Structural Plasticity of the SARS-CoV-2 3CL M^{pro} Active Site Cavity Revealed by Room Temperature X-ray Crystallography

Daniel W. Kneller
Neutron Scattering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN, 37831, USA

Gwyndalyn Phillips
Neutron Scattering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN, 37831, USA

Hugh M. O'Neill
Neutron Scattering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN, 37831, USA

Robert Jedrzejczak
Center for Structural Genomics of Infectious Diseases, Consortium for Advanced Science and Engineering, University of Chicago, Chicago, Illinois 60667, USA

Lucy Stols
Center for Structural Genomics of Infectious Diseases, Consortium for Advanced Science and Engineering, University of Chicago, Chicago, Illinois 60667, USA

Paul Langan
Neutron Scattering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN, 37831, USA

Andrzej Joachimiak
Center for Structural Genomics of Infectious Diseases, Consortium for Advanced Science and Engineering, University of Chicago, Chicago, Illinois 60667, USA

Leighton Coates (coatesl@ornl.gov)
Neutron Scattering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN, 37831, USA

Andrey Kovalevsky (kovalevskyay@ornl.gov)
Neutron Scattering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN, 37831, USA

Research Article

Keywords: X-ray Crystallography, SARS-CoV-2, 3CL Main Protease, 3CL Mpro

Posted Date: May 13th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-28669/v1

Version of Record: A version of this preprint was published at Nature Communications on June 24th, 2020. See the published version at https://doi.org/10.1038/s41467-020-16954-7.

Abstract

The COVID-19 disease caused by the SARS-CoV-2 Coronavirus has become a pandemic health crisis. An attractive target for antiviral inhibitors is the main protease 3CL M^{pro} due to its essential role in processing the polyproteins translated from viral RNA. Here we report the room temperature X-ray structure of unliganded SARS-CoV-2 3CL M^{pro}, revealing the resting structure of the active site and the conformation of the catalytic site cavity. Comparison with previously reported low-temperature ligand-free and inhibitor-bound structures suggest that the room temperature structure may provide more relevant information at physiological temperatures for aiding in molecular docking studies.

Main Text

A new coronavirus named Severe Accute Respitory Syndrom Coronavirus 2, or SARS-CoV–2, caused an outbreak of pulmonary disease in the city of Wuhan in China at the end of 2019 which has since spread into a world pandemic disease called COVID–19[1–4]. A significant research push is now underway to repurpose existing drugs and to design new therapeutic agents targeting various components of the virus.[5] The viral single-stranded RNA genome is 82% identical to the earlier SARS coronavirus (SARS-CoV) with some viral proteins being more than 90% homologous to SARS-CoV.[6] SARS-CoV–2, similar to many other single-stranded RNA viruses, employs a chymotrypsin-like protease (3CL main protease, or 3CL M^{pro}) to enable the production of non-structural proteins essential for viral replication.[7–9]

3CL M^{pro} cleaves two large overlapping polyproteins pp1a and pp1ab at at least 11 conserved sites, including its own N- and C-terminal autoprocessing sites. The enzyme has a recognition sequence of Leu-GlnSer-Ala-Gly, where marks the cleavage site, but shows sequence promiscuity. The absolute dependence of the virus on the correct function of this protease, together with the absence of a homologous human protease, makes 3CL M^{pro} an attractive, albeit difficult, target for the design of specific protease inhibitors.[10] Unfortunately, to date, no protease inhibitors targeting SARS-CoV 3CL Mpro have been FDA-approved, despite significant research effort during the past fifteen years.[11–17]

The 3CL M^{pro} structure is composed of three domains.[18,19] Domains I (residues 8–101) and II (residues 102–184) are composed of antiparallel β-barrel structures and are the catalytic domains. Domain III (residues 201–303) is composed of five α-helices and is responsible for the enzyme dimerization. This helical domain plays an essential role in the protease function as the monomeric enzyme is not catalytically active. Thus, 3CL M^{pro} forms a functional dimer through intermolecular interactions, mainly between the helical domains (Figure1a).

3CL M^{pro} is uniquely diversified to have an unconventional Cys catalytic residue. Unlike other chymotrypsin-like enzymes and many Ser (or Cys) hydrolases, it has a catalytic Cys-His dyad instead of a canonical Ser(Cys)-His-Asp(Glu) triad.[8] The catalytic residues Cys145 and His41 in 3CL M^{pro} are buried in

an active site cavity located on the surface of the protein. This cavity can accommodate four substrate residues in positions P1' through P4, and it is flanked by residues from both domains I and II (Figure 1b).

We present here new atomic details pertinent to the function and inhibitor binding to SARS-CoV–2 3CL Mpro. To gain these insights we determined a room temperature (293K) X-ray structure of the enzyme to 2.30 Å resolution by growing large crystals (Figure S1) that could be used on a home source to ensure minimal radiation damage. In our structure of ligand-free 3CL M^{pro}, the catalytic Cys145 S is 3.8 Å from His41 N2, which appears to be too long for the formation of a hydrogen bond (Figure 2). This is not surprising, taking into account the experimental pK_a values of 8.0 ± 0.3 for Cys145 and 6.3 ± 0.1 for His41 measured previously for the SARS 3CL M^{pro} that shares 96% homology with the SARS-CoV–2 enzyme[20,21] and the poor hydrogen bonding properties of thiols. Thus, in our crystallization conditions (see Mehtods) at the pH in the crystallization drop of 7.0, both catalytic residues are expected to be uncharged adopting the enzyme's resting state.

In this resting state, the thiol of Cys145 is protonated and the imidazole of His41 is neutral, and the catalytic dyad would be activated by a proton transfer from Cys145 to His41 possibly triggered by substrate binding or occurring in a transition state during the attack by the sulfur on the carbonyl carbon atom of the scissile peptide bond. Conversely, His41 makes a strong hydrogen bond with a water molecule (speculatively named H_2O_{cat}), which in turn is stabilized through hydrogen bonds of 2.9 and 3.0 Å with the side chains of Asp187 and His164, respectively. The position of Asp187 is further stabilized through a salt-bridge with the nearby residue Arg40.

H_2O_{cat} is involved in a complex network of interactions, mediating polar contacts between the catalytic His41, a conserved His164, and a conserved Asp187 located in the domain II-III junction. It is not unreasonable to suggest that this water may play a role of the third catalytic residue, completing the non-canonical catalytic triad in 3CL M^{pro} and acting to stabilize the positive charge on His41 by mediating its electrostatic interaction with the negatively charged Asp187 during catalysis. We note that in some X-ray structures of the ligand-free 3CL M^{pro} from SARS-CoV–2 (e.g., PDB ID 6M03) obtained at 100K, this potentially crucial water molecule is absent.

Unsurprisingly, a significant number of reports have now appeared in which 100K X-ray structures of the ligand-free 3CL M^{pro} have been used for molecular docking simulations of various small molecules, including many of the therapeutics approved to treat other diseases. We superimposed our room temperature structure of 3CL M^{pro} with one obtained at 100K (PDB ID 6Y2E).[18] While the overall structures are similar with an R. M. S. D. for C_α atoms of 0.32 Å (Figure 3a). The conformation of residues 192–198 differs between the room temperature and 100K structures (Figure 3b). The peptide bond of Ala194 is flipped in the room temperature structure pointing inwards into the P5 inhibitor binding pocket where it adopts a conformation similar to that seen in 3CL M^{pro} in complex with the with inhibitor N3 (PDB ID 6LU7).[19] Residues Thr196 and Asp197 also differ significantly in their conformations between the room temperature and 100K structures. The backbone carbonyl oxygen atom of Thr 196 differs in

position by 1.3 Å, the CG atoms of Asp197 are separated by 1.9 Å, and the position of backbone carbonyl oxygen atoms of Asp 197 differs in position by 2.6 Å. The conformations observed in the ligand-free enzyme at room temperature may be more relevant for screening of possible drug candidates.

It is also instructive to compare our room temperature structure of the protease with the structure of an inhibitor-bound complex. For this comparison, we chose the complex with a structurally long peptidomimetic inhibitor N3[19] because it has substituents spanning all substrate binding subsites, including substituents at positions P4 and P5, thus closely resembling an actual substrate. Figure 4 shows the superposition of the two structures. The structural comparison reveals significant structural plasticity of the enzyme in the vicinity of the active site. To accommodate the inhibitor several secondary-structure elements move by more than 1 Å away from their positions in the room temperature structure of the ligand-free form. Such conformational changes can be characterized as induced fit due to ligand binding.

On ligand binding, the small helix near P2 group containing residues 46–50 and the β-hairpin loop near P3-P4 substituents with residues 166–170 shift apart by 2.4 Å, whereas the P5 loop spanning residues 190–194 moves closer to the P3-P4 loop. Two methionines, Met49 and Met165, avoid clashing with the inhibitor's leucine at position P2 by altering their side-chain conformations in the structure of the complex. Further, the change in Met49 conformation cascades to changes in the side chain positions of Ser46 and Leu50. More dramatic conformational changes due to inhibitor binding occur at the enzyme's C-termini. Unexpectedly, the C-terminal tail consisting of residues Ser301 through Gln306 swings 180 from its position in the room temperature ligand-free structure and is situated above the helical domain in the N3 inhibitor-bound form (Figure S2).

The drastic flip in the C-terminal loop conformation eliminates several hydrogen bonds made as part of the dimer interface in the ligand-free form, which may destabilize the dimer in the inhibitor-bound form to a certain degree. To assess the flexibility of these enzyme regions we performed a 1 s molecular dynamics (MD) simulation of the ligand-free 3Cl Mpro. As shown in Figure S3, in our MD simulation the same regions, including the P2 helix (residues 45–50), the P5 loop (residues 190–194) and the C-terminal tail are the most dynamic, showing the largest root-mean square fluctuations (RMSF) (Figure S3). Therefore, these structural regions are quite malleable, possibly able to accommodate various chemical groups at the P2-P5 sites of inhibitors.

The conformational flexibility of the enzyme active site detected by comparisons between the room temperature ligand-free structure reported here with the low-temperature ligand-free and inhibitor-bound structures previously reported leads us to suggest that room-temperature structures of the 3CL M^{pro} ligand-free form may be the more physiologically relevant structure for performing molecular docking studies to estimate drug binding and enable drug design.

Declarations

Acknowledgments

This work was supported by the Laboratory Directed Research and Development Program at Oak Ridge National Laboratory (ORNL). Research at ORNL's Spallation Neutron Source and HFIR was sponsored by the Scientific User Facilities Division, Office of Basic Energy Sciences, U.S. Department of Energy. The Office of Biological and Environmental Research supported research at ORNL's Center for Structural Molecular Biology (CSMB), using facilities supported by the Scientific User Facilities Division, Office of Basic Energy Sciences, U.S. Department of Energy. Funding for this project was provided in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract HHSN272201700060C. We thank Swati Pant, Kevin Weiss, Yichong Fan, Qiu Zhang of ORNL for their help with the plasmid preparation and initial protein expression.

Contributions

L.C, A.K., H.O'N and P.L. designed the study. D.W.K. and G.P. expressed and purified the protein, D.W.K, and A.K. crystallized the protein. A.K. and D.W.K. collected the data. L.C. reduced the data and refined the structure. D.W.K., L.C., P.L. and A.K. wrote the paper with help from all co-authors.

Data Availability

The structure and corresponding structure factors have been deposited into the protein data bank with the PDB accession code 6WQF. Supporting information is available online.

Competing interests

There is NO Competing Interest.

References

1. Zhou, P.; Yang, X. L.; 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. D.; Liu, M. Q.; Chen, Y.; Shen, X. R.; Wang, X.; Zheng, X. S.; Zhao, K.; Chen, Q. J.; Deng, F.; Liu, L. L.; Yan, B.; Zhan, F. X.; Wang, Y. Y.; Xiao, G. F.; Shi, Z. L., A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020,** *579* (7798), 270-273.
2. 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., A new coronavirus associated with human respiratory disease in China. *Nature* **2020,** *579* (7798), 265-269.
3. 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., Author Correction: A new coronavirus associated with human respiratory disease in China. *Nature* **2020,** *580* (7803), E7.

4. Coronaviridae Study Group of the International Committee on Taxonomy of, V., The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* **2020,** *5* (4), 536-544.
5. Liu, C., Zhou, Q., Li, Yingzhu, Garner, L. V., Watkins, S. P., Carter, L. J., Smoot, J., Gregg, A. C., Daniels, A. D., Jervey, S., Albaiu, D. Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. ACS Central Sci. 2020, 6, 315-331.
6. Xu, J., Zhao, S., Teng, T., Abdalla, A. E., Zhu, W., Xie, L., Wang, Y., Guo, X. Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. Viruses 2020, 12, 244.
7. Hilgenfeld, R., From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J* **2014,** *281* (18), 4085-96.
8. Gorbalenya, A. E., Snijder, E. J. Viral cysteine proteases. Perspect. Drug Discov. Des. 1996, 6, 64-86.
9. Muramatsu, T., Takemoto, C., Kim, Y.-T., Wang, H., Nishli, W., Terada, T., Shirouzu, M., Yokoyama, S. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. Proc. Natl. Acad. Sci. U.S.A. 2016, 113, 12997-13002.
10. Dai, W., Zhang, B., Jiang, X.-M., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J., Liu, F., Li, C., Li, Y., Bai, F., Wang, H., Cheng, X., Cen, X., Hu, S., Yang, X., Wang, J., Liu, X., Xiao, G., Jiang, H., Rao, Z., Zhang, L.-K., Xu, Y., Yang, H., Liu, H. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science 2020 (10.1126/science.abb4489).
11. Ghosh, A. K., Xi, K., Ratia, K., Santarsiero, B. D., Fu, W., Harcourt, B. H., Rota, P. A., Baker, S. C., Johnson, M. E., Mesecar, A. D. Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors. J. Med. Chem. 2005, 48, 6767-6771.
12. Martina, E., Stiefl, N., Degel, B., Schulz, F., Breuning, A., Schiller, M., Vicik, R., Baumann, K., Ziebuhr, J., Schirmeister, T. Screening of electrophilic compounds yields an aziridinyl peptide as new active-site directed SARS-CoV main protease inhibitor. Bioorg. Med. Chem. Lett. 2005, 15, 5365-5369.
13. Thanigaimalai, P., Konno, S., Yamamoto, T., Koiwai, Y., Taguchi, A., Takayama, K., Yakushiji, F., Akaji, K., Chen, S.-E., Naser-Tavakolian, A., Schon, A., Freire, E., Hayashi, Y. Development of potent dipeptide-type SARS-CoV 3CL protease inhibitors with novel P3 scaffolds: Desing, synthesis, biological evaluation, and docking studies. Eur. J. Med. Chem. 2013, 68, 372-384.
14. Jacobs, J., Grum-Tokars, V., Zhou, Y., Turlington, M., Saldanha, S. A., Chase, P., Eggler, A., Dawson, E. S., Baez-Santos, Y. M., Tomar, S., Mielech, A. M., Baker, S. C., Lindsley, C. W., Hodder, P., Mesecar, A., Stauffer, S. R. Discovery, synthesis, and structure-based optimization of a series of N-(tert-butyl)-2-(N-arylamido)-2-(pyridin-3yl) acetamides (ML188) as potent noncovalent small molecule inhibitors of the severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL protease. J. Med. Chem. 56, 534-546.

15. Wang, L., Bao, B.-B., Song, G.-Q., Chen, C., Zhang, X.-M., Lu, W., Wang, Z., Cai, Y., Li, S., Fu, S., Song, F.-H., Yang, H., Wang, J.-G. Discovery of unsymmetrical aromtic disulfides as novel inhibitors of SARS-CoV main protease: Chemical synthesis, biological evaluation, molecular docking and 3D-QSAR study. Eur. J. Med. Chem. 2017, 137, 450-461.
16. Jo, S., Kim, S., Shin, D. H., Kim, M.-S. Inhibition of SARS-CoV 3CL protease by flavonoids. J. Enz. Inhibit. Med. Chem. 2020, 35, 145-151.
17. Pillayar, T., Manickam, M., Namasivayam, V., Hayashi, Y., Jung, S.-H. An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. J. Med. Chem. 2016, 59, 6595-6628.
18. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R., Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science* **2020** (10.1126/science.abb3405).
19. Jin, Z.; Du, 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.; Liu, 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., Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* **2020** (10.1038/s41586-020-2223-y).
20. Huang, C.; Wei, P.; Fan, K.; Liu, Y.; Lai, L., 3C-like Proteinase from SARS Coronavirus Catalyzes Substrate Hydrolysis by a General Base Mechanism. *Biochemistry* **2004,** *43* (15), 4568-4574.
21. Solowiej, J., Thomson, J. A., Ryan, K., Luo, C., He, M., Lou, J., Murray, B. W. Steady-state and pre-steady-state kinetic evaluation of severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro cysteine protease: development of an ion-pair model for catalysis. Biochemistry 2008, 47, 2617-2630.

Figures

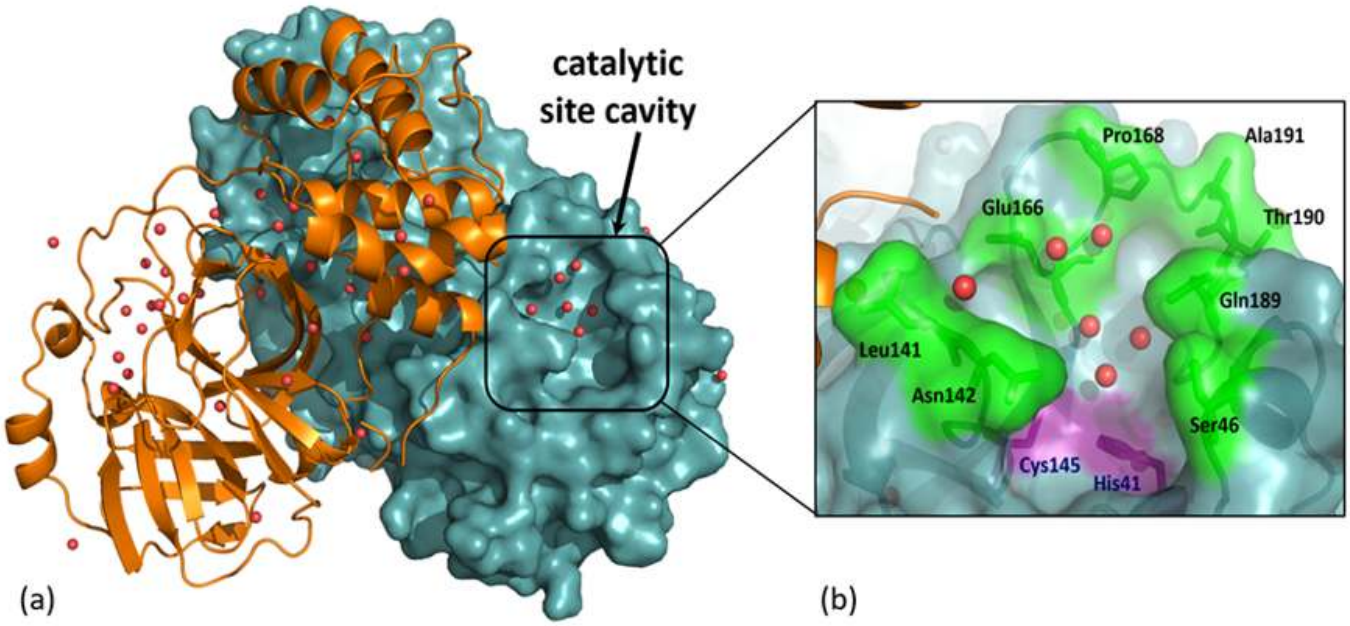

Figure 1

The three-dimensional structure of 3CL Mpro from SARS-CoV-2. (a) One monomer of the dimer is shown as an orange cartoon, while the other monomer is shown as a teal surface with the catalytic site cavity highlighted with water molecules shown as red spheres. (b) A closeup view of the catalytic site cavity in which the catalytic residues (Cys145 and His41) are highlighted in purple with the residues that flank the cavity highlighted in green with water molecules shown as red spheres.

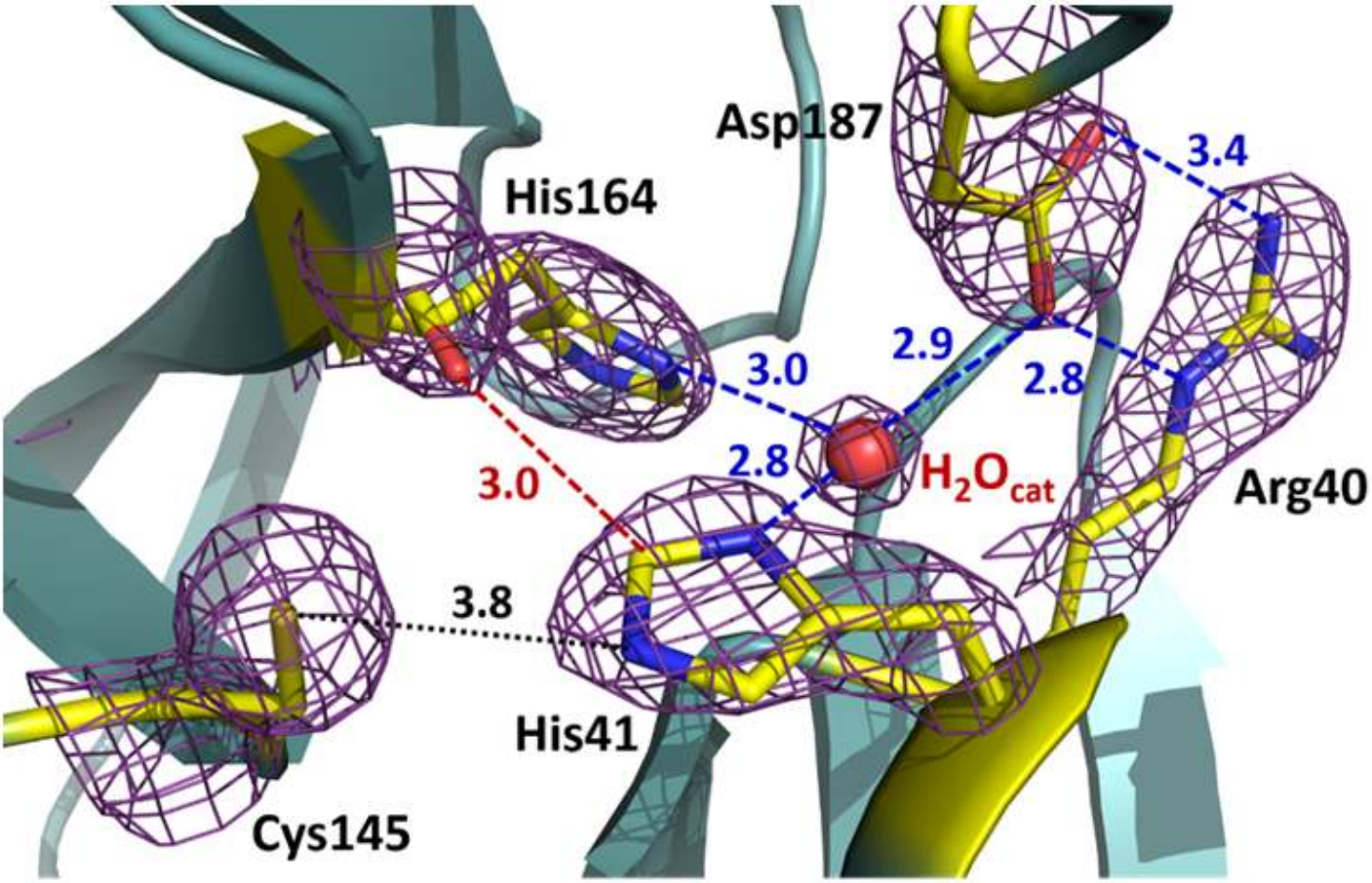

Figure 2

The catalytic site of 3CL Mpro from SARS-CoV-2. Hydrogen bonds are shown as blue dashed lines; the distance between Cys145 and His41 is shown as a black dotted line, the dashed red line indicates a strong C-H…O bond. All distances are given in Ångstroms.

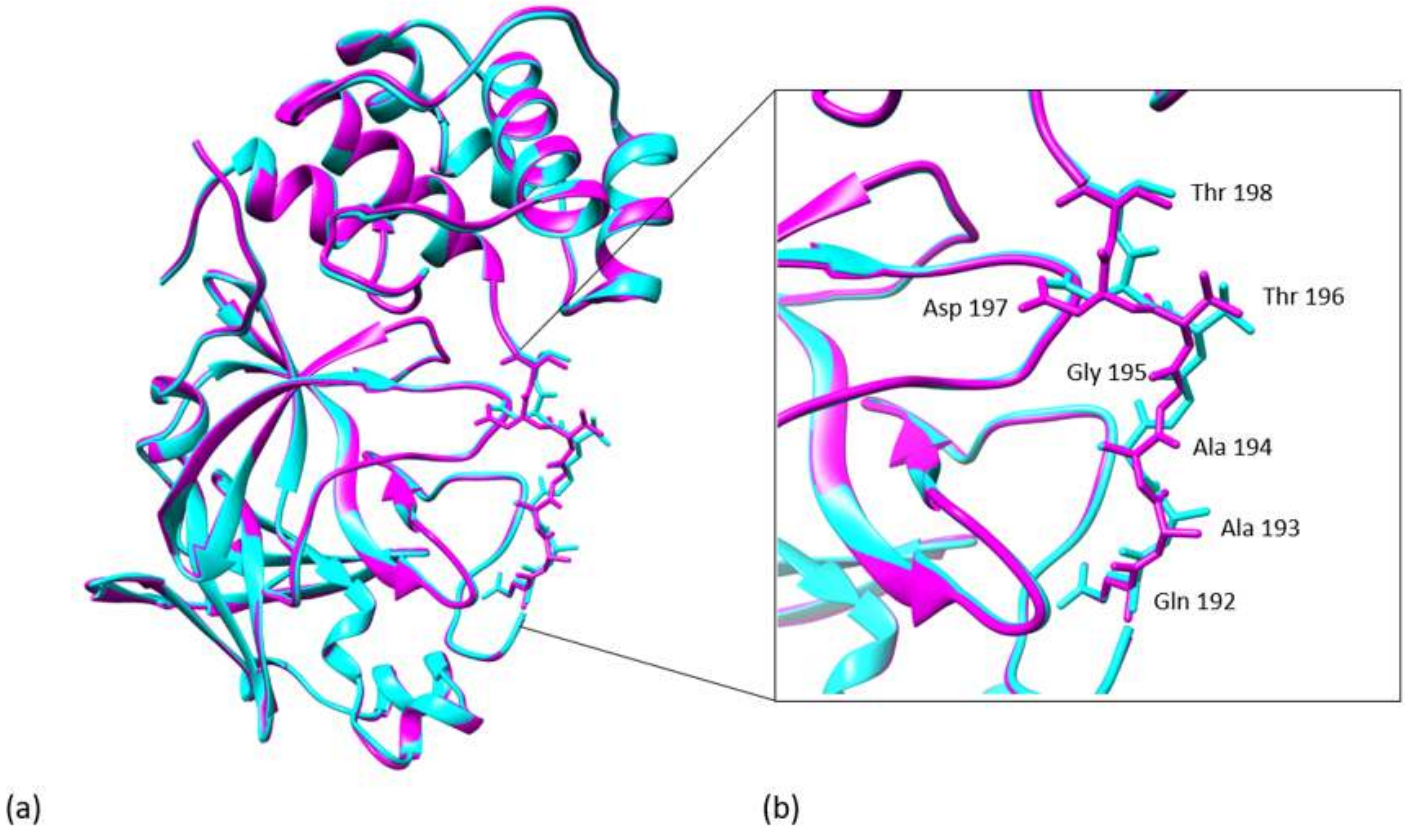

Figure 3

(a) A superposition of our room temperature ligand-free structure of 3CL Mpro (Magenta) with the ligand-free structure of 3CL Mpro (PDB ID 6Y2E) obtained at 100K (Cyan). (b) Residues 192-198 in the P5 binding pocket differ in conformation between the room temperature and 100K structures.

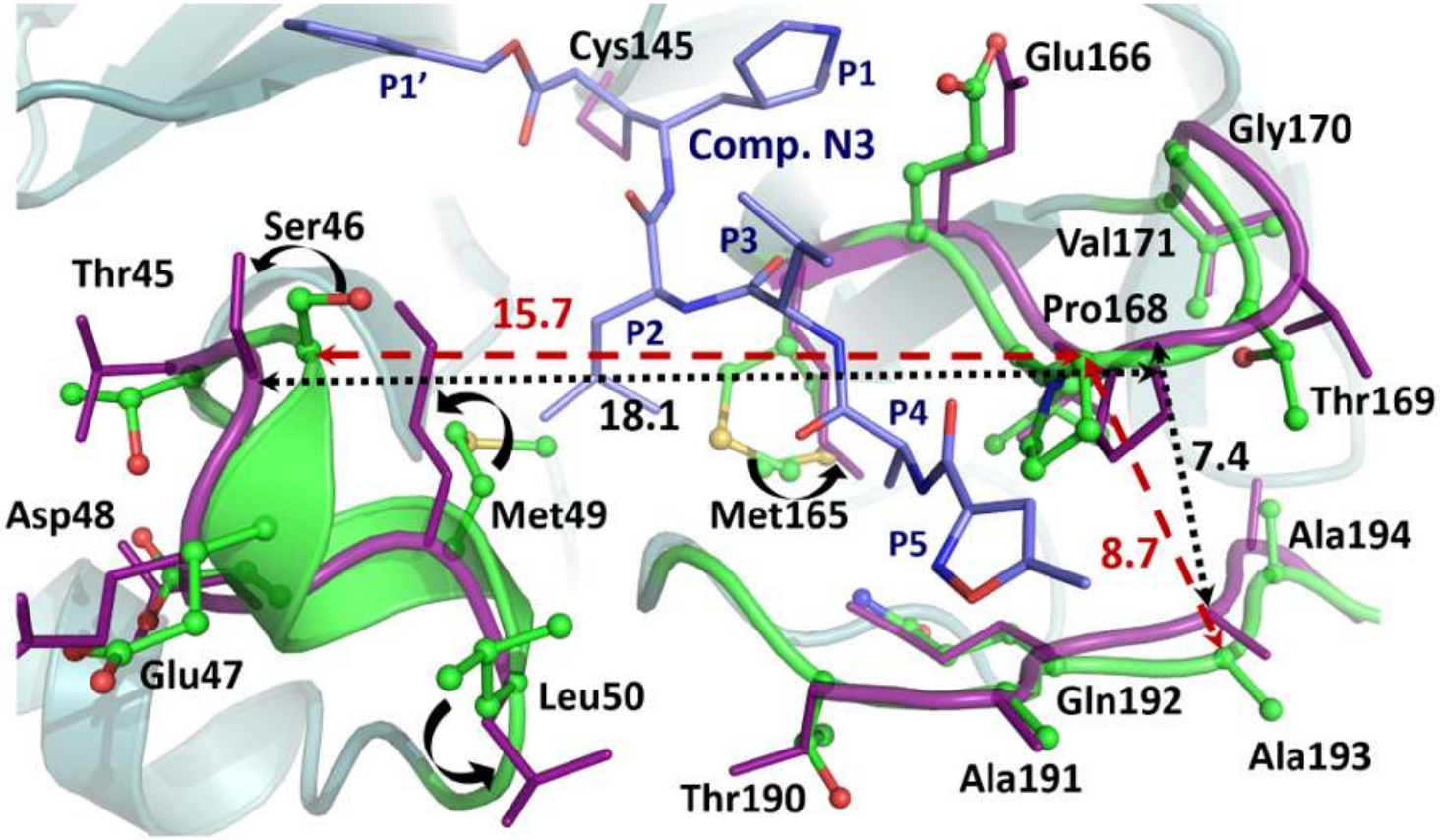

Figure 4

Superposition of the room temperature ligand-free structure of 3CL Mpro (green carbon atoms) in complex with inhibitor N3 (deep purple, PDB ID 6LU7) from SARS-CoV-2. Upon inhibitor binding, residues Met49, Leu50 and Met165 change their conformations (curved black arrows), whereas the small helix with residues 46-50 and the β-hairpin loop with residues 166-170 move apart, resulting in the loop with residues 190-194 which accommodates the inhibitor's P5 substituent to shift closer to the β-hairpin loop. All distances are given in Ångstroms.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 2562830supp4601904q9xhy4.docx