

1 **Lead compounds for the development of SARS-CoV-2 3CL protease inhibitors**

2

3 Sho Iketani^{1,2}, Farhad Forouhar³, Hengrui Liu⁴, Seo Jung Hong⁵, Fang-Yu Lin⁶, Manoj S. Nair¹, Arie
4 Zask⁷, Yaoxing Huang¹, Li Xing⁶, Brent R. Stockwell^{4,7*}, Alejandro Chavez^{5*}, David D. Ho^{1*}

5

6 ¹ Aaron Diamond AIDS Research Center, Columbia University Irving Medical Center, New York, NY
7 10032, USA

8 ² Department of Microbiology and Immunology, Columbia University Irving Medical Center, New
9 York, NY 10032, USA

10 ³ Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New
11 York, NY 10032, USA

12 ⁴ Department of Chemistry, Columbia University, New York, NY 10027, USA

13 ⁵ Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York,
14 NY 10032, USA

15 ⁶ WuXi AppTec, Cambridge, MA 02142, USA

16 ⁷ Department of Biological Sciences, Columbia University, New York, NY 10027, USA

17 * Corresponding authors

18

19 Correspondence to: dh2994@cumc.columbia.edu; ac4304@cumc.columbia.edu;

20 bstockwell@columbia.edu

21

22

23 **Abstract**

24

25 We report the identification of three structurally diverse compounds – compound 4, GC376, and
26 MAC-5576 – as inhibitors of the SARS-CoV-2 3CL protease. Structures of each of these compounds
27 in complex with the protease revealed strategies for further development, as well as general
28 principles for designing SARS-CoV-2 3CL protease inhibitors. These compounds may therefore
29 serve as leads for the basis of building effective SARS-CoV-2 3CL protease inhibitors.

30

31 **Main Text**

32

33 As the etiologic agent of COVID-19, SARS-CoV-2 has resulted in hundreds of thousands of deaths
34 and caused rampant economic damage worldwide^{1,2}. While some treatments have been identified,
35 their clinical efficacy is low, making continued research essential^{3,4}. Similar to other coronaviruses,
36 SARS-CoV-2 encodes an essential 3CL protease (M_{pro}) that processes its polyproteins, which has
37 garnered interest as a target for potential viral inhibitors^{5,6}. Here, we describe a series of compounds
38 with inhibitory activity against the SARS-CoV-2 3CL and determine their structures in complex with
39 the protease. These data provide general insights into the design of 3CL protease inhibitors, along
40 with potential avenues by which these classes of compounds can be further developed.

41

42 We hypothesized that previously identified SARS-CoV-1 3CL protease inhibitors may also be
43 effective against the SARS-CoV-2 3CL, given the conservation between the two proteases (96%
44 amino acid identity)^{1,2}. Using a biochemical assay to report SARS-CoV-2 3CL protease activity
45 (**Extended Data Fig. 1a, b**), we identified three diverse compounds of interest: compound 4⁷,
46 GC376⁸, and MAC-5576⁹, which had IC_{50} values (mean \pm s.d.) of $0.149 \pm 0.002 \mu\text{M}$, 0.139 ± 0.016
47 μM , and $0.088 \pm 0.003 \mu\text{M}$, respectively (**Fig. 1a, b**). Encouraged by these results, we then tested
48 these compounds for inhibition of SARS-CoV-2 viral replication. We found that compound 4 and
49 GC376 could block viral infection (EC_{50} values (mean \pm s.d.): $3.023 \pm 0.923 \mu\text{M}$ and 4.481 ± 0.529
50 μM , respectively), whereas MAC-5576 did not (**Fig. 1c**). Finally, we confirmed that these compounds
51 did not result in cytotoxicity to the cells at the tested concentrations (**Extended Data Fig. 2**).

52

53 As the three compounds exhibited inhibitory activity against the SARS-CoV-2 3CL, we
54 proceeded to solve the crystal structure of the *apo* 3CL protease alone and of each of these
55 compounds in complex with the protease to understand their mechanism of binding as well as to
56 guide future structure-based optimization efforts. We note that while MAC-5576 did not exhibit
57 activity in the cellular assay, its low molecular weight and reasonable biochemical activity prompted

58 us to pursue its crystallization as well, as our goal was to broadly investigate inhibitory scaffolds for
59 the SARS-CoV-2 3CL protease. Crystals were obtained (see **Methods** for detailed information) and
60 structures at 1.85 Å, 1.80 Å, 1.83 Å, and 1.73 Å resolution limits for *apo* 3CL and 3CL bound to
61 compound 4, GC376, and MAC-5576, respectively, were solved (**Fig. 2a, b, c, Extended Data Fig.**
62 **3, and Extended Data Fig. 4**; see **Supplementary Table 1** for statistics).

63
64 The X-ray crystal structures revealed that all three of the compounds bind covalently to the
65 catalytically active Cys145 residue within the substrate-binding pocket of the protease. We observed
66 distinct mechanisms by which these compounds acted on this residue. Compound 4 functioned in a
67 similar binding mode as other reported compounds, covalently modifying Cys145 through Michael
68 addition (**Fig. 2a**)⁵. For GC376, the bisulfite adduct was converted to an aldehyde as previously
69 reported, allowing it to then react with Cys145 through nucleophilic addition and hemithioacetal
70 formation (**Fig. 2b**)⁸. MAC-5576 also covalently modified Cys145 by nucleophilic linkage, as
71 expected (**Fig. 2c**).

72
73 As we solved the structures for multiple compounds, we hypothesized that general principles
74 for the design of SARS-CoV-2 3CL protease inhibitors could be identified. We first overlaid all four
75 crystal structures of the 3CL with or without inhibitors (**Extended Data Fig. 5**). We observed local
76 conformational changes, with Thr45 to Pro52 distinct from the *apo* 3CL in all three inhibitor-bound
77 structures, whereas Arg188 to Gln192 differed only in the compound 4 and GC376-bound, but not
78 MAC-5576-bound structures. We then overlaid each of the inhibitors in the substrate binding pocket
79 of the 3CL protease to find commonalities in their interactions (**Fig. 2d**). Most notably, we found that
80 all of these compounds occupied the S2 site, with compound 4 and GC376 further anchored in the
81 S1 and S3 sites. The backbone NH of Gly143 points toward the ligand binding pocket, forming
82 hydrogen bonds with the carbonyl oxygen of the ethyl ester of compound 4, and the hemithioacetal
83 of GC376 after the Cys145 addition to the original aldehyde, even though the former hydrogen bond
84 is stronger than the latter. In both structures, the γ -lactam groups occupy the S1 site, and are

85 strongly anchored by two hydrogen bonds with the side chains of His163 and Glu166. The isobutyl
86 groups are favorably embedded in the hydrophobic S2 site, surrounded by the alkyl portion of the
87 side chains of His41, Met49, His164, Met165, Asp187, and Gln189. Extending into the S3 pocket,
88 the amide bonds of compound 4 and GC376 are stabilized by hydrogen bond interactions with the
89 side chain of Gln189. Similar interactions are also observed in reports of related compounds,
90 suggesting that overall, the binding modes of this class of substrate mimetic inhibitors share
91 remarkable similarities^{5,6,10}. Specifically, they all have a γ -lactam occupying the S1 pocket,
92 preserving the dual hydrogen bonds with His163 and Glu166. Furthermore, they commonly contain a
93 hydrophobic moiety occupying the S2 site. As shown in a structural overlay of compound 4 and
94 GC376 with these related compounds (**Extended Data Fig. 6**), the segment of the inhibitors from S1
95 to S2 align closely on top of each other. Variations of binding start to emerge in the S3 and S4
96 region, which exhibits high degrees of freedom in terms of structural diversity as well as
97 conformational flexibility. In our experiments, the S3 and S4 sites displayed weaker electron density,
98 indicating flexibility in the inhibitor and/or the protease in these regions (**Extended Data Fig. 4**).
99 These observations suggest that development of 3CL protease inhibitors may benefit from first
100 establishing robust interactions within the S1, S2, and/or S1' sites, before extending into the S3 and
101 S4 sites. Possibly, compounds such as compound 4 and GC376 are not optimized for binding into
102 the S3 and S4 sites, and there are ample opportunities to improve the inhibitory potencies against
103 the 3CL by designing compounds that exploit the accessible contact points to strengthen the ligand-
104 protein interactions.

105
106 On the other hand, the binding of MAC-5576, as a non-peptidic small molecule, displays
107 unique features that differ from that of compound 4 or GC376. We observed that the thiophene
108 group forms π - π stacking with the His41 side chain imidazole, which undergoes a conformational
109 rotation around its beta-carbon to align parallel to the thiophene, as compared to the other peptide-
110 bound structures. Additionally, the side chain of Gln189 also shows notable conformational variation
111 compared to those in the compound 4 and GC376 crystal structures, possibly in response to the

112 specific hydrogen bond interactions induced by the respective ligands. Notably, the rotation of His41
113 has been reported previously in the crystal structure of a benzotriazole ester inhibitor (XP-59) in
114 complex with the SARS-CoV-1 3CL protease (PDB:2V6N)¹¹. An overlaid model of the crystal
115 structures of MAC-5576 bound to SARS-CoV-2 3CL and XP-59 bound to SARS-CoV-1 3CL shows
116 that both compounds have similar binding modes when covalently bound to Cys145, in which the
117 thiophene of MAC-5576 and the phenyl ring of XP-59 almost overlap with each other, both engaging
118 the His41 side chain via π - π stacking interactions (**Extended Data Fig. 7**).

119

120 In summation, we have identified compound 4, GC376, and MAC-5576 as inhibitors of the
121 SARS-CoV-2 3CL protease. Crystal structures of the compounds complexed to the protease
122 suggested their mechanisms of action, as well as portended guidelines for the development of
123 SARS-CoV-2 3CL protease inhibitors, which may aid in the future development of novel inhibitors to
124 combat this virus.

125 **Methods**

126

127 *Compounds*

128 Compound 4 was synthesized using the synthesis route previously described, with the exception of
129 using a sodium borohydride-cobaltous chloride reduction of the nitrile in the construction of the
130 lactam, thus avoiding the high pressure hydrogenation in the original route^{7, 12}. GC376 was
131 purchased from Aobious and MAC-5576 was purchased from Maybridge.

132

133 *Expression and purification of SARS-CoV-2 3CL protease.*

134 The SARS-CoV-2 3CL protease gene was codon optimized for bacterial expression and synthesized
135 (Twist Bioscience), then cloned into a bacterial expression vector (pGEX-5X-3, GE) that expresses
136 the protease as a fusion construct with a N-terminal GST tag, followed by a Factor Xa cleavage site.
137 After confirmation by Sanger sequencing, the construct was transformed into BL21 (DE3) cells.
138 These *E. coli* were inoculated and grown overnight as starter cultures, then used to inoculate larger
139 cultures at a 1:100 dilution, which were then grown at 37 °C, 220 RPM until the OD reached 0.6-0.7.
140 Expression of the protease was induced with the addition of 0.5 mM IPTG, and then the cultures
141 were incubated at 16 °C, 180 RPM for 10 h. Cells were pelleted at 4500 RPM for 15 min at 4 °C,
142 resuspended in lysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl), homogenized by sonication,
143 then clarified by centrifuging at 25000 x *g* for 1 h at 4 °C. The supernatant was mixed with
144 Glutathione Sepharose resin (Sigma) and placed on a rotator for 2 h at 4 °C. The resin was then
145 repeatedly washed by centrifugation at 3500 RPM for 15 min at 4 °C, discarding of the supernatant,
146 and then resuspension of the resin in fresh lysis buffer. After ten washes, the resin was resuspended
147 in lysis buffer, and Factor Xa was added and incubated for 18 h at 4 °C on a rotator. The resin was
148 centrifuged at 3500 RPM for 15 min at 4 °C, and then the supernatant was collected and
149 concentrated using a 10 kDa concentrator (Amicon) before being loaded onto a Superdex 10/300 GL
150 column for further purification by size exclusion chromatography. The appropriate fractions were

151 collected and pooled with a 10 kDa concentrator, and then the final product was assessed for quality
152 by SDS-PAGE and measurement of biochemical activity.

153

154 *Measurement of SARS-CoV-2 3CL protease biochemical activity.*

155 The *in vitro* biochemical activity of the SARS-CoV-2 3CL protease was measured as previously
156 described⁵. The fluorogenic peptide MCA-AVLQSGFR-Lys(DNP)-Lys-NH₂, corresponding to the
157 nsp4/nsp5 cleavage site in the virus, was synthesized (GL Biochem), then resuspended in DMSO to
158 use as the substrate. Different concentrations of this substrate, ranging from 5 μ M to 100 μ M, were
159 prepared in the assay buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA) in a 96 well-plate. The protease
160 was then added to each well at a concentration of 0.2 μ M, and then fluorescence was continuously
161 measured on a plate reader for 3 min. The catalytic efficiency of the protease was then calculated by
162 generating a double-reciprocal plot.

163

164 *Measurement of SARS-CoV-2 3CL protease inhibition.*

165 Inhibition of the biochemical activity of the SARS-CoV-2 3CL protease was quantified as previously
166 described with modifications⁵. Serial dilutions of the test compound were prepared in the assay
167 buffer, and then incubated with 0.2 μ M of the protease for 10 min at 37 °C. The substrate was then
168 added at 20 μ M per well, and then fluorescence was continuously measured on a plate reader for 3
169 min. Inhibition was then calculated by comparison to control wells with no inhibitor added. IC₅₀
170 values were determined by fitting an asymmetric sigmoidal curve to the data (GraphPad Prism).

171

172 *Measurement of SARS-CoV-2 viral inhibition.*

173 Stocks of SARS-CoV-2 strain 2019-nCoV/USA_WA1/2020 was propagated and titered in Vero-E6
174 cells. One day prior to the experiment, Vero-E6 cells were seeded at 30,000 cells/well in 96 well-
175 plates. Serial dilutions of the test compound were prepared in cell media (EMEM + 10% FCS +
176 penicillin/streptomycin), overlaid onto cells, and then virus was added to each well at an MOI of 0.2.
177 Cells were incubated at 37 °C under 5% CO₂ for 72 h and then viral cytopathic effect was scored in

178 a blinded manner. Inhibition was calculated by comparison to control wells with no inhibitor added.
179 EC₅₀ values were determined by fitting an asymmetric sigmoidal curve to the data (GraphPad Prism).
180 Cells were confirmed as mycoplasma negative prior to use. All experiments were conducted in a
181 biosafety level 3 (BSL-3) lab.

182

183 *Measurement of cellular cytotoxicity.*

184 Vero-E6 cells were incubated with the compound of interest for 48 h at 37 °C under 5% CO₂ and
185 then cellular cytotoxicity was determined with the XTT Cell Proliferation Assay Kit (ATCC) according
186 to the manufacturer's instructions.

187

188 *Crystallization, data collection, and structure determination.*

189 To generate the complex of SARS-CoV-2 3CL protease bound to compound 4, 50 µM of the 3CL
190 protease was incubated with 500 µM of compound 4 in a buffer comprised of 50 mM Tris-HCl (pH
191 7.5), 1 mM EDTA, and 5% (v/v) glycerol for 1 h at 4 °C. This complex was then concentrated to 8.5
192 mg/mL using a 10 kDa concentrator, and initially subjected to extensive robotic screening at the
193 High-Throughput Crystallization Screening Center¹³ of the Hauptman-Woodward Medical Research
194 Institute (HWI) (<https://hwi.buffalo.edu/high-throughput-crystallization-center/>). The most promising
195 crystal hits were then reproduced using the microbatch-under-oil method at 4 °C. Block-like crystals
196 of 3CL in complex with compound 4 appeared after a few days in the crystallization condition
197 comprised of 0.1 M potassium nitrate, 0.1 M sodium acetate (pH 5), and 20% (w/v) PEG 1000 with
198 protein to crystallization reagent at a 2:1 ratio. The crystals were subsequently transferred into the
199 same crystallization reagent supplemented with 15% (v/v) glycerol and flash-frozen in liquid nitrogen.

200

201 To obtain crystals of 3CL in complex with GC376, crystals of *apo* 3CL were initially grown by using
202 seeding method in a crystallization reagent comprised of 0.1 M sodium phosphate-monobasic, 0.1
203 MES (pH 6), and 20% (w/v) PEG 4000. The *apo* crystals were subsequently soaked with 15 mM

204 GC376, followed by flash-freezing of the crystals in the same reagent supplemented with 15%
205 ethylene glycol.

206
207 To generate the complex of SARS-CoV-2 3CL protease bound to MAC-5576, 50 μ M of the 3CL
208 protease was incubated with 500 μ M of compound 4 in a buffer comprised of 50 mM Tris-HCl (pH
209 7.5), 1 mM EDTA, and 5% (v/v) glycerol for 1 h at 4 °C. The complex was concentrated to 10 mg/mL
210 using a 10 kDa concentrator, and then crystallized in the same conditions as those used for
211 crystallization of *apo* 3CL.

212
213 A native dataset was collected on each crystal of 3CL, alone (*apo*), and in complex with compound 4
214 and GC376 at the NE-CAT24-ID-C beam line of Advanced Photon Source (APS) at Argonne
215 National Laboratory, and the NE-CAT 24-ID-E beam line of APS was used for data collection on
216 crystals of 3CL-MAC-5576. Crystals of *apo* 3CL and in complex with compound 4, GC376, and
217 MAC-5576 diffracted the X-ray beam to resolution 1.85 Å, 1.80 Å, 1.83 Å, 1.73 Å, respectively. The
218 images were processed and scaled in space group C2 using XDS¹⁴. The structure of 3CL-compound
219 4 was determined by molecular replacement method using program MOLREP¹⁵ and the crystal
220 structure of 3CL in complex with inhibitor N3 (PDB id: 6LU7)⁵ was used as a search model. The
221 geometry of each crystal structure was subsequently fixed and the corresponding inhibitor was
222 modeled in by XtalView¹⁶ and Coot¹⁷, and refined using PHENIX¹⁸. The mapping of electrostatic
223 potential surfaces was generated in PyMOL with the APBS plug-in¹⁹. There is one protomer of 3CL
224 complex in the asymmetric unit of each crystal. The crystallographic statistics are shown in

225 **Supplementary Table 1.**

226
227 *Reporting Summary*

228 Further information on research design is available in the Nature Research Reporting Summary
229 linked to this article.

230

231 **Data availability**

232
233 Structural data for the *apo* SARS-CoV-2 3CL protease and 3CL in complex with compound 4,
234 GC376, and MAC-5576 will be deposited in the Protein Data Bank (PDB) and made publicly
235 available upon publication. Source data for Fig. 1, Extended Data Fig. 1b, Extended Data Fig. 2, and
236 the unprocessed gel for Extended Data Fig. 1a are available with the paper online.

237

238 **Acknowledgements**

239
240 This work was supported by a grant from the Jack Ma Foundation to D.D.H. and A.C. and by grants
241 from Columbia Technology Ventures and the Columbia Translational Therapeutics (TRx) program to
242 B.R.S. A.C. is also supported by a Career Awards for Medical Scientists from the Burroughs
243 Wellcome Fund. S.I. is supported by NIH grant T32AI106711. We thank the staff of the High-
244 Throughput Crystallization Screening Center of the Hauptman-Woodward Medical Research Institute
245 for screening of crystallization conditions and the staff of the Advanced Photon Source at Argonne
246 National Laboratory for assistance with data collection.

247

248 **Author contributions**

249
250 D.D.H. conceived the project. S.I., B.R.S., A.C., and D.D.H. planned and designed the experiments.
251 S.I. and S.J.H. cloned, expressed, and purified proteins. A.Z. synthesized compound 4. S.I.
252 conducted the protease inhibition assay. M.N. and Y.H. conducted the antiviral assay and
253 cytotoxicity measurements. F.F. and H.L. crystallized the proteins. F.F. collected diffraction data and
254 solved the crystal structures. F.F., F.-Y.L., and L.X. analyzed the structural data. S.I., F.F., F.-Y.L.,
255 L.X., A.C., and D.D.H. wrote the manuscript with input from all authors.

256

257 **Competing interests**

258

259 S.I., H.L., A.Z., B.R.S., A.C., and D.D.H. are inventors on a patent application submitted based on

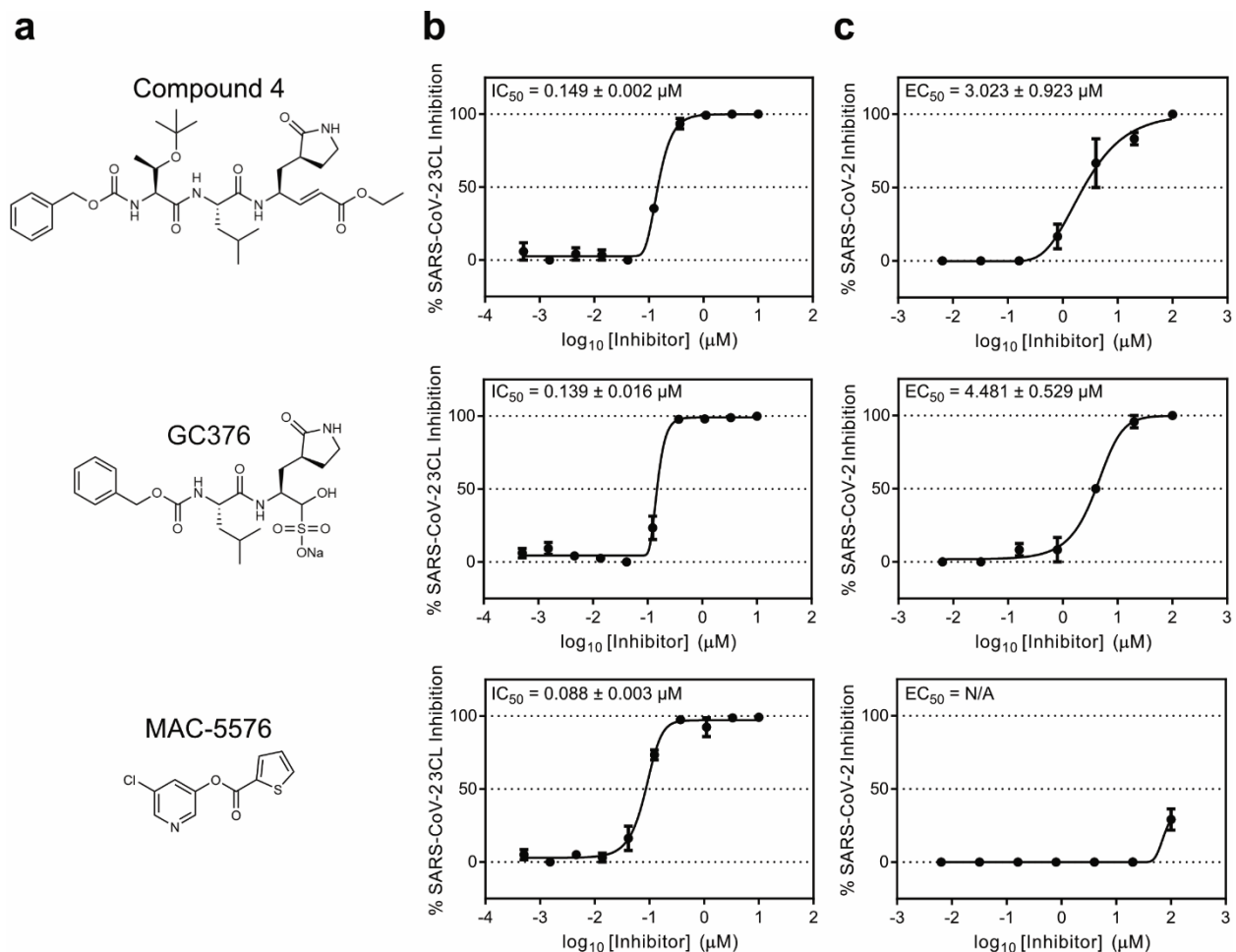
260 this work. B.R.S. is an inventor on additional patents and patent applications related to small

261 molecule therapeutics, and co-founded and serves as a consultant to Inzen Therapeutics and

262 Nevrox Limited.

263

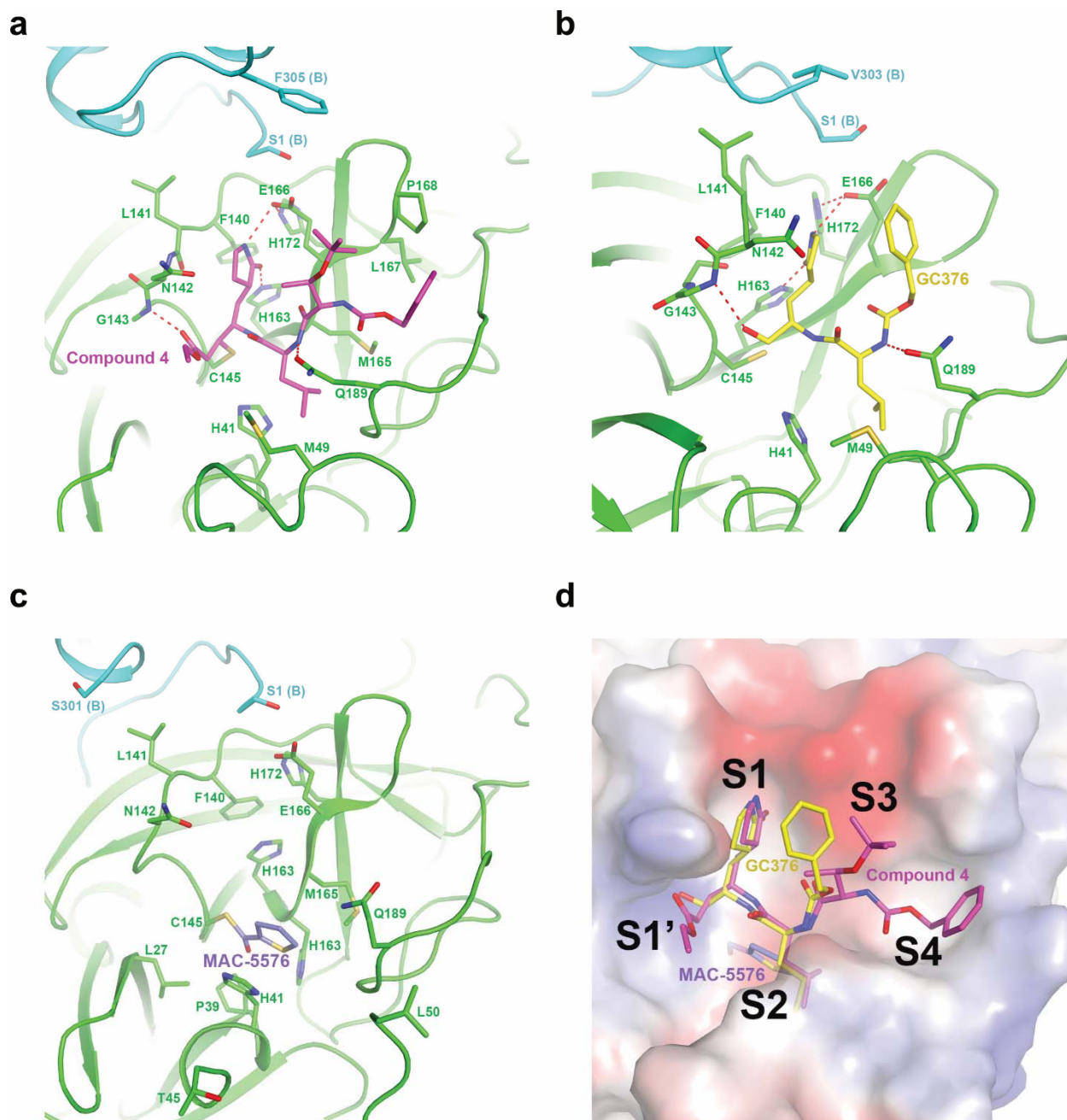
264 **Fig 1. Inhibition of SARS-CoV-2 3CL protease by compound 4, GC376, and MAC-5576.** **a,**
265 Chemical structures of the three compounds in this study. **b,** Inhibition of purified native SARS-CoV-
266 2 3CL protease by each compound. **c,** Inhibition of SARS-CoV-2 viral replication by each compound.
267 Data are shown as mean \pm s.e.m. for two or three technical replicates. IC₅₀ and EC₅₀ values are
268 indicated as mean \pm s.d. for two biological replicates.
269



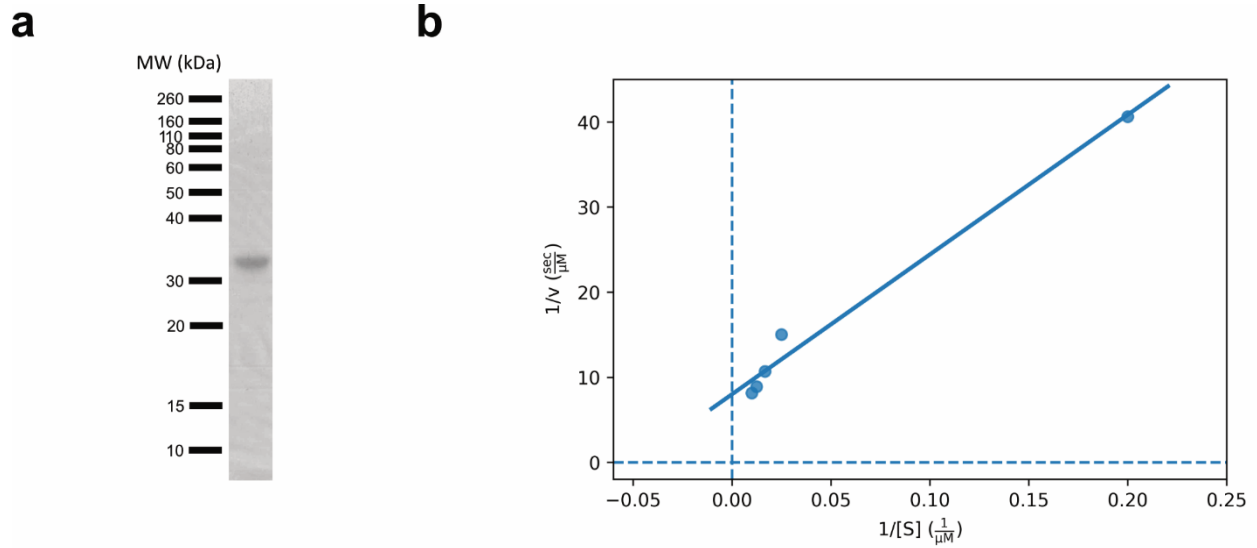
270

271

272 **Fig 2. Crystal structures of inhibitors in complex with the SARS-CoV-2 3CL protease. a-c,**
273 Structure of compound 4 (a), GC376 (b), or MAC-5576 (c) bound to SARS-CoV-2 3CL. Protomer A
274 is denoted in green and protomer B is denoted in cyan. **d,** Overlay of all three compounds in the
275 substrate binding pocket of the 3CL protease.
276



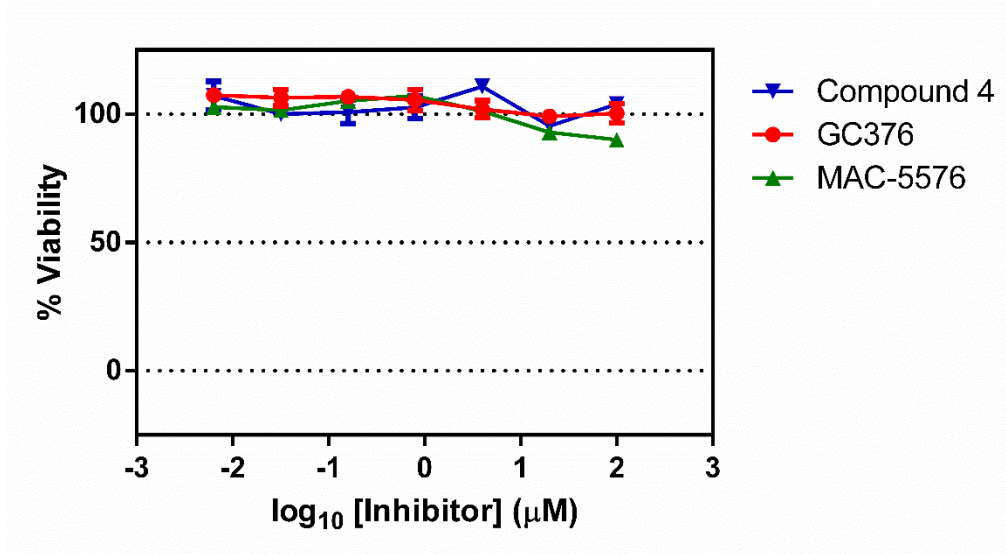
278 **Extended Data Fig. 1. Production of native SARS-CoV-2 3CL protease in E. coli.** **a**, The purified
279 protease was ran on SDS-PAGE to confirm size and purity. **b**, Confirmation of enzymatic activity of
280 SARS-CoV-2 3CL protease by quantification of cleavage of a fluorogenic peptide substrate.
281



282

283

284 **Extended Data Fig. 2. Cytotoxicity of compounds in Vero-E6 cells.** Vero-E6 cells were incubated
285 with serial dilutions of each compound for 48 h and then tested for cytotoxicity. Data are shown as
286 mean \pm s.e.m. for three experimental replicates.
287



288

289

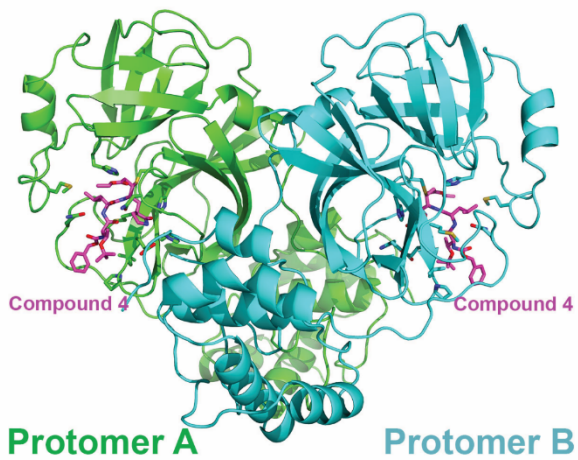
290 **Extended Data Fig. 3. Overall crystal structures of inhibitors in complex with the SARS-CoV-2**

291 **3CL protease dimer. a-c, Structures of compound 4 (a), GC376 (b), or MAC-5576 (c) bound to the**

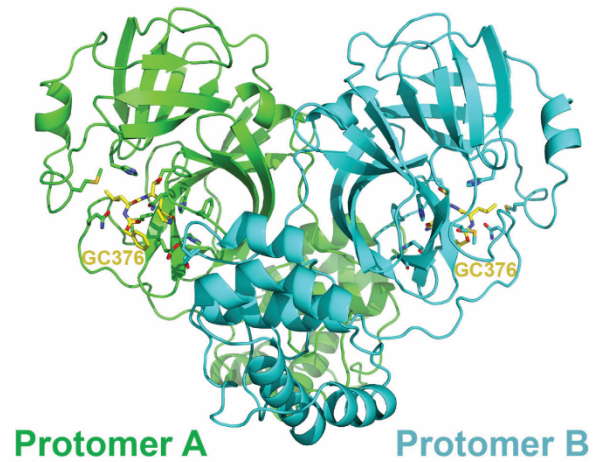
292 3CL dimer.

293

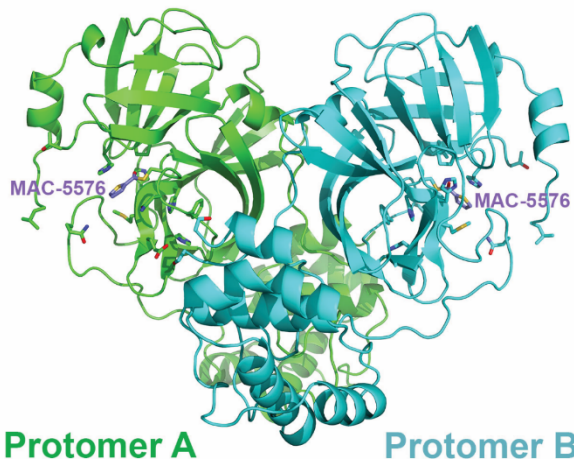
a



b

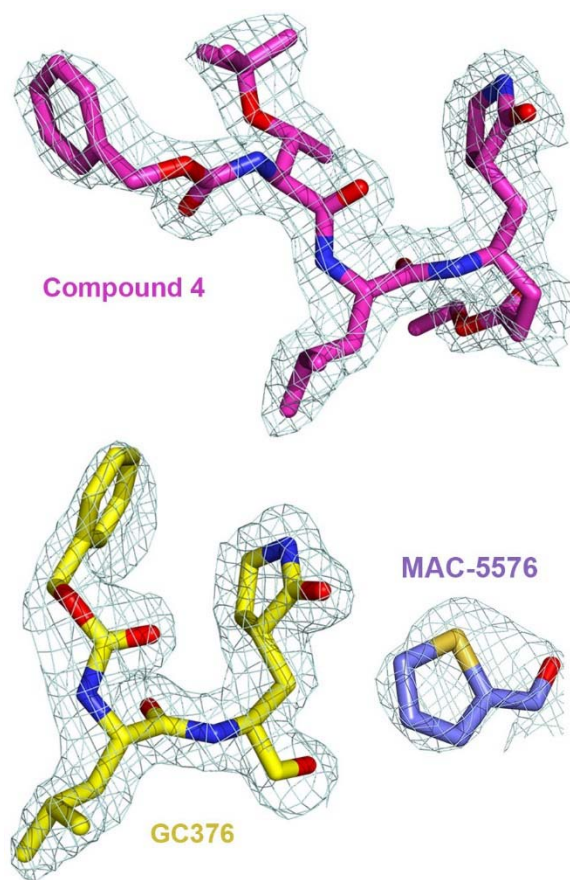


c



294

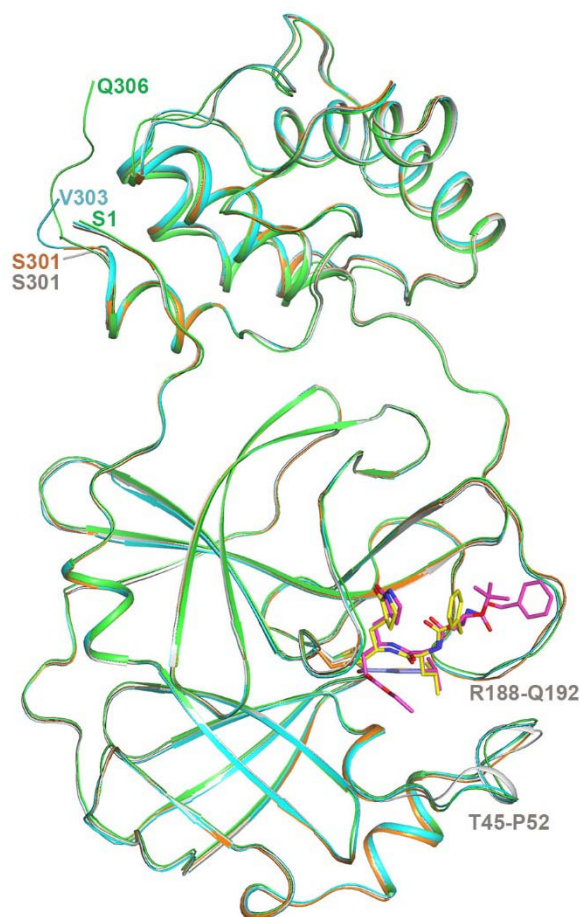
295 **Extended Data Fig. 4. 2Fo–Fc electron density map at 1 σ for the inhibitors in the crystal**
296 **structures.** Electron density mesh (light cyan) for compound 4, GC376, and MAC-5576 in complex
297 with 3CL.
298



299

300

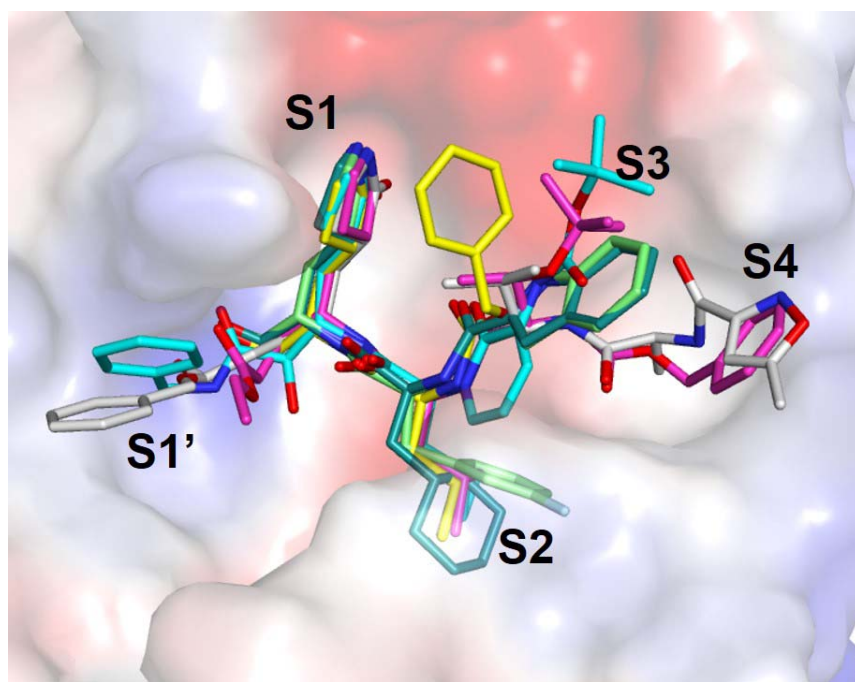
301 **Extended Data Fig. 5. Overlay of the four crystal structures of 3CL.** Crystal structure of 3CL in
302 *apo* form (gray), in complex with, compound 4 (green for 3CL and magenta for compound 4), GC376
303 (cyan for 3CL and yellow for GC376), and MAC-5576 (orange for 3CL and purple for MAC-5576
304 fragment). One protomer for each structure is shown, with the inhibitors shown with stick models.
305 The terminal residue of each structure, as well as two stretches of residues near the binding site that
306 exhibit local conformational change between the *apo* and inhibitor-bound structures are labeled.
307



308

309

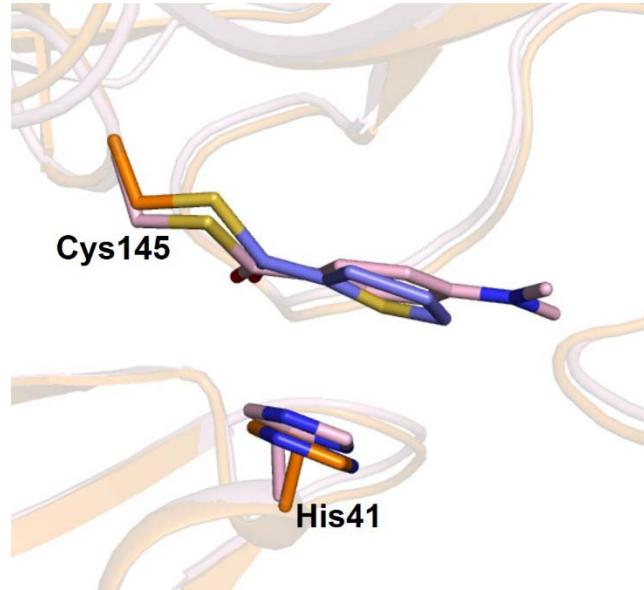
310 **Extended Data Fig. 6. Comparison of the binding modes of compound 4 and GC376 with**
311 **other peptide-like inhibitors.** Compound 4 (magenta) and GC376 (yellow) were overlaid with
312 previously identified compounds, compound 13b (cyan, PDB: 6Y2F), compound 11a (dark green,
313 PDB: 6LZE), compound 11b (light green, PDB: 6M0K), and N3 (white; PDB: 7BQY).
314



315

316

317 **Extended Data Fig. 7. Comparison of the binding modes of MAC-5576 with XP-59.** MAC-5576
318 (purple) bound to the SARS-CoV-2 3CL protease (orange) was overlaid with XP-59 bound to the
319 SARS-CoV-1 3CL protease (pink, PDB: 2V6N).



320

321

322 **Supplemental Table 1. Data collection and refinement statistics of apo SARS-CoV-2 3CL**
 323 **protease and 3CL protease bound to compound 4, GC376, and MAC-5576.**

324

	<i>apo</i> 3CL	3CL with Compound 4	3CL with GC376	3CL with MAC-5576
Data collection				
Space group	<i>C</i> 2	<i>C</i> 2	<i>C</i> 2	<i>C</i> 2
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	98.7, 82.0, 51.8	97.2, 81.9, 54.2	98.8, 80.2, 52.0	98.3, 82.5, 51.8
α , β , γ (°)	90, 114.9, 90	90, 117.1, 90	90, 114.3, 90	90, 114.83, 90
Resolution (Å)	60.4-1.85 (1.88-1.85)*	59.5-1.80 (1.82-1.80)*	59.9-1.83 (1.86-1.83)*	60.58-1.73 (1.76-1.73)*
R_{merge} (%)	7.2 (65.7)	18.7 (57.6)	14.7 (60.5)	4.1 (68.8)
$I/\sigma I$	13.4 (2.1)	10.2 (2.4)	12.7 (2.2)	23.5 (2.3)
Completeness (%)	98.9 (97.4)	98.8 (99.0)	98.5 (96.0)	99.2 (88.3)
Redundancy	6.8 (6.5)	6.7 (6.5)	6.9 (6.1)	6.8 (6.6)
CC1/2	99.8 (95.1)	0.99 (0.91)	0.99 (0.93)	100 (0.90)
Refinement				
Resolution (Å)	47.0-1.85 (1.88-1.85)*	48.2-1.80 (1.82-1.80)*	47.38-1.83 (1.86-1.83)	47.0-1.73 (1.75-1.73)
No. reflections	31,710 (3,219)	34,294 (3,497)	31,972 (3,134)	38,879 (3,899)
$R_{\text{work}}/R_{\text{free}}$ (%)	16.8 (25.2)/19.8 (29.3)	18.7 (25.7)/22.3 (30.7)	17.4 (25.0)/20.6 (28.7)	16.6 (26.8)/19.0 (27.6)
No. atoms				
Protein	2,329	2,370	2,340	2,298
Ligand/ion	5	44	28	16
Water	128	293	163	245
B-factors				
Protein	50.0	35.2	41.4	40.5
Ligand/ion	52.6	36.6	42.4	43.7
Water	53.5	46.7	48.2	50.7
R.m.s deviations				
Bond lengths (Å)	0.006	0.007	0.006	0.006
Bond angles (°)	0.8	0.8	0.8	0.8

325 *Highest resolution shell is shown in parenthesis.

326

327

328 **References (limit of 20)**

329

330 1 Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature*
331 **579**, 265-269, doi:10.1038/s41586-020-2008-3 (2020).

332 2 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat
333 origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).

334 3 Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 - Preliminary Report. *N Engl J*
335 *Med*, doi:10.1056/NEJMoa2007764 (2020).

336 4 Wang, Y. *et al.* Remdesivir in adults with severe COVID-19: a randomised, double-blind,
337 placebo-controlled, multicentre trial. *Lancet* **395**, 1569-1578, doi:10.1016/S0140-
338 6736(20)31022-9 (2020).

339 5 Jin, Z. *et al.* Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. *Nature*,
340 doi:10.1038/s41586-020-2223-y (2020).

341 6 Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design
342 of improved alpha-ketoamide inhibitors. *Science* **368**, 409-412, doi:10.1126/science.abb3405
343 (2020).

344 7 Yang, S. *et al.* Synthesis, crystal structure, structure-activity relationships, and antiviral
345 activity of a potent SARS coronavirus 3CL protease inhibitor. *J Med Chem* **49**, 4971-4980,
346 doi:10.1021/jm0603926 (2006).

347 8 Kim, Y. *et al.* Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses,
348 noroviruses, and coronaviruses. *J Virol* **86**, 11754-11762, doi:10.1128/JVI.01348-12 (2012).

349 9 Blanchard, J. E. *et al.* High-throughput screening identifies inhibitors of the SARS
350 coronavirus main proteinase. *Chem Biol* **11**, 1445-1453, doi:10.1016/j.chembiol.2004.08.011
351 (2004).

352 10 Dai, W. *et al.* Structure-based design of antiviral drug candidates targeting the SARS-CoV-2
353 main protease. *Science*, doi:10.1126/science.abb4489 (2020).

- 354 11 Verschueren, K. H. *et al.* A structural view of the inactivation of the SARS coronavirus main
355 proteinase by benzotriazole esters. *Chem Biol* **15**, 597-606,
356 doi:10.1016/j.chembiol.2008.04.011 (2008).
- 357 12 Luft, J. R. *et al.* A deliberate approach to screening for initial crystallization conditions of
358 biological macromolecules. *J Struct Biol* **142**, 170-179, doi:10.1016/s1047-8477(03)00048-0
359 (2003).
- 360 13 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta*
361 *Crystallogr D Biol Crystallogr* **66**, 133-144, doi:10.1107/S0907444909047374 (2010).
- 362 14 Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol*
363 *Crystallogr* **66**, 22-25, doi:10.1107/S0907444909042589 (2010).
- 364 15 McRee, D. E. XtalView/Xfit--A versatile program for manipulating atomic coordinates and
365 electron density. *J Struct Biol* **125**, 156-165, doi:10.1006/jsbi.1999.4094 (1999).
- 366 16 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot.
367 *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
- 368 17 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular
369 structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221,
370 doi:10.1107/S0907444909052925 (2010).
- 371 18 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of
372 nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**,
373 10037-10041, doi:10.1073/pnas.181342398 (2001).
- 374 19 Zhai, Y. *et al.* Cyanohydrin as an Anchoring Group for Potent and Selective Inhibitors of
375 Enterovirus 71 3C Protease. *J Med Chem* **58**, 9414-9420,
376 doi:10.1021/acs.jmedchem.5b01013 (2015).