the third catalytic residue replaced by a stable water molecule[10].

Figure 1. The comparison of SARS-CoV and SARS-CoV-2 Mpros. (A) The differing amino acids localisation. (B) The differences between the MAVs of apo structures of Mpros (SARS-CoV and SARS-CoV-2) and with removed N3 inhibitor (SARS-CoV-f and SARS-CoV-2-f). (C) The differences between flexibility of the loops regulating access to the binding cavity. (D) The localisation of global hot-spots: SARS-CoV – black, SARS-CoV-f – grey, SARS-CoV-2-f – white, SARS-CoV-2 – violet for selected cosolvents. (E) The localisation of local hot-spots of all 6 cosolvents (purple – urea, green – dimethylsulfoxide, yellow – methanol, orange – acetonitrile, pink – phenol, red – benzene) in the binding site cavity. (F) The distribution of two major groups of evolutionary correlated residues. (G) The potential mutability of the binding site residues reflected by the differences in Gibbs free energy of protein folding corresponding to the wild-type. On panels A-F the active site residues are shown as red sticks.

A total of 2 µs classical molecular dynamics (cMD) simulations of both SARS-CoV-2 and SARS-CoV Mpros with different starting points were run to examine the plasticity of their binding cavities. The maximal accessible volume (MAV) of the binding cavities during the simulation time and the overall distribution of solvent in the protein's interior were inspected using ligand tracking approach[11]. Surprisingly, despite their high similarity (Fig. 1A), the binding cavities of SARS-CoV and SARS-CoV-2 Mpros show significantly different MAV. Both proteins reduce their MAV upon inhibitor binding approximately 20%, but the maximal volume of SARS-CoV is over 50% larger than those of SARS-CoV-2 (Fig. 1B). It is also worth noting that, despite the differences in size and shape between both Mpros binding cavities, the local solvent distribution approach always detects the highest water density next to the H41 residue, reflecting the position of a catalytic water of Mpro replacing the missing third catalytic site amino acid (Fig. 1B). These findings[10] are supported by the movements of loops that regulate the active sites' accessibility in Mpros. We found that the C44-P52 loop of the SARS-CoV Mpro is more flexible than the corresponding loop of the SARS-CoV-2 Mpro, while the adjacent loops are mildly flexible (Fig. 1C). It is worth adding, that this loop is carrying the unique SARS-CoV-2 Mpro residue S46. These results indicate a

2

serious obstacle for Structure-Based Drug Design which focuses mostly on the similarities between the SARS-CoV and SARS-CoV-2 Mpros and might ignore the fact that the actual available binding space differs significantly. Therefore, repurposing SARS drugs against COVID-19 may not be successful due to major shape and size differences, and despite docking methods, the enhanced sampling should be considered.

We used 6 different cosolvents as molecular probes to investigate the most probable interaction sites in the Mpros' interior and run in total 600 ns of mixed-solvent molecular dynamics (MixMD) simulations. The combination of local solvent distribution and ligand tracking approach provided identification of those regions of the Mpros structures that attract molecular probes (so-called global hot-spots describing the intramolecular voids of the protein, and the local hot-spots describing the binding cavity). The largest number and the densest hot-spots are located within the binding cavity and the region essential for Mpros dimerisation[12], between the II and III domains. In both Mpros the binding cavity is particularly occupied by urea and benzene molecules, which is especially interesting, due to the fact that these solvents exhibit different chemical properties. The general distribution of the global hot-spots from particular cosolvents is quite similar and verifies specific interactions with the particular regions of the analysed proteins. A notable number of hot-spots are located around the amino acids that vary between the SARS-CoV-2 and SARS-CoV Mpros (Fig. 1D). The benzene hot-spots for the SARS-CoV-2 Mpro structure are localised also deep inside the active site cavity, while SARS-CoV Mpro features mostly benzene hot-spots at the cavity entrance (Fig. 1E). This is interesting since, in the absence of cosolvent molecules, the water accessible volume for SARS-CoV-2 Mpro was 50% smaller than in the case of SARS-CoV Mpro, underlining huge plasticity of the binding cavity and suggesting large conformational changes induced by interaction with a potential ligand. It is also interesting, that both global and local hot-spots of the SARS-CoV Mpro structure are located in the proximity of the C44-P52 loop, which potentially regulates the access to the active site.

All above-mentioned differences underline difficulties in antiviral drug discovery arising from evolution-driven virus variability. The future changes in Mpros structures can dash the effort applied for drug discovery. Indeed, the analysis of evolutionary-correlated residues shows that they are dispersed throughout the structure (Fig. 1E). Among them, we identified also those that differ between SARS-CoV-2 and SARS-CoV Mpros, located on the C44-P52 loop (regulating access to the binding site) and the F185-T201 linker loop (contributing to Mpro dimerization[13]). The correlated mutation analysis indicate that Q189 from the linker loop correlates with residues from the C44-P52 loop, whereas R188, A191, and A194 correlate with selected residues from all domains, but not with the C44-P52 loop. To further evaluate the effect of the 12 amino acid replacements on the SARS-CoV-2 Mpro structure stability, we calculated the differences in Gibbs free energies of protein folding corresponding to the wild-type. The replacements were found to stabilise the protein's folding (e.g., H134F: -0.85 kcal/mol) or have almost neutral character (e.g., R88K, S94A, T285A, I286L). Such a result was expected, since for successful virus transformation both stability and the performance of the Mpro have to be sustained. To examine the potential risk of further Mpro structure evolution we simulated the single nucleotide substitutions process to the SARS-CoV-2 main protease gene. If a substitution of a single nucleotide caused translation to a different amino acid than the corresponding residue in the wild-type structure, an appropriate mutation was proposed and the differences in Gibbs free energies of protein folding were calculated. A close look at residues located within 7 Å from the N3 inhibitor suggests that mutations of residues that contribute to ligand binding or access to the active site are energetically favourable, likely to occur (Fig. 1F). Some of the residues that are prone to mutate would provide the inactive enzyme (e.g., the residues forming the catalytic dyad), but others (e.g., amino acids from the C44-P52 loop, T45, S46, E47, L50) could significantly modify the inhibitors binding mode of Mpro.

3

In general, all the above-mentioned findings indicate potential difficulties in the identification of specific inhibitors toward Mpro proteins. The huge plasticity of the binding cavity provides substantial difficulties with drug repositioning and new inhibitors designing. The high potential mutability of the C44-P52 loop indicates that the future evolution of the Mpro can wipe out all our efforts when this protein is considered as a molecular target for COVID-19 treatment due to a probable development of drug resistance. However, our results indicate also residues that are energetically unfavourable to mutate (P39, R40, P52, G143, G146, L167) and they could provide an anchor for successful drug design that can outlast coronavirus Mpros variability in future. Alternatively, we would suggest targeting the region between II and III domains which contributes to the dimer formation.

Materials and Methods
A total of 2 µs of cMD simulations (AMBER18[14]) of SARS-CoV-2 and SARS-CoV Mpros starting from the apo structures and with a N3 inhibitor were performed (PDB IDs: 6lu7, 6y2e, 1q2w and 2amq); the inhibitor was removed from the structures before starting the simulation. The systems preparation protocol and simulation conditions were taken from the previous work[15] (pH set to 7.4). AQUA-DUCT software was employed to inspect the MAV of the binding cavities during the simulation time and the overall distribution of solvent in the protein's interior. The region of interest was defined as a 5Å sphere around the centre of geometry of the active site residues (H41, C145, H164, D187), and molecules of interest were traced within the interior of a convex hull of the protein's atoms. Various cosolvents: acetonitrile, benzene, dimethylsulfoxide, methanol, phenol, and urea were used as molecular probes to investigate the most probable interaction sites of both apo Mpro structures. A total of 600 ns of MixMD simulations (protocol as for the cMD, the heating stage extended to 40 ps) were performed and ligand tracking approach was used to identify regions in which cosolvent molecules are being trapped and or/caged, located within the protein itself (global hot-spots) and inside the binding cavity (local hot-spots).

To verify the potential threat of the further mutability, the correlated mutation analysis based on the multiple sequence alignment of 2643 sequences of viral Mpros was conducted. The differences in Gibbs free energies of protein folding of introduced 12 amino acids replacements into the SARS-CoV Mpro structure corresponding to the wild-type were calculated (FoldX).

Author Contributions
MB and **KM**: Resources, Calculations, Data analysis, Data curation, Writing-Original draft preparation, Writing - Review & Editing, Visualization. **AR**: Calculations, Data analysis, Writing - Review & Editing, Visualization. **AS**: Calculations, Data analysis. **JT**: Funding acquisition, Writing - Review & Editing. **AG**: Conceptualization, Supervision, Data analysis, Visualization, Funding acquisition, Project administration, Writing - Review & Editing.

References
1. Huang C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet (2020). doi:10.1016/S0140-6736(20)30183-5
2. Zhang L., Shen F., Chen F., Lin Z. Origin and evolution of the 2019 novel coronavirus. Clin. Infect. Dis. (2020). doi:10.1093/cid/ciaa112
3. Zumla A., Chan J. F. W., Azhar E. I., Hui D. S. C., Yuen K.-Y. Coronaviruses — drug discovery and therapeutic options. Nat. Rev. Drug Discov. 15, 327–347 (2016).
4. Liu, W. Morse J. S., Lalonde T., Xu S. Learning from the Past: Possible Urgent

Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019 - nCoV. ChemBioChem cbic.202000047 (2020). doi:10.1002/cbic.202000047

5. Ton A.-T., Gentile F., Hsing M., Ban F., Cherkasov A. Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. ChemRxiv (2020).

6. Xu Z. et al. Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. bioRxiv (2020). doi:10.1101/2020.01.27.921627

7. Liu X., Wang X.-J. Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines. bioRxiv (2020). doi:10.1101/2020.01.29.924100

8. Li Y. et al. Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening. bioRxiv (2020).

9. Bacha U., Barrila J., Velazquez-Campoy A., Leavitt S. A., Freire E. Identification of Novel Inhibitors of the SARS Coronavirus Main Protease 3CLpro. Biochemistry 43, 4906–4912 (2004).

10. Anand K. Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. Science 300, 1763–1767 (2003).

11. Magdziarz T. et al. AQUA-DUCT 1.0: structural and functional analysis of macromolecules from an intramolecular voids perspective. Bioinformatics (2019). doi:10.1093/bioinformatics/btz946

12. Li C., et al. Maturation Mechanism of Severe Acute RespiratorySyndrome (SARS) Coronavirus 3C-like Proteinase, J Biol Chem., 285(36):28134-28140 (2010) 10.1074/jbc.M109.095851

13. Tsai M.-Y. et al. Essential covalent linkage between the chymotrypsin-like domain and the extra domain of the SARS-CoV main protease. J. Biochem. 148, 349–358 (2010).

14. Case D. A. et al. AMBER 2018. (2018).

15. Mitusińska K., Magdziarz T., Bzówka, M. Stańczak A., Gora A. Exploring Solanum tuberosum Epoxide Hydrolase Internal Architecture by Water Molecules Tracking. Biomolecules 8, 143 (2018).