Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved \(\alpha\)-ketoamide inhibitors

Linlin Zhang\(^1,2\), Daizong Lin\(^1,3\), Xinyuanyuan Sun\(^1,2\), Ute Curth\(^4\), Christian Drosten\(^2\), Lucie Saußerhering\(^6,7\), Stephan Becker\(^6,7\), Katharina Rox\(^8,9\), Rolf Hilgenfeld\(^1,2\)*

\(^1\)Institute of Biochemistry, Center for Structural and Cell Biology in Medicine, University of Lübeck, 23562 Lübeck, Germany. \(^2\)German Center for Infection Research (DZIF), Hamburg-Lübeck-Borstel-Riems Site, University of Lübeck, 23562 Lübeck, Germany. \(^3\)Changchun Discovery Sciences Ltd., 789 Shunda Road, Changchun, Jilin 130012, China. \(^4\)Institute for Biophysical Chemistry, Hannover Medical School, 30625 Hannover, Germany. \(^5\)Institute of Virology, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany. \(^6\)Institute of Virology, University of Marburg, 35043 Marburg, Germany. \(^7\)German Center for Infection Research (DZIF), Marburg-Gießen-Langen Site, University of Marburg, 35043 Marburg, Germany. \(^8\)Department of Chemical Biology, Helmholtz Center for Infection Research (HDI), Inhoffenstraße 7, 38124 Braunschweig, Germany. \(^9\)German Center for Infection Research (DZIF), Hannover-Braunschweig Site, Helmholtz Center for Infection Research, 38124 Braunschweig, Germany.

*Corresponding author. Email: rolf.hilgenfeld@uni-luebeck.de

The COVID-19 pandemic caused by SARS-CoV-2 is a global health emergency. An attractive drug target among coronaviruses is the main protease (M\(^{\text{pro}}\), 3CL\(^{\text{pro}}\)), due to its essential role in processing the polyproteins that are translated from the viral RNA. We report the X-ray structures of the unliganded SARS-CoV-2 M\(^{\text{pro}}\) and its complex with an \(\alpha\)-ketoamide inhibitor. This was derived from a previously designed inhibitor but with the P3-P2 amide bond incorporated into a pyridone ring to enhance the half-life of the compound in plasma. Based on the structure, we developed the lead compound into a potent inhibitor of the SARS-CoV-2 M\(^{\text{pro}}\). The pharmacokinetic characterization of the optimized inhibitor reveals a pronounced lung tropism and suitability for administration by the inhalative route.

In December 2019, a new coronavirus caused an outbreak of pulmonary disease in the city of Wuhan, the capital of Hubei province in China, and has since spread globally (\(^1\), \(^2\)). The virus has been named SARS-CoV-2 (\(^3\)), because the RNA genome is about 82% identical to the SARS coronavirus (SARS-CoV); both viruses belong to clade b of the genus Betacoronavirus (\(^1\), \(^2\)). The disease caused by SARS-CoV-2 is called COVID-19. Whereas at the beginning of the outbreak, cases were connected to the Huanan seafood and animal market in Wuhan, efficient human-to-human transmission led to exponential growth in the number of cases. On March 11, the World Health Organization (WHO) declared the outbreak a pandemic. As of March 15, there are >170,000 cumulative cases globally, with a ~3.7% case-fatality rate.

One of the best characterized drug targets among coronaviruses is the main protease (M\(^{\text{pro}}\), also called 3CL\(^{\text{pro}}\)) (\(^4\)). Along with the papain-like protease(s), this enzyme is essential for processing the polyproteins that are translated from the viral RNA (\(^5\)). The M\(^{\text{pro}}\) operates at no less than 11 cleavage sites on the large polyprotein 1ab (replicase 1ab, ~790 kDa); the recognition sequence at most sites is Leu-Gln (\(\text{Ser,Ala,Gly}\); 1 marks the cleavage site). Inhibiting the activity of this enzyme would block viral replication. Since no human proteases with a similar cleavage specificity are known, inhibitors are unlikely to be toxic.

Previously, we designed and synthesized peptidomimetic \(\alpha\)-ketoamides as broad-spectrum inhibitors of the main proteases of betacoronaviruses and alphacoronaviruses as well as the 3C proteases of enteroviruses (\(^6\)). The best of these compounds (\(^1\)r; Fig. 1) showed an EC\(_{50}\) of 400 picomolar against MERS-CoV in Huh7 cells as well as low micromolar EC\(_{50}\) values against SARS-CoV and a whole range of enteroviruses in various cell lines, although the antiviral activity seemed to depend to a great extent on the cell type used in the experiments (\(^6\)). In order to improve the half-life of the compound in plasma, we modified \(^1\)r by hiding the P3-P2 amide bond within a pyridone ring (Fig. 1, green circles), in the expectation that this might prevent cellular proteases from accessing this bond and cleaving it. Further, to increase the solubility of the compound in plasma and to reduce its binding to plasma proteins, we replaced the hydrophobic cinnamoyl Boc group (Fig. 1, red circles) to give \(^1\)3a (see scheme S1 for synthesis).

In order to examine whether the introduced pyridone ring is compatible with the three-dimensional structure of
the target, we determined the crystal structure, at 1.75 Å resolution, of the M\textsuperscript{pro} of SARS-CoV-2 (Fig. 2). The three-dimensional structure is highly similar to that of the SARS-CoV M\textsuperscript{pro}, as expected from the 96% sequence identity (see fig. S7); the r.m.s. deviation between the two free-enzyme structures is 0.53 Å for all Cα positions (comparison between SARS-CoV-2 M\textsuperscript{pro} structure and SARS-CoV M\textsuperscript{pro}, PDB entry 2BX4 (7)). The chymotrypsin- and picornavirus 3C protease-like domains I and II (residues 10-99 and 100-182, respectively) are six-stranded antiparallel β-barrels that harbor the substrate-binding site between them. Domain III (residues 198-303), a globular cluster of five helices, is involved in regulating dimerization of the M\textsuperscript{pro}, mainly through a salt-bridge interaction between Glu\textsuperscript{199} of one protomer and Arg\textsuperscript{4} of the other (8). The tight dimer formed by SARS-CoV-2 M\textsuperscript{pro} has a contact interface, predominantly between domain II of molecule A and the NH\textsubscript{2}-terminal residues (“N-finger”) of molecule B, of ≈1394 Å\textsuperscript{2}, with the two molecules oriented perpendicular to one another (Fig. 2). Dimerization of the enzyme is necessary for catalytic activity, because the N-finger of each of the two protomers interacts with Glu\textsuperscript{166} of the other protomer and thereby helps shape the S1 pocket of the substrate-binding site (9). To reach this interaction site, the N-finger is squeezed in between domains II and III of the parent monomer and domain II of the other monomer. Interestingly, in the SARS-CoV but not in the SARS-CoV-2 M\textsuperscript{pro} dimer, there is a polar interaction between the two domains III involving a 2.60-Å hydrogen bond between the side-chain hydroxyl groups of residue Thr\textsuperscript{285} of each protomer, and supported by a hydrophobic contact between the side-chain of Ile\textsuperscript{286} and Thr\textsuperscript{285} Cγ2. In SARS-CoV-2, the threonine is replaced by alanine (indicated by the black sphere in Fig. 2), and the isoleucine by leucine (see fig. S7). It has previously been shown that replacing Ser\textsuperscript{284}, Thr\textsuperscript{285}, and Ile\textsuperscript{286} by alanine residues in SARS-CoV M\textsuperscript{pro} leads to a 3.6-fold enhancement of the catalytic activity of the protease, concomitant with a slightly closer packing of the two domains III of the dimer against each other (10). This was accompanied by changes in enzyme dynamics that transmit the effect of the mutation to the catalytic center. Indeed, the Thr\textsuperscript{285}Ala replacement observed in the SARS-CoV-2 M\textsuperscript{pro} also allows the two domains III to approach each other a little closer (the distance between the Cα atoms of residues 285 in molecules A and B is 6.77 Å in SARS-CoV M\textsuperscript{pro} and 5.21 Å in SARS-CoV-2 M\textsuperscript{pro} and the distance between the centers of mass of the two domains III shrinks from 33.4 Å to 32.1 Å). However, the catalytic efficiency of SARS-CoV-2 M\textsuperscript{pro} (k\textsubscript{cat}/K\textsubscript{m} = 3426.1 ± 416.9 s\textsuperscript{-1}M\textsuperscript{-1}) is only slightly higher, if at all (k\textsubscript{cat}/K\textsubscript{m} = 3242.1 ± 416.9 s\textsuperscript{-1}M\textsuperscript{-1}) than that of SARS-CoV M\textsuperscript{pro} (k\textsubscript{cat}/K\textsubscript{m} = 3011.3 ± 294.6 s\textsuperscript{-1}M\textsuperscript{-1}). Further, the estimated K\textsubscript{D} of dimer dissociation is the same (~2.5 µM) for the two enzymes, as determined by analytical ultracentrifugation (fig. S8).

We used this crystal structure to dock the α-ketoamide 13a; this suggested that the pyridone ring might have some steric clash with the side-chain of Gln189. However, in our previous work (6), we had found Gln189 to be quite flexible and therefore we went ahead with 13a as a lead. The plasma half-life of this compound in mice was increased ~3-fold compared to 11r (from 0.3 hours to 1.0 hours), the in-vitro kinetic plasma solubility was improved by a factor of ~19 (from 6 µM for 11r to 112 µM for 13a) and the thermodynamic solubility by a factor of ~13 (from 41 µM to 530 µM). Binding to mouse plasma protein was reduced from 99% to 97% (many drugs have plasma protein binding of >90%; (II)). However, compared to 11r (IC\textsubscript{50} = 0.18 ± 0.02 µM), the structural modification led to some loss of inhibitory activity against the main protease of SARS-CoV-2 (IC\textsubscript{50} = 2.39 ± 0.63 µM) as well as the 3C proteases (3C\textsuperscript{pro}) of enteroviruses. 11r was designed for broad-spectrum activity, with the P2 cyclohexyl moiety intended to fill a pocket in the enterovirus 3C\textsuperscript{pro}. The S2 pocket of the betacoronavirus M\textsuperscript{pro} (see Fig. 3) features substantial plasticity enabling it to adapt to the shape of smaller inhibitor moieties (6). To enhance the anti-viral activity against betacoronaviruses of clade b (SARS-CoV-2 and SARS-CoV), we sacrificed the goal of broad-spectrum activity and replaced the P2 cyclohexyl moiety of 13a by the smaller cyclopropyl in 13b (Fig. 1, blue circles). Here we present X-ray crystal structures in two different crystal forms, at 1.95 and 2.20 Å resolution, of the complex between α-ketoamide 13b and the M\textsuperscript{pro} of SARS-CoV-2 (Fig. 3). One structure is in space group C2, where both protomers of the M\textsuperscript{pro} dimer are bound by crystal symmetry to have identical conformations, the other is in space group P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1}, where the two protomers are independent of each other and free to adopt different conformations. Indeed, we find that in the latter crystal structure, the key residue Glu\textsuperscript{166} adopts an inactive conformation in protomer B (as evidenced by its distance from His\textsuperscript{272} and the lack of H-bonding interaction with the P1 moiety of the inhibitor), even though compound 13b is bound in the same mode as in molecule A. This phenomenon has also been observed with the SARS-CoV M\textsuperscript{pro} (12) and is consistent with the half-site activity described for this enzyme (13). In all copies of the inhibited SARS-CoV-2 M\textsuperscript{pro}, the inhibitor binds to the shallow substrate-binding site at the surface of each protomer, between domains I and II (Fig. 3).

Through the nucleophilic attack of the catalytic Cys\textsuperscript{345} onto the α-keto group of the inhibitor, a thiohemiketal is formed in a reversible reaction. This is clearly reflected in the electron density (Fig. 3 inset); the stereochemistry of this chiral moiety is S in all copies of compound 13b in these structures. The oxanion (or hydroxyl) group of this thiohemiketal is stabilized by a hydrogen bond from His\textsuperscript{341},
whereas the amide oxygen of 13b accepts a hydrogen bond from the main-chain amides of Gly143, Cys145, and partly Ser144, which form the canonical “oxyanion hole” of the cysteine protease. It is an advantage of the α-ketoamides that their warhead can interact with the catalytic center of the target proteases through two hydrogen bonding interactions (δ), rather than only one as with other warheads such as aldehydes (I4) or Michael acceptors (I5).

The P1 γ-lactam moiety, designed as a glutamine surrogate (I5, I6), is deeply embedded in the S1 pocket of the protease, where the lactam nitrogen donates a three-center (bifurcated) hydrogen bond to the main-chain oxygen of Phe140 (3.20/3.10/3.28 Å; values for the structure in space group C2/space group P212121 molecule A/space group P212121 molecule B) and to the Glu166 carboxylate (3.35/3.33/(3.55) Å), and the carbonyl oxygen accepts a 2.57/2.51/2.81-Å H-bond from the imidazole of His160. The P2 cyclopropyl methyl moiety fits snugly into the S2 subsite, which has shrunk by 28 Å² compared to the complex between compound 13a with P2 = cyclohexyl methyl and the SARS-CoV Mpro (I7). The pyridone in the P3 - P2 position of the inhibitor occupies the space normally filled by the substrate’s main chain, its carbonyl oxygen accepts a 2.89/2.99/3.00-Å hydrogen bond from the main-chain amide of residue Glu166. Further, the P3 amide donates a 2.83/2.96/2.87-Å H-bond to the main-chain oxygen of Glu166. Embedded within the pyridone, the P2 nitrogen can no longer donate a hydrogen bond to the protein (the H-bond prevented from forming would connect the P2 nitrogen and the side-chain oxygen of Gln189; these two atoms are highlighted in fig. S8). However, our previous crystal structures showed that the P2 main-chain amide of the linear α-ketoamides does not make a hydrogen bond with the protein in all cases, so this interaction does not seem to be critical (δ). The protecting Boc group on P3 does not occupy the canonical S4 site of the protease (in contrast to the protecting groups of other inhibitors in complex with the SARS-CoV Mpro (I8)), but is located near Pro166 (3.81/4.17/3.65 Å; Fig. 3); due to this interaction, the latter residue moves outward by more than 2 Å (compared to the structure of the free enzyme). This contact explains why removing the Boc group as in compound 14b (Fig. 1, purple circles) weakens the inhibitory potency of this compound by a factor of about 2. Interestingly, there is a space between the pyridone ring of 13b, the main chain of residue Thr189, and the side-chain of Gln189 (smallest distance: 3.6 Å) which is filled by a DMSO molecule in the C2 crystal structure and a water molecule in the P2;2;2; structure. This suggests that P3 moieties more bulky than pyridone may be accepted here.

Compound 13b inhibits the purified recombinant SARS-CoV-2 Mpro with IC50 = 0.67 ± 0.18 μM. The corresponding IC50 values for inhibition of the SARS-CoV Mpro and the MERS-CoV Mpro are 0.90 ± 0.29 μM and 0.58 ± 0.22 μM, respectively. In a SARS-CoV replicon (I9), RNA replication is inhibited with EC50 = 1.75 ± 0.25 μM. In human Calu3 cells infected with the novel coronavirus, SARS-CoV-2, an EC50 of 4 - 5 μM is observed, whereas compound 14b lacking the Boc group is almost inactive (Fig. 4). This suggests that the hydrophobic and bulky Boc group is necessary to cross the cellular membrane and that an even more hydrophobic moiety might be advantageous here, although this may again lead to increased plasma protein binding as observed for the cinnamoyl-containing 11r.

To assess the absorption - distribution - metabolism - excretion (ADME) properties of the pyridone-containing α-ketoamides, we first investigated compound 13a. Metabolic stability in mouse and human microsomes was good, with intrinsic clearance rates Clint,mouse = 32.0 μL/min/mg protein and Clint,human = 21.0 μL/min/mg protein. This means that after 30 min, around 80% for mouse and 60% for humans, respectively, of residual compound remained metabolically stable. Pharmacokinetic studies in CD-1 mice using the subcutaneous route at 20 mg/kg showed that 13a stayed in plasma for up to only 4 hours, but was excreted via urine for up to 24 hours. The Cmax was determined at 334.5 ng/mL and the mean residence time was about 1.6 hours. Although 13a seemed to be cleared very rapidly from plasma, it was found at 24 hours at 135 ng/g tissue in the lung and at 52.7 ng/mL in broncho-alveolar lavage fluid (BALF) suggesting that it was mainly distributed to tissue. Next, we investigated 13b for its pharmacokinetic properties in CD-1 mice using the subcutaneous route as well, but at 3 mg/kg. ADME parameters of 13b were similar to 13a; in addition, the binding to human plasma proteins was found to be 90%. The Cmax of 13b was determined at 126.2 ng/mL. This is around 37% of the Cmax detected for 13a, although 13b dosage was approximately 7-times lower. The mean residence time for 13b was extended to 2.7 hours and the plasma half-life in mice was 1.8 hours. In addition, 13b showed a less rapid clearance compared to 13a (table S3). During the pharmacokinetic study with 13b, we monitored its lung tissue levels. After 4 hours, around 13 ng/g 13b were still found in lung tissue. This lung tropism of 13a and 13b is beneficial given that COVID-19 affects the lungs. In addition to subcutaneous administration, 13b was nebulized using an inhalation device at 3 mg/kg. After 24 hours, 33 ng/g 13b were found in lung tissue. Inhalation was tolerated well and mice did not show any adverse effects, suggesting that this way, direct administration of the compound to the lungs would be possible. Given these favorable pharmacokinetic results, our study provides a useful framework for development of the pyridone-containing inhibitors toward anticonvoviral drugs.
We thank the German Center for Infection Research (DZIF) for financial support (projects TTU01, grant #8011801806, and TTU09, grant #8004709710). The plasmid encoding the SARS-CoV-2 M pro will be freely available. The available amounts of inhibitors are limited.

SUPPLEMENTARY MATERIALS

Data and materials availability: Crystallographic coordinates and structure factors are available from the PDB under accession codes 6Y2E (unliganded M pro), 6Y2F (complex with 13b in space group C2), and 6Y2G (complex with 13b in space group P2_12_1). The plasmid encoding the SARS-CoV-2 M pro will be freely available. The available amounts of inhibitors are limited.

ACKNOWLEDGMENTS

The authors are grateful to Yuri Kusov and Guido Hansen as well as Aws Aljnabi for determining the inhibitory activities of compounds in a SARS-CoV replicon and against recombinant MERS-CoV M pro, respectively, and to Thorsten Biet for recording 2D NMR spectra. We thank Andrea Ahlers, Janine Schreiber, and Lidia Litz for excellent technical assistance, Kuilan Chen for continuous organizational support, and the staff at beamline 14.2 of BESSY II, Berlin, Germany, for help with diffraction data collection. Funding: We thank the German Center for Infection Research (DZIF) for financial support (projects TTU01, grant #8011801806, and TTU09, grant #8004709710).

Author contributions: Conceptualization: LLZ, DZL, RH; Investigation: LLZ, DZL, XYYS, UC, LS, SB, KR, RH; Contribution of research materials: CD; Writing - original draft preparation: RH, DZL, KR; Writing - review and editing: LLZ, DZL, UC, LS, KR, RH; Visualization: LLZ, LS; Supervision: RH; Funding acquisition: RH, SB, KR.

Competing interests: The University of Lübeck has filed a patent application covering compounds 13a and 13b as well as related compounds with a pyridone structure in the P3-P2 position, with LLZ, DZL, and RH as inventors.

First release: 20 March 2020

www.sciencemag.org  (Page numbers not final at time of first release)
**Fig. 1.** Chemical structures of α-ketoamide inhibitors 11r, 13a, 13b, and 14b. Colored circles highlight the modifications from one development step to the next (see text).

**Fig. 2.** Three-dimensional structure of SARS-CoV-2 M\textsuperscript{pro}, in two different views. One protomer of the dimer is shown in light blue, the other one in orange. Domains are labeled by Roman numbers. Amino-acid residues of the catalytic site are indicated as yellow and blue spheres, for Cys\textsuperscript{145} and His\textsuperscript{41}, respectively. (An asterisk marks a residue from protomer B (orange)). Black spheres indicate the positions of Ala\textsuperscript{285} of each of the two domains III (see text). Chain termini are labeled N and C for molecule A (light blue) and N\textsuperscript{*} and C\textsuperscript{*} for molecule B (orange).
Fig. 3. Compound 13b in the substrate-binding cleft located between domains I and II of the Mprα, in the monoclinic crystal form (space group C2). Fo-Fc density is shown for the inhibitor (contouring level: 3σ). Carbon atoms of the inhibitor are magenta, except in the pyridone ring, which is black; oxygen atoms are red, nitrogens blue, and sulfur yellow. Light-blue symbols S1, S2, S3, S4 indicate the canonical binding pockets for moieties P1, P2, P3, P4 (red symbols) of the peptidomimetic inhibitor. Hydrogen bonds are indicated by dashed red lines. Note the interaction between the N-terminal residue of chain B, Ser1*, and Glu166 of chain A, which is essential for keeping the S1 pocket in the right shape and the enzyme in the active conformation. Inset: Thiohemiketal formed by the nucleophilic attack of the catalytic cysteine onto the ω-carbon of the inhibitor in its Fo-Fc density (contoured at 3σ). The stereochemistry of the ω-carbon is S. See fig. S8 for more details.
Fig. 4. Compound 13b inhibits SARS-CoV-2 replication in human Calu3 lung cells. (A) Calu-3 cells were infected with SARS-CoV-2 using an MOI of 0.05 and stimulated with DMSO (black bar) or different amounts (5, 10, 20, or 40 µM) of 13b (blue bars) or 14b (orange bars) and analyzed at 24 hours p.i.. In (A), total RNA was isolated from cell lysates and viral RNA content was analyzed by qPCR. (B) For the estimation of the EC_{50} value of compound 13b against SARS-CoV-2, a dose-response curve was prepared (GraphPad). (A) represents means ± SD of two biological experiments with two technical replicates each.
Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors
Linlin Zhang, Daizong Lin, Xinyuanyuan Sun, Ute Curth, Christian Drosten, Lucie Sauerhering, Stephan Becker, Katharina Rox and Rolf Hilgenfeld

published online March 20, 2020