Structure-based drug design, virtual screening and high-throughput screening rapidly identify antiviral leads targeting COVID-19

Zhenming Jin[1,2,9], Xiaoyu Du[2,9], Yechun Xu[3,9], Yongqiang Deng[4,9], Meiqin Liu[5,9], Yao Zhao[1], Bing Zhang[1], Xiaofeng Li[4], Leike Zhang[5], Yinkai Duan[1], Jing Yu[1], Lin Wang[1], Kailin Yang[6], Fengjiang Liu[1], Tian You[1], Xiaoce Liu[1], Xiuna Yang[1], Fang Bai[1], Hong Liu[3], Xiang Liu[7], Luke W. Guddat[8], Gengfu Xiao[5], Chengfeng Qin[4], Zhengli Shi[5], Hualiang Jiang[1,3*], Zihe Rao[1,2,7*] & Haitao Yang[1*]

[1]Shanghai Institute for Advanced Immunochemical Studies and School of Life Science and Technology, ShanghaiTech University, Shanghai, China.

[2]School of Life Sciences, Tsinghua University, Beijing, China.

[3]Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China.

[4]Department of Virology, State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences, Beijing, China.

[5]CAS Key Laboratory of Special Pathogens, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, China.

[6]Taussig Cancer Center, Cleveland Clinic, Cleveland, USA.

[7]State Key Laboratory of Medicinal Chemical Biology, Frontiers Science Center for Cell Response, College of Life Sciences, College of Pharmacy, Nankai University, Tianjin, China.

[8]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia.

[9]These authors contributed equally: Zhenming Jin, Xiaoyu Du, Yechun Xu, Yongqiang Deng, Meiqin Liu.

*e-mail: hljiang@simm.ac.cn; raozh@tsinghua.edu.cn; yanght@shanghaitech.edu.cn

Abstract

A coronavirus identified as 2019 novel coronavirus (COVID-19) is the etiological agent responsible for the 2019-2020 viral pneumonia outbreak that commenced in Wuhan[1-4]. Currently there is no targeted therapeutics and effective treatment options remain very limited. In order to rapidly discover lead compounds for clinical use, we initiated a program of combined structure-assisted drug design, virtual drug screening, and high-throughput screening to identify new drug leads that target the COVID-19 main protease (M^{pro}). M^{pro} is a key coronavirus enzyme, which plays a pivotal role in mediating viral replication and transcription, making it an attractive drug target for this virus[5,6]. Here, we identified a mechanism-based inhibitor, N3, by computer-aided drug design and subsequently determined the crystal structure of COVID-19 M^{pro} in complex with this compound. Next, through a combination of structure-based virtual and high-throughput screening, we assayed over 10,000 compounds including approved drugs, drug candidates in clinical trials, and other pharmacologically active compounds as inhibitors of M^{pro}. Seven of these inhibit M^{pro} with IC_{50} values ranging from 0.48 to 16.62 μM. Ebselen, thiadiazolidinone-8 (TDZD-8) and N3 also exhibited strong antiviral activity in cell-based assays. Our results demonstrate the efficacy of this screening strategy, and establishes a new paradigm for the rapid discovery of drug leads with clinical potential in response to new infectious diseases where no specific drugs or vaccines are available.

An epidemic of contagious pneumonia has been spreading rapidly from its likely origin in Central China since December 2019[1,2]. The infectious pneumonia is now found in all provinces of China, with cases also confirmed in other 35 countries. As of February 25, 2020, around 80,000 people were infected and over 2,500 have succumbed to the epidemic[7,8]. Symptoms associated with this disease include fever, cough, myalgia or fatigue and dyspnea[9,10]. The World Health Organization announced that the outbreak of this deadly and fast-spreading infectious pneumonia constitutes a global health emergency[8]. A coronavirus named 2019 novel coronavirus (COVID-19), has been identified as the etiological agent responsible for the outbreak[3,4]. At present, there are no targeted therapeutics and effective treatment options remain very limited.

CoVs infect humans and other animal species, causing a variety of highly prevalent and severe diseases, including Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS)[11]. CoVs are positive-sense, single-stranded RNA viruses featuring the largest viral RNA genomes known to date[12]. The COVID-19 genome is comprised of ~30,000 nucleotides; its replicase gene encodes two overlapping polyproteins, pp1a and pp1ab, required for viral replication and transcription[3,4]. The functional polypeptides are released from the polyproteins by extensive proteolytic processing, predominantly by a 33.8-kDa main protease (M^{pro}), also referred to as the 3C-like protease. M^{pro} digests the polyprotein at no less than 11 conserved sites, starting with the autolytic cleavage of this enzyme itself from pp1a and pp1ab[5]. The functional importance of M^{pro} in the viral life cycle, together with the absence of closely related homologues in humans, identify the M^{pro} as an attractive target for antiviral drug design[6].

To facilitate the rapid discovery of antiviral compounds with clinical potential, we

71 developed a strategy combining structure-assisted drug design, virtual drug screening and
72 high-throughput screening to repurpose existing drugs to target M^{pro}, which is a key
73 enzyme for the viral life cycle. Our strategy has successfully identified several new
74 compounds that showed both strong inhibition of the target enzyme and antiviral activity
75 in cell-based assays. These studies pave the way for the further development of novel
76 antiviral drugs against this CoV.

77 **Establishing a high-throughput activity assay**

78 Recombinant COVID-19 M^{pro} with native N and C termini was expressed in *Escherichia*
79 *coli* and subsequently purified (Extended Data Fig. 1a, b). The molecular weight of
80 COVID-19 M^{pro} as determined by mass spectroscopy is 33797.0 Da, consistent with its
81 theoretical molecular weight 33796.8 Da. In order to characterize its enzymatic activity
82 and to carry out high throughput screening of inhibitors, we developed a fluorescence
83 resonance energy transfer (FRET) assay. To do this, a fluorescently labeled substrate,
84 MCA-AVLQ↓SGFR-Lys(Dnp)-Lys-NH_2, derived from the N-terminal auto-cleavage
85 sequence of the viral protease was designed and synthesized for time-dependent kinetic
86 analysis (Extended Data Fig. 1e). The catalytic efficiency (k_{cat}/K_m) for COVID-19 M^{pro} was
87 measured to be 28,500 $M^{-1}s^{-1}$ which is slightly higher than that for SARS-CoV M^{pro}
88 (k_{cat}/K_m=26,500 $M^{-1}s^{-1}$)[13], but over 30-fold higher than that of human rhinovirus 3C
89 protease (k_{cat}/K_m=920 $M^{-1}s^{-1}$)[14].

90 **N3 is a potent irreversible inhibitor of COVID-19 M^{pro}**

91 In a previous study, we designed a Michael acceptor inhibitor N3 using computer-aided
92 drug design (CADD) (Extended Data Fig. 1c), which can specifically inhibit multiple CoV
93 M^{pro}s, including those from SARS-CoV and MERS-CoV[15-18]. It also has displayed potent

antiviral activity against infectious bronchitis virus in an animal model[16]. Next, we constructed a homology model for COVID-19 M^{pro} and used molecular docking to see if N3 could target this new CoV M^{pro}. A docking pose showed that it could fit inside the substrate-binding pocket. To assess the efficacy of N3 for COVID-19 M^{pro} , kinetic analysis was performed[15]. A progress curve showed that it is time-dependent irreversible inhibitor of this enzyme. Further, the shape of this curve supports the mechanism of two-step irreversible inactivation. The inhibitor first associates with COVID-19 M^{pro} (EI) with a dissociation constant K_i; then, a stable covalent bond is formed between N3 and M^{pro} (E−I). The evaluation of this time-dependent inhibition requires both the equilibrium-binding constant K_i (designated as k_2/k_1) and the inactivation rate constant for covalent bond formation k_3 (Extended Data Fig. 1d). However, N3 exhibits very potent inhibition of COVID-19 M^{pro} (Extended Data Fig. 1e), such that measurement of K_i and k_3 proved difficult. When very rapid inactivation occurs, k_{obs}/[I] was utilized to evaluate the inhibition as an approximation of the pseudo second-order rate constant (k_3/K_i)[15]. The value of k_{obs}/[I] of N3 for COVID-19 M^{pro} was determined to be 11,300 ± 880 $M^{-1}s^{-1}$, suggesting this Michael acceptor has potent inhibition.

The crystal structure of COVID-19 M^{pro} in complex with N3

In order to elucidate the inhibitory mechanism of this compound, we determined the crystal structure of COVID-19 M^{pro} in complex with N3 to 2.1-Å resolution. The asymmetric unit contains only one polypeptide (Extended Data Table 1). However, two of these associate to form a dimer by a crystallographic 2-fold symmetry axis (the two molecules are designated protomer A and B) (Fig. 1b). All residues (residues 1–306) are visible in electron density maps. Each protomer is composed of three domains (Fig. 1a). Domains I (residues

8–101) and II (residues 102–184) have an antiparallel β-barrel structure. Domain III (residues 201–303) contains five α-helices arranged into a largely antiparallel globular cluster, and is connected to domain II by means of a long loop region (residues 185–200). COVID-19 M^{pro} has a Cys–His catalytic dyad, and the substrate-binding site is located in a cleft between Domain I and II. These features are similar to those of other M^{pro}s reported previously[5,16,18-20]. The electron density map shows that N3 binds in the substrate-binding pocket in an extended conformation (Fig. 1c, Extended Data Fig. 2), with the inhibitor backbone atoms forming an antiparallel sheet with residues 164–168 of the long $strand_{155-168}$ on one side, and with residues 189–191 of the loop linking domains II and III.

Here we detail the specific interactions of N3 with M^{pro} (Fig. 1c). The electron density shows that the Sγ atom of C145-A forms a covalent bond (1.8-Å) with the Cβ of the vinyl group, confirming that the Michael addition has occurred (Fig. 1c, d). The S1 subsite has an absolute requirement for Gln at the P1 position. The side chains of F140-A, N142-A, E166-A, H163-A, H172-A, S1-B (from protomer B), and main chains of F140-A and L141-A are involved in S1 subsite formation, which also includes two ordered water molecules (named W1 and W2). The lactam at P1 inserts into the S1 subsite and forms a hydrogen bond with H163-A. The side chain of Leu at P2 site deeply inserts into the hydrophobic S2 subsite, which consists of the side chains of H41-A, M49-A, Y54-A, M165-A, and the alkyl portion of the side chain of D187-A. The side chain of Val at P3 is solvent-exposed, indicating that this site can tolerate a wide range of functional groups. The side chain of Ala at P4 side is surrounded by the side chains of M165-A, L167-A, F185-A, Q192-A and the main chain of Q189-A, all of which form a small hydrophobic pocket. P5 makes van der Waals contacts with P168-A and the backbone of residues 190–191. The bulky benzyl

group extends into the S1′ site, possibly forming van der Waals interactions with T24-A and T25-A. In addition, N3 forms multiple hydrogen bonds with the main chain of the residues in the substrate-binding pocket, which also helps lock the inhibitor inside the substrate-binding pocket (Fig. 1c, Extended Data Fig. 2).

Virtual screening

The structure of COVID-19 M^{pro} in complex with N3 provides a model for rapidly identifying lead inhibitors to target COVID-19 M^{pro} through *in silico* screening. To achieve this, an in-house database of potential binding compounds was docked using Glide (version 8.2) of Schrödinger 2019-1. The results show that cinanserin fits snugly into the substrate-binding pocket, by interacting with H41 and E166 of M^{pro} by cation-π. Subsequently we determined this compound has an IC_{50} value of 124.93 μM for M^{pro}, which validated our virtual screening approach (Extended Data Fig. 3). Moreover, cinanserin is a well-characterized serotonin antagonist, which has undergone preliminary clinical testing in humans in the 1960s[21] and has previously been shown to inhibit SARS-CoV M^{pro} [22]. Thus, it has potential for further optimization as an anti-viral drug lead.

High-throughput screening

Next, we used our FRET assay, to screen a library of ~10,000 compounds consisting of approved drugs, clinical trial drug candidates and natural products. Primary hits included seven compounds that are either FDA-approved drugs or clinical trial/preclinical drug candidates. We then determined their IC_{50} values, which are in the range from 0.48 to 16.62 μM (Fig. 2). Amongst them, disulfiram and carmofur are FDA-approved drugs, whilst ebselen, TDZD-8, shikonin, tideglusib, and PX-12 are currently in clinical trials or preclinical studies. Ebselen has the strongest inhibition of M^{pro} activity with an IC_{50} of 0.48

163 μM. Since our structural data is only for N3, we investigated if molecular docking could
164 predict how these inhibitors bind to this protein. In all cases, reasonable docking poses
165 were found, demonstrating that they could fit inside the substrate-binding pocket (Extended
166 Data Fig. 4).

167 **Antiviral activity assay**

168 To further substantiate the enzymatic inhibition results *in vitro*, we then evaluated the
169 ability of these compounds to prevent cells from being infected by COVID-19. As shown
170 in Fig. 3, ebselen, TDZD-8 and N3 showed the most potent antiviral effects at a
171 concentration of 10 μM, entirely preventing the cytopathic effect (CPE) in COVID-19-
172 infected Vero cells. Quantitative real-time RT-PCR (qRT-PCR) demonstrated that treatment
173 of ebselen reduced the amounts of COVID-19 20.3-fold compared with the treatment in
174 the absence of inhibitor, and TDZD-8 and N3 exhibited 10.9-fold and 8.4-fold reduction in
175 COVID-19 growth, respectively. Ebselen is an organoselenium compound with anti-
176 inflammatory, anti-oxidant and cytoprotective properties. This compound has been
177 investigated for the treatment of multiple diseases, such as bipolar disorders and hearing
178 loss. The toxicity of ebselen in humans has been evaluated in a number of clinical trials[23-
179 25]. These data strongly suggest the clinical potential of ebselen for CoV treatment. TDZD-
180 8 was originally designed as a lead compound for the treatment of Alzheimer's disease,
181 which has been shown to be effective in animal models[26]. It is also interesting that
182 cinanserin displayed moderate inhibition against COVID-19 with an IC_{50} value of 20.61
183 μM (Extended Data Fig. 3), which is superior to that in the enzymatic inhibition assay,
184 suggesting that cinanserin might have multi-drug targets in preventing viral infection.

185 **Discussion**

186 In the last twenty years, new infectious agents have emerged to cause epidemics, such as
187 SARS and MERS[11]. The timely development of effective antiviral agents for clinical use
188 is extremely challenging because conventional drug development approaches normally
189 take years of investigations and cost billions of dollars [27]. The repurposing of approved
190 pharmaceutical drugs and drug candidates provides an alternative approach to rapidly
191 identify potential drug leads to manage rapidly emerging viral infections[28]. Cell-based
192 phenotypic screening has proven to be valuable[29], but the corresponding methodological
193 complexity is not readily compatible with high-throughput pipelines, and fails to inform
194 the identity of the molecular target or mechanism of action[30]. In this regard, we combined
195 structure-based *ab initio* drug design, virtual screening and high-throughput screening to
196 discover new drugs for treating COVID-19 infection. The convergence of these three
197 approaches proved to be an efficient strategy in finding three lead compounds with good
198 antiviral activity. Ebselen is of particular interest for further investigation since its safety
199 in humans has been supported by different clinical trials[23-25]. Our methodology establishes
200 a paradigm of rapid discovering drug leads with clinical potential in response to new
201 emerging infectious diseases lacking specific drugs and vaccines.

202 The crystal structural and docking data have shown that these drug leads can bind to the
203 substrate-binding pocket of COVID-19 M^{pro}. In our previous study, we proposed that all
204 the CoV M^{pro}s share a highly conservative substrate-recognition pocket, which serves as a
205 drug target for broad-spectrum inhibitor design[15]. Recent discovery of new CoVs and
206 accumulation of structural data for CoV M^{pro}s from various species provided the
207 opportunity to further examine our previous hypothesis. Superposition of the 12 crystal

structures of M^{pro}s[15,16,18,20,31-35] have shown that the most variable regions were the helical domain III and surface loops, but the substrate-binding pockets located in a cleft between domains I and II are still highly conserved among all CoV M^{pro}s, suggesting the antiviral inhibitors targeting at this site should have wide-spectrum anti-CoV activity (Fig. 4). This strongly supports our hypothesis that development of a single antiviral agent could provide an effective first line of defense against all CoV-associated diseases.

References

1 Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine* (2020).

2 Qun Li *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine* (2020).

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

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

5 Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. Coronavirus main proteinase (3CL(pro)) structure: Basis for design of anti-SARS drugs. *Science* **300**, 1763-1767, doi:10.1126/science.1085658 (2003).

6 Pillaiyar, 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. *Journal of Medicinal Chemistry* **59**, 6595-6628, doi:10.1021/acs.jmedchem.5b01461 (2016).

7 National Health Commission, PRC. *Novel Coronavirus(2019-nCoV) Situation Report of China*, <http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml> (2020).

8 WHO. Novel Coronavirus(2019-nCoV) Situation Report. (2020).

9 Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, doi:10.1016/s0140-6736(20)30183-5 (2020).

10 Michelle L. Holshue, C. D., Scott Lindquist, Kathy H. Lofy, John Wiesman, Hollianne Bruce, Christopher Spitters, for the Washington State 2019-nCoV Case Investigation Team. First Case of 2019 Novel Coronavirus in the United States. (2020).

11 de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nature Reviews Microbiology* **14**, 523-534, doi:10.1038/nrmicro.2016.81 (2016).

12 Perlman, S. & Netland, J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nature Reviews Microbiology* **7**, 439-450, doi:10.1038/nrmicro2147 (2009).

13 Xue, X. Y. *et al.* Production of authentic SARS-CoV M-pro with enhanced activity: Application as a novel tag-cleavage endopeptidase for protein overproduction. *Journal of Molecular Biology* **366**, 965-975, doi:10.1016/j.jmb.2006.11.073 (2007).

14 Wang, Q. M., Johnson, R. B., Cox, G. A., Villarreal, E. C. & Loncharich, R. J. A Continuous Colorimetric Assay for Rhinovirus-14 3C Protease Using Peptidep-Nitroanilides as Substrates. *Analytical biochemistry* **252**, 238-245 (1997).

15 Yang, H. T. *et al.* Design of wide-spectrum inhibitors targeting coronavirus main proteases. *Plos Biology* **3**, 2044-2044, doi:10.1371/journal.pbio.0030428 (2005).

16 Xue, X. Y. *et al.* Structures of two coronavirus main proteases: Implications for substrate binding

and antiviral drug design. *Journal of Virology* **82**, 2515-2527, doi:10.1128/jvi.02114-07 (2008).

17 Ren, Z. L. *et al.* The newly emerged SARS-Like coronavirus HCoV-EMC also has an "Achilles' heel": current effective inhibitor targeting a 3C-like protease. *Protein & Cell* **4**, 248-250, doi:10.1007/s13238-013-2841-3 (2013).

18 Wang, F. H. *et al.* Michael Acceptor-Based Peptidomimetic Inhibitor of Main Protease from Porcine Epidemic Diarrhea Virus. *Journal of Medicinal Chemistry* **60**, 3212-3216, doi:10.1021/acs.jmedchem.7b00103 (2017).

19 Yang, H. T. *et al.* The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 13190-13195, doi:10.1073/pnas.1835675100 (2003).

20 Zhao, Q. *et al.* Structure of the main protease from a global infectious human coronavirus, HCoV-HKU1. *Journal of virology* **82**, 8647-8655, doi:10.1128/JVI.00298-08 (2008).

21 Rubin, B. & Waugh, M. H. Antiphlogistic effects of antiserotonin (SQ 10,643) and aminopyrine in rats versus endotoxin and other agents. *Proceedings of the Society for Experimental Biology and Medicine* **119**, 438-443 (1965).

22 Chen, L. *et al.* Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication in vitro. *Journal of virology* **79**, 7095-7103 (2005).

23 Lynch, E. & Kil, J. in *Seminars in Hearing.* 047-055 (© Thieme Medical Publishers).

24 Masaki, C. *et al.* Effects of the potential lithium-mimetic, ebselen, on impulsivity and emotional processing. **233**, 2655-2661, doi:10.1007/s00213-016-4319-5 (2016).

25 Kil, J. *et al.* Safety and efficacy of ebselen for the prevention of noise-induced hearing loss: a randomised, double-blind, placebo-controlled, phase 2 trial. *The Lancet* **390**, 969-979 (2017).

26 Koehler, D., Shah, Z. A. & Williams, F. E. The GSK3β inhibitor, TDZD-8, rescues cognition in a zebrafish model of okadaic acid-induced Alzheimer's disease. *Neurochemistry international* **122**, 31-37 (2019).

27 Bedford, J. *et al.* A new twenty-first century science for effective epidemic response. *Nature* **575**, 130-136, doi:10.1038/s41586-019-1717-y (2019).

28 Farha, M. A. & Brown, E. D. Drug repurposing for antimicrobial discovery. *Nature Microbiology* **4**, 565-577, doi:10.1038/s41564-019-0357-1 (2019).

29 Xu, M. *et al.* Identification of small-molecule inhibitors of Zika virus infection and induced neural cell death via a drug repurposing screen. *Nature Medicine* **22**, 1101-1107, doi:10.1038/nm.4184 (2016).

30 Aulner, N., Danckaert, A., Ihm, J., Shum, D. & Shorte, S. L. Next-Generation Phenotypic Screening in Early Drug Discovery for Infectious Diseases. *Trends in parasitology* (2019).

31 Wang, F., Chen, C., Tan, W., Yang, K. & Yang, H. Structure of Main Protease from Human Coronavirus NL63: Insights for Wide Spectrum Anti-Coronavirus Drug Design. *Sci Rep* **6**, 22677-22677, doi:10.1038/srep22677 (2016).

32 Lee, C.-C. *et al.* Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds. *J Biol Chem* **284**, 7646-7655, doi:10.1074/jbc.M807947200 (2009).

33 Wang, F. H. *et al.* Crystal Structure of Feline Infectious Peritonitis Virus Main Protease in Complex with Synergetic Dual Inhibitors. *Journal of Virology* **90**, 1910-1917, doi:10.1128/jvi.02685-15 (2016).

34 Cui, W. *et al.* The crystal structure of main protease from mouse hepatitis virus A59 in complex with an inhibitor. *Biochemical and Biophysical Research Communications* **511**, 794-799, doi:10.1016/j.bbrc.2019.02.105 (2019).

35 Needle, D., Lountos, G. T. & Waugh, D. S. Structures of the Middle East respiratory syndrome coronavirus 3C-like protease reveal insights into substrate specificity. *Acta Crystallographica Section D-Structural Biology* **71**, 1102-1111, doi:10.1107/s1399004715003521 (2015).

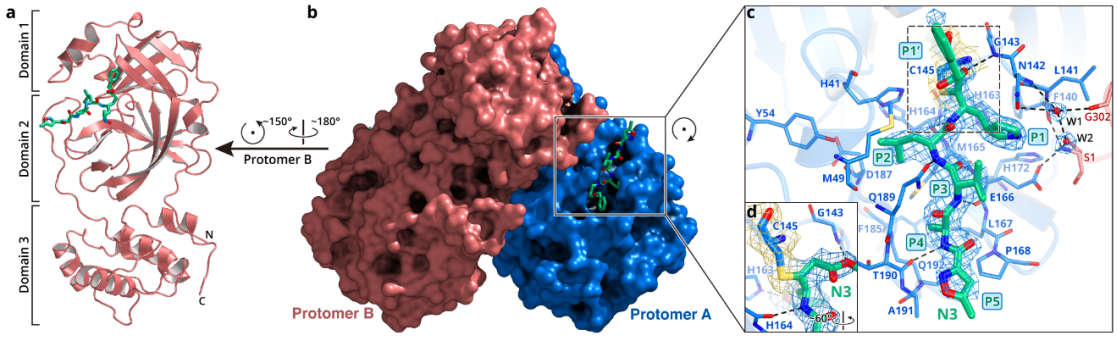


Fig. 1 | The crystal structure of COVID-19 M^{pro} in complex with N3.

a, Cartoon representation of one protomer of the dimeric M^{pro}-inhibitor complex. N3 is shown in green sticks. **b,** Surface representation of the homodimer of M^{pro}. Protomer A is in blue while Protomer B in salmon; N3 is presented as green sticks. **c,** A zoomed view of the substrate-binding pocket. The key residues forming the binding pocket are shown in sticks; the two waters, assigned as W1 and W2, are shown as red spheres; the $2F_o$-F_c density map contoured at 1.2 σ is shown around N3 molecule (blue mesh), C145-A (yellow mesh), and the two waters (blue mesh). **d,** The C-S covalent bond.

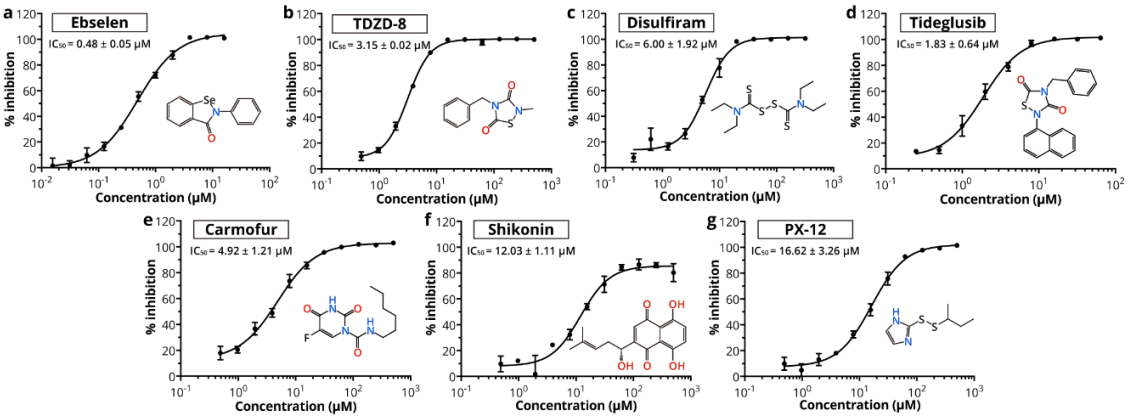


Fig. 2 | Plots showing the inhibitory activity for the seven drug leads. The hydrolytic activity of COVID-19 M^{pro} is measured in the presence of varying concentrations of compounds. Curves represent best fits for calculating the IC_{50} values, error bars represent mean ± SEM based on three independent measurements.

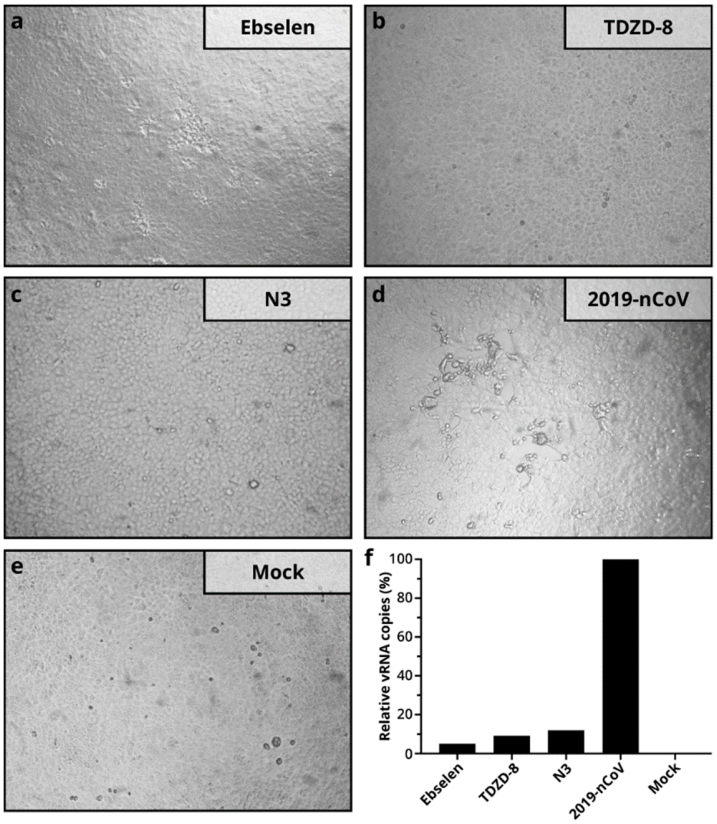


Fig. 3 | Antiviral activity assay. Virus-induced cytopathic effect in Vero cells in the presence of M^{Pro} inhibitors. **a-c,** Treated with Ebselen, TDZD-8 and N3 respectively. **d,** Cells infected with COVID-19. **e,** Mock-infection with DMSO (0.1%). **f,** The relative concentration of vRNA present in the supernatant at 72 h post infection (p.i.) determined by qRT-PCR analysis.

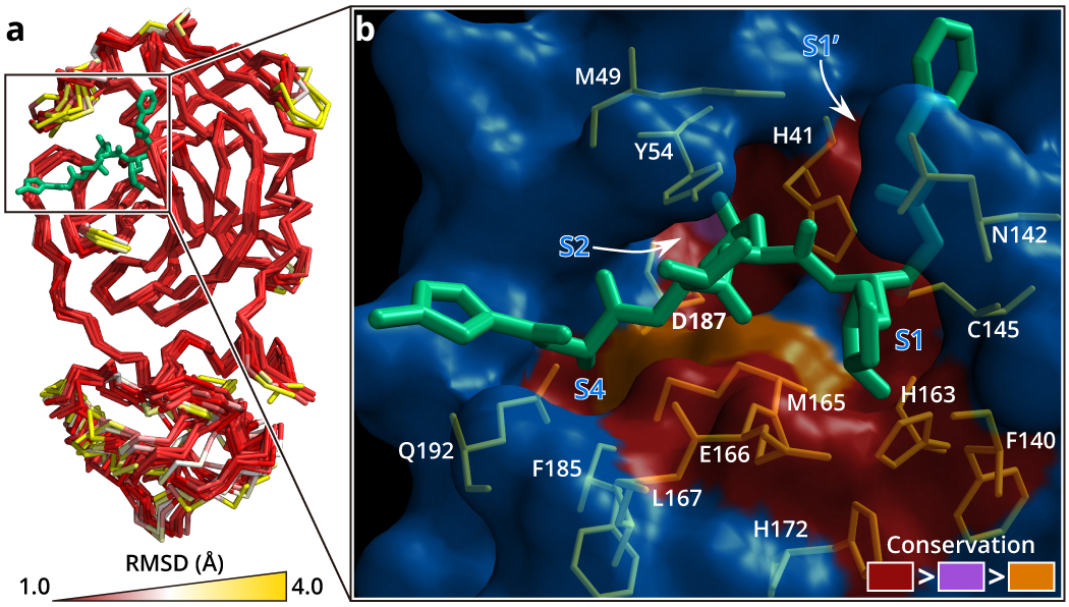


Fig. 4 | The substrate-binding pocket of M^{pro}s is conserved among 12 different CoVs. **a,** Superposition of crystal structures of M^{pro}s (Cα 1-300) from 12 CoVs, including COVID-19, SARS-CoV, MERS-CoV, HCoV-HKU1, BtCoV-HKU4, MHV-A59, PEDV, FIPV, TGEV, HCoV-NL63, HCoV-229E and IBV. The color spectrum represents the root-mean-square deviation (RMSD) of the aligned Cα atoms. **b,** Surface presentation of conserved substrate-binding pockets of 12 CoV M^{pro}s. N3 is shown as green sticks; the background (blue) is the surface of COVID-19 M^{pro}. Red: residues are entirely identical among all 12 M^{pro}s; violet: conserved substitution in one CoV M^{pro}; orange: conserved substitution in more than one CoV M^{pro}s. S1, S2, S4, and S1′ subsites are indicated.

335 **Methods**

336 **Cloning, protein expression and purification**

337 The full-length gene encoding COVID-19 M^{pro} was optimized and synthesized for
338 *Escherichia coli* expression (GENEWIZ). Cloning strategy for producing authentic viral
339 M^{pro} has been reported previously[13]. The expression plasmid was transformed into
340 *Escherichia coli* BL21 (DE3) cells and then cultured in Luria broth (LB) media containing
341 100 μg/mL ampicillin at 37°C, 220 rpm. When the cells were grown to OD_{600} of 0.6-0.8,
342 0.5 mM IPTG was added to the cell culture to induce the expression of the recombinant
343 COVID-19 M^{pro} at 16°C, 180 rpm. After 10 hours, the cells were harvested by
344 centrifugation at 3,000×g for 20 min.

345 The cell pellets were resuspended in lysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl),
346 lysed by high-pressure homogenization, and then centrifuged at 25,000×g for 40 min. The
347 supernatant was loaded onto Ni-NTA affinity column (QIAGEN), and washed by the
348 resuspension buffer containing 20 mM imidazole. The His tagged M^{pro} was eluted by
349 cleavage buffer (50 mM Tris-HCl, pH 7.0, 150 mM NaCl) including 300 mM imidazole.
350 Human rhinovirus 3C protease was added to remove the C-terminal His tag. The COVID-
351 19 M^{pro} was further purified by anion-exchange chromatography and size-exclusion
352 chromatography (Superdex 200, GE Healthcare). The purified COVID-19 M^{pro} was stored
353 in 50 mM Tris-HCl, pH 7.3, 1 mM EDTA.

354 **Crystallization**

355 COVID-19 M^{pro} was incubated with 10 mM N3 for 30 min and the complex (5 mg/ml) was
356 crystallized by hanging drop vapor diffusion method at 20°C. The best crystals were grown
357 with well buffer containing 2% polyethylene glycol (PEG) 6000, 3% DMSO, 1 mM DTT,

358 0.1 M MES (pH 6.0). The cryo-protectant solution contained 30% PEG 400, 0.1 M MES
359 (pH 6.0).

360 **Data collection and structure determination**

361 X-ray data were collected on beamline BL17U1 at Shanghai Synchrotron Radiation
362 Facility (SSRF) at 100 K and at a wavelength of 1.07180 Å using an Eiger X 16M image
363 plate detector. Data integration and scaling were performed using the program Xia2. The
364 structure was determined by molecular replacement (MR) with the Phaser module[36] in
365 CCP4[37] using the SARS-CoV M^{pro} (PDB ID: 2H2Z) as a search template. The output model
366 from MR was subsequently subjected to iterative cycles of manual model adjustment with
367 Coot[38] and refinement was finished with Phenix[39]. The inhibitor N3 was built according to
368 the omit map. The phasing and refinement statistics are summarized in Extended Data
369 Table 1. The R_{work}/R_{free} values are 0.202/0.235, respectively. 97.3% residues are in most
370 favored regions of the Ramachandran plot, and no residues are found in disallowed regions.
371 Coordinates and structure factors for COVID-19 M^{pro} in complex with the inhibitor N3
372 have been deposited in Protein Data Bank with accession number 6LU7.

373 **Enzymatic activity and inhibition assays**

374 The enzyme activity assays have been described previously[13]. Briefly, the activity of
375 COVID-19 M^{pro} was measured by continuous kinetic assay, with the substrate MCA-
376 AVLQSGFR-Lys(Dnp)-Lys-NH2 (GL Biochem Shanghai Ltd), using wavelengths of 320
377 and 405 nm for excitation and emission, respectively. The assays started by immediately
378 mixing 0.2 μM COVID-19 M^{pro} with different concentrations of substrate (2.5-100 μM).
379 Fluorescence intensity was monitored with EnVision multimode plate reader
380 (PerkinElmer). Initial rates were obtained by fitting the linear portion of the curves (usually

381 the first 3 min of the assay curves) to a straight line. The kinetic parameters K_m and k_{cat}
382 were calculated from a double-reciprocal plot. As N3 is a mechanism-based irreversible
383 inhibitor for COVID-19 M^{pro}, k_{obs}/[I] was used as an approximation of the pseudo second-
384 order rate constant to evaluate the inhibition effect of the inhibitor N3[15]. In this case, the
385 measurement was carried out with 0.2 μM of enzyme, 20 μM of substrate and inhibitor at
386 6 different concentrations (0-1 μM).

387 **Virtual screening**

388 The virtual screening was performed using our in-house database via a workflow
389 application of Glide (v8.2)[40], which is a module of molecular simulation software package,
390 Maestro (Schrodinger 2019-1a). All compounds in the database were considered to be at
391 pH 7.4 ± 0.2 to estimate their protonation state using the program EpiK[41]. Their three
392 dimensional conformations were generated by the ligPrep module of Maestro. The newly
393 obtained complex structure of COVID-19 M^{pro} (PDB ID: 6LU7) was used to generate
394 receptor grid for docking simulations. The center of active site of the grid was determined
395 according to the N3 in the structure. The flexibility of the receptor hydroxyl and thiol
396 groups in side chains of C145, S46 and Y54 were considered. At the very beginning, a
397 relatively fast but raw screening was performed by using the glide standard precision model,
398 and the top 20% of compounds were kept. Then covalent docking simulations were
399 performed by choosing Michael Addition as the reaction type for these top 20% compounds
400 that contained Michael acceptors, and C145 was set as the source of Michael donor. Finally,
401 the candidate molecules were picked out by analysing the predicted binding modes and
402 their scores.

403 **High-throughput drug screen and IC_{50} measurement**

404 Potential inhibitors against COVID-19 M^{pro} were screened by enzymatic inhibition assay.
405 When different compounds were added into the enzymatic reaction, the change of initial
406 rates was calculated to evaluate their inhibitory effect. Five drug libraries, Approved Drug
407 Library (TargetMol, USA), Clinic Compound Library (TargetMol, USA), FDA-approved
408 Drug Library (Selleck, USA), Natural Product Library (Selleck, USA), and Anti-virus drug
409 library (Shanghai Institute for Advanced Immunochemical Studies, SIAIS), which includes
410 ~ 10,000 compounds, were used. The preliminary screening reaction mixture included 0.2
411 μM protein, 20 μM substrate and 50 μM compounds. The compounds of interest were
412 defined as those with a percentage of inhibition over 60% compared with the reaction in
413 the absence of inhibitor. IC_{50} values were measured at 12 concentrations and three
414 independent experiments were performed. All experimental data was analyzed using
415 GraphPad Prism software.

416 **Molecular docking**

417 To understand the binding interaction of these molecules with COVID-19 M^{pro}, two
418 different molecular docking methods, i.e., Glide (v8.2)[40] and iFitDock[42] were used to
419 predict their binding poses. Then a 3D molecular similarity calculation method, SHAFTs[43],
420 was used for molecular alignment poses enumeration by matching the critical
421 pharmacophore and volumetric overlay between the N3 molecule within the M^{pro} structure
422 and the seven drug candidates. However, the selenium atom of ebselen could not be treated
423 by any of these above methods, so sulfur was used to replace it in the calculations. Then
424 the obtained optimal superposition of these molecules were used to assist the reasonability
425 of the predicted binding poses from the two docking methods, and only the binding
426 orientations which were consistent among different methods were kept for constructing the

427 initial complexes. Finally, these complexes were further optimized (refined binding poses
428 are illustrated in Extended Data Fig. 4) and re-scored (shown in Extended Data Table 2) by
429 using MM-GBSA module[44] of Schrödinger, and the residues within 5 Å around the ligand
430 were refined.

431 **Antiviral assay for compounds identified from high-throughput screening**

432 Vero (African green monkey kidney) cells were incubated at 37°C in Dulbecco's modified
433 Eagle's medium (DMEM; Invitrogen) with 10% fetal bovine serum (FBS), 100 U/mL
434 penicillin, and 100 μg/mL streptomycin.

435 The COVID-19 was originally isolated from a throat swab specimen from a pneumonia
436 patient and amplified in Vero cells. The *in vitro* antiviral effects of the selected compounds
437 in Vero cells were assayed as previously described[45]. Briefly, cultured Vero cells were
438 infected with COVID-19 in the presence of each compounds at the concentration of 10 μM.
439 Cells were then cultured for 72 hours, and the CPE was observed daily and viral RNA copy
440 numbers in supernatants were quantified by in the COVID-19-specific qRT-PCR.

441 **Antiviral assay for cinanserin**

442 Vero E6 cell line was obtained from American Type Culture Collection (ATCC) and
443 maintained in Dulbecco's Modified Eagle Medium (DMEM; Gibco Invitrogen)
444 supplemented with 10% fetal bovine serum (FBS; Gibco Invitrogen) at 37 °C in a
445 humidified atmosphere of 5% CO_2. A clinical isolate COVID-19 [3] was propagated in Vero
446 E6 cells, and viral titer was determined as described previously[45]. All the infection
447 experiments were performed at biosafety level-3 (BLS-3).

448 Pre-seeded Vero E6 cells (5×10^4 cells/well) were pre-treated with the different
449 concentration of the indicated compounds for 1 hour and the virus was subsequently added

(MOI of 0.05) to allow infection for 2 hours. Then, the virus-drug mixture was removed and cells were further cultured with fresh drug containing medium. At 24 h p.i., the cell supernatant was collected and viral RNA in supernatant was subjected to qRT-PCR analysis.

36 McCoy, A. J. *et al.* Phasercrystallographic software. *Journal of Applied Crystallography* **40**, 658-674, doi:10.1107/s0021889807021206 (2007).

37 Potterton, L. *et al.* CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D-Structural Biology* **74**, 68-84, doi:10.1107/s2059798317016035 (2018).

38 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**, 486-501 (2010).

39 Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D-Structural Biology* **68**, 352-367, doi:10.1107/s0907444912001308 (2012).

40 Friesner, R. A. *et al.* Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* **47**, 1739-1749, doi:10.1021/jm0306430 (2004).

41 Greenwood, J. R., Calkins, D., Sullivan, A. P. & Shelley, J. C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *Journal of computer-aided molecular design* **24**, 591-604 (2010).

42 Bai, F. *et al.* Free energy landscape for the binding process of Huperzine A to acetylcholinesterase. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 4273-4278, doi:10.1073/pnas.1301814110 (2013).

43 Liu, X. F., Jiang, H. L. & Li, H. L. SHAFTS: A Hybrid Approach for 3D Molecular Similarity Calculation. 1. Method and Assessment of Virtual Screening. *Journal of Chemical Information and Modeling* **51**, 2372-2385, doi:10.1021/ci200060s (2011).

44 Guimaraes, C. R. W. & Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *Journal of Chemical Information and Modeling* **48**, 958-970, doi:10.1021/ci800004w (2008).

45 Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Research*, doi:10.1038/s41422-020-0282-0 (2020).

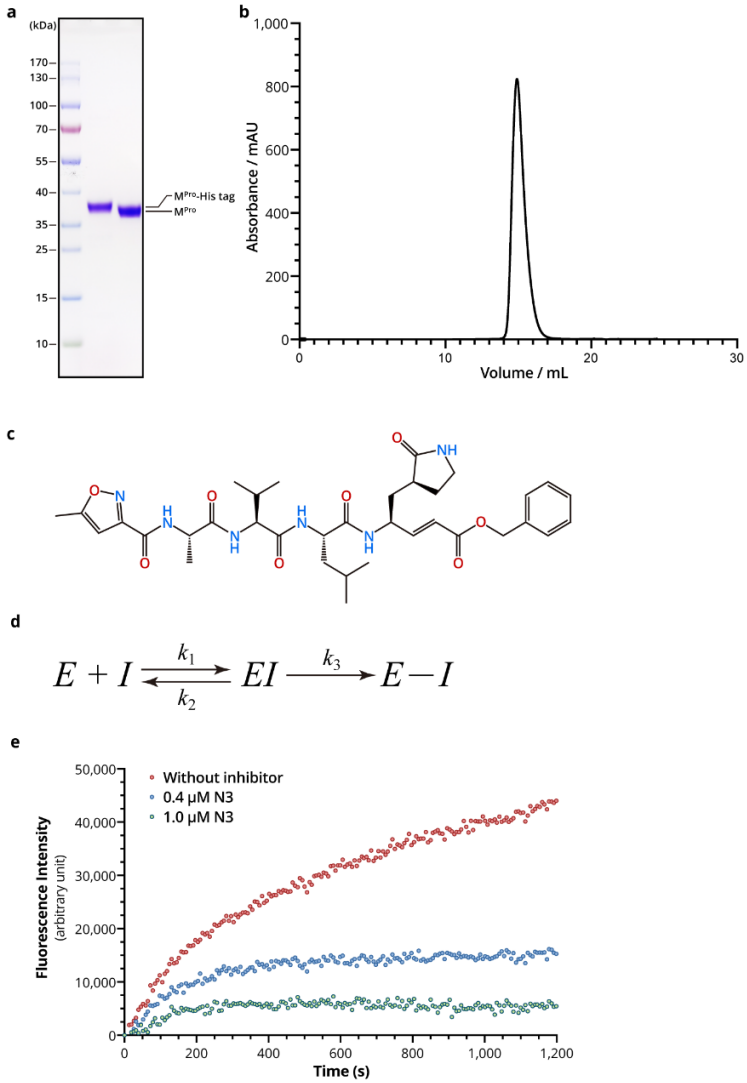
Acknowledgments We would like to thank Ying Lei and Juan Kong from High Throughput Platform, staff from Analytical Chemistry Platform at Shanghai Institute for Advanced Immunochemical Studies, for their technical support. We are grateful to National Centre for Protein Science Shanghai and The Molecular and Cell Biology Core Facility of the School of Life Science and Technology, ShanghaiTech University for use of their instrumentation and technical assistance. We thank Prof. Wenqing Xu, Prof. Zhijie Liu, and Haixia Su for discussion. We also thank the staff from beamlines BL17U1, BL18U1 and BL19U1 at Shanghai Synchrotron Radiation Facility (China). This work was supported by grants from National Key R&D Program of China (grant No. 2017YFC0840300) and

Project of International Cooperation and Exchanges NSFC (grant No. 81520108019) to Z.R..

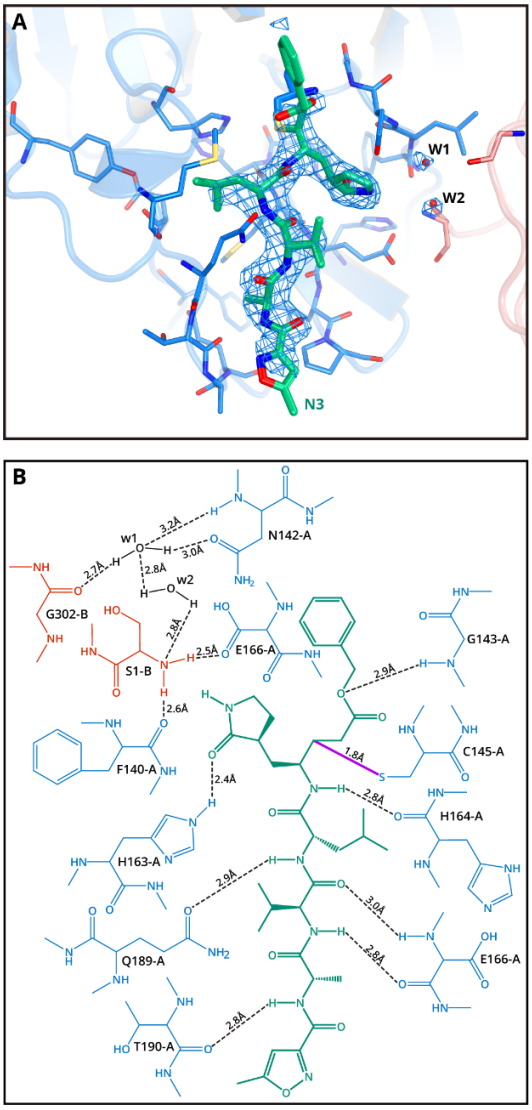
Author contributions Z.R. and H.Y. conceived the project. Z.J., H.J., Z.R., and H.Y. designed the experiments; Z.J., X.D., Y.X., Y.D., M.L., Y.Z., B.Z., X. Li, L.Z., Y.D., J.Y., F.L., T.Y., Xiaoce Liu and X.Y. performed the experiments. Z.J., B.Z. and F.L. collected the diffraction data. B.Z. and Xiang Liu solved the crystal structure. Z.J., X.D., Y.X., K.Y., F.B., H.L., Xiang Liu, L.G., G.X., C.Q., Z.S., H.J., Z.R. and H.Y. analyzed and discussed the data. Z.J., X.D., F.B., Xiang Liu, L.G., G.X., C.Q., Z.S., H.J., Z.R. and H.Y wrote the manuscript.

Competing interests The authors declare no competing interests.

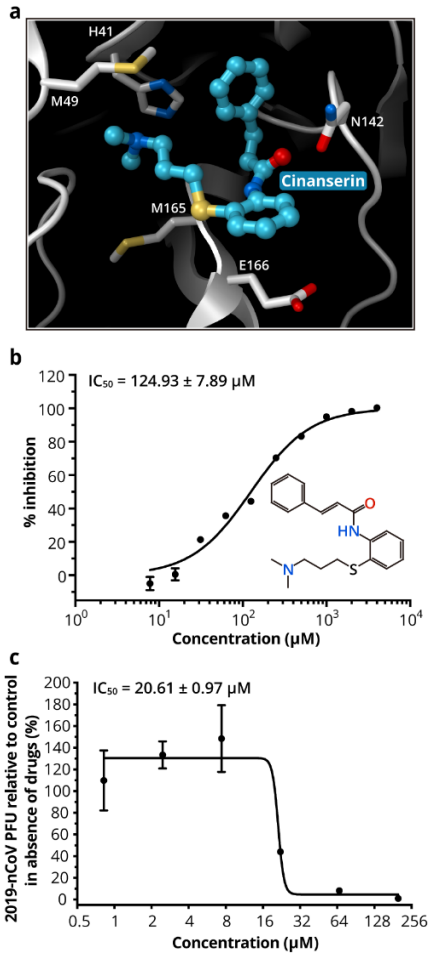
Data and materials availability: The PDB accession No. for the coordinates of COVID-19 M^{pro} in complex with N3 is 6LU7 (Deposited: 2020-01-26; Released: 2020-02-05).



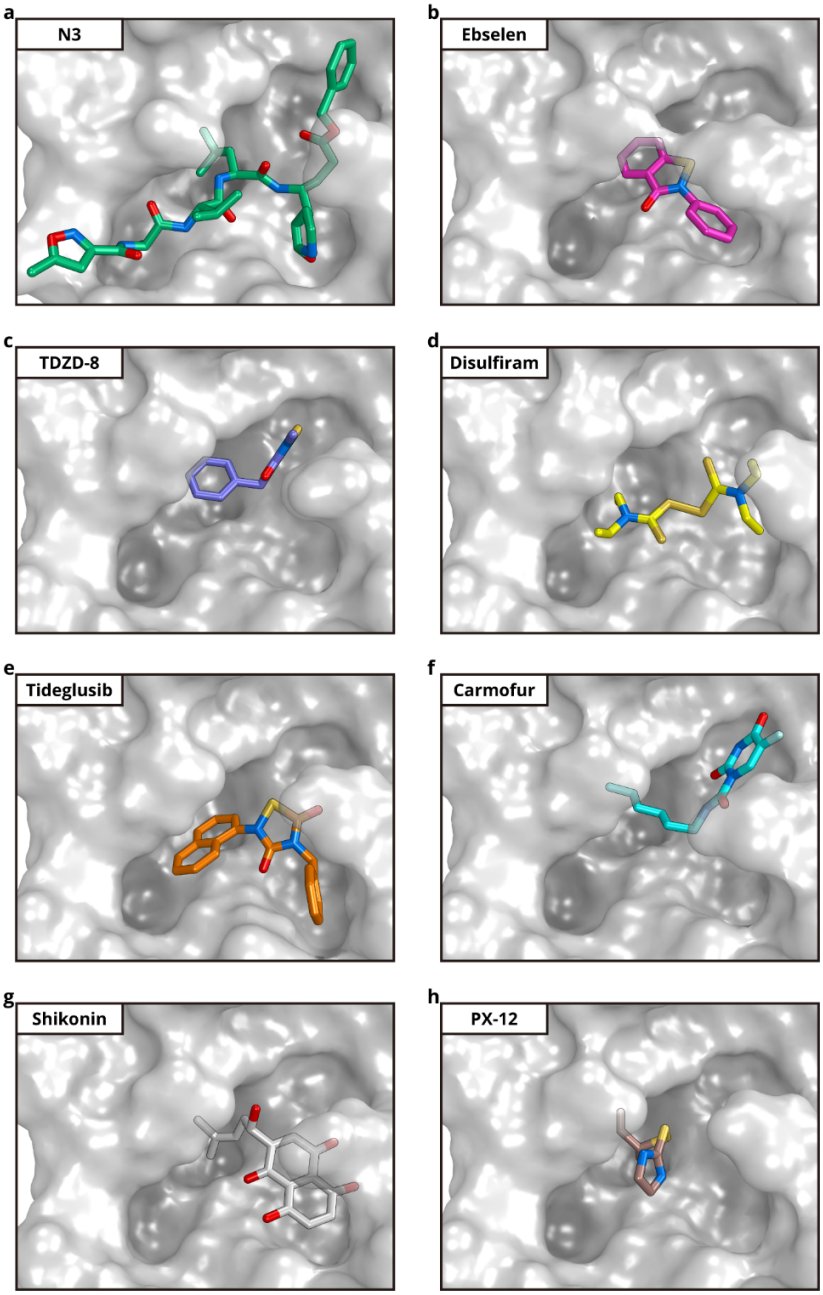
Extended Data Fig. 1 | The purification of COVID-19 M^{pro} and the inhibitory assay of N3 compound. a, the SDS-PAGE gel of COVID-19 M^{pro}. The first lane: marker; the second lane: M^{pro} before treating with rhinovirus 3C protease; third lane: the M^{pro} after the cleavage of C-terminal tag. **b,** Size-exclusion chromatography profile of M^{pro}. **c,** The chemical structure of N3 inhibitor. **d,** Inhibition mechanism for N3. **e,** Typical inhibition curves for N3.



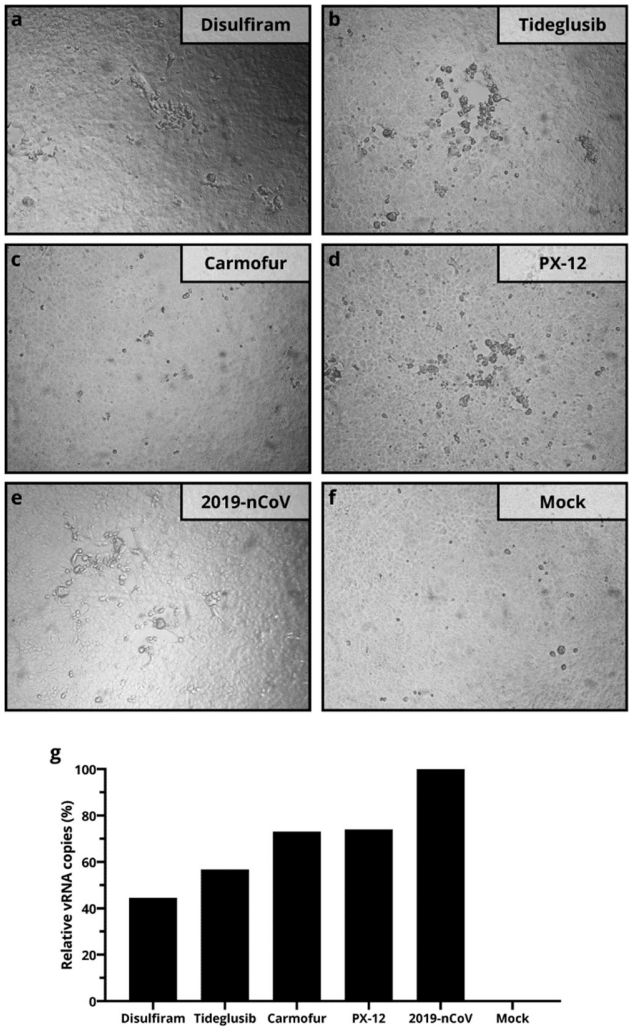
Extended Data Fig. 2 | The interactions between M^{pro} and N3. a, The F_o-F_c omit map (contour level = 3 σ, shown as the blue mesh). **b,** Detailed view of the interactions between the N3 and COVID-19 M^{pro}. The hydrogen bonds are shown as black dashed lines. The covalent bond between N3 and C145-A is in purple.



Extended Data Fig. 3 | Cinanserin is an inhibitor for M^{pro}. a, The docking result of cinanserin. The structure of COVID-19 M^{pro} is shown as a white cartoon; cinanserin is shown as cyan sticks and balls; residues predicted to be interacting with cinanserin are shown as sticks. **b,** Inhibiton of cinanserin on M^{pro}. **c,** Antiviral activity of cinanserin.



Extended Data Fig. 4 | Docking Poses of different COVID-19 M^{pro} inhibitors. The docking results show that the inhibitors identified through the high-throughput screening are likely to occupy the same pocket as N3.



Extended Data Fig. 5 | Virus-induced CPE in Vero cells treated with M^{pro} inhibitors. a-d, Treated with disulfiram, tideglusib, carmofur and PX-12 respectively; **e,** Cells infected with COVID-19; **f,** Mock-infection with DMSO (0.1%). **g,** The relative concentration of vRNA present in the supernatant at 72h p.i. determined by qRT-PCR analysis. Because shikonin showed cellular toxicity at the test concentration, its antiviral activity assay did not further proceed.

536 **Extended Data Table 1 | Data collection and refinement statistics**

537

	PDB code: 6LU7
Data collection	
Space group	*C*2
Cell dimensions	
a, *b*, *c* (Å)	97.931, 79.477, 51.803
α, β, γ (°)	90, 114.55, 90
Resolution (Å)	50.00-2.16 (2.22-2.16)
R_{merge}	18.9 (147.2)
$I / \sigma I$	6.3 (3.2)
Completeness (%)	100.0 (100.0)
Redundancy	6.6 (6.1)
Refinement	
Resolution (Å)	50.00-2.16
No. reflections	19455 (1431)
R_{work} / R_{free}	0.2020/0.2350
No. atoms	
Protein	2367
Ligand/ion	49
Water	84
B-factors	
Protein	42.7
Ligand/ion	46.3
Water	44.2
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.474

538 *Values in parentheses are for highest-resolution shell.

539

Extended Data Table 2 | Predicted binding affinities for the drug leads to COVID-19 M^{pro} by using MM-GBSA module integrated in Schrödinger

Drug	Predicted ΔG_{bind} (kcal/mol)
PX-12	-50.84
Shikonin	-67.95
TDZD-8	-43.77
Tideglusib	-61.79
Carmofur	-47.12
Ebselen	-53.83
Disulfiram	-46.16