Structure of M^{pro} from COVID-19 virus and discovery of its inhibitors

Zhenming Jin[1,2,10], Xiaoyu Du[2,10], Yechun Xu[3,10], Yongqiang Deng[4,10], Meiqin Liu[5,10], Yao Zhao[1], Bing Zhang[1], Xiaofeng Li[4], Leike Zhang[5], Chao Peng[6], Yinkai Duan[1], Jing Yu[1], Lin Wang[1], Kailin Yang[7], Fengjiang Liu[1], Rendi Jiang[5], Xinglou Yang[5], Tian You[1], Xiaoce Liu[1], Xiuna Yang[1], Fang Bai[1], Hong Liu[3], Xiang Liu[8], Luke W. Guddat[9], Wenqing Xu[1,6], Gengfu Xiao[5], Chengfeng Qin[4], Zhengli Shi[5], Hualiang Jiang[1,3*], Zihe Rao[1,2,8*] & Haitao Yang[1*]

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

[2]Laboratory of Structural Biology, School of Life Sciences and School of Medicine, 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]National Facility for Protein Science in Shanghai, Zhangjiang Lab, Shanghai Advanced Research Institute, Chinese Academy of Science, Shanghai, China.

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

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

[9]School of Chemistry and Molecular Biosciences, the University of Queensland, Brisbane, Australia.

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

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

SUMMARY

A new coronavirus (CoV) identified as COVID-19 virus 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 virus main protease (M^{pro}). M^{pro} is a key CoV 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 virus 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}. Six of these inhibit M^{pro} with IC_{50} values ranging from 0.67 to 21.4 μM. Ebselen also exhibited promising antiviral activity in cell-based assays. Our results demonstrate the efficacy of this screening strategy, which can lead to the rapid discovery of drug leads with clinical potential in response to new infectious diseases where no specific drugs or vaccines are available.

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)[7]. The COVID-19 virus 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[8]. 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[9].

To facilitate the rapid discovery of antiviral compounds with clinical potential, we developed a strategy combining structure-assisted drug design, virtual drug screening and high-throughput screening to repurpose existing drugs to target COVID-19 virus M^{pro}.

Establishing a high-throughput activity assay

Recombinant COVID-19 virus M^{pro} with native N and C termini was expressed in *Escherichia coli* and subsequently purified (Extended Data Fig. 1a, b). The molecular weight of COVID-19 virus M^{pro} as determined by mass spectroscopy is 33797.0 Da, consistent with its theoretical molecular weight 33796.8 Da. In order to characterize its enzymatic activity and to carry out high-throughput screening of inhibitors, we developed a fluorescence resonance energy transfer (FRET) assay. To do this, a fluorescently labeled substrate, MCA-AVLQ↓SGFR-Lys(Dnp)-Lys-NH_2, derived from the N-terminal auto-cleavage sequence of the viral protease was designed and synthesized for time-dependent

kinetic analysis (Extended Data Fig. 1e). The catalytic efficiency (k_{cat}/K_m) for COVID-19 virus M^{pro} was measured to be 28,500 $M^{-1}s^{-1}$ which is slightly higher than that for SARS-CoV M^{pro} (k_{cat}/K_m=26,500 $M^{-1}s^{-1}$)[10], but over 30-fold higher than that of human rhinovirus 3C protease (k_{cat}/K_m=920 $M^{-1}s^{-1}$)[11].

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

In a previous study, we designed a Michael acceptor inhibitor N3 using computer-aided drug design (CADD) (Extended Data Fig. 1c), which can specifically inhibit multiple CoV M^{pro}s, including those from SARS-CoV and MERS-CoV[12-15]. It also has displayed potent antiviral activity against infectious bronchitis virus in an animal model[13]. The CC_{50} of N3 is >133 μM (Extended Data Fig. 1f). Next, we constructed a homology model for COVID-19 virus 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 virus M^{pro}, kinetic analysis was performed. A progress curve showed that it is a 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 virus 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. However, N3 exhibits very potent inhibition of COVID-19 virus M^{pro}, such that measurement of K_i and k_3 proved not feasible (Extended Data Fig. 1d, e). 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)[12]. The value of k_{obs}/[I] of N3 for COVID-19 virus 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 virus M^{pro} in complex with N3

In order to elucidate the inhibitory mechanism of this compound, we determined the crystal structure of COVID-19 virus 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 virus 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[5,6,13-15]. 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, d). 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. 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 inserts deeply 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, 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.

An overlay of the structures of COVID-19 virus M^{pro}-N3 and SARS-CoV M^{pro}-N1[12] shows that N3 and N1 bind to M^{pro}s in a similar mode (Fig. 2a, Extended Data Fig. 3). The major difference lies in the P1′ site. Compared with the benzyl ester portion of N3 in the COVID-19 virus M^{pro} structure, the ethyl ester portion in N1 adopts a slightly different conformation. This may be attributed to an ordered water (W1) in SARS-CoV M^{pro}-N1 structure, which makes a long-distance hydrogen bond to the carboxylate oxygen of the ester and also forms two hydrogen bonds from the backbone NH of G143 and the side chain of N142. In our previous study, we proposed that all the CoV M^{pro}s share a highly conserved substrate-recognition pocket, which could serve as a drug target for the design of broad-spectrum inhibitors[12]. The recent discovery of new CoVs and accumulation of structural data for CoV M^{pro}s from various species provided the opportunity to further examine this hypothesis. Superposition of the 12 crystal structures of M^{pro}s[12-21] have shown

141 that the most variable regions were the helical domain III and surface loops, but the
142 substrate-binding pockets located in a cleft between domains I and II are still highly
143 conserved among all CoV M^{pro}s, suggesting the antiviral inhibitors targeting this site should
144 have wide-spectrum anti-CoV activity (Fig. 2b, c).

145 **Virtual screening**

146 The structure of COVID-19 virus M^{pro} in complex with N3 provides a model for identifying
147 lead inhibitors to target COVID-19 virus M^{pro} through *in silico* screening. To achieve this,
148 an in-house database of potential binding compounds was docked using Glide (v8.2). The
149 results show that cinanserin fits snugly into the substrate-binding pocket, by interacting
150 with H41 and E166 of M^{pro} by cation-π. Subsequently we determined this compound has
151 an IC_{50} value of 125 μM for M^{pro}. Moreover, cinanserin is a well-characterized serotonin
152 antagonist, which underwent preliminary clinical testing in humans in the 1960s[22] and has
153 previously been shown to inhibit SARS-CoV M^{pro}[23]. The CC_{50} of cinanserin is > 200 μM
154 (Extended Data Fig. 4). Thus, it has potential for optimization as an anti-viral drug lead.

155 **High-throughput screening**

156 Next, we used our FRET assay, to screen a library of ~10,000 compounds consisting of
157 approved drugs, clinical trial drug candidates and natural products. Primary hits included
158 seven compounds that are either FDA-approved drugs or clinical trial/preclinical drug
159 candidates. We then determined their IC_{50} values, which are in the range from 0.67 to 21.4
160 μM (Fig. 3). Amongst them, disulfiram and carmofur are FDA-approved drugs, whilst
161 ebselen, shikonin, tideglusib, PX-12 and TDZD-8 are currently in clinical trials or
162 preclinical studies. Ebselen has the strongest inhibition of M^{pro} activity with an IC_{50} of 0.67
163 μM. However, in a detergent-based assay[24], TDZD-8 was found to be an aggregate-based

164 inhibitor, which might non-specifically inhibit M^{pro} (Extended Data Fig. 5) and was
165 therefore not considered for further investigation. Next, we set out to identify the potential
166 covalent inhibitors among these compounds through tandem MS/MS analysis. The MS/MS
167 data shows that ebselen, PX-12 and carmofur are all able to covalently bind to C145 of the
168 catalytic dyad in COVID-19 virus M^{pro}. However, while PX-12 and carmofur completely
169 modified M^{pro}, ebselen could only partially modify this viral cysteine protease (Extended
170 Data Fig. 6). Since ebselen has even stronger inhibition than the others, there is a possibility
171 that ebselen could also inhibit M^{pro} through non-covalent binding. It is likely that a portion
172 of the hits identified by screening are covalently bonded to the catalytic cysteine of M^{pro}
173 through their sulfhydryl groups. In general, such molecules are expected to be promiscuous
174 binders and therefore, as they stand, may have limited potential as drug leads. Since our
175 structural data is based on N3, we investigated if molecular docking could predict how
176 disulfiram, tideglusib and shikonin bind to this protein. In all cases, reasonable docking
177 poses were found, demonstrating that they could fit inside the substrate-binding pocket
178 (Extended Data Fig. 7).

179 **Antiviral activity assay**

180 To further substantiate the enzymatic inhibition results *in vitro*, we evaluated whether these
181 compounds could prevent viral replication in cell-based assays. As shown in Fig. 4a,
182 quantitative real-time RT-PCR (qRT-PCR) demonstrated that ebselen and N3 showed the
183 strongest antiviral effects among them at a concentration of 10 μM treatment in COVID-
184 19 virus infected Vero cells. A plaque-reduction assay (Extended Data Fig. 8) was
185 performed to further assess the efficacy of these two compounds in protecting cells.
186 Ebselen and N3 displayed inhibition against COVID-19 virus with individual EC_{50} values

of 4.67 μM and 16.77 μM, respectively (Fig. 4b, c). The dose-response curves suggest that both of them could be able to penetrate cellular membrane to access their targets. Ebselen is an organoselenium compound with anti-inflammatory, anti-oxidant and cytoprotective properties. This compound has been investigated for the treatment of multiple diseases, such as bipolar disorders[25] and hearing loss[26,27]. Ebselen has extremely low cytotoxicity (LD_{50} in rats > 4,600 mg/kg, per os)[28] and its safety in humans has been evaluated in a number of clinical trials[26,27,29]. These data strongly suggest the clinical potential of ebselen for CoV treatment. It is also interesting that cinanserin displayed moderate inhibition against COVID-19 virus with an EC_{50} value of 20.61 μM from qRT-PCR analysis (Extended Data Fig. 4), which is superior to that in the enzymatic inhibition assay, suggesting that cinanserin might have multi-drug targets in preventing viral infection. In further studies, selection and characterization of drug-resistant mutants will help clarify the mode of cinanserin's action.

Discussion

Our crystal structural and docking data have shown that the drug leads identified can bind to the substrate-binding pocket of COVID-19 virus M^{pro}, which is highly conserved among all CoV M^{pro}s. This strongly supports our hypothesis that development of a single antiviral agent targeting M^{pro} or in combination with other potential therapies could provide an effective first line of defense against all CoV-associated diseases.

In the last twenty years, new infectious agents have emerged to cause epidemics, such as SARS and MERS[7]. The timely development of effective antiviral agents for clinical use is extremely challenging because conventional drug development approaches normally take years of investigations and cost billions of dollars. The repurposing of approved

pharmaceutical drugs and drug candidates provides an alternative approach to rapidly identify potential drug leads to manage rapidly emerging viral infections. Cell-based phenotypic screening has proven to be valuable[30], but the complexity of this approach is not readily compatible with high-throughput pipelines, and it cannot identify the molecular target or mechanism of action[31]. In this study, the convergence of structure-based *ab initio* drug design, virtual screening and high-throughput screening proved to be an efficient strategy to find antiviral leads against COVID-19 virus. The methods presented here can greatly assist in the rapid discovery of drug leads with clinical potential in response to new emerging infectious diseases that currently lack specific drugs and vaccines.

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. *et al.* Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. *The EMBO Journal* **21**, 3213-3224 (2002).

6 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).

7 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).

8 Hegyi, A. & Ziebuhr, J. Conservation of substrate specificities among coronavirus main proteases. *Journal of general virology* **83**, 595-599 (2002).

9 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).

10 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).

11 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).

12 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).

13 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).

14 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).

15 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).

16 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).

17 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).

18 Ma, Q., Xiao, Y., Hilgenfeld, R. *Inhibitor for the Main Protease of Coronavirus Hku4* (2012).

19 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).

20 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).

21 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).

22 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).

23 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).

24 Feng, B. Y. & Shoichet, B. K. A detergent-based assay for the detection of promiscuous inhibitors. *Nature protocols* **1**, 550-553 (2006).

25 Singh, N. *et al.* A safe lithium mimetic for bipolar disorder. *Nature communications* **4**, 1-7 (2013).

26 Lynch, E. & Kil, J. Development of ebselen, a glutathione peroxidase mimic, for the prevention and treatment of noise-induced hearing loss. *Seminars in Hearing* **30**, 047-055 (2009).

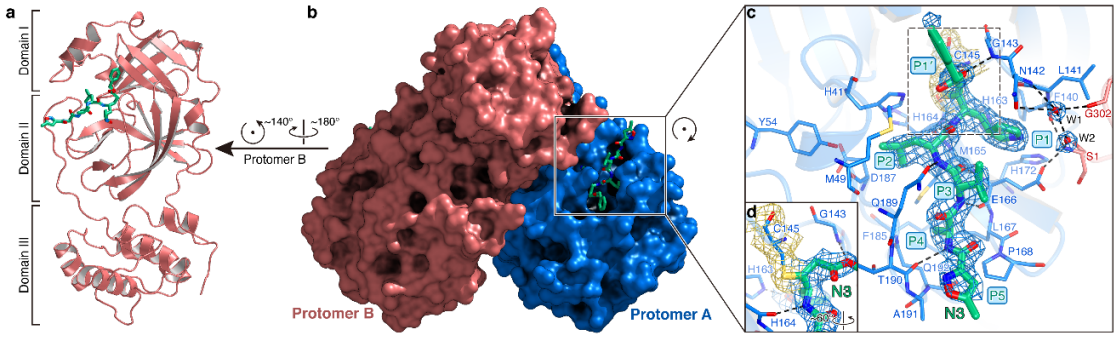
27 Kil, J. *et al.* Safety and efficacy of ebselen for the prevention of noise-induced hearing loss: a randomised, double-blind, placebo-controlled, phase 2 trial. *The Lancet* **390**, 969-979 (2017).

28 Renson, M., Etschenberg, E. & Winkelmann, J. 2-Phenyl-1, 2-benzisoselenazol-3 (2H)-one containing pharmaceutical preparations and process for the treatment of rheumatic diseases. (1982).

29 Masaki, C. *et al.* Effects of the potential lithium-mimetic, ebselen, on impulsivity and emotional processing. *Psychopharmacology* **233**, 2655-2661 (2016).

30 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).

31 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).



296

297 **Fig. 1 | The crystal structure of COVID-19 virus M^{pro} in complex with N3. a,** Cartoon
298 representation of one protomer of the dimeric M^{pro}-inhibitor complex. **b,** Surface representation of
299 the homodimer of M^{pro}. Protomer A is in blue, protomer B is in salmon, N3 is presented as green
300 sticks. **c,** A zoomed view of the substrate-binding pocket. The key residues forming the binding
301 pocket are shown in sticks, the two waters, assigned as W1 and W2, are shown as red spheres. P1,
302 P1′, P2, P3, P4 and P5 sites of N3 are indicated. Hydrogen bonds that help to lock the inhibitor are
303 shown in black dashed lines. The $2F_o$-F_c density map contoured at 1.2 σ is shown around N3
304 molecule (blue mesh), C145-A (yellow mesh), and the two waters (blue mesh). **d,** The C-S covalent
305 bond.

306

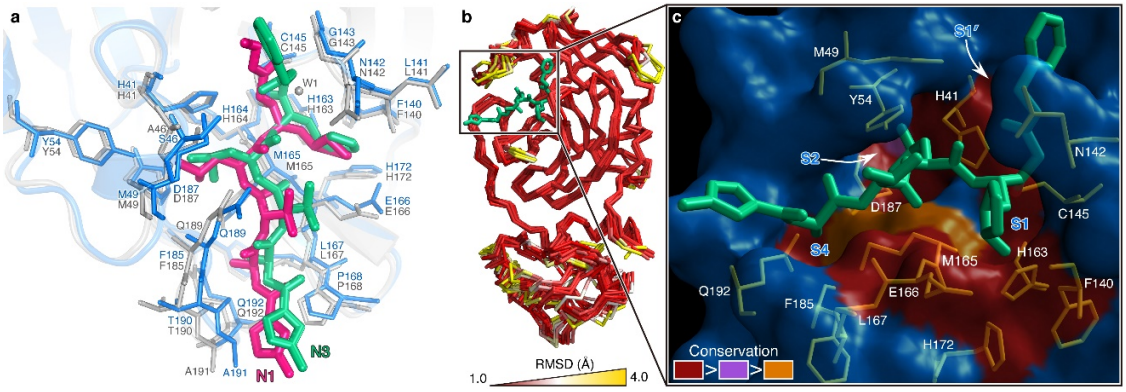


Fig. 2 | The substrate-binding pockets of CoV M^{pro}s across different species. a, Comparison of inhibitor binding mode between the structures of COVID-19 virus M^{pro}-N3 and SARS-CoV M^{pro}-N1. COVID-19 virus M^{pro} is shown in marine cartoon; SARS-CoV M^{pro} in grey; N3 in green sticks; N1 in hot pink. **b,** Superposition of crystal structures of M^{pro}s (Cα 1-300) from 12 CoVs, including COVID-19 virus, 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. **c,** Surface presentation of conserved substrate-binding pockets of 12 CoV M^{pro}s. 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.

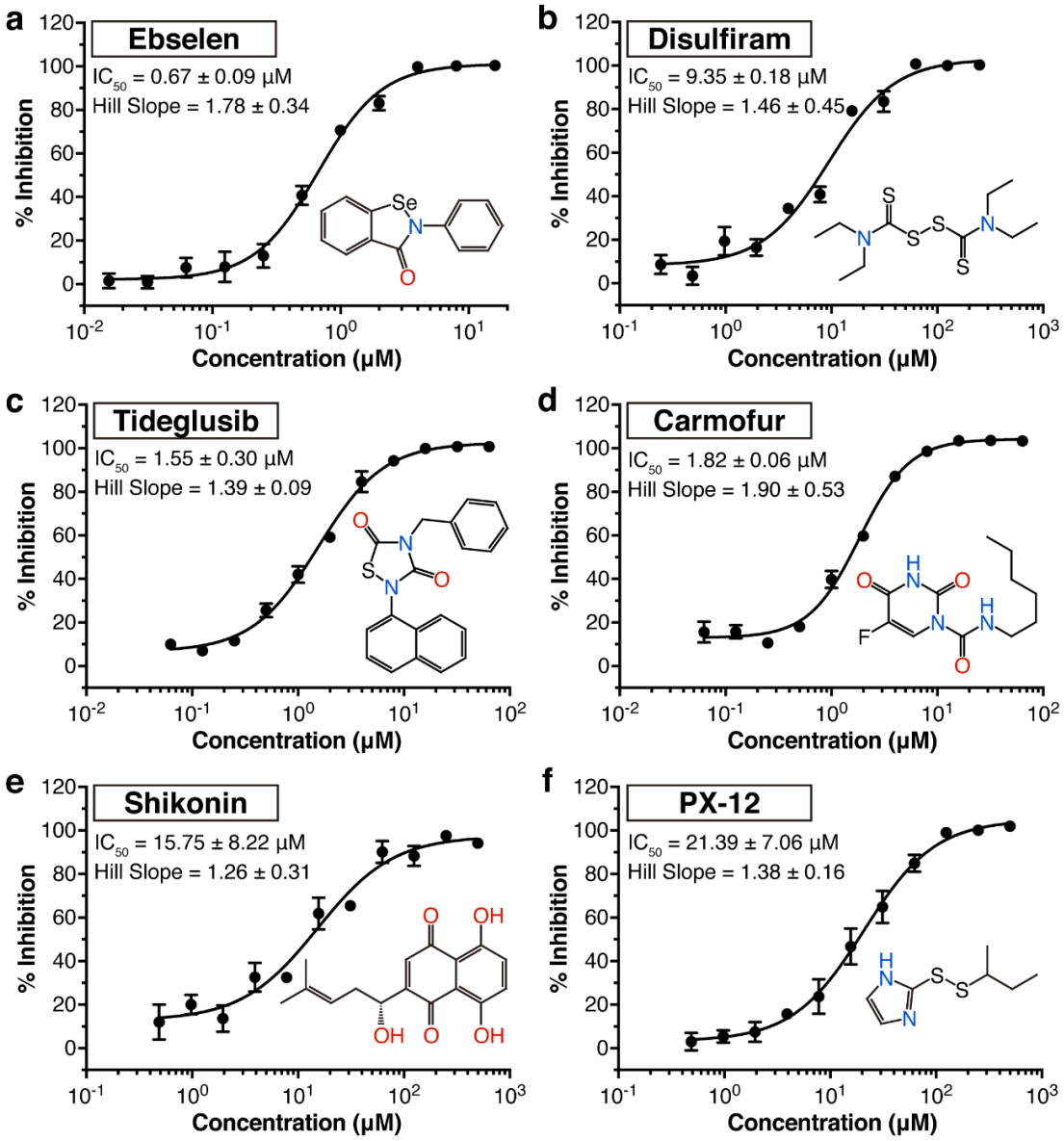


Fig. 3 | Drug leads inhibit the activity of COVID-19 virus M^{pro}. a-f, The hydrolytic activity of COVID-19 virus M^{pro} was measured in the presence of varying concentrations of the drug candidates. Dose-response curves for half-maximum inhibitory concentration (IC_{50}) values were determined by nonlinear regression. All data are shown as mean ± s.e.m., n = 3 biological replicates.

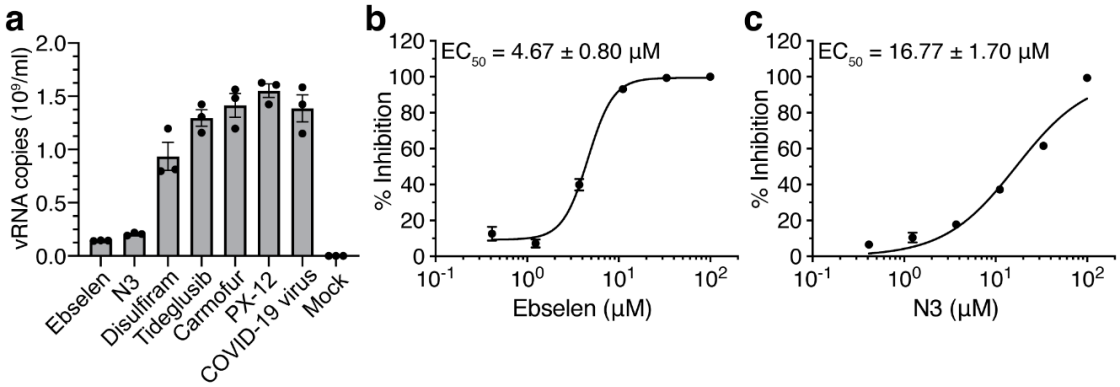


Fig. 4 | Antiviral activities of the drug leads against COVID-19 virus. a, The quantification of absolute viral RNA copies (per ml) in the supernatant at 72 h post infection (p.i.) determined by qRT-PCR analysis. Data are mean ± s.e.m., n = 3 biological replicates. **b, c,** Dose-response curves for ebselen and N3 in the plaque-reduction assay, respectively; all data are shown as mean ± s.e.m., n = 4 biological replicates.

332 **Methods**

333 **Cloning, protein expression and purification of COVID-19 virus M^{pro}**

334 The full-length gene encoding COVID-19 virus M^{pro} (NC_045512) was optimized and
335 synthesized for *Escherichia coli* expression (Genewiz, USA). Cloning strategy for
336 producing authentic viral M^{pro} has been reported previously[10]. The expression plasmid was
337 transformed into *Escherichia coli* BL21 (DE3) cells and then cultured in Luria Broth
338 medium containing 100 μg/ml ampicillin at 37 °C. When the cells were grown to OD_{600} of
339 0.6-0.8, 0.5 mM IPTG was added to the cell culture to induce the expression at 16 °C. After
340 10 h, the cells were harvested by centrifugation at 3,000*g*. The cell pellets were
341 resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl), lysed by high-
342 pressure homogenization, and then centrifuged at 25,000*g* for 40 min. The supernatant was
343 loaded onto Ni-NTA affinity column (Qiagen, Germany), and washed by the resuspension
344 buffer containing 20 mM imidazole. The His tagged M^{pro} was eluted by cleavage buffer
345 (50 mM Tris-HCl pH 7.0, 150 mM NaCl) including 300 mM imidazole. Human rhinovirus
346 3C protease was added to remove the C-terminal His tag. The M^{pro} was further purified by
347 ion exchange chromatography and size exclusion chromatography. CoV M^{pro}s exist as a
348 mixture of monomers and dimers in solutions[32]. The purified M^{pro} was stored in 50 mM
349 Tris-HCl pH 7.3, 1 mM EDTA.

350 **Crystallization, data collection and structure determination**

351 COVID-19 virus M^{pro} was incubated with 10 mM N3 for 30 min and the complex (5 mg/ml)
352 was crystallized by hanging drop vapor diffusion method at 20 °C. The best crystals were
353 grown with well buffer containing 0.1 M MES pH 6.0, 2% polyethylene glycol (PEG) 6000,
354 3% DMSO, 1 mM DTT. The cryo-protectant solution contained 0.1 M MES pH 6.0, 30%

355 PEG 400.

356 X-ray data were collected on beamline BL17U1 at Shanghai Synchrotron Radiation
357 Facility (SSRF) at 100 K and at a wavelength of 1.07180 Å using an Eiger X 16M image
358 plate detector. Data integration and scaling were performed using the program Xia2[33]. The
359 structure was determined by molecular replacement (MR) with the Phaser module[34] in
360 CCP4[35] using the SARS-CoV M^{pro} (PDB ID: 2H2Z) as a search template. The output model
361 from MR was subsequently subjected to iterative cycles of manual model adjustment with
362 Coot[36] and refinement was finished with Phenix[37]. The inhibitor N3 was built according to
363 the omit map. The phasing and refinement statistics are summarized in Extended Data
364 Table 1. The R_{work}/R_{free} values are 0.202/0.235, respectively. 97.3% residues are in most
365 favored regions of the Ramachandran plot, and no residues are found in disallowed regions.
366 Coordinates and structure factors for COVID-19 virus M^{pro} in complex with the inhibitor
367 N3 have been deposited in Protein Data Bank (PDB) with accession number 6LU7. While
368 this work was under review, we solved the complex structure at a higher resolution (1.7 Å).
369 The relevant coordinates and structure factors have been deposited in PDB with accession
370 number 7BQY.

371 **Enzymatic activity and inhibition assays**

372 The enzyme activity assays have been described previously[10]. Briefly, the activity of
373 COVID-19 virus M^{pro} was measured by a continuous kinetic assay, with the substrate
374 MCA-AVLQSGFR-Lys(Dnp)-Lys-NH2 (GL Biochem, Shanghai), using wavelengths of
375 320 nm and 405 nm for excitation and emission, respectively. The assay started by
376 immediately mixing 0.2 μM COVID-19 virus M^{pro} with different concentrations of
377 substrate (2.5-100 μM). Fluorescence intensity was monitored with an EnVision multimode

378 plate reader (Perkin Elmer, USA). Initial rates were obtained by fitting the linear portion
379 of the curves to a straight line. The kinetic parameters K_m and k_{cat} were calculated from a
380 double-reciprocal plot. As N3 is a mechanism-based irreversible inhibitor for COVID-19
381 virus M^{pro}, k_{obs}/[I] was used as an approximation of the pseudo second-order rate constant
382 to evaluate the inhibition effect of the inhibitor N3[12]. In this case, the measurement was
383 carried out with 0.2 μM of enzyme, 20 μM of substrate and inhibitor at 6 different
384 concentrations (0-1 μM).

385 **Virtual screening**

386 The virtual screening was performed using our in-house database via a workflow
387 application of Glide (v8.2), Maestro (Schrödinger 2019-1a)[38]. All compounds in the
388 database were considered to be at pH 7.4 ± 0.2 to estimate their protonation state using the
389 program EpiK[39]. Their three dimensional conformations were generated by the ligPrep
390 module of Maestro. The structure of COVID-19 virus M^{pro} (PDB ID: 6LU7) was used to
391 generate receptor grid for docking simulations. The center of active site of the grid was
392 determined according to the position of N3 in the structure. The flexibility of the receptor
393 hydroxyl and thiol groups in side chains of C145, S46 and Y54 were considered. At the
394 very beginning, a relatively fast but raw screening was performed by using the glide
395 standard precision model, and the top 20% of compounds were kept. Finally, the candidate
396 molecules were picked by analysing the predicted binding modes and their scores.

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

398 Potential inhibitors against COVID-19 virus M^{pro} were screened by an enzymatic inhibition
399 assay. When the different compounds were added into the enzymatic reaction mixture, the
400 change of initial rates was calculated to evaluate their inhibitory effect. Five drug libraries,

401 Approved Drug Library (Target Mol, USA), Clinic Compound Library (Target Mol, USA),
402 FDA-approved Drug Library (Selleck, USA), Natural Product Library (Selleck, USA), and
403 Anti-virus Drug Library (Shanghai Institute for Advanced Immunochemical Studies,
404 SIAIS), which includes ~10,000 compounds, were used. The preliminary screening
405 reaction mixture included 0.2 μM protein, 20 μM substrate and 50 μM compounds. The
406 compounds of interest were defined as those with a percentage of inhibition over 60%
407 compared with the reaction in the absence of inhibitor. IC_{50} values of seven drug leads were
408 measured using 0.2 μM protein, 20 μM substrate and 11 different inhibitor concentrations.
409 In order to exclude inhibitors possibly acting as aggregators, detergent-based control was
410 performed by adding 0.001% or 0.01% freshly made up Triton X-100 to the reaction at the
411 same time[24]. All experimental data was analyzed using GraphPad Prism. All experiments
412 were performed in triplicate.

413 **Molecular docking**

414 To understand the binding interaction of these molecules with COVID-19 virus M^{pro}, two
415 different molecular docking methods, i.e., Glide (v8.2)[38] and iFitDock[40] were used to
416 predict their binding poses. Then a 3D molecular similarity calculation method, SHAFTS[41],
417 was used for molecular alignment poses enumeration by matching the critical
418 pharmacophore and volumetric overlay between the N3 molecule within the M^{pro} structure
419 and the six drug candidates. However, the selenium atom of ebselen could not be treated
420 by any of these above methods, so sulfur was used to replace it in the calculations. Then
421 the obtained optimal superposition of these molecules was used to assess the reasonability
422 of the predicted binding poses from the two docking methods, and only the binding
423 orientations which were consistent among different methods were kept for constructing the

424 initial complexes. Finally, these complexes were further optimized and re-scored by using
425 MM-GBSA module[42] of Schrödinger, and the residues within 5 Å around the ligand were
426 refined.

427 **Antiviral and cytotoxicity assays for compounds from high-throughput screening**

428 The *in vitro* antiviral efficacy of the drug candidates on Vero cells were determined by qRT-
429 PCR. About 1×10^4 Vero cells were seeded into a 96-well plate and incubated for 20-24 h
430 at 37 °C. All the infection experiments were performed at biosafety level-3 (BSL-3). Cells
431 were pre-treated with the drug candidates (10 μM) for 1 h; the COVID-19 virus (MOI of
432 0.01) was subsequently added to allow infection for 2 h. Then, the virus-drug mixture was
433 removed and cells were further cultured with fresh drug-containing medium. At 72 h post
434 infection (p.i.), viral RNA (vRNA) was extracted from the culture supernatant using
435 QIAamp viral RNA mini kit (Qiagen, Germany) according to the manufacturer's
436 recommendation and detected by qRT-PCR assay using the COVID-19 virus-specific
437 primers. Because shikonin showed cellular toxicity at the test concentration, its antiviral
438 activity assay did not further proceed. Viral RNA copies per milliliter were determined
439 using a synthetic RNA fragment to amplify the target region. The linearized plasmid
440 containing S gene of COVID-19 virus was subjected to *in vitro* transcription. The resulting
441 RNA transcripts were purified and then quantified using spectrophotometry on Nanodrop
442 2000 (Thermo Fisher Scientific, USA). The purified RNA was diluted 10-fold serially
443 using RNase-free water and was detected using qRT-PCR. Threshold cycle (Ct) values for
444 the known concentrations of the RNA were plotted against the log of the number of genome
445 equivalent copies. The resultant standard curve was used to determine the number of
446 genome equivalents of vRNA in the samples. The determination of the detection limit was

447 based on the lowest level at which vRNA was detected and remained within the range of
448 linearity of a standard curve (Ct value of 38). TaqMan primers for COVID-19 virus are 5'-
449 TCCTGGTGATTCTTCTTCAGG-3' and 5'-TCTGAGAGAGGGTCAAGTGC-3' with
450 COVID-19 virus probe 5'-FAM-AGCTGCAGCACCAGCTGTCCA-BHQ1-3'. The
451 cytotoxicity of the tested drugs on Vero cell were determined by MTS cell proliferation
452 assays (Promega, USA). 1×10^4 cells were seeded into a 96-well plate and incubated for 20-
453 24 h at 37 °C. After that, the medium was removed, and 100 μl of medium containing
454 decreasing concentrations of antiviral compounds were added to the wells. After 4 days
455 incubation at 37 °C, MTS assays were performed according to manufacturer's protocols.
456 All experiments were performed in triplicate. Vero cells were obtained from ATCC
457 （American Type Culture Collection）with authentication service. All cell lines were tested
458 negative for mycoplasma contamination. No commonly misidentified cell lines were used.

459 **Antiviral and cytotoxicity assays for cinanserin**

460 For the antiviral assay, a clinical isolate COVID-19 virus[3] was propagated in Vero E6 cells,
461 and viral titer was determined as described previously[43]. All the infection experiments were
462 performed at BSL-3. Pre-seeded Vero E6 cells (5×10^4 cells/well) were pre-treated with the
463 different concentrations of cinanserin for 1 h and the virus was subsequently added (MOI
464 of 0.05) to allow infection for 2 h. Then, the virus-drug mixture was removed and cells
465 were further cultured with fresh drug containing medium. At 24 h p.i., the cell supernatant
466 was collected and vRNA in supernatant was subjected to qRT-PCR analysis. For
467 cytotoxicity assays, Vero E6 cells were suspended in growth medium in 96-well plates. The
468 next day, appropriate concentrations of cinanserin were added to the medium. After 24 h,
469 the relative numbers of surviving cells were measured by CCK8 (Beyotime, China) assay

470 in accordance with the manufacturer's instructions. All experiments were performed in
471 triplicate. Vero E6 cells were obtained from ATCC with authentication service. All cell lines
472 were tested negative for mycoplasma contamination. No commonly misidentified cell lines
473 were used.

474 **Plaque-reduction assays**

475 1×10^5 Vero E6 cells were seeded in a 24-well plate and treated with different doses of the
476 inhibitors. All the infection experiments were performed at BSL-3. Inhibitors with different
477 dilution concentrations were mixed with COVID-19 virus (100 PFU), 200 μl mixtures were
478 inoculated onto monolayer Vero E6 cells for 1 h. After removing the supernatant, the plate
479 was washed twice with DMEM medium, cells were incubated with 0.9% agarose
480 containing appropriate concentrations of inhibitors. The overlay was discarded at 4 days
481 p.i. and cells were fixed for 30 min in 4% polyoxymethylene and stained with crystal violet
482 working solution. The plaque forming units were determined. All experiments were
483 performed in four biological replicates.

484 **Intact protein analysis**

485 2.5 μl of compounds (10 mM in DMSO) were added into 50 μl of COVID-19 virus M^{pro}
486 (10 mg/ml). The mixtures were kept in room temperature for 30 min. Liquid
487 chromatography-mass spectrometry (LC-MS) analyses were performed in positive-ion
488 mode with a quadrupole-time-of-flight (QTOF) mass spectrometer (Agilent 6550, USA)
489 coupled with a high-performance liquid chromatograph (HPLC, Agilent 1260, USA) for
490 detecting the molecular weight of intact proteins. The samples were eluted from a
491 Phenomenex Jupiter C4 300Å LC Column (2×150 mm, 5 μm) over a 15 min gradient from
492 5% to 100% acetonitrile containing 0.1% formic acid at a flow rate of 0.5 ml/min. The

493 acquisition method in positive-ion mode with Dual Agilent Jet Stream electrospray voltage
494 used a capillary temperature of 250 °C, a fragmentor of 175 V, a capillary voltage of 3000
495 V. Mass deconvolution was performed using Agilent MassHunter Qualitative Analysis
496 B.06.00 software with BioConfirm Workflow.

497 **Tandem MS/MS analysis**

498 The samples were precipitated and redissolved by 8 M urea, and then digested for 16 h at
499 25 °C by chymotrypsin at an enzyme-to-substrate ratio of 1:50 (w/w). The digested
500 peptides were desalted and loaded onto a homemade 30 cm-long pulled-tip analytical
501 column (ReproSil-Pur C18 AQ 1.9 μm particle size, Dr. Maisch GmbH, 75 μm ID× 360
502 μm OD) connected to an Easy-nLC1200 UHPLC (Thermo Fisher Scientific, USA) for
503 mass spectrometry analysis. The elution gradient and mobile phase constitution used for
504 peptide separation were as follows: 0-1 min, 4%-8% B; 1-96 min, 8-35% B; 96-104 min,
505 35-60% B; 105-120min, 60-100% B (mobile phase A: 0.1% formic acid in water; mobile
506 phase B: 0.1% formic acid in 80% acetonitrile) at a flow rate of 300 nl /min. Peptides eluted
507 from the LC column were directly electro-sprayed into the mass spectrometer with the
508 application of a distal 1.8-kV spray voltage. Survey full-scan MS spectra (from m/z 300–
509 1800) were acquired in the Orbitrap analyzer (Q Exactive, Thermo Fisher Scientific, USA)
510 with resolution r = 70,000 at m/z 400. The top 20 MS/MS events were sequentially
511 generated and selected from the full MS spectrum at a 30% normalized collision energy.
512 The dynamic exclusion time was set at 10 seconds. One acquisition cycle includes one full-
513 scan MS spectrum followed by top 20 MS/MS events, sequentially generated on the first
514 to the twentieth most intense ions selected from the full MS spectrum at a 28% normalized
515 collision energy. The acquired MS/MS data were analyzed UniProtKB *E.coli* database

(database released on Nov. 11, 2016) containing nsp5 using Protein Discoverer 2.1. In order to accurately estimate peptide probabilities and false discovery rates (FDR), we used a decoy database containing the reversed sequences of all the proteins appended to the target database. FDR was set at 0.01. Mass tolerance for precursor ions was set at 20 ppm. Chymotrypsin was defined as cleavage enzyme and the maximal number of missed cleavage sites was set at 4. Protein N-terminus acetylation, methionine oxidation and compounds covalent bindings were set as variable modifications. The modified peptides were manually checked and labeled.

32 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).

33 Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography* **43**, 186-190 (2010).

34 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674 (2007).

35 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).

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

37 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).

38 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).

39 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).

40 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).

41 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).

42 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).

43 Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Research* **30**, 269-271, doi:10.1038/s41422-020-0282-0 (2020).

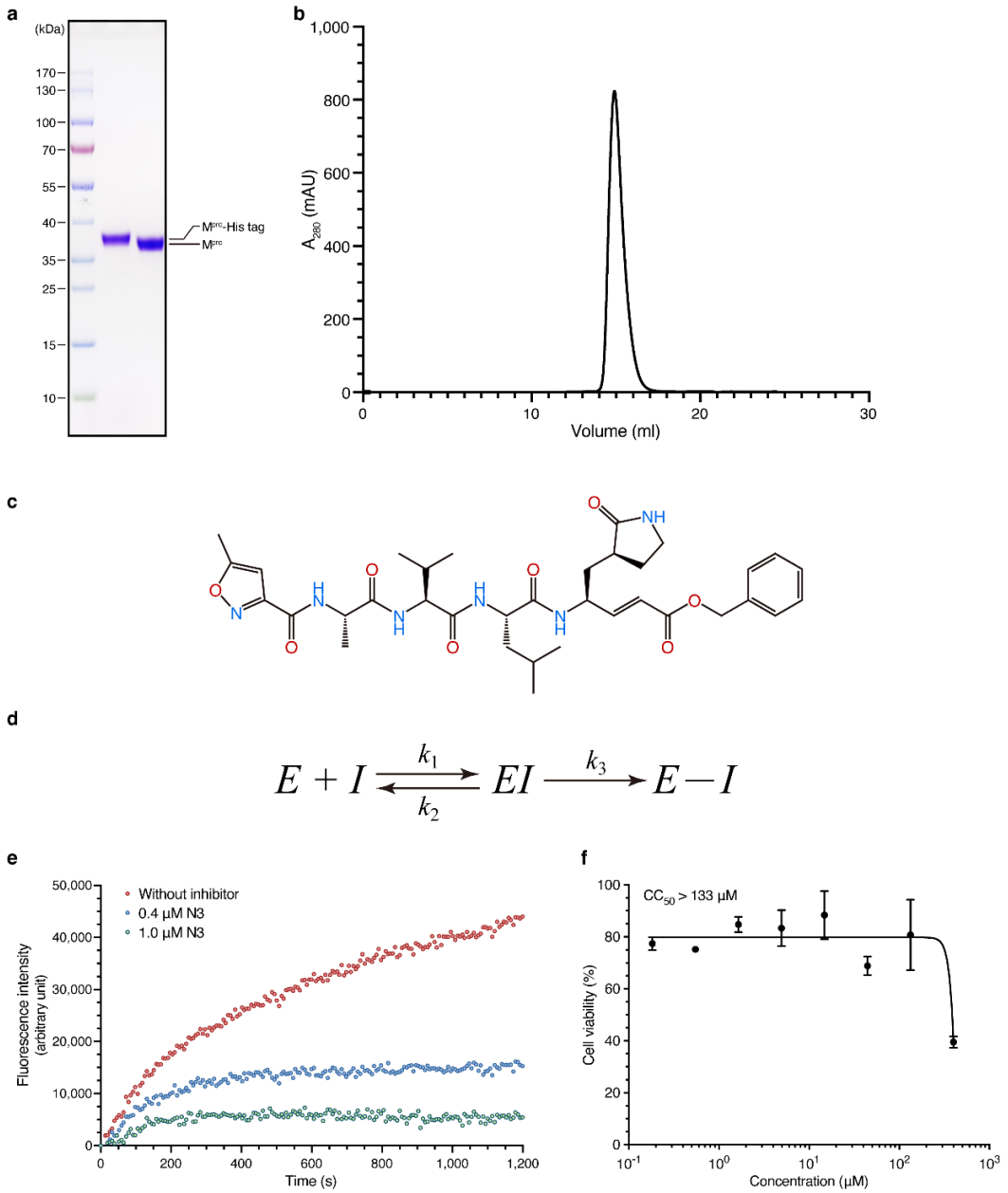
Acknowledgments We would like to thank Ying Lei and Juan Kong from High Throughput Platform, staff from Analytical Chemistry Platform at Shanghai Institute for Advanced

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. 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 to Z.R.), Project of International Cooperation and Exchanges NSFC (grant No. 81520108019 to Z.R.), Science and Technology Commission of Shanghai Municipality (grant No. 20431900200) and Department of Science and Technology of Guangxi Zhuang Autonomous Region (grant No. 2020AB40007).

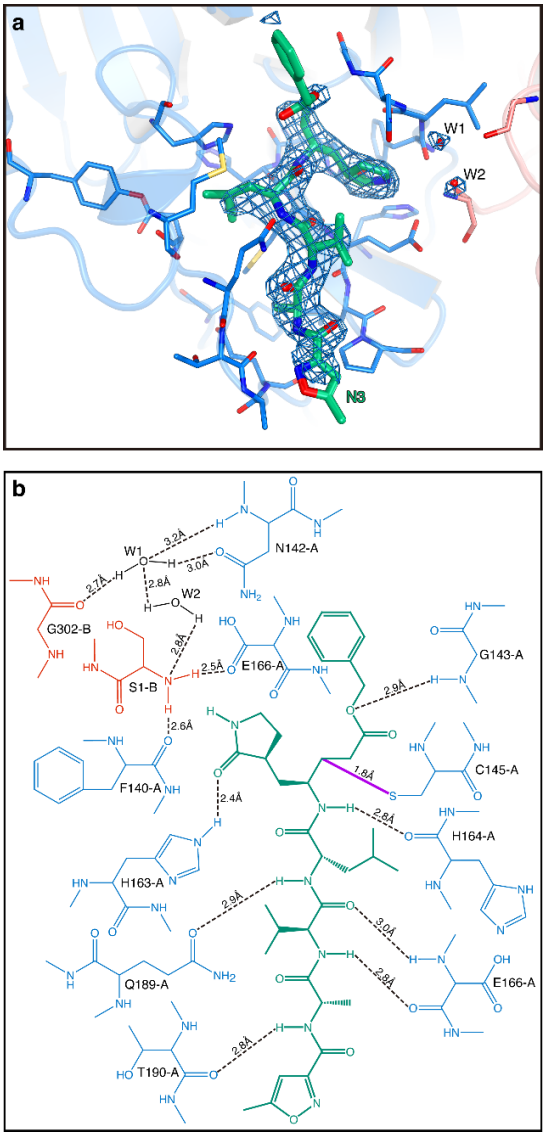
Author contributions Z.R. and H.Y. conceived the project; Z.J., H.J., Z.R., and H.Y. designed the experiments; Z.J., X.D., Y.Duan., J.Y., T.Y., Xiaoce Liu and Xiuna Yang cloned, expressed, purified and crystallized proteins; Z.J., Y.Z., B.Z. and F.L. collected the diffraction data; B.Z. and Xiang Liu solved the crystal structure; Z.J., X.D., Y.Duan. and J.Y. performed enzymatic activity and inhibition assay, high-throughput drug screen and IC50 measurement; L.W. and F.B. performed virtual screening and molecular docking; Y.X., L.Z. and H.L. performed enzymatic inhibition, cell-based antiviral and cytotoxicity assay for cinanserin; Y.Deng. and X. Li performed qRT-PCR analysis and cytotoxicity assay of N3; M.L., R.J. and Xinglou Yang performed plaque-reduction assay; C.P. performed intact protein and tandem MS/MS analyses; Z.J., X.D., Y.X., Y.Deng., C.P., F.B., H.L., Xiang Liu, K.Y., L.G., W.X., 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.

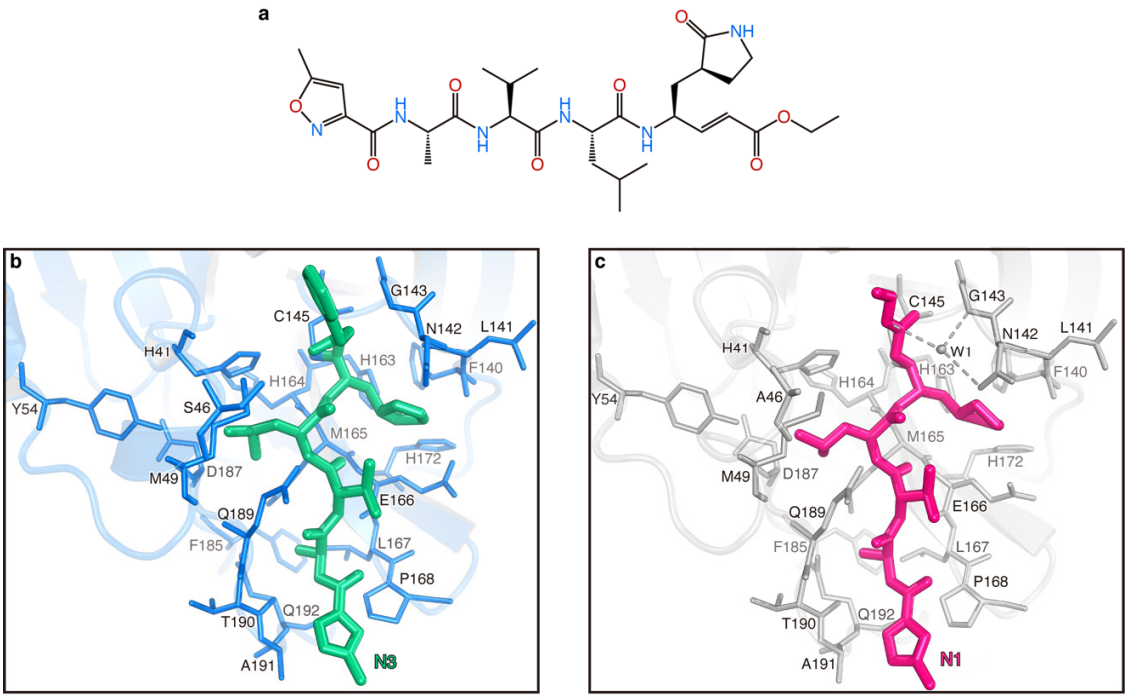
Data and materials availability: The PDB accession No. for the coordinates of COVID-19 virus M^{pro} in complex with N3 is 6LU7 (Deposited: 2020-01-26; Released: 2020-02-05).



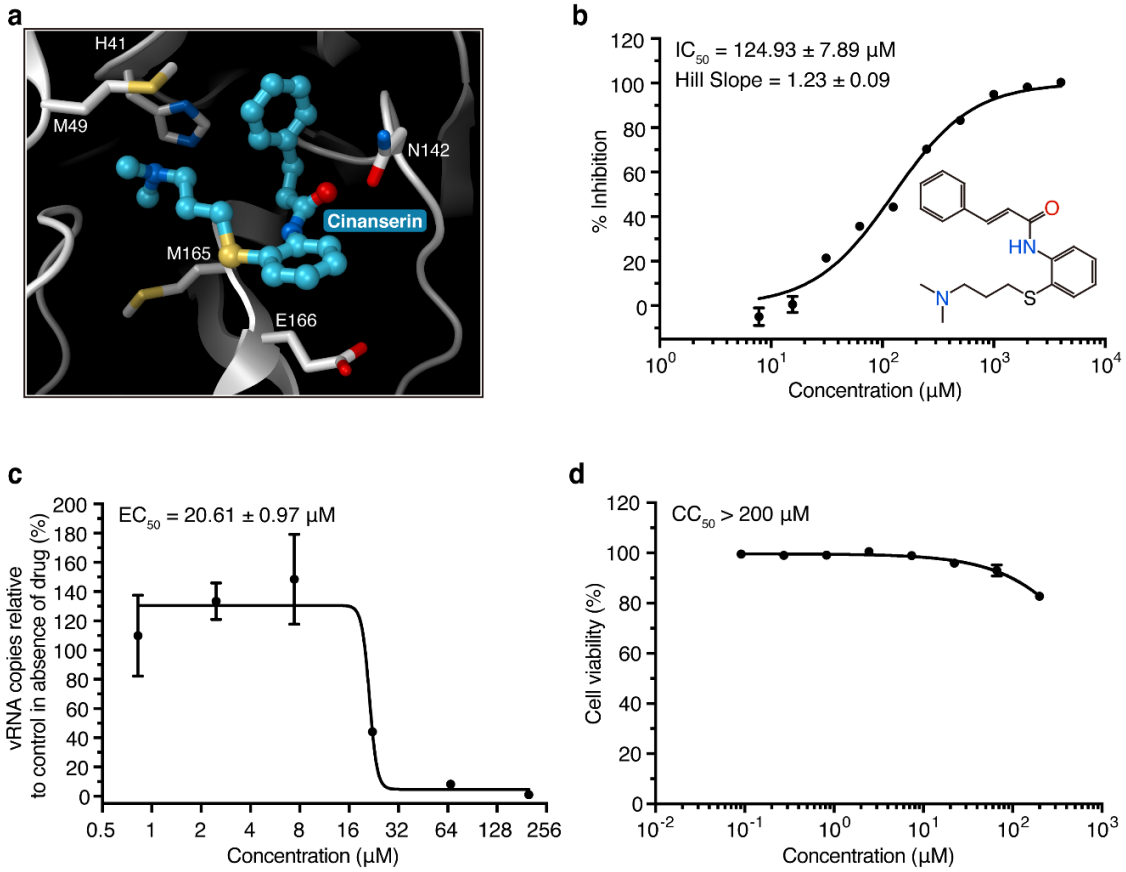
Extended Data Fig. 1 | The purification of COVID-19 virus M^{pro} and the inhibitory assay of N3 compound. a, The SDS-PAGE gel of COVID-19 virus M^{pro}. The first lane: marker; the second lane: M^{pro} before treating with rhinovirus 3C protease; third lane: M^{pro} after the cleavage of C-terminal His tag. For gel source data, see Supplementary Figure 1. **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. **f,** Cytotoxicity assay of N3 on Vero cells, data are shown as mean ± s.e.m., $n = 3$ biological replicates. The data (**a, b, f**) are representative of three independent experiments with similar results.



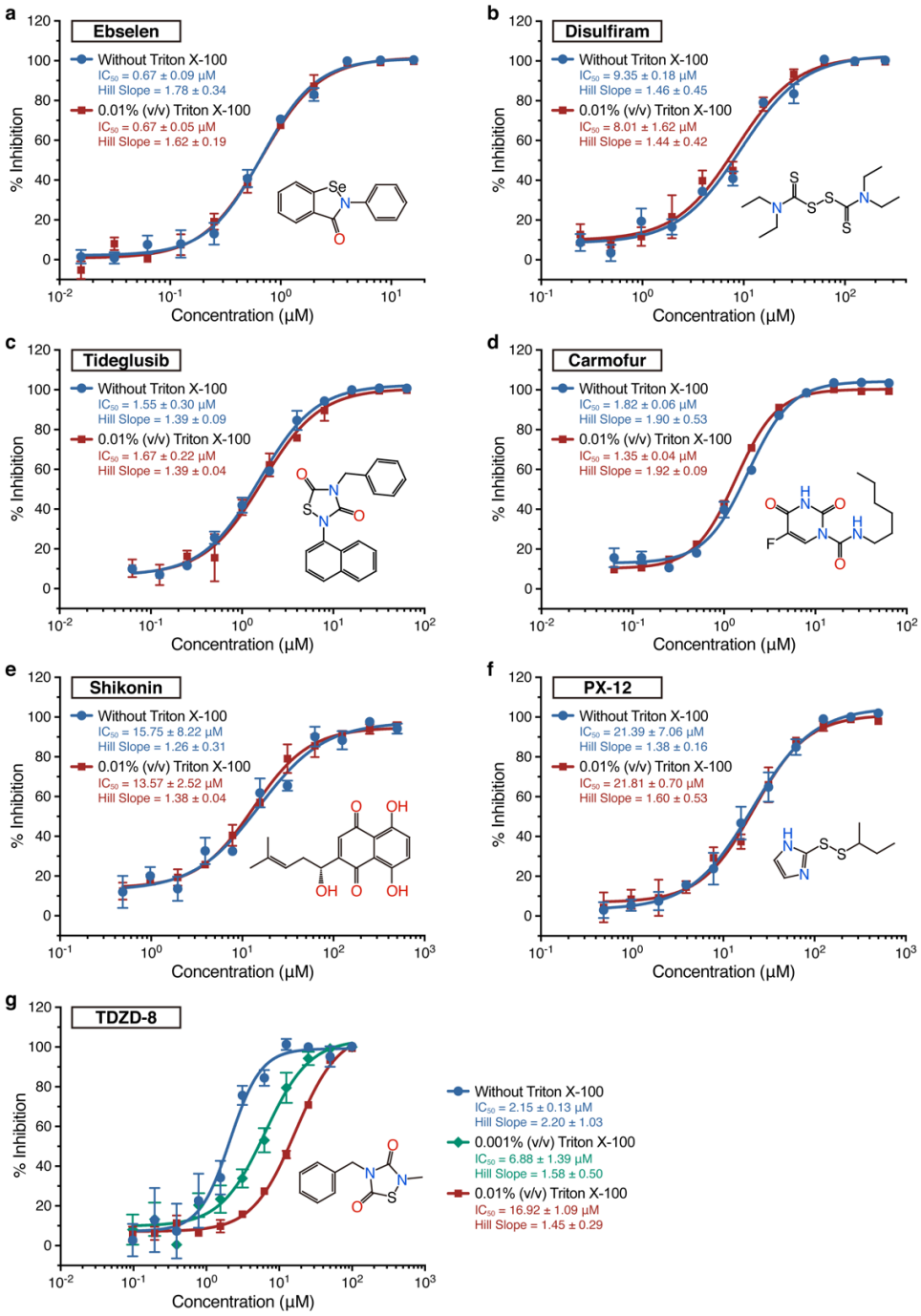
Extended Data Fig. 2 | The interactions between COVID-19 virus M^{pro} and N3. a, The F_o-F_c omit map (contour level = 3 σ, shown as the blue mesh). **b,** Detailed view of the interactions between the inhibitor N3 and COVID-19 virus M^{pro}. M^{pro} residues are shown in blue (Protomer A) and salmon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



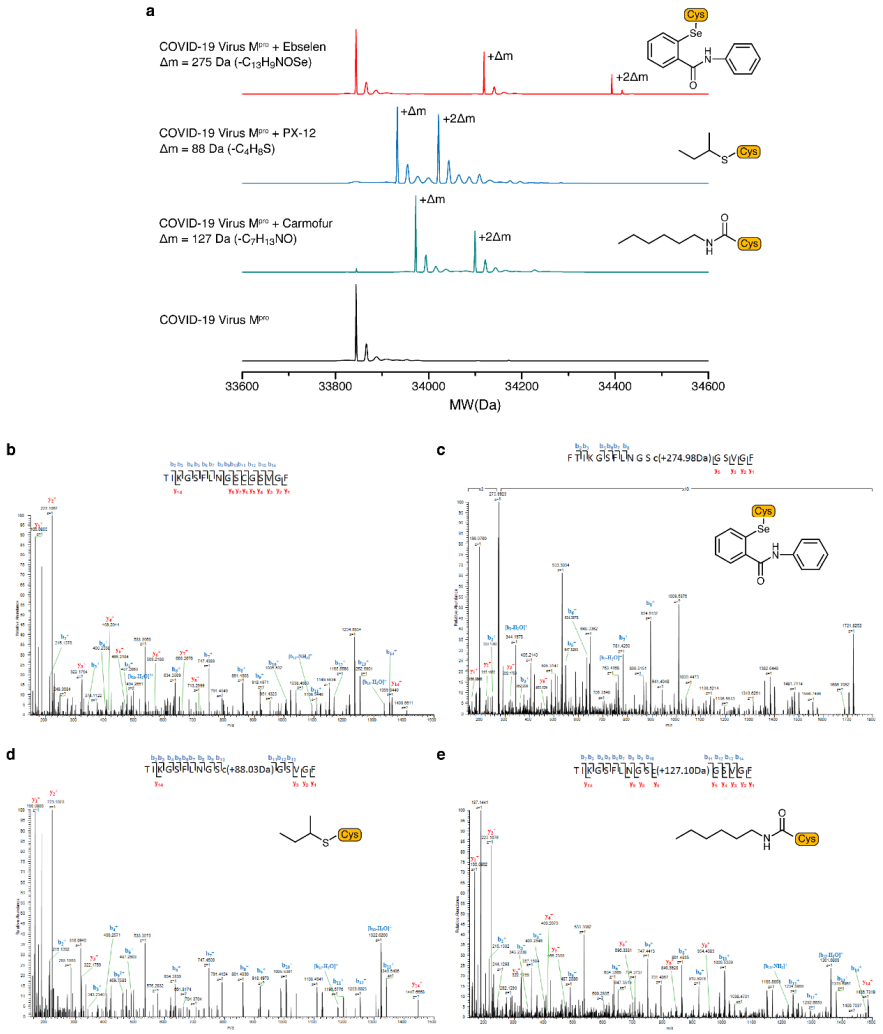
Extended Data Fig. 3 | Comparison of the binding modes between COVID-19 virus M^{pro}-N3 and SARS-CoV M^{pro}-N1. a, The chemical structure of N1 inhibitor. **b,** The binding mode of COVID-19 virus M^{pro} (blue sticks) with N3 (green sticks). **c,** The binding mode of SARS-CoV M^{pro} (grey sticks) with N1 (pink sticks). The hydrogen bonds formed by water (W1) are indicated by the dashed lines.



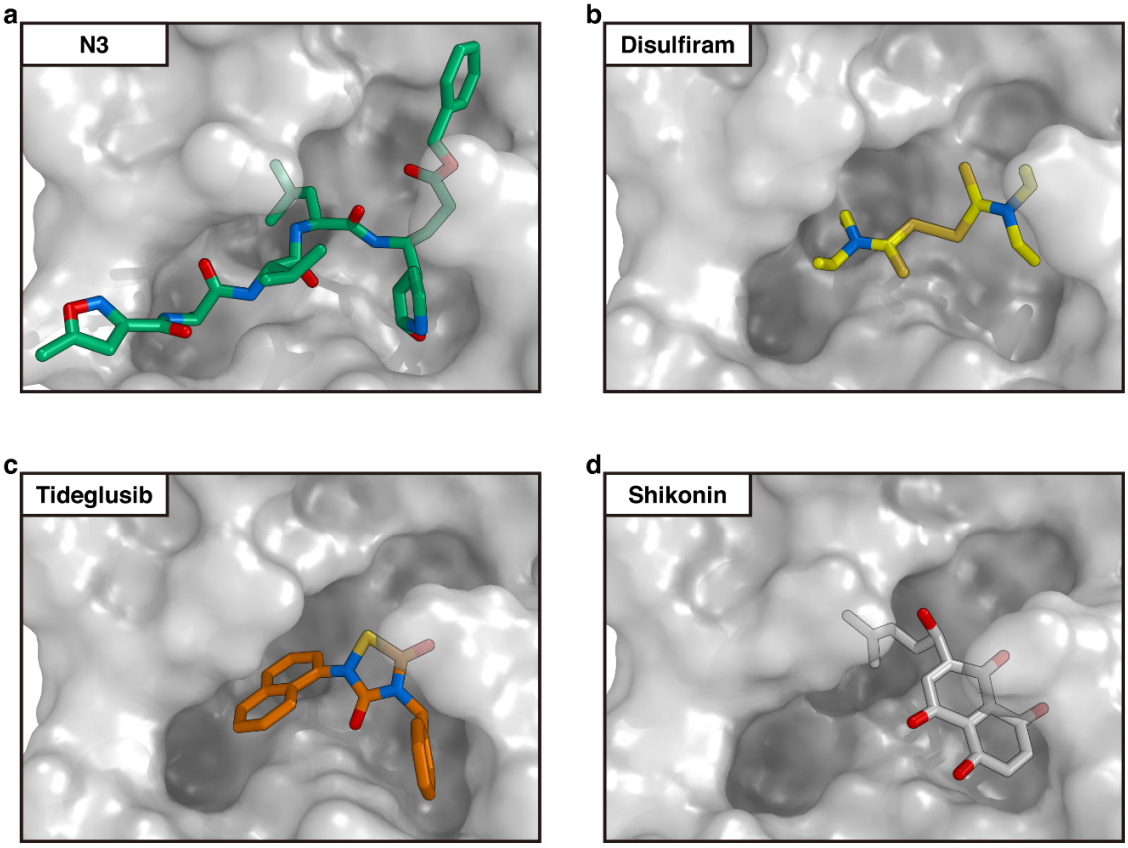
Extended Data Fig. 4 | Cinanserin is an inhibitor for COVID-19 virus M^{pro}. a, The docking result of cinanserin. The structure of COVID-19 virus M^{pro} is shown as a white cartoon, cinanserin is shown as cyan balls and sticks, residues predicted to be interacting with cinanserin are shown as sticks. **b,** Inhibitory activity of cinanserin on M^{pro}. **c,** Antiviral activity of cinanserin determined by qRT-PCR. **d,** Cytotoxicity assay of cinanserin on Vero E6 cells. All data are shown as mean ± s.e.m., $n = 3$ biological replicates.



Extended Data Fig. 5 | The detergent-based assay for drug leads. a-f, The IC_{50} values determined by in the presence or absence of 0.01% Triton X-100, which showed that detergent did not affect the results. **g,** Different concentrations of Triton X-100 notably affected IC_{50} curves for TDZD-8. All data are shown as mean ± s.e.m., n = 3 biological replicates.



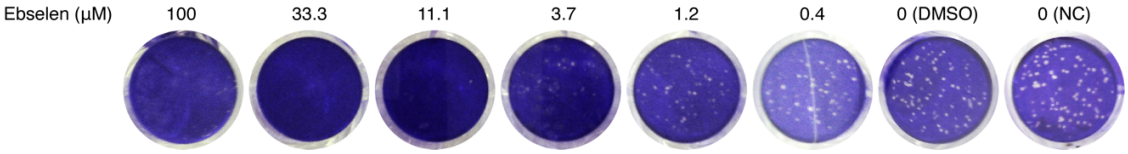
Extended Data Fig. 6 | Tandem MS/MS analysis reveals that ebselen, PX-12 and carmofur are able to covalently bind to C145 of COVID-19 virus M^{pro}. a, Molecular weight of apo COVID-19 virus M^{pro} and compounds treated M^{pro}. The mass shifts (Δm) of the proteins indicate that more than one molecular of the compounds can be covalently bonded to one molecular of M^{pro}. **b-e,** A higher-energy collisional dissociation (HCD) MS/MS spectrum recorded on the $[M+H]^{2+}$ ion **b,** at m/z 787.3852 of the M^{pro} unmodified peptide TIKGSFLNGSCGSVGF, **c,** at m/z 998.4152 of the M^{pro} modified peptide FTIKGSFLNGSCGSVGF harboring a modification ($-C_{13}H_9NOSe$) induced by ebselen on C145, **d,** at m/z 831.4080 of the M^{pro} modified peptide TIKGSFLNGSCGSVGF harboring a modification($-C_4H_8S$) induced by PX-12 on C145, **e,** at m/z 850.9414 of the M^{pro} modified peptide TIKGSFLNGSCGSVGF harboring a modification($-C_7H_{13}NO$) induced by carmofur on C145. Predicted b- and y-type ions (not including all) are listed above and below the peptide sequence, respectively. The experiment was performed once.



e

Drug	Predicted ΔG_{bind} (kcal/mol)
Disulfiram	-46.16
Tideglusib	-61.79
Shikonin	-47.35

Extended Data Fig. 7 | Docking Poses of different COVID-19 virus M^{pro} inhibitors. a, The crystal structure of COVID-19 virus M^{pro}-N3 complex**. b-d,** The docking results of three drug leads. M^{pro} is shown as grey background, inhibitors are in different colors. The inhibitors identified through the high-throughput screening are likely to occupy the same pocket as N3. **e,** Predicted binding affinities for the drug leads to COVID-19 virus M^{pro} by using MM-GBSA module integrated in Schrödinger.



Extended Data Fig. 8 | Images for the plaque-reduction assay (ebselen). As the concentration of the inhibitor (ebselen) increases, there is a significant reduction in the numbers of the plaques by comparison with NC (negative control) and DMSO. Results are shown as representative of four biological replicates. For image source data, see Supplementary Figure 2.

Extended Data Table 1 | Data collection and refinement statistics

	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

*A single crystal was used for data collection and structure determination. †Values in parentheses are for highest-resolution shell.

651