SARS-CoV-2 M^{pro} inhibitors and activity-based probes for patient-sample imaging

Wioletta Rut[1*], Katarzyna Groborz[1,9], Linlin Zhang[2,3,4,9], Xinyuanyuan Sun[2,3,4], Mikolaj Zmudzinski[1], Bartlomiej Pawlik[6,7], Xinyu Wang[8], Dirk Jochmans[8], Johan Neyts[8], Wojciech Młynarski[6], Rolf Hilgenfeld[2,4,5], Marcin Drag[1*]

[1]Department of Chemical Biology and Bioimaging, Wroclaw University of Science and Technology, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland
[2]Institute of Molecular Medicine, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany
[3]Institute of Biochemistry, Center for Structural and Cell Biology in Medicine, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany
[4]German Center for Infection Research (DZIF), Hamburg-Lübeck-Borstel-Riems Site, University of Lübeck, 23562 Lübeck, Germany
[5]Center for Brain, Behavior, and Metabolism (CBBM), University of Lübeck, Lübeck, Germany
[6]Department of Pediatrics, Oncology & Hematology, Medical University of Lodz, Sporna 36/50, 91-738 Lodz, Poland
[7]Postgraduate School of Molecular Medicine, Medical University of Warsaw, Trojdena 2a, 02-091 Warsaw, Poland
[8]Laboratory of Virology and Chemotherapy, Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, 3000, Leuven, Belgium
[9]Equal contribution

- corresponding authors: wioletta.rut@pwr.edu.pl, marcin.drag@pwr.edu.pl

Abstract

In December 2019, the first cases of infection with a novel coronavirus, SARS-CoV-2, were diagnosed. Currently, there is no effective antiviral treatment for COVID-19. To address this emerging problem, we focused on the SARS-CoV-2 main protease which constitutes one of the most attractive antiviral drug targets. We have synthesized a combinatorial library of fluorogenic substrates with glutamine in the P1 position. We used it to determine the substrate preferences of the SARS-CoV and SARS-CoV-2 main proteases. On the basis of these findings, we designed and synthesized a potent SARS-CoV-2 inhibitor (Ac-Abu-DTyr-Leu-Gln-VS, EC_{50} = 3.7 μM) and two activity-based probes, for one of which we determined the crystal structure of its complex with the SARS-CoV-2 M^{pro}. We visualized active SARS-CoV-2 M^{pro} within nasopharyngeal epithelial cells of patients suffered from COVID-19 infection. The results of our work provide a structural framework for the design of inhibitors as antiviral agents and/or diagnostic tests.

Introduction

In December 2019, a severe respiratory disease of unknown origin emerged in Wuhan, Hubei province, China[1,2]. Symptoms of the first patients were flu-like and included fever, cough and myalgia, but with a tendency to develop a potentially fatal dyspnea and acute respiratory distress syndrome[2]. Genetic analysis confirmed a betacoronavirus as the causing agent. The virus was initially named 2019 novel coronavirus (2019-nCoV)[1-3], but shortly thereafter, it was renamed to SARS-CoV-2[4]. By August 28, 2020, the WHO had registered >24 million cumulative cases of coronavirus disease 2019 (COVID-19), with 836 475 deaths[5].

Currently, there is no approved vaccine or treatment for COVID-19. Efforts are being made to characterize molecular targets, pivotal for the development of anti-coronaviral therapies[6]. The main protease (M^{pro}, also known as $3CL^{pro}$), is one of the coronavirus non-structural proteins (Nsp5) designated as a potential target for drug development[7,8]. M^{pro} cleaves the viral polyproteins, generating twelve non-structural proteins (Nsp4-Nsp16), including the RNA-dependent RNA polymerase (RdRp, Nsp12) and the helicase (Nsp13). Inhibition of M^{pro} would prevent the virus from replication and therefore constitutes one of the potential anti-coronaviral strategies[7-9].

Due to the close phylogenetic relationship between SARS-CoV-2 and SARS-CoV[3,10,11], their main proteases share many structural and functional features. From the perspective of the design and synthesis of new M^{pro} inhibitors, a key feature of both the enzymes is their ability to cleave the peptide bond following Gln. The SARS-CoV M^{pro} cleaves polyproteins mainly within the Leu-Gln↓(Ser, Ala, Gly) sequence (↓ indicates the cleavage site), which appears to be a preferred pattern of this protease[7,9,12]. The specificity for peptide bond hydrolysis after Gln residues is also observed for main proteases of other coronaviruses[13,14] but is rare for human enzymes. This observation, along with further studies on the M^{pro}, can potentially lead to new broad-spectrum anti-coronaviral inhibitors with minimum side effects[15].

In the present study, we applied the HyCoSuL (Hybrid Combinatorial Substrate Library) approach to determine the full substrate specificity profile of SARS-CoV and SARS-CoV-2 M^{pro}s. The use of natural and a large number of unnatural amino acids with diverse chemical structures allowed an in-depth characterization of the residue preference of the binding pockets within the substrate-binding site of the proteases. The results from library screening enabled us to design and synthesize ACC-labeled substrates with improved catalytic efficiency in comparison to a substrate containing

only natural amino acids. Moreover, results from our studies clearly indicate that SARS-CoV M^{pro} and SARS-CoV-2 M^{pro} exhibit highly overlapping substrate specificities. We have used this knowledge to design activity-based probes (ABPs) specific for the SARS-CoV main proteases as well as tetrapeptide inhibitors. Further, we present a crystal structure of the SARS-CoV-2 M^{pro} in complex with one of the ABPs. Finally, using the ABP, we were able to visualize active SARS-CoV-2 M^{pro} within nasopharyngeal epithelial cells of patients with active COVID-19 infection. These data provide a useful basis for the design of chemical compounds for effective diagnosis and therapy of COVID-19.

Results

Substrate specificity of SARS-CoV and SARS-CoV-2 M^{pros}

To determine the SARS-CoV M^{pro} and SARS-CoV-2 M^{pro} substrate preferences, we applied a hybrid combinatorial substrate library (HyCoSuL) approach. The library consists of three sublibraries, each of them comprising a fluorescent tag – ACC (7-amino-4-carbamoylmethylcoumarin), two fixed positions and two varied positions containing an equimolar mixture of 19 amino acids (Mix) (P2 sublibrary: Ac-Mix-Mix-X-Gln-ACC, P3 sublibrary: Ac-Mix-X-Mix-Gln-ACC, P4 sublibrary: Ac-X-Mix-Mix-Gln-ACC, X = 19 natural and over 100 unnatural amino acids, **Extended Data Fig. 1**). We incorporated glutamine at the P1 position, because the available crystal structures of SARS-CoV M^{pro} revealed that only glutamine (and, at only one cleavage site, histidine) can occupy the S1 pocket of this enzyme[12,16]. The imidazole of His163, located near the bottom of the S1 pocket, is suitably positioned to interact with the Gln side chain. The Gln is also involved in two other interactions, i.e. with the main chain of Phe140 and the side-chain of Glu166. The library screen revealed that SARS-CoV and SARS-CoV-2 M^{pro} display very similar substrate specificity (**Figure 1**). The most preferred amino acid at the P2 position is leucine in case of both proteases. SARS-CoV M^{pro} exhibits lower activity toward other tested amino acids at this position (<30%). Leu selectivity for SARS-CoV M^{pro} is in high agreement with a previous report by Zhu et al.[17]. However, we have noticed some discrepancies in the level of selectivity for other natural amino acids. This is the result of differences in the substrate specificity profiling approach, where we are using a combinatorial mixture of natural amino acids, while Zhu et al. were using a defined peptide sequence. Nevertheless, in both approaches,

preferences for the same natural amino acids was observed. The S2 pocket of both investigated enzymes can accommodate other hydrophobic residues, such as 2-Abz, Phe(4-NO_2), 3-Abz, β-Ala, Dht, hLeu, Met, and Ile (amino acid structures are presented in **Supplementary Table 1**). At the P3 position, both enzymes prefer hydrophobic D and L amino acids and also positively charged residues; the best are: Tle, D-Phe, D-Tyr, Orn, hArg, Dab, Dht, Lys, D-Phg, D-Trp, Arg, and $Met(O)_2$. At the P4 position, SARS-CoV and SARS-CoV-2 M^{pro} possess broad substrate specificity. The most preferred are small aliphatic residues such as Abu, Val, Ala, and Tle, but other hydrophobic amino acids are also accepted. These findings can be partly explained by the available crystal structures of SARS-CoV M^{pro} in complex with inhibitors[8,12,16]. The hydrophobic S2 subsite of SARS-CoV M^{pro} is more flexible compared to alphacoronavirus M^{pro}s, which explains less stringent specificity[15]. The S2 pocket can form hydrophobic interactions with P2 residues that are not only limited to leucine. The S3 pocket of SARS-CoV M^{pro} is not well defined, which is also reflected in our P3 substrate specificity profile. The S4 pocket can be occupied by small residues due to the crowded cavity formed by Pro168, Leu167 at the bottom and Thr190, Ala191 at the top wall.

To validate the results from library screening, we designed and synthesized ACC-labeled substrates containing the most preferred amino acids in each position. Then, we measured the rate of substrate hydrolysis (chemical compounds 1-8) relevant to each protease (**Extended Data Fig. 2**). The data clearly demonstrate that SARS-CoV M^{pro} and SARS-CoV-2 M^{pro} exhibit the same activity toward tested substrates. The results are consistent with the HyCoSuL screening data. The most preferred substrate, Ac-Abu-Tle-Leu-Gln-ACC (1), is composed of the best amino acids in each position **(Supplementary Table 2)**. Kinetic parameters were determined for the two best substrates (Ac-Abu-Tle-Leu-Gln-ACC, 1; Ac-Thz-Tle-Leu-Gln-ACC, 2) and one containing the best recognized natural amino acids (Ac-Val-Lys-Leu-Gln-ACC, 3) (**Table 1, Extended Data Fig. 3**) toward SARS-CoV-2 M^{pro}. Due to substrate precipitation because of high concentration needed in the assay, kinetic parameters toward SARS-CoV M^{pro} could not be determined. Analysis of kinetic parameters revealed that these three substrates differ in the k_{cat} value, while K_M values are comparable. Since SARS-CoV-2 M^{pro} exhibits low activity toward tetrapeptide substrates, we synthesized internally quenched fluorescent substrates (IQF)

containing the same amino acids at the P4-P1 positions as substrates QS1, QS2, and QS4. The general sequence of IQF substrates was derived from the N-terminal autocleavage sequence of the viral protease reported previously (MCA-AVLQSGFRK(Dnp)K)[18]; instead of the MCA fluorophore, ACC was used. We also incorporated a small linker (glycine residue) between ACC and the amino acid at the P4 position for efficient synthesis. Specificity constant values (k_{cat}/K_M) for these IQF substrates were much higher than for tetrapeptide substrates. IQF substrates containing the same amino acids at P4-P1 positions as substrates QS1 (ACC-Gly-Abu-Tle-Leu-Gln-Ser-Gly-Phe-Arg-Lys(dnp)-Lys-NH_2, 9) and QS4 (ACC-Gly-Thz-Tle-Leu-Gln-Ser-Gly-Phe-Arg-Lys(dnp)-Lys-NH_2, 10) were better recognized than the IQF substrate possessing the same peptide sequence as the QS2 (ACC-Gly-Val-Lys-Leu-Gln-Ser-Gly-Phe-Arg-Lys(dnp)-Lys-NH_2, 11) substrate (**Supplementary Table 3**). This is in line with kinetic results obtained for tetrapeptide substrates.

Inhibitors and activity-based probes of SARS-CoV-2 M^{pro}

In the next step, the best substrate QS1 was converted to an inhibitor and activity-based probes. The inhibitor contained the acetylated peptide sequence and vinyl sulfone as an irreversible reactive group (Ac-Abu-Tle-Leu-Gln-VS, Ac-QS1-VS, 12, **Figure 2a**). The probes included an N-terminal biotin (B) tag or the Cyanine 5 dye (Cy5), polyethylene glycol (PEG(4)) as a linker, the best peptide sequence, and vinyl sulfone (**Figure 2a, Extended Data Fig. 4a**). To evaluate the sensitivity of the designed probes, we performed SDS-PAGE analysis followed by protein transfer onto membranes and ABP visualization. We observed SARS-CoV-2 M^{pro} (100 nM) labelling by B-QS1-VS, 13 at a concentration of 200 nM and by Cy5-QS1-VS, 14 at 100 nM (**Figure 2b, Extended Data Fig. 4b**) which reflected the results of the k_{obs}/I analysis (**Table 2, Extended Data Fig. 5**). In our previous studies, we showed that an activity-based probe with low potency can be applied to detect protease activity in cell assays[19]. To determine the probe selectivity, we performed cell lysate assays. A HeLa lysate was incubated with different probe concentrations (50, 100, and 200 nM) (**Extended Data Fig. 4c**). The cell lysate experiment confirmed probe selectivity. To verify that unknown bands (about 30 kDa and between 49 and 62 kDa) were due to unspecific Cy5 labelling, we incubated the cell lysate with inhibitor (Ac-QS1-VS) for 30 min at 37°C and then with Cy5-QS1-VS for 15 min at 37°C (lanes 8-10 on the membrane, **Extended Data Fig. 4c**). The same protein bands were observed on the

membrane when cell lysates were incubated with and without Ac-QS1-VS, which confirmed unspecific protein labelling by the Cy5 dye. Due to this unspecific protein labelling by Cy5-QS1-VS, we decided to slightly modify the probe structure. Since amino acids with D-stereochemistry are not recognized by many human proteases[20-23], we changed Tle at the P3 position to D-Tyr, which is well recognized by SARS-CoV-2 M^{pro} at this position, in order to improve probe selectivity. The Cy5 fluorophore was exchanged to the Bodipy FL fluorophore. The designed inhibitor (Ac-Abu-DTyr-Leu-Gln-VS, Ac-QS5-VS, 15) and the activity-based probe (Bodipy-PEG(4)-Abu-DTyr-Leu-Gln-VS, Bodipy-QS5-VS, 16) were tested on the recombinant enzyme and cell lysates (**Figure 2**). Kinetic analysis revealed that Ac-QS5-VS exhibited the same potency toward SARS-CoV-2 M^{pro} as Ac-QS1-VS (**Table 2**). The designed ABP, Bodipy-QS5-VS, was slightly less active toward the main protease than Cy5-QS1-VS. To assess Bodipy-QS5-VS sensitivity, we incubated SARS-CoV-2 M^{pro} (at three different concentrations: 5, 10, and 50 nM) with different probe concentrations for 30 minutes. We observed the detection of SARS-CoV-2 M^{pro} at 5 nM by 2.5 μM Bodipy-QS5-VS (**Figure 2c**). The enzyme labelling by the activity-based probe is noticeable after 5 minutes of incubation (**Figure 2d**). Cell lysate assays (HeLa and HEK-293T) confirmed Bodipy-QS5-VS selectivity (**Figure 2e**). Ac-Abu-DTyr-Leu-Gln-VS was also tested for its antiviral activity using a fully replication-competent SARS-CoV-2 strain in Huh7 cell culture. Here the inhibitor showed selective inhibition (EC_{50} = 3.7 μM; interquartile range (IQR) 3.5-4.0 μM) with no significant toxicity (CC_{50} > 100 μM). As a comparison, the parent nucleoside of remdesivir (GS-441524) showed under the same conditions an EC_{50} of 1.15 μM (IQR 1.1-1.2 μM) and a CC_{50} of 87 μM (IQR 82-92 μM) (**Figure 2f, g**).

Crystal structure of SARS-CoV-2 M^{pro} with B-QS1-VS

To visualize the steric details of the interactions between the M^{pro} and the activity-based probe, Biotin-PEG(4)-Abu-Tle-Leu-Gln-VS (B-QS1-VS), we determined the X-ray crystal structure of the complex between the two components. The probe was cocrystallized with the recombinant and highly purified SARS-CoV-2 M^{pro}. Crystals diffracted to 1.7 Å resolution and were of space group $P6_122$, with one ABP-M^{pro} monomer per asymmetric unit (see **Supplementary Table 4** for crystallographic details). The structure was refined to reasonable R factors and good geometry.

Surprisingly, most of the atoms of the activity-based probe were clearly seen in the $2F_o$-F_c electron density at a contour level of 0.5σ. The reason for this is that the flexible tail of the ABP, the PEG(4) chain and the terminal biotin label, are in contact with a neighboring M^{pro} dimer in the crystal lattice (**Extended Data Fig. 6a**). This is of course of little relevance for the situation in solution; hence we discuss only the interaction of the P1-P4 residues with the parent M^{pro} molecule (**Fig. 3**) here. The oxygen atoms of the vinylsulfone group point towards the oxyanion hole of the protease and accept hydrogen bonds (2.81 Å and 3.26 Å) from the main-chain amide groups of Gly143 and Cys145, respectively. Also, the side-chain amide-nitrogen of Asn142 donates a 2.90-Å H-bond to one of these oxygens. The methyl group attached to the sulfone makes hydrophobic contacts with the Cγ2 atom of Thr25 and the side-chain of Leu27, within the S1' subsite. The catalytic cysteine residue is covalently linked to the Cβ atom of the vinyl group, at a distance of 1.83 Å.

The P1-glutamine side-chain makes the expected interaction with His163 (2.72 Å, through the Oε1 atom) and the Phe140 main-chain oxygen (3.22 Å, through Nε2) and Glu166 Oε1 (3.25 Å). The Nε2 of the P1-Gln thus donates a three-center (bifurcated) hydrogen bond to these two acceptors. This is also reflected in inhibitors that carry a γ-lactam as Gln surrogate in the P1 position, such as compound 13b[9]. The Leu residue in the S2 pocket makes the canonical interactions previously observed[17], with residues Met49, Met165, His41, and the Cα atom of Arg188. Also, the main-chain amide of the P2-Leu donates a 2.98-Å H-bond to the side-chain Oε1 of Gln189.

There is no well-defined pocket for the P3 moiety, which is therefore mostly solvent-exposed. There may be some weak interaction of the Tle side-chain with the hydrophobic portion of the neighboring Glu166 side-chain. The polar main-chain atoms of the P3 residue form hydrogen-bonds with the protein main chain at Glu166. The aminobutyric acid in P4 makes weak hydrophobic interactions with Leu167 and Gln189, and its main-chain NH group donates a H-bond to Thr190 O.

We have previously noticed that compared to the SARS-CoV M^{pro}, Thr285 has been replaced by Ala in SARS-CoV-2 M^{pro}, and the neighboring Ile286 by Leu[9]. In the SARS-CoV M^{pro}, the Thr285 makes a hydrogen bond with its symmetry-mate across the two-fold axis creating the M^{pro} dimer. The loss of this H-bond in the SARS-CoV-2 M^{pro} enables the monomers of the dimer to approach each other more closely. Interestingly, in the crystal structure presented here, the space between the two

protomers generated by the mutation is filled by a chloride ion adopted from the crystallization buffer (**Extended Data Fig. 6b**). This might be taken as a hint for this region around residues 284 - 286 being a hotspot for mutations, due to non-ideal packing of the two protomers of the dimer at this point.

SARS-CoV-2 M^{pro} detection and imaging in patient samples

Fluorescent-tagged ABPs currently represent the classic standard in terms of application for labelling of biological samples and have been successfully used for visualization of many proteases in the past[24-26]. We wanted to see if the Cy5-QS1-VS developed by us can be used for detection of SARS-CoV-2 M^{pro} in human samples. Thus, we recruited one patient with mild symptoms of COVID-19, who was positive for SARS-CoV-2 RNA in two independent quantitative RT-PCR assays validated for diagnostic purposes. We incubated the probe at 1 μM final concentration with cells collected from nasopharyngeal swabs of the patient on the first and fifth day after diagnosis and subjected them to confocal laser scanning microscopy. In parallel, we carried out the same experiment with a healthy donor (COVID-19-negative control). These serial measurements revealed that 10-15% of the cells were positive for staining with the Cy5-QS1-VS probe in the COVID-19-positive patient (**Extended Data Fig. 7**). A particularly strong signal from SARS-CoV-2 M^{pro} was observed on the 5^{th} day after diagnosis. No signal from SARS-CoV-2 M^{pro} labelling was observed in the sample from the healthy donor (**Extended Data Fig. 7**). Due to a noticeable background fluorescence signal from the Cy5 dye in the control sample (negative for COVID-19), we changed the probe structure and performed the same experiments using a new probe, Bodipy-QS5-VS, among additional six patients positive for SARS-CoV-2 RNA with clinical signs of COVID-19 and three negative, healthy individuals (control samples). We observed labelling of SARS-CoV-2 M^{pro} in the COVID-19-positive patients by Bodipy-QS5-VS and no signal from the probe in the negative patients (**Figures 4**, **Extended Data Fig. 8a, c**). When the cells (samples positive for COVID-19) were incubated with Ac-QS1-VS inhibitor prior to the Bodipy-QS5-VS probe, SARS-CoV-2 M^{pro} staining was reduced (**Extended Data Fig. 8b**). Thus, we were able

to show for the first time that human cells collected *ex vivo* contain active SARS-CoV-2 M^{pro} during SARS-CoV-2 infection.

Discussion

SARS-CoV-2, first observed in Wuhan, China, caused the current pandemic of COVID-19 disease. The lack of a vaccine or approved medications for direct treatment of the disease has led to strenuous efforts to find therapies to stop the pandemic. Among the promising therapeutic targets are two viral proteases - SARS-CoV-2 M^{pro} and SARS-CoV-2 PL^{pro}. The first of these, SARS-CoV-2-M^{pro}, is used by the virus for protein maturation and its structure has already been described recently[9]. Moreover, the results of retargeting about 10,000 drugs, drug candidates in clinical trials, and other bioactive compounds resulted in selection of several candidates as potential inhibitors of this enzyme[18]. In our research, we decided to thoroughly examine SARS-CoV-2 M^{pro} to find the optimal chemical tools in the form of substrates, inhibitors and activity-based probes. First, we obtained a targeted library of fluorogenic substrates (HyCoSuL) towards this protease and determined the substrate specificity at the P4-P2 positions. We directly compared the substrate specificity of the main protease with the same protease from the previous SARS coronavirus, which caused an epidemic in 2003. Our data clearly demonstrate that these enzymes have very similar preferences for natural and unnatural amino acids at the P4-P2 positions. They tolerate many different amino acids at P4 and P3, and have a strong preference for Leu at P2. This information is certainly crucial for the aspect of drug retargeting, but so is the use of information obtained in previous years for SARS-CoV M^{pro} to be used for current research. In the next step, we created potent inhibitors and activity-based probes, whose sequences were based on the HyCoSuL screening results. This strategy had been successfully applied in our previous studies to design selective inhibitors and activity-based probes for cysteine[24,27], serine[22], and threonine[23] proteases. In the case of B-QS1-VS, we also obtained a crystal structure of its complex with SARS-CoV-2 M^{pro}, which accurately shows the binding mechanism in the P4-P1` pockets. This knowledge certainly complements the information already obtained for other inhibitor molecules published for this enzyme. In turn, we used the fluorescent activity-based probe to visualize the SARS-CoV-2 M^{pro} activity in patient samples with COVID-19,

thus confirming that it is an excellent tool that can be used to detect this enzyme, and also be used as a diagnostic tool.

In conclusion, our research allowed the exact characterization of SARS-CoV-2 M^{pro} in terms of both amino-acid preferences as well as the design of targeted inhibitors and activity-based probes. The reagents described here can be used to optimize structures that lead to anti-COVID-19 drugs, as well as for further drug retargeting.

Acknowledgments

The Drag laboratory is supported by the National Science Centre in Poland and the "TEAM/2017-4/32" project, which is conducted within the TEAM programme of the Foundation for Polish Science cofinanced by the European Union under the European Regional Development Fund. Work in the Hilgenfeld laboratory was supported by the SCORE project of the European Union (grant agreement # 101003627), by the German Center for Infection Research (DZIF), and by the Government of Schleswig-Holstein through its Structure and Excellence Fund, as well as by a close partnership between the Possehl Foundation (Lübeck) and the University of Lübeck. Work in the Neyts laboratory was similarly supported by the SCORE project of the European Union (grant agreement # 101003627). X.W. received funding of the China Scholarship Council (CSC) (grant number 201806170087). L.Z. is supported by a stipend from the German Center for Infection Research (DZIF). W.R. is a beneficiary of a START scholarship from the Foundation for Polish Science. The Mlynarski laboratory (B.P., W.M.) is supported by the FIXNET project, which is conducted within the TEAM-NET programme of the Foundation for Polish Science cofinanced by the European Union under the European Regional Development Fund (no POIR.04.04.00-00-1603/18). We thank Dr. Anna Piekarska (Medical University of Lodz, Poland) for an expert consultation in clinical assessments of COVID-19 patients.

Author contributions

M. D. and W.R. designed the research; W.R., K.G., L.Z., X.W., D.J. and M.Z. performed the research and collected data; R.H., X.S. and L.Z. contributed enzymes; L.Z. and R.H. solved and refined the crystal structure; B.P. and W.M. carried out labelling experiments in patient samples and confocal imaging; W.R., L.Z., W.M., X.W., D.J.,

J.N., R.H. and M.D. analyzed and interpreted the data; W.R., L.Z., R.H. and M.D. wrote the manuscript; and all authors critically revised the manuscript.

Competing interest

Wroclaw University of Science and Technology has filed a patent application covering compounds: Ac-Abu-Tle-Leu-Gln-VS, Ac-Abu-DTyr-Leu-Gln-VS, Biotin-PEG(4)-Abu-Tle-Leu-Gln-VS and Cy5-PEG(4)-Abu-Tle-Leu-Gln-VS as well as related compounds with W.R. and M.D. as inventors.

References:

1. Wang, C., Horby, P.W., Hayden, F.G. & Gao, G.F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470-473 (2020).
2. Huang, C.L. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497-506 (2020).
3. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269 (2020).
4. Jiang, S.B. et al. A distinct name is needed for the new coronavirus. *Lancet* **395**, 949-949 (2020).
5. WHO. Coronavirus disease (COVID-2019) situation reports 5 August 2020. *https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200805-covid-19-sitrep-198.pdf?sfvrsn=f99d1754_2*.
6. Li, G.D. & De Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nature Reviews Drug Discovery* **19**, 149-150 (2020).
7. Hilgenfeld, R. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *Febs Journal* **281**, 4085-4096 (2014).
8. 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 (2016).
9. Zhang, L.L. et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science* **368**, 409-412 (2020).
10. Wu, A.P. et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host & Microbe* **27**, 325-328 (2020).

11. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
12. 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 (2003).
13. Wu, A.D. et al. Prediction and biochemical analysis of putative cleavage sites of the 3C-like protease of Middle East respiratory syndrome coronavirus. *Virus Research* **208**, 56-65 (2015).
14. Chuck, C.P., Chow, H.F., Wan, D.C.C. & Wong, K.B. Profiling of Substrate Specificities of 3C-Like Proteases from Group 1, 2a, 2b, and 3 Coronaviruses. *Plos One* **6** DOI: 10.1371/journal.pone.0027228 (2011).
15. Zhang, L.L. et al. alpha-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *Journal of Medicinal Chemistry* **63**, 4562-4578 (2020).
16. 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 (2003).
17. Zhu, L.L. et al. Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease. *Antiviral Research* **92**, 204-212 (2011).
18. Jin, Z.M. et al. Structure of M-pro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289-293 (2020).
19. Aleshin, A.E. et al. Activity, Specificity, and Probe Design for the Smallpox Virus Protease K7L. *Journal of Biological Chemistry* **287**, 39470-39479 (2012).
20. Poreba, M. et al. Unnatural amino acids increase sensitivity and provide for the design of highly selective caspase substrates. *Cell Death and Differentiation* **21**, 1482-1492 (2014).
21. Kasperkiewicz, P. et al. Design of ultrasensitive probes for human neutrophil elastase through hybrid combinatorial substrate library profiling. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 2518-2523 (2014).
22. Kasperkiewicz, P., Altman, Y., D'Angelo, M., Salvesen, G.S. & Drag, M. Toolbox of Fluorescent Probes for Parallel Imaging Reveals Uneven Location of Serine

Proteases in Neutrophils. *Journal of the American Chemical Society* **139**, 10115-10125 (2017).

23. Rut, W., Poreba, M., Kasperkiewicz, P., Snipas, S.J. & Drag, M. Selective Substrates and Activity-Based Probes for Imaging of the Human Constitutive 20S Proteasome in Cells and Blood Samples. *Journal of Medicinal Chemistry* **61**, 5222-5234 (2018).
24. Poreba, M. et al. Selective imaging of cathepsin L in breast cancer by fluorescent activity-based probes. *Chemical Science* **9**, 2113-2129 (2018).
25. Poreba, M. et al. Small Molecule Active Site Directed Tools for Studying Human Caspases. *Chemical Reviews* **115**, 12546-12629 (2015).
26. Maluch, I., Czarna, J. & Drag, M. Applications of Unnatural Amino Acids in Protease Probes. *Chemistry-an Asian Journal* **14**, 4103-4113 (2019).
27. Poreba, M. et al. Potent and selective caspase-2 inhibitor prevents MDM-2 cleavage in reversine-treated colon cancer cells. *Cell Death and Differentiation* **26**, 2695-2709 (2019).

Figure legends

Figure 1. Substrate specificity profiles of SARS-CoV M^{pro} and SARS-CoV-2 M^{pro} presented as heat maps. The average relative activity is presented as a percentage of the best-recognized amino acid (n=2, where n represents the number of independent experiments).

Figure 2. SARS-CoV-2 M^{pro} detection by activity-based probes.
(a) Structure of inhibitors and activity-based probes. SARS-CoV-2 M^{pro} labelling by probes B-QS1-VS (b) and Bodipy-QS5-VS) (c, d). (d) Time-point analysis of Bodipy-QS5-VS binding to the enzyme. (e) SARS-CoV-2 M^{pro} probe selectivity in HeLa and HEK-293T lysates (asterisk shows the SARS-CoV-2 M^{pro}, which was spiked into the cell lysates). The cell lysate was incubated with or without the inhibitor Ac-QS5-VS for 30 min at 37°C; next, different probe concentrations were added and the samples were incubated for 30 min at 37°C. The biotinylated activity-based probe was detected with a fluorescent streptavidin Alexa Fluor 647 conjugate (1:10000) in TBS-T with 1% BSA using an Azure Biosystems Sapphire Biomolecular Imager (laser 658 nm and 488 nm for Bodipy FL dye) and the Azure Spot Analysis Software. Each experiment was repeated independently two times with similar results. Antiviral activity on SARS-CoV-

2 (f) and cellular toxicity (g) in Huh7 cells of Ac-Abu-DTyr-Leu-Gln-VS and GS-441524. Antiviral activity and toxicity was determined by quantification of viability of infected (f) or uninfected (g) cell cultures (see methods). Two independent dose-response curves were investigated for each compound (indicated by the circles and triangels). Non-linear fitting of the dose response curves results in EC_{50}s of 1.2 and 1.3 μM for GS-441524 and 3.5 and 4.0 μM for Ac-Abu-DTyr-Leu-Gln-VS. Calculated CC_{50}s are 69 and 95 μM for GS-441524 and > 100 μM for Ac-Abu-DTyr-Leu-Gln-VS.

Figure 3. Three-dimensional structure of the P1' - P4 residues of the activity-based probe Biotin-PEG(4)-Abu-Tle-Leu-Gln-VS (B-QS1-VS) in the substrate-binding site of the SARS-CoV-2 M^{pro}. The F_o-F_c electron density for residues P1' - P4 of the ABP is shown at a contour level of 2.5σ. All parts of the ABP beyond the P4 residue would be outside the parent M^{pro} dimer and have been omitted from this figure. However, the rest of the ABP interacts with a neighboring M^{pro} dimer in the crystal (cf. **Extended Data Fig. 6a**) and is seen in the electron density maps.

Figure 4. SARS-CoV-2 M^{pro} detection by the activity-based probe Bodipy-QS5-VS in nasopharyngeal epithelial cells from patients positive for SARS-CoV-2 RNA (P01DZ, P02JB, P03KB, scale bars, 25 μm) and from a healthy control (CTRL 1, scale bars, 50 μm). Confocal microscopy of the epithelial cells of nasopharyngeal swabs co-stained with Bodipy-QS5-VS SARS-CoV-2 M^{pro} probe and anti-ACE2 antibody (ACE2 - angiotensin-converting enzyme 2) with Cy5 AffiniPure donkey anti-Goat secondary antibody and DAPI (4′,6-diamidino-2-phenylindole). Due to ethical concerns one nasopharyngeal swab was collected from each patient.

Tables

Table 1. Kinetic parameters of selected substrates for SARS-CoV-2 M^{pro} (Data presented as mean values ± SD; n=3, where n represents the number of independent experiments).

Substrate	K_M, μM	k_{cat}, s^{-1}	k_{cat}/K_M, $M^{-1}s^{-1}$
Ac-Abu-Tle-Leu-Gln-ACC, QS1 (**1**)	207.3 ± 12	0.178 ± 0.016	859 ± 57
Ac-Thz-Tle-Leu-Gln-ACC, QS4 (**2**)	189.5 ± 2.7	0.144 ± 0.006	760 ± 50
Ac-Val-Lys-Leu-Gln-ACC, QS2 (**3**)	228.4 ± 9.9	0.050 ± 0.002	219 ± 3

Table 2. Inhibition rate constants of SARS-CoV-2 M^{pro} for inhibitors and activity-based probes (Data presented as mean values ± SD**;** n=3, where n represents the number of independent experiments)**.**

Compound	k_{obs}/I, $M^{-1}s^{-1}$
Ac-Abu-Tle-Leu-Gln-VS, Ac-QS1-VS (**12**)	730 ± 17
Biotin-PEG(4)-Abu-Tle-Leu-Gln-VS, B-QS1-VS (**13**)	200 ± 11
Cy5-PEG(4)-Abu-Tle-Leu-Gln-VS, Cy5-QS1-VS (**14**)	591 ± 45
Ac-Abu-DTyr-Leu-Gln-VS, Ac-QS5-VS (**15**)	812 ± 70
Bodipy-PEG(4)-Abu-DTyr-Leu-Gln-VS, Bodipy-QS5-VS (**16**)	439 ± 34

Online Methods

Reagents

The reagents used for solid-phase peptide synthesis were as follows: Rink Amide (RA) resin (particle size 100-200 mesh, loading 0.74 mmol/g), all Fmoc-amino acids, *O*-benzotriazole-*N*,*N*,*N*`,*N*`-tetramethyl-uronium-hexafluoro-phosphate (HBTU), 2-(1-H-7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluranium hexafluorophosphate (HATU), piperidine, diisopropylcarbodiimide (DICI), and trifluoroacetic acid (TFA), purchased from Iris Biotech GmbH (Marktredwitz, Germany); anhydrous *N*-hydroxybenzotriazole (HOBt) from Creosauls, Louisville, KY, USA; 2,4,6-collidine (2,4,6-trimethylpyridine), HPLC-grade acetonitrile, triisopropylsilane (TIPS) from Sigma-Aldrich (Poznan, Poland); and *N,N*-diisopropylethylamie (DIPEA) from VWR International (Gdansk, Poland). *N,N*-dimethylformamide (DMF), dichloromethane (DCM), methanol (MeOH), diethyl ether (Et_2O), acetic acid (AcOH), and phosphorus pentoxide (P_2O_5), obtained from Avantor (Gliwice, Poland). Designed substrates were purified by HPLC on a Waters M600 solvent delivery module with a Waters M2489 detector system using a semipreparative Wide Pore C8 Discovery column. The solvent composition was as follows: phase A (water/0.1% TFA) and phase B (acetonitrile/0.1% TFA). The purity of each compound was confirmed with an analytical HPLC system using a Jupiter 10 µm C4 300 Å column (250 x 4.6 mm). The solvent composition was as follows: phase A (water/0.1% TFA) and phase B (acetonitrile/0.1% TFA); gradient, from 5% B to 95% B over a period of 15 min. The purity of all compounds was ≥95%. The molecular weight

of each substrate was confirmed by high-resolution mass spectrometry using a Waters LCT premier XE with electrospray ionization (ESI) and a time-of-flight (TOF) module.

Enzyme preparation

Gene cloning and recombinant production of the SARS-CoV and SARS-CoV-2 M^{pro} are described elsewhere[9,17].

Combinatorial library synthesis

Synthesis of H_2N-ACC-resin. ACC synthesis was carried out according to Maly et al.[28]. To a glass reaction vessel, 1 eq (9.62 mmol, 13 g) of Rink AM resin was added and stirred gently once per 10 min in DCM for 1 h, then filtered and washed 3 times with DMF. Fmoc-group deprotection was performed using 20% piperidine in DMF (three cycles: 5, 5, and 25 min), filtered and washed with DMF each time (six times). Next, 2.5 eq of Fmoc-ACC-OH (24.05 mmol, 10.64 g) was preactivated with 2.5 eq HOBt monohydrate (24.05 mmol, 3.61 g) and 2.5 eq DICI (24.05 mmol, 3.75 mL) in DMF and the slurry was added to the resin. The reaction mixture was shaked gently for 24 hours at room temperature. After this time, the resin was washed four times with DMF and the reaction was repeated using 1.5 eq of above reagents in order to improve the yield of ACC coupling to the resin. After 24 hours, the resin was washed with DMF and the Fmoc protecting group was removed using 20% piperidine in DMF (5, 5, and 25 min), filtered and washed with DMF (six times).

Synthesis of H_2N-Gln(Trt)-ACC-resin: 2.5 eq Fmoc-Gln(Trt)-OH (24.05 mmol, 14.69 g) with 2.5 eq HATU (24.05 mmol, 9.15 g), 2.5 eq collidine (24.05 mmol, 3.18 mL) in DMF were activated for 2 min and added to filter cannula with 1 eq (9.62 mmol) H_2N-ACC-resin and the reaction was carried out for 24 h. Next, the resin was washed four times with DMF and the same reaction was performed again using 1.5 eq of above reagents. After four DMF washes, the Fmoc protecting group was removed using 20% piperidine in DMF (5, 5, and 25 min). Subsequently, the resin was washed with DCM (3 times) and MeOH (3 times) and dried over P_2O_5.

Library synthesis: The synthesis of P2, P3, and P4 sublibraries is exemplified in detail with the P2 sublibrary. The P2 library consisted of 137 compounds where all of the natural amino acids (omitting cysteine) and a pool of unnatural amino acids were used at a defined position (in this case, the P2 position) and an isokinetic mixture of 19 amino acids (without cysteine; plus norleucine mimicking methionine) was coupled in

the remaining positions (in case of the P2 sublibrary, positions P3 and P4 were occupied by an isokinetic mixture). Equivalent ratios of amino acids in the isokinetic mixture were defined based on their reported coupling rates. A fivefold excess (over the resin load) of the mixture was used. For fixed positions, 2.5 eq of single amino acid was used. All reactions were performed with the use of the coupling reagents DICI and HOBt. For P2 coupling, the synthesis of the library was performed using a MultiChem 48-well synthesis apparatus (FlexChem from SciGene, Sunnyvale, CA, USA). To each well of the reaction apparatus, 1 eq of dry H_2N-Gln(Trt)-ACC-resin (0.059 mmol, 80 mg) was added and stirred gently for 30 minutes in DCM, and then washed four times with DMF. In separate Eppendorf tubes, 2.5 eq (0.15 mmol) Fmoc-P2-OH was preactivated with 2.5 eq HOBt (0.15 mmol, 22.5 mg) and 2.5 eq DICI (0.15 mmol, 23.55 μL) in DMF. Next, preactivated amino acids were added to wells of the apparatus containing H_2N-Gln(Trt)-ACC-resin, followed by 3 h of agitation at room temperature. Then, the reaction mixture was filtered, washed with DMF (4 times), and the ninhydrin test was carried out in order to confirm P2-amino acid coupling. Subsequently, Fmoc protecting groups were removed with the use of 20% piperidine in DMF (5, 5, and 25 min). For P3 and P4 position coupling, an isokinetic mixture for 48 portions was prepared from 18 Fmoc-protected natural amino acids (omitting cysteine; plus norleucine mimicking methionine; 19 amino acids in total). Next, 5 eq of isokinetic mixture, 5 eq HOBt (14.16 mmol, 2.13 g), and 5 eq DICI (14.16 mmol, 2.22 mL) were diluted in DMF and preactivated for 3 min. The activated isokinetic mixture was added to each of 48 wells containing 1 eq of H_2N-P2-Gln(Trt)-ACC-resin. After 3 h of gentle agitation, the slurry was filtered off and washed with DMF (4 times). A ninhydrin test was carried out and the Fmoc protecting group was removed using 20% piperidine in DMF (5, 5, and 25 min). The same procedure was applied for the remaining compounds. The isokinetic mixture was added to prepare the P4 position in the same manner as for the P3 position. In the last step of the synthesis, *N*-terminal acetylation was performed; to prepare the mixture for 48 compounds, 5 eq of AcOH (14.16 mmol, 807 μL), 5 eq HBTU (14.16 mmol, 5.37 g), and 5 eq DIPEA (14.16 mmol, 2.44 mL) in ~45 mL of DMF were added to a 50-mL falcon tube. After gentle stirring for 1 min, the mixture (~800 μL) was added to each well in the reaction apparatus, containing the H_2N-Mix-Mix-P2-Gln(Trt)-ACC-resin, followed by gentle agitation for 30 min. Next, the resin was washed six times with DMF, three times with DCM, three times with MeOH, and dried over P_2O_5. After completing the synthesis, peptides were cleaved from the

resin with a mixture of cold TFA:TIPS:H_2O (%, v/v/v 95:2.5:2.5; 2 mL/well; 2 hours, shaking once per 15 min). The solution from each well was collected separately and the resin was washed once with a portion of fresh cleavage solution (1 mL), followed by addition of diethyl ether (Et_2O, 14 mL) into falcons with peptides in solution. After precipitation (30 min at -20°C), the mixture was centrifuged and washed again with Et_2O (5 mL). After centrifugation, the supernatant was removed and the remaining white precipitate was dissolved in ACN/H_2O (v/v, 3/1) and lyophilized. The products were dissolved in DMSO to a final concentration of 10 mM and used without further purification. The synthesis of P3 and P4 sublibraries was performed in the same manner as described above; P3 and P4 sublibraries were synthesized by coupling fixed amino-acid residues to P3 (isokinetic mixture coupled to P2 and P4) and P4 position (isokinetic mixture coupled to P2 and P3).

Library screening

Hybrid combinatorial substrate library screening was performed using a spectrofluorometer (Molecular Devices Spectramax Gemini XPS) in 384-well plates (Corning). The assay conditions were as follows: 1 µL of substrate and 49 µL of enzyme, which was incubated at 37°C for 10 min in assay buffer (20 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.3). The final substrate concentration was 100 µM and the final enzyme concentration was 1 µM SARS-CoV and 0.6 µM SARS-CoV-2 M^{pro}, respectively. The release of ACC was measured for 45 min (λ_{ex} = 355 nm, λ_{em} = 460 nm) and the linear part of each progress curve was used to determine the substrate hydrolysis rate. Substrate specificity profiles were established by setting the highest value of relative fluorescence unit per second (RFU/s) from each position as 100% and others were adjusted accordingly.

Kinetic analysis of substrates

Substrate screening was carried out in the same manner as the library assay. Substrate concentration was 5 µM, SARS-CoV M^{pro} was 0.3 µM and SARS-CoV-2 M^{pro} was 0.3 µM. Substrate hydrolysis was measured for 30 min using the following wavelengths: λ_{ex} = 355 nm, λ_{em} = 460 nm. The experiment was repeated three times. Results were presented as mean values with standard deviations. Kinetic parameters were assayed in 96-well plates (Corning). Wells contained 80 µL of enzyme in assay buffer (0.016-0.1 µM SARS-CoV-2 M^{pro}) and 20 µL of substrate at eight different

concentrations ranging from 58.5 µM to 1200 µM (0.44-7 µM for IQF substrates). ACC liberation was monitored for 30 min (λ_{ex} = 355 nm, λ_{em} = 460 nm). Each experiment was repeated at least three times. Kinetic parameters were determined using the Michaelis-Menten equation and GraphPad Prism software. Due to the precipitation of internally quenched fluorescence substrates in assay buffer, only the specificity constant (k_{cat}/K_M) was determined. When $[S_0] << K_M$, the plot of v_i (the initial velocities) versus $[S_0]$ yields a straight line with the slope representing V_{max}/K_M, k_{cat}/K_M = slope/E (E – total enzyme concentration).

Determination of inhibition kinetics (k_{obs}/I) for inhibitors and activity-based probes

SARS-CoV-2 M^{pro} (75 nM) was preincubated in assay buffer (20 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.3) for 10 min at 37°C. Then, the enzyme was added to wells containing seven different concentrations of inhibitor or probe (ranging from 3.9 µM to 20 µM) and 50 µM of substrate (QS1). The measurement was conducted for 30 minutes and repeated at least three times; k_{obs}/I was calculated as previously described[21].

SARS-CoV-2 M^{pro} labelling

SARS-CoV-2 M^{pro} (5, 10, 50, 100, or 200 nM) was incubated with different probe concentrations (50-2500 nM) in assay buffer (20 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.3) for 15 or 30 min at 37°C. Then 3x SDS/DTT was added, and the samples were boiled for 5 min at 95°C and resolved on 4-12% Bis-Tris Plus 12-well gels at 30 µL sample/well. Electrophoresis was performed at 200 V for 29 min. Next, the proteins were transferred to a nitrocellulose membrane (0.2 µm, Bio-Rad) for 60 min at 10 V. The membrane was blocked with 2% BSA in Tris-buffered saline with 0.1% (v/v) Tween 20 (TBS-T) for 60 min at RT. The biotinylated activity-based probe was detected with a fluorescent streptavidin Alexa Fluor 647 conjugate (1:10000) in TBS-T with 1% BSA using an Azure Biosystems Sapphire Biomolecular Imager (laser 658 nm and 488 nm for Bodipy FL dye) and Azure Spot Analysis Software. To perform a time-point analysis, SARS-CoV-2 M^{pro} (10 nM) was incubated with Bodipy-QS5-VS (2.5 µM) for different durations of time in the range from 1 to 60 min. Then the samples were boiled, and run on a gel. Electrophoresis, protein transfer to a nitrocellulose

membrane, and probe visualization were conducted in the same manner as described above.

SARS-CoV-2 M^{pro} labelling in cell lysates

HeLa and HEK-293T cells were cultured in DMEM supplemented with 10% fetal bovine serum, 2 mM L-glutamine, and antibiotics (100 U/mL penicillin, 100 µg/mL streptomycin) in a humidified 5% CO_2 atmosphere at 37°C. Approximately 1,200,000 cells were harvested and washed three times with PBS. The cell pellet was lysed in buffer containing 20 mM Tris, 150 mM NaCl, and 5 mM DTT, pH 8.0, using a sonicator. The cell lysate was centrifuged for 10 min, and the supernatant was collected. Twenty microliters of cell lysate was incubated with or without 30 µL of inhibitor Ac-QS1-VS and with or without SARS-CoV-2 M^{pro} (100 or 25 nM) for 30 min at 37°C. Next, 50 µL of probes at different concentrations was added to the samples and they were incubated for 15 or 30 min at 37°C. Then the samples were combined with 50 µL 3xSDS/DTT, boiled, and run on a gel. Electrophoresis, protein transfer to a nitrocellulose membrane, and probe visualization were conducted in the same manner as described above. β-actin was used as a loading control (Source Data Fig 1).

Antiviral activity of Ac-Abu-DTyr-Leu-Gln-VS toward a SARS-CoV-2 strain in Huh7 cell culture

The human hepatoma cell line Huh7 (kindly provided by Ralf Bartenschlager, University of Heidelberg, Germany) was maintained in Dulbecco's modified Eagle's medium (DMEM) (Gibco cat no 41965-039) supplemented with 10% fetal bovine serum (FBS), 2% HEPES 1M (Gibco cat no 15630106), 5 mL sodium bicarbonate 7.5% (Gibco 25080-060) 1% non-essential amino acids (NEAA Gibco cat no 11140050) and 1% Penicillin-Streptomycin 10,000 U/mL (Gibco cat no 15140148) in a humidified 5% CO2 incubator at 37°C. Assay medium, used for producing virus stocks and antiviral testing, was prepared by supplementing DMEM with 4% FBS, 2% HEPES 1M, 5 mL sodium bicarbonate 7.5% and 1% NEAA.

In order to quantify antiviral activity on Huh7 cells, we selected a SARS-CoV-2 virus strain that produces sufficient CPE on this cell line. We started from passage 6 of the SARS-CoV-2 strain BetaCov/Belgium/GHB-03021/2020 (EPI ISL 407976|2020-02-03) that has been described previously[29], and passaged this 3 additional times on Huh7 cells while selecting those cultures that showed most CPE. This resulted in a virus

stock (passage 9) that confers full CPE on Huh7 ($5.6x10^4$ $TCID_{50}$/mL) as well as on VeroE6 cells ($1.8x10^7$ $TCID_{50}$/mL). The genotype of this virus stock shows four nucleotide changes as compared with the mother virus stock (P6) and these are currently being analyzed. None of the nucleotide changes occur in the part of the genome that encodes the 3C-like protease, validating this virus stock for testing protease inhibitors.

For antiviral testing, Huh7 cells were seeded in 96-well plates (Corning® CellBIND® 96-well Microplate cat no 3300) at a density of 6000 cells per well in assay medium. After overnight growth, cells were treated with the indicated compound concentrations and infected with an MOI of 0.005 $TCID_{50}$/cell of the P9 virus (final volume 200 µL / well in assay medium). On day 4 p.i., differences in cell viability caused by virus-induced CPE or by compound-specific side effects were analyzed using MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, inner salt). For this, an MTS/Phenazine methosulphate (PMS) stock solution (2 mg/mL MTS (Promega, Leiden, The Netherlands) and 46 µg/mL PMS (Sigma–Aldrich, Bornem, Belgium) in PBS at pH 6–6.5) was diluted 1/20 in MEM without phenol red (Gibco cat no 51200038). Medium was aspirated from wells of the test plates and 70 µL of MTS/PMS solution was added. After 0.5-1 h incubation at 37°C, absorbance was measured at 498 nm. Cytotoxic effects caused by compound treatment alone were monitored in parallel plates containing mock-infected cells.

All virus-related work was conducted in the high-containment BSL3+ facilities of the KU Leuven Rega Institute (3CAPS) under licenses AMV 30112018 SBB 219 2018 0892 and AMV 23102017 SBB 219 2017 0589 according to institutional guidelines.

Crystallization of SARS-CoV-2 M^{pro} in complex with Biotin-PEG(4)-Abu-Tle-Leu-Gln-VS

The purified SARS-CoV-2 M^{pro} was concentrated to 23 mg/mL, mixed with Biotin-PEG(4)-Abu-Tle-Leu-Gln-VS at a molar ratio of 1:5, and the mixture was incubated at 4ºC overnight. The next day, the mixture was clarified by centrifugation at 12,000 x *g*, 4°C, and the supernatant was set for crystallization screening by using commercially available screening kits (PEGRx™ 1 & 2 (Hampton Research) and Morpheus HT-96 (Molecular Dimensions)). A Gryphon LCP crystallization robot (Art Robbins) was used for setting up the crystallization screens with the sitting-drop vapor-diffusion method at

18°C, where 0.15 µL of protein solution and 0.15 µL of reservoir were mixed to equilibrate against 40 µL reservoir solution.

Crystals were observed under several conditions in the two 96-well plates. The crystals were fished directly from the basic screening plates. The cryo-protectant consisted of mother liquor plus varied concentrations (5% to 20%) of glycerol, and 2 mM of the activity-based probe. Subsequently, liquid nitrogen was used for flash-cooling the crystals prior to data collection.

Diffraction data collection and determination of the structure

Several diffraction data sets were collected at 100 K at the P11 beamline of PETRA III (DESY, Hamburg, Germany), using synchrotron radiation of wavelength 1.0332 Å and a Pilatus 6M detector (Dectris). For structure determination, a data set was used that was collected using a crystal fished from the condition No. E7 of Morpheus HT-96 (0.12 M ethylene glycols (0.3 M diethylene glycol, 0.3 M triethylene glycol, 0.3 M tetraethylene glycol, 0.3 M pentaethylene glycol), 0.1 M buffer system 2 (1.0 M sodium HEPES, MOPS (acid), pH 7.5), pH 7.5, 30% precipitant mix 3 (20% v/v glycerol, 10% PEG 4000)).

XDSapp[30]*, Pointless*[31,32]*,* and *Scala*[31] (the latter two from the CCP4 suite[33]) were used for processing and scaling the dataset. The space group was determined as $P6_122$ and the resolution limit was set at a Bragg spacing of 1.70 Å. The molecular replacement method was employed for phase determination using the *Molrep* program[33,34] and the free-enzyme crystal structure of SARS-CoV-2 M^{pro} (PDB entry 6Y2E[9]) as the search model. The geometric restraints for the activity-based probe were generated using the *Jligand* programe from the CCP4 suite[33,35]; the ABP was built into the F_o-F_c density by using the *Coot* software[36]. Structure refinement was performed with *Refmac5*[33,36,37]*.* Ramachandran statistics after structure refinement are as follows: preferred region 97.0%, allowed region 3.0%, and outlier region 0.0%. Statistics of diffraction data processing and model refinement are shown in Supplementary Table 4.

Patient sample preparation

The study was approved by the Bioethics Committee of the Medical University of Lodz, Poland (# RNN/114/20/KE). The study was conducted in compliance with good clinical practice guidelines and under the principles of the Declaration of Helsinki. The study

participants or their parents provided written informed consent. Patients were hospitalized at the Central Medical University Hospital in Lodz (CUHL), Poland, and from the Department of Infectious Diseases and Hepatology, Medical University of Lodz (DIDH), Poland, and were positive for RNA of SARS-CoV-2 in serial measurements of nasopharyngeal swabs (at least 2 consecutive measurements) with clinical symptoms of infection.

For the current study, two nasopharyngeal swabs were collected from each patient at the same time, one was subjected for RNA SARS-CoV-2 testing and the other was used for smear preparation for confocal laser scanning microscopy. Smears were performed on glass slides covered with polylysine (Thermo Fisher Scientific, MA, USA).

Clinical characteristics of one patient (P01) recruited at CUHL

P01, a 13-year old boy, presented with mild symptoms of COVID-19 infection and no need for respiration support or mechanical ventilation; no specific treatment was applied. Two nasopharyngeal swabs for M^{pro} activity were collected during routine clinical examination, at day 1 and 5 after COVID-19 diagnosis. Tests positive for SARS-CoV-2 RNA were obtained on days 1 and 5, the first negative test was on day 7.

Clinical characteristics of six patients recruited at DIDH

P01DZ, a 59-year old woman, presented with moderate symptoms of COVID-19 infection, was hospitalized on day 3 after diagnosis, treated with respiration support (oxygen 3 L/min, for 10 days) and no mechanical ventilation was needed. She was treated with dexamethasone, hydroxychloroquine, lopinavir and ritonavir. One nasopharyngeal swab for M^{pro} activity was collected on day 12 after COVID-19 diagnosis during routine clinical examination. Tests for SARS-CoV-2 RNA were positive on days 1, 3, 12 and the first negative test was on day 20.

P02JB, a 34-year old woman, presented with moderate to severe COVID-19 pneumonia, was hospitalized on day 1 after diagnosis, treated with respiration support (oxygen 3 L/min, for 20 days) and no mechanical ventilation was needed. She was treated with dexamethasone, hydroxychloroquine, remdesivir, and vancomycin because of concomitant *Clostridium difficile* infection. One nasopharyngeal swab for M^{pro} activity was collected on day 10 after COVID-19 diagnosis during routine clinical

examination. Tests for SARS-CoV-2 RNA were positive on days 1, 7, 10, and 14, the first negative test was obtained on day 21.

P03KB, a 36-year old man, presented with moderate symptoms of COVID-19 pneumonia, was hospitalized on day 1 after diagnosis, treated with respiration support (oxygen 3 L/min, for 6 days) and no mechanical ventilation was needed. He was treated with dexamethasone, hydroxychloroquine, lopinavir and ritonavir. One nasopharyngeal swab for M^{pro} activity was collected on day 10 after COVID-19 diagnosis during routine clinical examination. Tests for SARS-CoV-2 RNA were positive on days 1, 7, 10 and the first negative test was on day 12.

P04WA, a 69-year old woman, presented with moderate symptoms of COVID-19 pneumonia, was hospitalized on day 1 after diagnosis, with respiration support (oxygen 3 L/min, required at the time of manuscript submission). She was treated with hydroxychloroquine, dexamethasone and ceftriaxone. One nasopharyngeal swab for M^{pro} activity was collected on day 5 after COVID-19 diagnosis during routine clinical examination. Tests for SARS-CoV-2 RNA were positive on day 1 and 5.

P05WM, a 72-year old man, presented with moderate symptoms of COVID-19 pneumonia, was hospitalized on day 1 after diagnosis, with respiration support (oxygen 3 L/min, required at the time of manuscript submission). He was treated with hydroxychloroquine and dexamethasone. One nasopharyngeal swab for M^{pro} activity was collected on day 5 after COVID-19 diagnosis during routine clinical examination. Tests for SARS-CoV-2 RNA were positive on day 1 and 5.

P06JM, a 34-year old man with moderate symptoms of COVID-19 pneumonia, was hospitalized on day 1 after diagnosis, no needs for respiration support or mechanical ventilation. He was treated with hydroxychloroquine and ceftriaxone. One nasopharyngeal swab for M^{pro} activity was collected on day 5 after COVID-19 diagnosis during routine clinical examination. Tests for SARS-CoV-2 RNA were positive on day 1 and 5.

Detection of SARS-CoV-2 M^{pro} in epithelial cells by immunofluorescence.

The nasopharyngeal swab was used for smear preparation for confocal laser scanning microscopy. Smears were performed on glass slides covered with polylysine (Thermo Fisher Scientific, MA, USA). These slides were treated with 1 or 2.5 μM probe of M^{pro} (Cy5-QS1-VS or Bodipy-QS5-VS) and incubated for 30 min at 37°C. Slides were fixed in 70% ethanol for 30 min. For further staining overnight at 4°C, incubation with recombinant goat antiACE2 antibody (R&D Systems, MN, USA, 1:20) was performed. The next day, antibodies were aspirated, and cells were washed twice with PBS and labeled with a secondary antibody (Donkey anti-Goat IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, AlexaFluor Plus 488 (Catalog A32814, 1:200 Invitrogen, ThermoFisher Scientific, USA) or Cy5 AffiniPure donkey anti-Goat IgG (H+L) (705-175-147, 1:200, Jackson ImmunoResearch) in PBS for 1 hour at room temperature. Next, the cells were washed twice with PBS, coverslips were mounted with Vectashield fluorescence mounting medium containing DAPI (Vector Lab. H-1000) and sealed with nail polish. Slides were stored at 4°C until use. Cells were then subjected to confocal microscopy analysis using a Leica TCS SP8. DAPI was detected by the DAPI channel (405 nm), the M^{pro} Cy5-QS1-VS probe was detected with the Cy5 filter single photon laser (638 nm) (Bodipy-QS5-VS 488 nm), and ACE2/secondary antibodies were read using the FITC filter (single photon laser: 458 nm) or Cy5 filter laser 638 nm. All images were acquired in .tiff format using Leica Application Suite X software. Images shown are representative views of cells from two coverslips.

Slides with inhibitor (Ac-QS1-VS) were prepared in the same manner as described above. Cells were treated with 50 μM Ac-QS1-VS and incubated for 30 min at 37°C. Then cells were washed twice with PBS and incubated with 2.5 μM Bodipy-QS5-VS for 30 min at 37°C.

Data availability

The authors declare that the main data supporting the findings of this study are available within the article and its Supplementary Information files. The expression plasmid for the SARS-CoV-2 main protease is available from Rolf Hilgenfeld (rolf.hilgenfeld@uni-luebeck.de). The atomic coordinates and structure factors for the complex between the SARS-CoV-2 M^{pro} and the activity-based probe have been deposited in the Protein Data Bank under code 6Z2E.

Abbreviations used

Abu, 2-aminobutanoic acid; 2-Abz, 2-(amino)benzoic acid; 3-Abz, 3-(amino)benzoic acid; ACC, 7-amino-4-carbamoylmethylcoumarin; Dab, 2,4-diaminobutyric acid; Dht, dihydrotryptophan; HyCoSuL, Hybrid Combinatorial Substrate Library; Orn, ornithine; RFU, relative fluorescence unit; D-Phg, D-phenylglycine; Thz, thiazolidine-4-carboxylic acid; Tle, tert-leucine;

Methods-only References:

28. Maly, D.J. et al. Expedient solid-phase synthesis of fluorogenic protease substrates using the 7-amino-4-carbamoylmethylcoumarin (ACC) fluorophore. *Journal of Organic Chemistry* **67**, 910-915 (2002).
29. Boudewijns, R. et al. STAT2 signaling as double-edged sword restricting viral dissemination but driving severe pneumonia in SARS-CoV-2 infected hamsters. *bioRxiv*, https://doi.org/10.1101/2020.04.23.056838 (2020).
30. Krug, M., Weiss, M.S., Heinemann, U. & Mueller, U. XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS. *Journal of Applied Crystallography* **45**, 568-572 (2012).
31. Evans, P. Scaling and assessment of data quality. *Acta Crystallographica Section D-Structural Biology* **62**, 72-82 (2006).
32. Evans, P.R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D-Biological Crystallography* **67**, 282-292 (2011).
33. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D-Structural Biology* **67**, 235-242 (2011).
34. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallographica Section D-Biological Crystallography* **66**, 22-25 (2010).
35. Lebedev, A.A. et al. JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallographica Section D-Structural Biology* **68**, 431-440 (2012).
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. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D-Structural Biology* **67**, 355-367 (2011).