

## Homology Models of the Papain-Like Protease PL<sup>pro</sup> from Coronavirus 2019-nCoV

Martin J. Stoermer\*

Division of Chemistry and Structural Biology, Institute for Molecular Bioscience, The University of Queensland, St. Lucia, 4072, Queensland, Australia.

\*Corresponding author

Dr. Martin J. Stoermer

Division of Chemistry and Structural Biology, Institute for Molecular Bioscience  
The University of Queensland

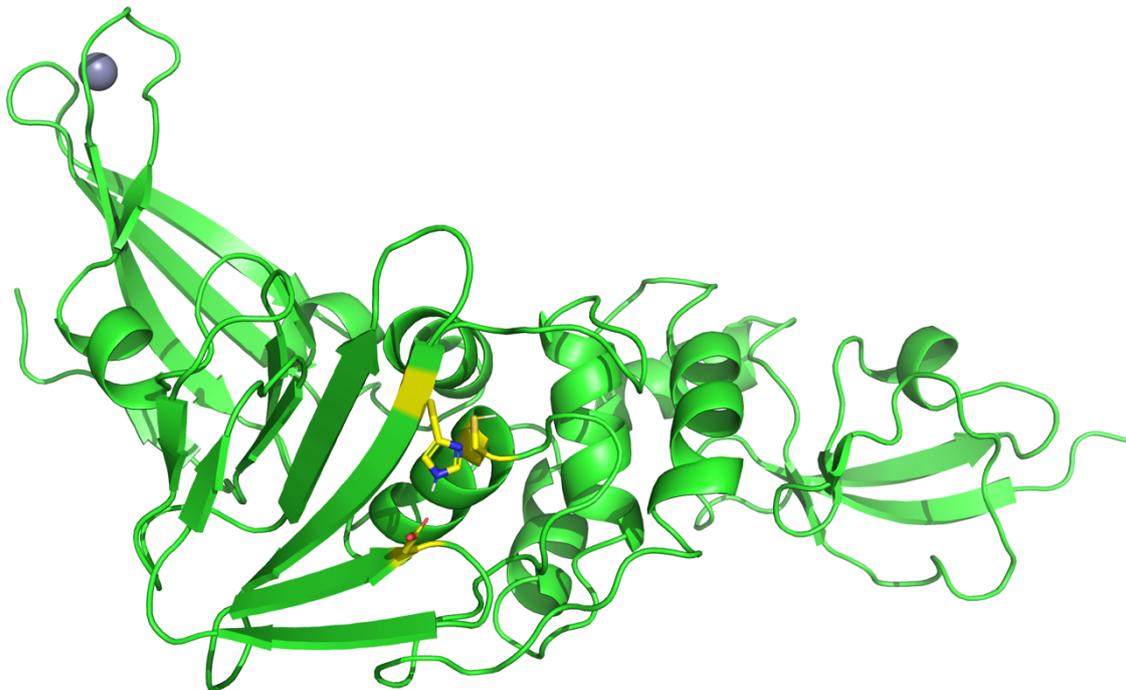
St. Lucia, 4072, Queensland, Australia

E-mail: martin.stoermer@uq.edu.au

Orcid: <https://orcid.org/0000-0003-3445-2104>

Twitter: @MartinStoermer

### Abstract.



The December 2019 outbreak of pneumonia in Wuhan, Hubei Province of China was rapidly linked to a novel coronavirus 2019-nCoV. The rapid spread and severity of the virus has led the World Health Organization to declare it a Public Health Emergency of International Concern. We recently described the first homology models of the main 3CL protease from 2019-nCoV, and now present models of the other viral protease, the papain-like protease or PL<sup>pro</sup>. Whilst the overall viral genome is most closely associated with bat coronaviruses, no bat PL<sup>pro</sup> crystal structures are known. Wuhan 2019-nCoV PL<sup>pro</sup> is most closely related to a bat coronavirus PL<sup>pro</sup> (97% identity), then SARS (80 %) and MERS (29%) and the most promising models presented here are prepared from SARS crystal structure templates.

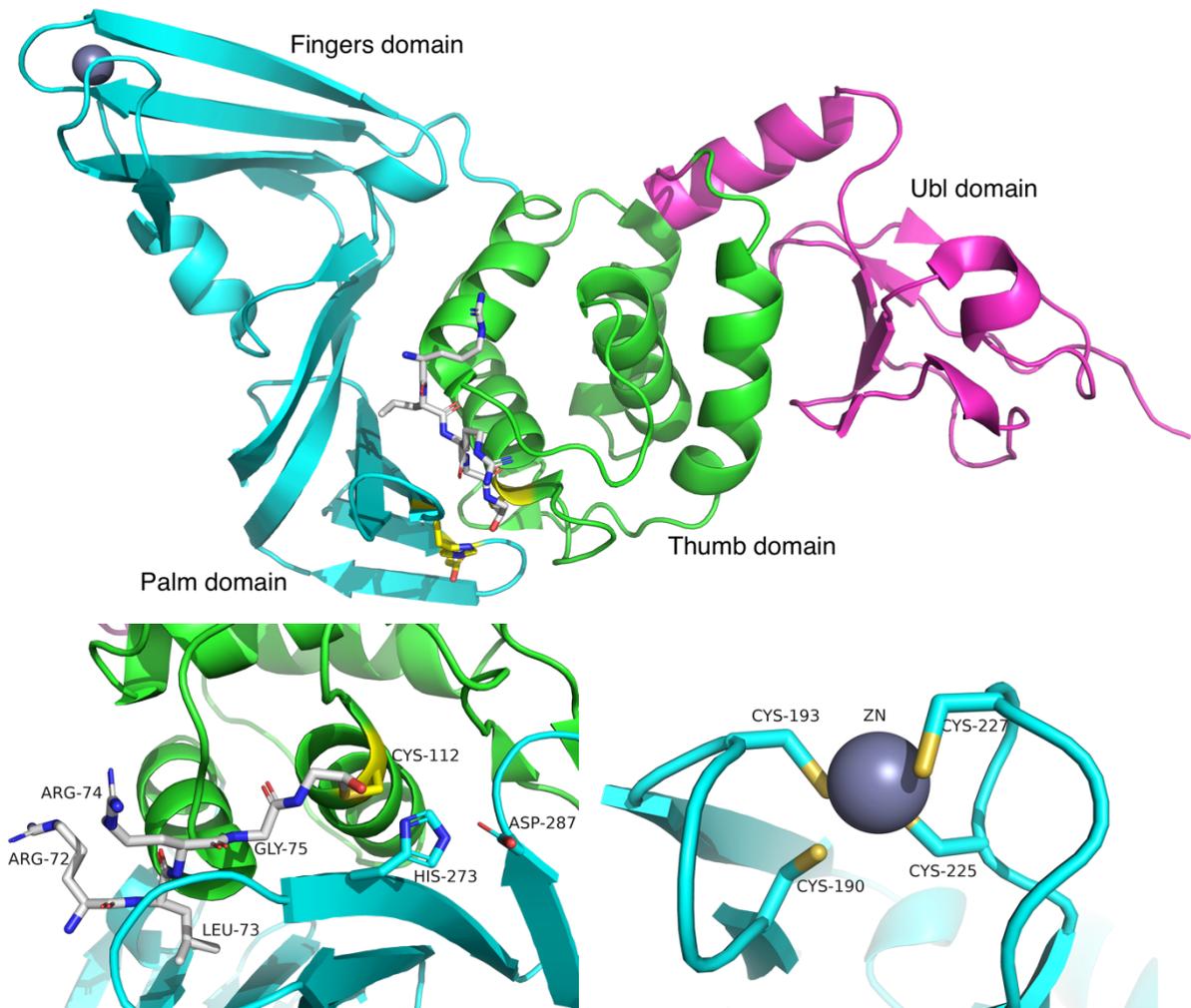
Keywords: Wuhan coronavirus, PL<sup>pro</sup>, protease, homology modelling

## Introduction.

The detection in December 2019 of a cluster of pneumonia cases in Wuhan City, China, with no known cause prompted a rapid response from health authorities.<sup>1,2</sup> Within a month a novel coronavirus 2019-nCoV, also named Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, had been identified as the causative agent and a preliminary genome sequence had been released to the research community,<sup>3</sup> followed by 3 more accurate iterations lodged in Genbank.<sup>4,5</sup> At the time of writing there have been >9000 confirmed cases and the virus has spread to 20 countries (Australia, Cambodia, Canada, China, Finland, France, Germany, India, Italy, Japan, Republic of Korea, Malaysia, Nepal, Philippines, Singapore, Sri Lanka, Thailand, United Arab Emirates, United States of America, Vietnam), culminating in the World Health Organization declaring the outbreak of 2019-nCoV to be a Public Health Emergency of International Concern (PHEIC).<sup>6</sup>

Whilst efforts to understand the biology of the virus have begun with a view to producing an early vaccine, global efforts to contain the outbreak are currently concentrated on rapid development of diagnostics, detection of infected persons, their isolation and treatment. In the next phase, work will begin on finding new specific drugs to treat 2019-nCoV infections, and structural biology of the viral proteins will play a key role. In a previous communication we reported initial homology models for the 3CLpro encoded by the Wuhan novel coronavirus (2019-nCoV 3CLpro).<sup>7</sup> Beta coronaviruses such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) coronavirus generally contain a second, papain-like protease (PL<sup>pro</sup>), which in concert with 3CL<sup>pro</sup> process the two viral polyproteins encoded by two open reading frames orf1a and orf1ab. In addition to their normal processing of viral polyproteins, the PL<sup>pro</sup> from SARS and MERS have also been shown to have de-ubiquitination and de-ISGylation functions. MERS PL<sup>pro</sup> also affects host innate immune response by interfering with the interferon regulatory factor 3 (IRF3) pathway.<sup>8-10</sup> SARS and MERS PL<sup>pro</sup> enzymes have a preference for cleaving substrates after two glycine residues.<sup>11,12</sup>

In contrast to 3CLpro which utilises a Cys/His catalytic dyad, papain-like proteases from MERS and SARS possess a classical Asp/His/Cys catalytic triad. They also possess a C-terminal zinc-finger like domain consisting of two  $\beta$ -hairpins bearing four cysteine residues which coordinate a Zn<sup>2+</sup> ion with tetrahedral geometry (Figure 1). The zinc-binding region has been shown through mutagenesis of the cysteine residues be crucial for both structural integrity and catalytic function.<sup>8</sup> The N-terminal region is a ubiquitin-like (Ubl) domain which is not directly involved in the proteolytic function of PL<sup>pro</sup>. The central “thumb” domain is primarily comprised of  $\alpha$ -helices and contributes the catalytic Cys112 to the active site. The C-terminal region is predominantly  $\beta$ -sheet and contains the fingers domain and also the “palm” region which bears the other two residues of the catalytic triad His275, Asp287.



**Figure 1.** Panel A: SARS PL<sup>pro</sup> covalently bound to ubiquitin aldehyde (C-terminal RLRGG-aldehyde region only shown as grey sticks) N-terminal Ubiquitin-like domain shown in magenta, Fingers domain in cyan ribbons. Panel B: Active site with catalytic triad Cys112, His273, Asp287 shown as yellow sticks. Panel C: Zinc-binding domain showing 4 coordinating cysteine residues Cys190, Cys193, Cys225, Cys227.

In this current work we now report the preparation of homology models of 2019-nCoV PL<sup>pro</sup> based on SARS and MERS X-ray crystallographic templates.

## Methods.

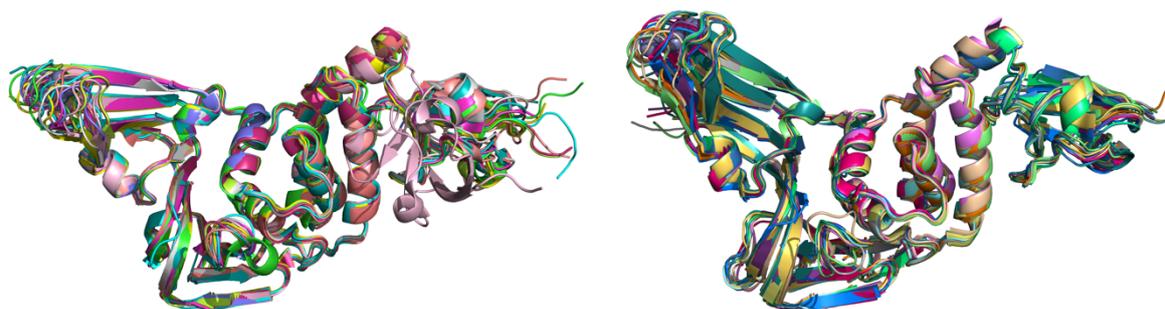
The Wuhan 2019-nCoV genome was obtained from Genbank/NCBI (release MN908947.1).<sup>4</sup> Sequences from crystal structure sequences were obtained from the Protein Data Bank (rcsb.org). Sequence analysis was carried out using the web interface at the European Bioinformatics Institute (EMBL-EBI).<sup>13</sup> Pairwise sequence alignments were performed using the EMBOSS-Needle method with default EBLOSUM62 matrices. Multiple sequence alignments were carried out using Clustal Omega using default HMM profile-profile parameters. Blast searches were carried out against the full UniProt Knowledgebase using blastp 2.9.0+ and default parameters.

Homology models were prepared using the publicly accessible online Swissmodel<sup>14-18</sup> programs, searching the SWISS-MODEL template library (SMTL version 2020-01-02, and PDB release 2019-12-27). Model 6 included an allosteric SARS PL<sup>pro</sup> inhibitor carried through from the PDB:3e9s template. 3 Templates did not contain the structural zinc atom, resulting in models 3,7,10 also having no zinc atoms. These were included in the study however to examine the effect of lack of zinc on model stability in the fingers region in molecular dynamics simulations. Model 9 was produced as a homodimer.

The initial heavy atom-only models were further refined using the Protein Preparation module in Schrödinger Suite 2019-2<sup>19</sup> to add hydrogen atoms and optimise the internal hydrogen bonding network. Finally, the models were energy minimised using the OPLS3e force field with charges from the force field and implicit water solvent, and the Polak-Ribier Conjugate Gradient (PRCG) method to gradient  $<0.05\text{\AA}$ . The final minimised models were then analysed using MolProbity<sup>20</sup> and visualised in Pymol v2.1.<sup>21</sup> Molecular dynamics simulations were performed with the Desmond Molecular Dynamics System (D. E. Shaw Research, New York, NY) by using the tools incorporated in the Schrödinger Suite 2019-2.<sup>19</sup> For full details of molecular dynamics simulation conditions see Supporting Information.

## Results and Discussion

There are a total of around 26 SARS and MERS PL<sup>pro</sup> crystal structures in the PDB<sup>22</sup> (see Supporting Information Table S3 for the full list), but none from bat coronaviruses. An overlay of the available structures reveals a general conservation of fold in the zinc-binding and catalytic domains, but a flexible linker region between the catalytic and fingers domains results in some differences in the overall shape. In general the SARS structures are slightly more elongated than in MERS (Figure 2), and they also display more structural diversity in the ubiquitin-like domain compared to MERS.



**Figure 2.** Overlay of 26 coronavirus PL<sup>pro</sup> proteins in the PDB. *Panel A:* 12 known SARS PL<sup>pro</sup> crystal structures. *Panel B:* 14 MERS structures. Zinc atoms shown as grey spheres.

An initial search of the Protein Data Bank was performed for PL<sup>pro</sup> templates and these were aligned with Clustal Omega<sup>13</sup> against the full length polyprotein provided for the Wuhan 2019-nCoV (Genbank ID:MN908947),<sup>4</sup> and this produced a 326 amino acid region of interest. Although most of the MERS and SARS PDB hits were of slightly shorter (316-322) length the longer sequence was used as it partially matched two longer Avian Infectious Bronchitis Virus structures. The 326aa sequence was trimmed from the polyprotein and

annotated as MN908947\_PLpro (Ser1561-Gly1896 polyprotein numbering). At the time of writing there were 17 2019-nCoV polyprotein sequences in the NCBI database, all 100% identical in the PLpro region.<sup>23</sup> Running a BLAST search on the trimmed WH\_PLpro.fasta against the full UniProt Knowledgebase indicated that the nCoV PLpro was most closely related (97% identity) to a bat SARS-like coronavirus (UniProtKB:A0A2R3SUX5), followed by another bat coronavirus and a cluster of SARS sequences all at 83% identity (see Supporting Information Table S4). Pairwise Emboss-Needle alignments of MN908947\_PLpro against individual Bat coronavirus, MERS and SARS sequences (see Supporting Information Figures S1-3), showed that the nCoV PL<sup>pro</sup> had only 29% sequence identity (48% similarity) to MERS PL<sup>pro</sup>, 80% identity (88% similarity) to SARS PL<sup>pro</sup>, but was 97% identical (99% similarity) to a bat coronavirus (UniProtKB: A0A2R3SUX5). A Clustal Omega multiple sequence alignment of MN908947\_PLpro against bat A0A2R3SUX5 and representative MERS and SARS PLpro sequences is shown in Figure 3. This suggested a catalytic triad for the novel coronavirus consisting of Cys114, His275, and Asp289.



**Figure 3.** Clustal Omega alignment of the extracted PL<sup>pro</sup> sequence of Wuhan coronavirus 2019-nCoV against Bat (A0A2R3SUX5) MERS coronavirus (PDB:4rna) and SARS (PDB:5tl6). Catalytic triads shown in red bold.

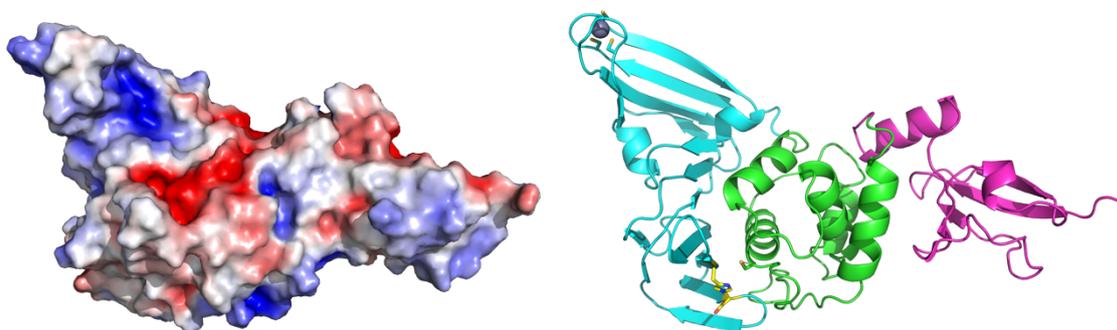
The trimmed MN908947\_PLpro sequence was then used for Swissmodel homology model generation. The default Swissmodel template search algorithm initially only suggested 8 high

quality SARS templates, 3 of which were semi-redundant as they arise from multiple non-identical protease monomers in the unit cell. Four of the five crystal structures had ubiquitin bound. In spite of their lower sequence identity, 2 MERS templates were manually added to the template search and the 10 models were built (Table 1). Model9 was obtained as a homodimer, although this may be as a result a crystallisation artefact in the template, or a curiosity of the Swissmodel template library. Nevertheless all 10 models were taken through the refinement protocol in Schrödinger Suite 2019, and analysed via manual inspection in Pymol, and semi-quantitatively using Molprobity (See Supporting Information Figures S5-S14) Taking both QMQE and QMEAN scoring schema into consideration, Models 1,2,4 &5 were likely of higher quality, narrowed to just model 1 when including Molprobity scores.

| Model | PDB Template  | Sequence identity % | Template Resolution (Å) | Swissmodel QMQE Score | Swissmodel QMEAN Score |
|-------|---------------|---------------------|-------------------------|-----------------------|------------------------|
| 1     | SARS 5tl6.1.A | 82.9                | 2.6Å                    | 0.92                  | -0.22                  |
| 2     | SARS 5tl7.1.B | 82.9                | 2.4Å                    | 0.92                  | 0.08                   |
| 3     | SARS 5e6j.1.A | 82.9                | 2.8Å                    | 0.88                  | -1.51                  |
| 4     | SARS 4m0w.1.A | 82.3                | 1.4Å                    | 0.91                  | -0.03                  |
| 5     | SARS 5tl7.2.B | 82.9                | 2.4Å                    | 0.9                   | -0.39                  |
| 6     | SARS 3e9s.1.A | 82.9                | 2.5Å                    | 0.9                   | -0.29                  |
| 7     | SARS 5e6j.2.A | 82.9                | 2.8Å                    | 0.89                  | -0.78                  |
| 8     | SARS 5tl6.2.A | 82.9                | 2.6Å                    | 0.88                  | -1.16                  |
| 9     | MERS 4rna.1.A | 30.4                | 1.79Å                   | 0.69                  | -1.58                  |
| 10    | MERS 5ko3.1.A | 30.0                | 1.95Å                   | 0.56                  | -2.91                  |

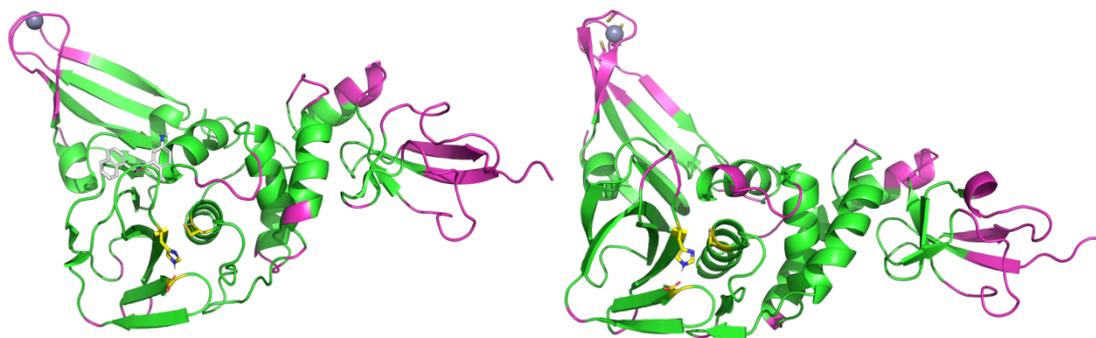
**Table 1.** Swissmodel templates and results statistics.

The best model 1 based on a SARS template PDB:5tl6 shows very high similarity to other coronaviral papain-like proteases, with a catalytic triad composed of Cys114-His275-Asp289 and a conventional zinc-binding domain of 4 cysteine residues Cys192, 194, 227, 229. The other models align very closely to model 1 in their catalytic domains with template-derived diversity in the zinc finger and Ubl domains (See Supporting Information Figures S15a-c).



**Figure 4.** Refined Model 4 of Wuhan 2019-nCoV papain-like protease (PL<sup>pro</sup>). *Panel A:* Electrostatic potential surface prepared in APBS/Pymol. *Panel B:* PL<sup>pro</sup> displayed as ribbons and strands except catalytic triad Cys114, His275, Asp289 as yellow sticks, the zinc-binding domain showing 4 coordinating cysteine residues Cys192, Cys194, Cys227, Cys229 as magenta sticks and zinc as a grey sphere.

The refined models were each also analysed by a 100ns molecular dynamics simulation to identify mobile regions and overall model stability. In all cases the ubiquitin-like and zinc-binding  $\beta$ -hairpins were the most mobile regions. The presence or absence of a bound allosteric inhibitor did not dramatically influence the mobile regions (Figure 5). By contrast, models without an included zinc atom displayed a very high degree of mobility of the four cysteine-containing  $\beta$ -sheets with a downstream effect on the position of the loop bearing catalytic His275, which may be responsible for the observed requirement of zinc for catalytic function of the coronaviral papain-like proteases (see Supporting information Figures S24, S37, S46).<sup>8</sup>



**Figure 5.** Molecular dynamics studies of bound (*Panel A*: model 5) and unbound (*Panel B*: model 1) Wuhan CoV PL<sup>pro</sup> homology models shown as green ribbons, except catalytic triad (Cys114, His275, Asp289 - yellow sticks), and magenta ribbons for regions found to be mobile ( $C\alpha$  RMSD > 1.5Å) in 100 ns molecular dynamics simulations. Zinc-binding cysteine residues (Cys192, 195, 227, 229) shown as magenta sticks, Zn atom - grey sphere.

## Conclusion.

The papain-like protease encoded by the 2019-nCoV Wuhan coronavirus is very highly homologous to bat and SARS coronaviral PL<sup>pro</sup>. We used SARS CoV PL<sup>pro</sup> crystallographic templates to prepare homology models of the 2019-nCoV PL<sup>pro</sup> in both ligand-bound and apo forms. Molecular dynamics simulations indicate most molecular motion is contained in the N-terminal Ubl and C-terminal zinc finger domains, and also support the thesis that zinc is required for structural integrity of the protease.

## Supporting Information

EMBOSS and Clustal Omega sequence alignments, MN908947\_PLpro fasta file, Molecular Dynamics simulations analysis, MolProbity statistics and Ramachandran plots for refined monomer and dimer models. PDB files for raw Swissmodel and individual refined models.

## Acknowledgements

For the rapid provision of 2019-nCoV genomic data the author thanks Professor Yong-Zhen Zhang, and all the researchers listed in Reference 4 at Fudan University, Shanghai, the Shanghai Public Health Clinical Center & School of Public Health, the Central Hospital of Wuhan, Huazhong University of Science and Technology, the Wuhan Center for Disease

Control and Prevention, the National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control, and Professor Edward Holmes, Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, Australia.

### Supporting Information

EMBOSS and Clustal Omega sequence alignments, MolProbity statistics and Ramachandran plots for refined models. PDB files for raw Swissmodel and refined models. Molecular Dynamics simulation analysis of homology models.

### References.

1. <https://www.who.int/westernpacific/emergencies/pneumonia-in-wuhan-china>. Accessed 12<sup>th</sup> January 2020.
2. Hui DS, I Azhar E, Madani TA, et al. 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*. 2020;91: 264-266. (DOI: 10.1016/j.ijid.2020.01.009)
3. Holmes EC on behalf of Zhang Y-Z. <http://virological.org/t/initial-genome-release-of-novel-coronavirus/319>. Accessed 12<sup>th</sup> January 2020.
4. Zhang Y-Z, Wu F, Chen Y-M, Pei Y-Y, Xu L, Wang W, Zhao S, Yu B, Hu Y, Tao Z-W, Song Z-G, Tian J-H, Zhang Y-L, Liu Y, Zheng J-J, Dai F-H, Wang Q-M, She J-L, and Zhu T-Y Nucleotide Accession No. MN908947.1, Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]. Accessed 13<sup>th</sup> January 2020. Available from: <https://www.ncbi.nlm.nih.gov/nuccore/MN908947>
5. Wu F, Zhao S, Yu B, et al. Complete genome characterisation of a novel coronavirus associated with severe human respiratory disease in Wuhan, China. *bioRxiv*. 2020: 2020.2001.2024.919183. (DOI:10.1101/2020.01.24.919183)
6. [https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)). Accessed 31<sup>st</sup> January 2020.
7. Stoermer MJ. Homology Models of Wuhan Coronavirus 3CLpro Protease, *ChemRxiv*, 2020, 11637294. (DOI:10.26434/chemrxiv.11637294.v1)
8. Barretto N, Jukneliene D, Ratia K, Chen Z, Mesecar AD, Baker SC. The Papain-Like Protease of Severe Acute Respiratory Syndrome Coronavirus Has Deubiquitinating Activity. *Journal of Virology*. 2005;79(24): 15189. (DOI:10.1128/JVI.79.24.15189-15198.2005)

9. Mielech AM, Kilianski A, Baez-Santos YM, Mesecar AD, Baker SC. MERS-CoV papain-like protease has deISGylating and deubiquitinating activities. *Virology*. 2014;450-451: 64-70. (DOI:10.1016/j.virol.2013.11.040)
10. Yang X, Chen X, Bian G, et al. Proteolytic processing, deubiquitinase and interferon antagonist activities of Middle East respiratory syndrome coronavirus papain-like protease. *Journal of General Virology*. 2014;95(3): 614-626. (DOI:10.1099/vir.0.059014-0)
11. Han Y-S, Chang G-G, Joo C-G, et al. Papain-Like Protease 2 (PLP2) from Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV): Expression, Purification, Characterization, and Inhibition. *Biochemistry*. 2005;44(30): 10349-10359. (DOI:10.1021/bi0504761)
12. Lei J, Mesters JR, Drosten C, Anemüller S, Ma Q, Hilgenfeld R. Crystal structure of the papain-like protease of MERS coronavirus reveals unusual, potentially druggable active-site features. *Antiviral Research*. 2014;109: 72-82. (DOI:10.1016/j.antiviral.2014.06.011)
13. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*. 2019;47(W1): W636-W641. (DOI:10.1093/nar/gkz268)
14. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 2006;22(2): 195-201. (DOI:10.1093/bioinformatics/bti770)
15. Kopp J, Schwede T. The SWISS-MODEL Repository: new features and functionalities. *Nucleic Acids Research*. 2006;34(suppl\_1): D315-D318. (DOI:10.1093/nar/gkj056)
16. Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 2014;42(Web Server issue): W252-258. (DOI:10.1093/nar/gku340)
17. Bienert S, Waterhouse A, de Beer Tjaart AP, et al. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research*. 2017;45(D1): D313-D319. (DOI:10.1093/nar/gkw1132)
18. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*. 2018;46(W1): W296-W303. (DOI:10.1093/nar/gky427)
19. Schrodinger L. Small-Molecule Drug Discovery Suite 2019-3. 2017-2 ed. New York, NY: Schrodinger, LLC; 2019.

20. Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 2007;35(Web Server issue): W375-383. (DOI:10.1093/nar/gkm216)
21. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
22. Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) <https://www.rcsb.org>. Accessed 18<sup>th</sup> January 2020.
23. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]. <https://www.ncbi.nlm.nih.gov> Accessed 1<sup>st</sup> February 2020. NCBI protein accession IDs: YP\_009724389, QHN73809, QHD43415, QHN73794, QHO62876, QHO60603, QHR84448, QHR63249, QHQ71962, QHQ82463, QHR63259, QHQ71972, QHR63289, QHO62106, QHR63269, QHR63279, QHO62111.