Structure of SARS-CoV-2 main protease in the apo state reveals the inactive conformation

Xuelan Zhou^{a,1}, Fangling Zhong^{b,c,1}, Cheng Lin^{d,e}, Xiaohui Hu^a, Yan Zhang^f, Bing Xiong^g,
Xiushan Yin^{h,i}, Jinheng Fu^j, Wei He^b, Jingjing Duan^k, Yang Fu^l, Huan Zhou^m, Qisheng Wang^{m,*},
Jian Li^{b,c,*}, Jin Zhang^{a,*}

^a School of Basic Medical Sciences, Nanchang University, Nanchang, Jiangxi, 330031, China.

^b College of Pharmaceutical Sciences, Gannan Medical University, Ganzhou, 341000, Jiangxi, PR, China.

^c Laboratory of Prevention and treatment of cardiovascular and cerebrovascular diseases, Ministry of Education, Gannan Medical University, Ganzhou 341000, PR China

^d Shenzhen Crystal Biopharmaceutical Co., Ltd, Shenzhen, Guangdong, 518118, China

^e Jiangxi Jmerry Biopharmaceutical Co., Ltd, Ganzhou, Jiangxi, 341000, China.

^f The Second Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, 330031, China

^g Department of Medicinal Chemistry, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China.

^h Applied Biology Laboratory, Shenyang University of Chemical Technology, 110142, Shenyang, China

ⁱ Biotech & Biomedicine Science (Jiangxi) Co. Ltd, Ganzhou, 341000, China

^j Jiangxi-OAI Joint Research Institution, Nanchang University, Nanchang, 330047, China.

^k Human Aging Research Institute (HARI), School of Life Sciences, Nanchang University, Nanchang, Jiangxi, 330031, China

^l School of Medicine, Southern University of Science and Technology, Shenzhen, Guangdong, China 518055

^m Shanghai Synchrotron Radiation Facility, Shanghai Advanced Research Institute, Chinese Academy of Sciences 239 Zhangheng Road, Pudong District, Shanghai 201204, P.R.China

29 * Corresponding author.

30 E-mail address: zhangxiaokong@hotmail.com (J. Zhang.); rmsl_2040@163.com (J. Li.) ;

31 wangqisheng@zjlab.org.cn (Q. Wang.)

32 ¹ These authors contributed equally to this work.

33

34 Abstract

35 M^{pro} is of considerable interest as a drug target in the treatment of COVID-19 since
 36 the proteolytic activity of this viral protease is essential for viral replication. Here we
 37 report the first insight of the structure M^{pro} for SARS-CoV-2 in the inactive
 38 conformation under conditions close to the physiological state (pH 7.5) to an overall
 39 resolution of 1.9 Å. The comparisons of M^{pro} in different states reveal that substrate
 40 binding site and the active site are more flexible in the inactive conformation than that
 41 in the active conformations. Notably, compared with the active conformation of the
 42 apo state structure in pH7.6 of SARS, the SARS-CoV-2 apo state is in the inactive
 43 conformation under condition close to physiological state (pH7.5). Two water
 44 molecules are present in the oxyanion hole in our apo state structure, whereas in the
 45 ligand-bound structure, water molecular is absence in the same region. This structure
 46 provides novel and important insights that have broad implications for understanding
 47 the structural basis underlying enzyme activity, and can facilitate rational,
 48 structure-based, approaches for the design of specific SARS-CoV-2 ligands as new
 49 therapeutic agents.

50

51 **Introduction**

52 Coronavirus Disease 2019 (COVID-19) caused by SARS-CoV-2 has been a global
 53 pandemic that severely threatens to the global health and economy. However, there is
 54 currently no clinically approved vaccines or drugs against COVID-19 [1].
 55 SARS-CoV-2 particles contain single, positive stranded RNA genome with a length
 56 of about 30 kb, with a 5' cap structure and a 3' poly (a) bundle. The SARS-CoV-2
 57 genome encodes for replicase, spike glycoprotein (S), envelope protein (E),
 58 membrane protein (M) and nucleocapsid protein (N). SARS-CoV-2 main protease
 59 (M^{pro} , also called 3C-like protease, 3CL $^{\text{pro}}$) mediates the proteolytic processing of
 60 large replicase polyprotein 1a (pp1a) and pp1ab into non-structural proteins (NSPs) at
 61 11 conservative sites. [2]. Thus, M^{pro} is of considerable interest as a drug target in the
 62 treatment of COVID-19 since the proteolytic activity of this viral protease is essential
 63 for viral replication. Mutational and structural studies have identified substrate
 64 binding site and active site of M^{pro} that confers specificity for the Gln-P1 substrate
 65 residue in the active conformation [3-5]. Structures of M^{pro} for SARS-CoV-2 has been
 66 solved in complexes with Chinese herbal and novel inhibitors very recently [6,7].
 67 However, a structural description of these sites in the inactive conformation has
 68 remained elusive.

69 **Results**

70 **Overall structure of M^{pro} for SARS-CoV-2**

71 Here we report the first insight of the structure M^{pro} for SARS-CoV-2 in the inactive
 72 conformation under conditions close to the physiological state (pH 7.5) to an overall
 73 resolution of 1.9 Å (Table 1), guiding specific drug discovery and functional studies.
 74 The M^{pro} forms a dimer in the crystal and has two distinct dimer interfaces, which are
 75 located in the N-terminal domain (residues 1-11) and the oxyanion loop (residues
 76 137-145) (Fig 1a). Comparison of our M^{pro} structure in the apo state to the previously
 77 reported M^{pro} structure in complex with inhibitor revealed a backbone (C α) RMSD of
 78 0.92 Å showing a similar overall structure [5-7] (Fig. 2a). As in ligand-bound M^{pro}

79 structures, the protein consists of N-finger and other three domains that bind inhibitor
80 at the cleft between domains I and II. N-finger (residues 1-7) is a loop located in the
81 dimer interface and involved in the N-terminal auto cleavage. The domain I (residues
82 8-101) is comprised of three small α -helices and six β -strands. The domain II
83 (residues 102-184) consists of six β -strands. The domain III is composed of five
84 α -helices, which are closely related to the proteolytic activity.

85 **Structural comparisons between M^{pro} in the apo states and other M^{pro} structures**

86 There were, however, several notable local differences between the apo and
87 ligand-bound structures. Electron density of the N-finger (residues 1-2), oxyanion
88 loop (residues 141-142), C-terminal domain (residues 299-306) were insufficient for
89 backbone tracing, suggesting the flexibility of this region in the apo state. In addition,
90 electron densities of side chains Phe140 and Glu166, which are key residues involved
91 in the substrate binding are missing at this high resolution that may reflect different
92 conformation of the apo state (Fig. 2b).

93 The oxyanion hole composed of backbone amides or positively charged residues is
94 directly related to the enzyme activity and substrate binding. In ligand-bound
95 structures of M^{pro}, the oxyanion hole consists of loop (residues 140-145), negatively
96 charged residues Glu166, positively charged residues His41, His163 and His172
97 remains in an active conformation (Fig. 2c) [6]. A π - π stacking interaction
98 (Phe140/His163) is found in the oxyanion hole. A hydrogen bond and salt bridge
99 involving Glu166 with water and His172 at the domain II further stabilize oxyanion
100 hole. However, the oxyanion loop (residues 137-145) is less well ordered and the side
101 chains of Glu166 and Phe140 cannot be fit well due to poor density in our apo state
102 structure. The salt bridge and π - π stacking interactions between Glu166/His172 and
103 Phe140/His163 are broken, resulting in rearrangements in this region and further
104 collapses of the oxyanion hole (Fig. 2d).

105 **M^{pro} in the apo states at pH 7.5 is in an inactive conformation**

106 We propose that M^{pro} in the apo state is in the inactive conformation in the
 107 physiological condition, which is different from the active conformation of
 108 ligand-bound structures of M^{pro} [5]. The disordered N-finger is another feature of the
 109 inactive conformation. N-finger plays an important role in the formation of the active
 110 site and auto cleavage activity of M^{pro} [4]. Gly2 has interactions with Gly143 in the
 111 oxyanion loop in the neighboring protomer, stabilizing the active site and dimer in the
 112 active conformation, while the electron density of Gly2 is completely missing in the
 113 apo state. Inactive conformation in the apo state of our structure is further supported
 114 by the flexibility of N-finger in the apo state. It is consistent with the proof that lack
 115 of N-finger in TGEV M^{pro} is almost completely inactive [8]. Interestingly, His163
 116 forms hydrogen bonds with water molecular (Water 1) in our structure, which is not
 117 observed in the active conformation. Another unprecedented water molecular (Water 2)
 118 is found at Cys145-His41 catalytic dyad in the active site, working as bridge for
 119 proton transfer. We speculate that these water molecules may affect negatively
 120 charged oxygen of the substrate or inhibitor, which suffers from steric hindrance,
 121 making rational drug design more difficult (Fig. 2b and d).

122 Discussion

123 In summary, we determined the apo state structure of M^{pro} for SARS-CoV-2 in the
 124 inactive conformation. The comparisons of M^{pro} in different states reveal that
 125 substrate binding site and the active site are more flexible in the inactive conformation
 126 than that in the active conformations. Notably, compared with the active conformation
 127 of the apo state structure in pH7.6 of SARS, the SARS-CoV-2 apo state is in the
 128 inactive conformation under condition close to physiological state (pH7.5). The
 129 instable and disordered regions of oxyanion hole and the active site in the inactive
 130 conformation will raise the activation energy of the protease necessary for the
 131 reaction, slow down catalysis and finally extend the replication cycle of the virus.
 132 These structural differences may reveal the underlying reasons of why some patients
 133 infected with SARS-CoV-2 have longer virus latency than that of SARS. Further

134 studies of detailed molecular mechanisms of SARS-CoV-2 pathogenesis are needed.
 135 For the drug design based on the structure, water molecules imbedded in the oxyanion
 136 hole and corresponding interactions should be taken into more consideration. Two
 137 water molecules are present in the oxyanion hole in our apo state structure, whereas in
 138 the ligand-bound structure, water molecular is absence in the same region. The water
 139 molecules, which is found near His163 and His41 in the occluded pocket, stabilizes
 140 the positively charged His residues, increasing the steric hindrance that may slow
 141 down the enzyme reaction and decrease the catalytic efficiency of the enzyme.
 142 Altogether, the apo state structure of M^{pro} for SARS-CoV-2 is an important
 143 complementary to the available structures. This structure provides novel and
 144 important insights that have broad implications for understanding the structural basis
 145 underlying enzyme activity, and can facilitate rational, structure-based, approaches
 146 for the design of specific SARS-CoV-2 ligands as new therapeutic agents.

147

148 **Materials and Methods**

149 **Protein purification and crystallization**

150 The cDNA of full length COVID-2019 main protease 3CL (NC_045512) were was
 151 optimized and synthesized (Generay, China) connected into vector pET28a to obtain
 152 the wanted plasmid. The plasmid was transformed into competent cell E.coli Rosetta
 153 DE3. The bacteria were grown in 800mL of LB (Luria-Bertani) broth at 37°C. When
 154 the OD600 reach 0.6-0.8, 500μM IPTG was added to induce the E. coli expression
 155 and then incubated 3-5h at 30°C. Centrifuge the cells at 10000g for 10min at 4 °C,
 156 discard the supernatant, collect the precipitate and add buffer A(100mM Tris/HCl
 157 buffer,pH7.5,300mM NaCl 10mM imidazole and 5% glycerol) to blend the collected
 158 cells, which were broken up by JNBIO 3000 plus(JNBI). The supernatant containing
 159 needed protein was acquired by centrifugation at 30000g, 4°C for 30min. Transfer
 160 the supernatant into 5ml Ni-NTA(Ni²⁺-nitrilotriacetate)column (GE healthcare) and
 161 the protein wanted was loaded onto the column. Add buffer B(100mM Tris/Hcl buffer ,
 162 pH7.5 , 300mM NaCl ,100mM imidazole , and 5% glycerol) into beads which use

163 imidazole to wash. The His tagged protein was eluted by buffer C (50 mM Tris-HCl
164 pH 7.5, 300 mM NaCl and 300 mM imidazole). Superdex 200 PG gel filtration
165 column (GE healthcare) can more purify protein and remove imidazole, while need to
166 change the buffer to buffer C(25mM HEPES buffer ,pH7.5 ,300mM NaCl, 2mM DTT
167 and 5% glycerol). Collect postive peak protein and use tiny part test by SDS-PAGE.
168 In the end, the protein was flash-frozen in liquid nitrogen and stored at -80°C.
169 Thaw the protein and concentrate it to 5mg/ml in Amicon Ultra-15,10000Mr cut-off
170 centrifugal concentrator (Millipore). The hanging drop vapor diffusion method was
171 useful to gain crystal at 4°C. The crystals were grown with buffer containing 0.1M
172 HEPES sodium 7.5, 10% Propanol and 20% PEG4000 in 3-5 days.

173 **Data collection, structure determination and refinement.**

174 The crystals were tailored with cryo-loop (Hampton research,America)and then
175 flash-frozen in liquid nitrogen to collect better X-ray data. All data sets were collected
176 at 100 K on macromolecular crystallography beamline17U1 (BL17U1) at Shanghai
177 Synchrotron Radiation Facility (SSRF,Shanghai,China). All collected data were
178 handled by the HKL 2000 software package. The structures of COVID-2019 main
179 protease 3CL were determined by molecular replacement with PHENIX software.
180 The of 3T0H was referred as a model. The program Coot was used to rebuild the
181 initial model. The models were refined to resolution limit 1.93Å by using the
182 PHENIX software. The superimposed data was analyzed with PyMOL software
183 package. The complete wanted data collection and statistics of refinement are shown
184 in Table 1. The structure has been deposited in PDB (PDB code 7C2Q).

185

186 **Conflict of interest**

187 The authors declare that they have no conflict of interest.

188

189 **Acknowledgments**

190 Jin Zhang was supported by the Thousand Young Talents Program of China, the

191 National Natural Science Foundation of China (grant no. 31770795; grant no.
192 81974514), and the Jiangxi Province Natural Science Foundation (grant no.
193 20181ACB20014). Jian Li was supported by the Open Project of Key Laboratory of
194 Prevention and treatment of cardiovascular and cerebrovascular diseases, Ministry of
195 Education (No. XN201904), Gannan Medical University (QD201910) and
196 Jiangxi "Double Thousand Plan". This work was also supported by Ganzhou
197 COVID-19 Emergency Research Project.

198

199 Author contributions

200 Jin Zhang and Jian Li designed the project. Xuelan Zhou, Fanglin Zhong, Xiaohui Hu
201 and Cheng Lin made constructs for expression and determined the conditions used to
202 enhance protein stability. Huan Zhou and Qisheng Wang carried out X-ray
203 experiments, including data acquisition and processing. Jian Li and Jin Zhang built
204 the atomic model. Jin Zhang, Jian Li and Jingjing Duan drafted the manuscript. All
205 authors contributed to structure analysis/interpretation and manuscript revision. Jin
206 Zhang and Jian Li initiated the project, planned and analyzed the experiments, and
207 supervised the research.

208

209 References

- 210 [1] Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019
211 novel coronavirus: implications for virus origins and receptor binding. *Lancet*
212 2020;395(10224):565-74.
- 213 [2] Zumla A, Chan JF, Azhar EI, et al. Coronaviruses - drug discovery and therapeutic
214 options. *Nat Rev Drug Discov* 2016;15(5):327-47.
- 215 [3] Jin Z, Du X, Xu Y, et al. Structure of M(pro) from COVID-19 virus and discovery
216 of its inhibitors. *Nature* 2020.
- 217 [4] Hilgenfeld R. From SARS to MERS: crystallographic studies on coronaviral
218 proteases enable antiviral drug design. *Febs J* 2014;281(18):4085-96.

- 219 [5] Yang H, Yang M, Ding Y, et al. The crystal structures of severe acute respiratory
220 syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U*
221 *S A* 2003;100(23):13190-5.
- 222 [6] Dai W, Zhang B, Su H, et al. Structure-based design of antiviral drug candidates
223 targeting the SARS-CoV-2 main protease. *Science* 2020.
- 224 [7] Jin Z, Zhao Y, Sun Y, et al. Structural basis for the inhibition of SARS-CoV-2
225 main protease by antineoplastic drug carmofur. *Nat Struct Mol Biol* 2020.
- 226 [8] Anand K, Palm GJ, Mesters JR, et al. Structure of coronavirus main proteinase
227 reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *The*
228 *EMBO Journal* 2002;21(13):3213-24.
- 229

230 **Figure legends**

231 **Fig. 1.** The apo state structure of M^{pro} of SARS-CoV-2 in the inactive conformation.

232 (a) The structure of the M^{pro} dimer is shown in stereo. Individual protomers are shown
233 in red and green. (b) Two dimer interfaces of M^{pro} of SARS-CoV-2. Dimer interface 1
234 and 2 are located in the oxyanion hole and N-terminal domain, respectively. (c)
235 Comparison of M^{pro} structures in the apo state of (green) in SARS-CoV-2, with
236 inhibitor N3 in SARS-CoV-2 (red, PDB: 6LU7), with inhibitor 11b in SARS-CoV-2
237 (orange, PDB: 6LZE) and in the apo state of SARS (gray, PDB:1UJ1). N3 and 11b are
238 shown in pink and cyan, respectively. (d) Comparison of substrate binding site and
239 active site in the apo state (green) and in the ligand-bound state in M^{pro} of
240 SARS-CoV-2. 11b are shown in cyan. (e) Structure of M^{pro} bound with 11b in an
241 active conformation. (f) Structure of M^{pro} in an inactive conformation. Water 1 and 2
242 are shown in red spheres.

243 **Fig. 2.** (a) Comparison of M^{pro} structures in the apo state of (green) in SARS-CoV-2,
244 with inhibitor N3 in SARS-CoV-2 (red, PDB: 6LU7), with inhibitor 11b in
245 SARS-CoV-2 (orange, PDB: 6LZE) and in the apo state of SARS (gray, PDB:1UJ1).
246 N3 and 11b are shown in pink and cyan, respectively. (b) Comparison of substrate
247 binding site and active site in the apo state (green) and in the ligand-bound state in
248 M^{pro} of SARS-CoV-2. 11b are shown in cyan. (c) Structure of M^{pro} bound with 11b in
249 an active conformation. (d) Structure of M^{pro} in an inactive conformation. Water 1 and
250 2 are shown in red spheres.

251 **Table.1.** Statistics for data processing and model refinement of COVID-19 M^{pro}.

252

253 **Fig.1.**

258

259 **Table 1**

PDB code	7C2Q
Synchrotron	SSRF
Beam line	BL17U1
Wavelength (Å)	0.97918
Space group	P2 ₁ 2 ₁ 2 ₁
<i>a</i> , <i>b</i> , <i>c</i> (Å)	67.95, 102.66, 103.54
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90.00, 90.00, 90.00
Total reflections	716573
Unique reflections	55189
Resolution (Å)	1.93(1.98-1.93)
R-merge (%)	11.8(149.3)
Mean <i>I</i> / <i>σ</i> (<i>I</i>)	15.4/2.2
Completeness (%)	99.9(99.9)
Redundancy	13.0(12.8)
Resolution (Å)	72.90-1.93
<i>R</i> _{work} / <i>R</i> _{free} (%)	22.63/26.61
Atoms	4917
Mean temperature factor (Å ²)	36.7
Bond lengths (Å)	0.008
Bond angles (°)	0.88

260 Values in parentheses are for the highest-resolution shell.