

## Article

# Crystal Structures of Inhibitor-Bound Main Protease from Delta- and Gamma-Coronaviruses

Sarah N. Zvornicanin <sup>1</sup>, Ala M. Shaqra <sup>1</sup>, Qiuyu J. Huang <sup>1</sup>, Elizabeth Ornelas <sup>2</sup>, Mallika Moghe <sup>3</sup>, Mark Knapp <sup>2</sup>, Stephanie Moquin <sup>2</sup>, Dustin Dovala <sup>2</sup>, Celia A. Schiffer <sup>1</sup> and Nese Kurt Yilmaz <sup>1,\*</sup>

<sup>1</sup> Department of Biochemistry and Molecular Biotechnology, University of Massachusetts Chan Medical School, Worcester, MA 01605, USA

<sup>2</sup> Novartis Institutes for Biomedical Research, Emeryville, CA 94608, USA

<sup>3</sup> Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

\* Correspondence: nese.kurtyilmaz@umassmed.edu

**Abstract:** With the spread of SARS-CoV-2 throughout the globe causing the COVID-19 pandemic, the threat of zoonotic transmissions of coronaviruses (CoV) has become even more evident. As human infections have been caused by alpha- and beta-CoVs, structural characterization and inhibitor design mostly focused on these two genera. However, viruses from the delta and gamma genera also infect mammals and pose a potential zoonotic transmission threat. Here, we determined the inhibitor-bound crystal structures of the main protease (M<sup>Pro</sup>) from the delta-CoV porcine HKU15 and gamma-CoV SW1 from the beluga whale. A comparison with the apo structure of SW1 M<sup>Pro</sup>, which is also presented here, enabled the identification of structural arrangements upon inhibitor binding at the active site. The cocrystal structures reveal binding modes and interactions of two covalent inhibitors, PF-00835231 (active form of lufotrelvir) bound to HKU15, and GC376 bound to SW1 M<sup>Pro</sup>. These structures may be leveraged to target diverse coronaviruses and toward the structure-based design of pan-CoV inhibitors.



**Citation:** Zvornicanin, S.N.; Shaqra, A.M.; Huang, Q.J.; Ornelas, E.; Moghe, M.; Knapp, M.; Moquin, S.; Dovala, D.; Schiffer, C.A.; Kurt Yilmaz, N. Crystal Structures of Inhibitor-Bound Main Protease from Delta- and Gamma-Coronaviruses. *Viruses* **2023**, *15*, 781. <https://doi.org/10.3390/v15030781>

Academic Editors:

Luis Martinez-Sobrido and  
Fernando Almazan Toral

Received: 31 January 2023

Revised: 15 March 2023

Accepted: 16 March 2023

Published: 18 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** coronaviruses; main protease; GC376; crystal structure; lufotrelvir; HKU15; gammacoronavirus; beluga whale; direct-acting antivirals; drug design; pan-coronavirus inhibitor

## 1. Introduction

Viral species from the coronavirus (CoV) family can cause severe respiratory, enteric and hepatic disease, both in animals and humans. Wildlife, especially bats, harbor a diverse pool of coronavirus species that crossover to other animals, including mammals [1]. Further evolution of the virus to adapt to mammalian hosts increases the likelihood of zoonotic transmission to humans. Three such transmissions caused major outbreaks in humans, namely SARS-CoV-1, MERS, and most recently, SARS-CoV-2, which are thought to have originated in bats but transmitted through civets, camels, and possibly pangolins as intermediate hosts to humans [2–4]. Thus, investigating coronaviruses from various hosts, especially other mammals, is important to evaluate the natural diversity and evolution of these viruses toward readiness for future cross-species transmissions and potential outbreaks.

Despite the diversity and rapid evolution of coronavirus species, as exemplified in SARS-CoV-2 during the COVID-19 pandemic, certain elements of the viral genome are highly conserved. While the spike protein shows extreme diversity and quick adaptation to thwart antibody responses and vaccine effectiveness, the 3C-like main protease (M<sup>Pro</sup>) has remained relatively conserved. In fact, M<sup>Pro</sup> from SARS-CoV-1 and -2 are 96% identical. SARS-CoV-2 M<sup>Pro</sup> also shares the same overall fold and active site with 3C and 3C-like proteases from various coronavirus species and even other viral families [5,6]. This similarity prompted the exploration of potential pan-3C inhibitors that leverage the conserved substrate specificity at the P1 and P2 positions to target the conserved features of the active

site. These similarities have enabled the repurposing of 3C protease and M<sup>Pro</sup> inhibitors to inhibit SARS-CoV-2 and the unprecedented rapid clinical development of COVID-19 antivirals [7,8]. One such broad-spectrum inhibitor is GC376, which inhibits many picornavirus-like species, including rhinoviruses, enteroviruses, and coronaviruses [9]. GC376 is a bisulfite derivative prodrug that converts to the active aldehyde form GC373 and covalently attaches to the catalytic Cys of the protease. This dipeptidyl compound was first reported to inhibit feline alpha-CoV FIPV, and its cocrystal structure was solved with M<sup>Pro</sup> of FIPV as well as the porcine alpha-CoV TGEV [9,10] (PDB IDs: 4F49, 7SMV). GC376 also inhibits the M<sup>Pro</sup> of beta-CoVs, including SARS-CoV-2, with nanomolar K<sub>i</sub> values. The cocrystal structure for this complex was determined by multiple groups during the COVID-19 pandemic [11–13] (PDB IDs: 6WTJ, 6WTT, 7JSU), and M<sup>Pro</sup> inhibition by GC376 was further characterized to be covalent but reversible [14]. Another repurposed compound that progressed into clinical trials as lufotrelvir is the hydroxymethylketone-based covalent inhibitor PF-00835231 [7]. This phosphate prodrug converts to its active form PF-07304814 and attaches irreversibly to the active Cys (PDB ID: 6XHM). The two inhibitors, PF-00835231 and GC376, share the same *g*-lactam glutamine mimic in the P1 position and a leucine group at the P2 position while incorporating cyclic P3 groups (Figure S1). Both these inhibitors were reported to be highly potent and structurally characterized against M<sup>Pro</sup> of alpha- and beta-CoVs, but their binding mode to more divergent delta- and gamma-CoVs M<sup>Pro</sup> variants has not been described.

Viral species from the delta genus of coronaviruses have mostly been isolated from birds, but they can also infect mammals, as exemplified by the porcine HKU15. The gamma-CoVs were more recently identified as a genus, with the discovery of a novel and highly divergent coronavirus (named SW1) in the liver of a captive beluga whale [15]. Thus, delta- and gamma-CoVs have been shown to infect mammals that live in proximity to humans. The M<sup>Pro</sup> amino acid sequence of HKU15 and SW1 are less than 50% identical to those from beta-CoVs as well as to each other (Figure S2). The crystal structure of HKU15 M<sup>Pro</sup> has recently been determined with a relatively large peptidomimetic inhibitor [16]. However, how potential pan-CoV inhibitors bind to HKU15 is not known. Additionally, apart from the avian infectious bronchitis virus (IBV) [17], structural characterization of M<sup>Pro</sup> from gamma-CoVs is severely lacking.

In this work, we report the crystal structures of M<sup>Pro</sup> from the gamma-CoV SW1 in both apo and inhibitor-bound form, as well as the crystal structure of the delta-CoV HKU15 M<sup>Pro</sup> in complex with PF-00835231. These structures enabled the characterization and comparison of the main viral protease from these divergent mammal-infecting coronaviruses, as well as the identification of active site features that might be critical for the design of pan-coronaviral inhibitors for future potential outbreaks.

## 2. Materials and Methods

### 2.1. Protein Expression and Purification of M<sup>Pro</sup>

M<sup>Pro</sup> sequences were cloned into the pETite expression plasmid using standard techniques. The constructs included an N-terminal polyhistidine—SUMO tag, which was used for affinity purification and to prevent the toxic activity of the protein from impeding bacterial growth.

HI-Control™ BL21(DE3) (LGC Lucigen, Middleton, WI, USA) cells were transformed with plasmid, and single colonies were picked to start overnight cultures in LB medium supplemented with 50 µg/mL kanamycin at 37 °C. These cultures were used to inoculate 1 L cultures in a TB medium containing 50 mM sodium phosphate (pH 7.0) and kanamycin (50 µg/mL). When the OD<sub>600</sub> value reached ~1.5–2.0, 0.5 mM IPTG was added to induce protein expression, and the cell culture was further incubated overnight at 19 °C. Cells were harvested by centrifugation at 5000 rpm for 30 min, resuspended in Lysis Buffer (50 mM Tris-HCl (pH 8.0), 400 mM NaCl, 1 mM TCEP) and lysed by a cell disruptor. The lysate was clarified by centrifugation at 45,000 × *g* for 30 min. The supernatant was then subjected to IMAC purification by loading onto a HisTrap FF column (GE Healthcare,

Chicago, IL, USA) equilibrated with lysis buffer, washed with lysis buffer until the A280 stabilized, and followed by elution using a linear gradient of Lysis Buffer supplemented with 0–500 mM imidazole over 40 column volumes. The fractions containing M<sup>PRO</sup> were pooled and treated with ULP1 to remove the His-SUMO tag and produce an authentic N-terminus. Cleavage occurred overnight at room temperature while dialyzing into 3 L of Lysis Buffer (0 mM imidazole). We found prolonged exposure to 4 °C often led to substantial precipitation of the protein, which is why dialysis occurred at room temperature. Full cleavage was confirmed by ESI-LC/MS, and the cleaved protein was again subjected to IMAC purification by flowing through 5 mL of Ni-NTA resin pre-equilibrated with Lysis Buffer. The supernatant was collected and concentrated to approximately 5 mL before further purification by size-exclusion chromatography (SEC). SEC purification was performed on a HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated with 25 mM HEPES pH 7.5, 150 mM NaCl, and 1 mM TCEP. Pure fractions were pooled, concentrated to 10–15 mg/mL, aliquoted, and stored at –70 °C. The final yield from 1 L of bacteria expressing M<sup>PRO</sup> varied between 40 and 70 mg.

## 2.2. Protein Crystallization

For cocrystallization, protease and small molecule complexes were assembled by incubating 6 mg/mL of each M<sup>PRO</sup> with a 10-fold molar excess of ligand for 1 h at room temperature. Large crystals were obtained with 10–20% (*w/v*) PEG 3350, 0.2 M NaCl, and 0.1 M Bis-Tris Methane pH 5.5 by hanging drop vapor diffusion in pre-greased VDX trays (Hampton Research, Aliso Viejo, CA, USA) at room temperature. Varying the protein-to-mother liquor ratios (1 mL:2 mL, 2 mL:2 mL, 3 mL:2 mL) per well helped obtain large, diffraction-quality crystals. Cocrystals of M<sup>PRO</sup> with PF-00835231 appeared overnight and grew large enough for in-house data collection within 1 week. Crystals of SW1 and HKU15 M<sup>PRO</sup> complexes required 2–4 weeks of growth time to reach sufficient size for data collection.

For apo crystals, SW1 M<sup>PRO</sup> was screened against multiple crystallization conditions at 14.3 mg/mL (25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP). SW1 M<sup>PRO</sup> protein was supplemented with 1 mM EDTA and incubated for 10 min before drop setup. The best crystals obtained grew at 18 °C from JCSG+ Screen (NeXtal Biotechnologies, Holland, OH, USA) composed of 25% Peg3350, 0.2 M Ammonium Sulfate and 0.1 M Bis-Tris pH 5.5. The crystals reproduced easily and were grown by hanging drop vapor diffusion, with a 1:1 mL volume ratio. After 24 h, crystals were harvested and cryo-protected using 20% glycerol and well solution, flash-frozen in liquid nitrogen for data collection. Diffraction data were collected for apo SW1 M<sup>PRO</sup> crystals at the Advance Light Source beamline 5.0.2.

## 2.3. Data Collection and Structure Determination

Data collection for cocrystals was performed using a MicroMax-007HF X-ray generator equipped with a HyPix-6000HE detector (Rigaku Americas Corporation, The Woodlands, TX, USA) at the University of Massachusetts Chan Medical School, Crystallography and Structure Based Drug Design Core Facility. As data were collected under cryogenic conditions (100 K), crystals were protected from freeze damage by a quick soak in crystallization solution supplemented with 25% glycerol. Diffraction data were indexed, integrated, and scaled using CrysALIS<sup>PRO</sup>PX (Rigaku Americas Corporation, The Woodlands, TX, USA). Prior to analysis, data quality assessment was completed using Xtriage [18]. The structures were solved by molecular replacement using PHASER [19] with 7L0D [20] for the SARS-CoV-2 M<sup>PRO</sup>–PF-00835231 structure, SW1 apo M<sup>PRO</sup> (8FWX) to solve the SW1 M<sup>PRO</sup>–GC376 structure, and 7WKU [16] to solve the HKU15 M<sup>PRO</sup>–PF-00835231 structure. In each case, models for molecular replacement were prepared by removing ligands, structural and bulk solvent waters, and cryoprotectant molecules. SDF files of PF-00835231 (PDB: V2M), N-[(2S)-1-((2S,3S)-3,4-dihydroxy-1-[(3S)-2-oxopyrrolidin-3-yl]butan-2-yl)amino)-4-methyl-1-oxopentan-2-yl]-4-methoxy-1H-indole-2-carboxamide, and GC376 (PDB: UED), N~2~-[(benzyloxy)carbonyl]-N-[(2S)-1-hydroxy-3-[(3S)-2-oxopyrrolidin-3-

yl]propan-2-yl]-L-leucinamide were obtained from the Protein Data Bank. These models reflect the ligands in the covalently bound state. Using V2M and UED SDF files as inputs, eLBOW [21] generated ligand atomic position and constraint CIF files and corresponding PDB coordinates necessary to fit the small molecules into their respective densities. COOT [22] was used for modeling building, while iterative rounds of refinement were performed using PHENIX [18]. Due to merohedral twinning, the HKU15 M<sup>Pro</sup>-PF-00835231 structure was refined using twin law  $-h-k, k, -l$ . In order to limit bias throughout the refinement process, five percent of the data was reserved for the free R-value calculation [23]. Model quality assurance was assessed with MolProbity [24] prior to PDB deposition [25,26].

Structure analysis, superposition and figure generation were completed using PyMOL [27]. Hydrogen bonds and van der Waals contacts were determined as described previously [28]. Briefly, hydrogen bonds were determined according to the default criteria of the show\_contacts PyMOL Plugin, with a bond angle between 63 and 180 degrees and a distance of less than 4.0 Å between the proton and the heavy atom. For estimation of the van der Waals contacts, the interaction energies between the inhibitor atoms and the protease were calculated using a modified Lennard-Jones potential and summed up for each active site residue. X-ray data collection and crystallographic refinement statistics are presented in the Supporting Information (Table S1).

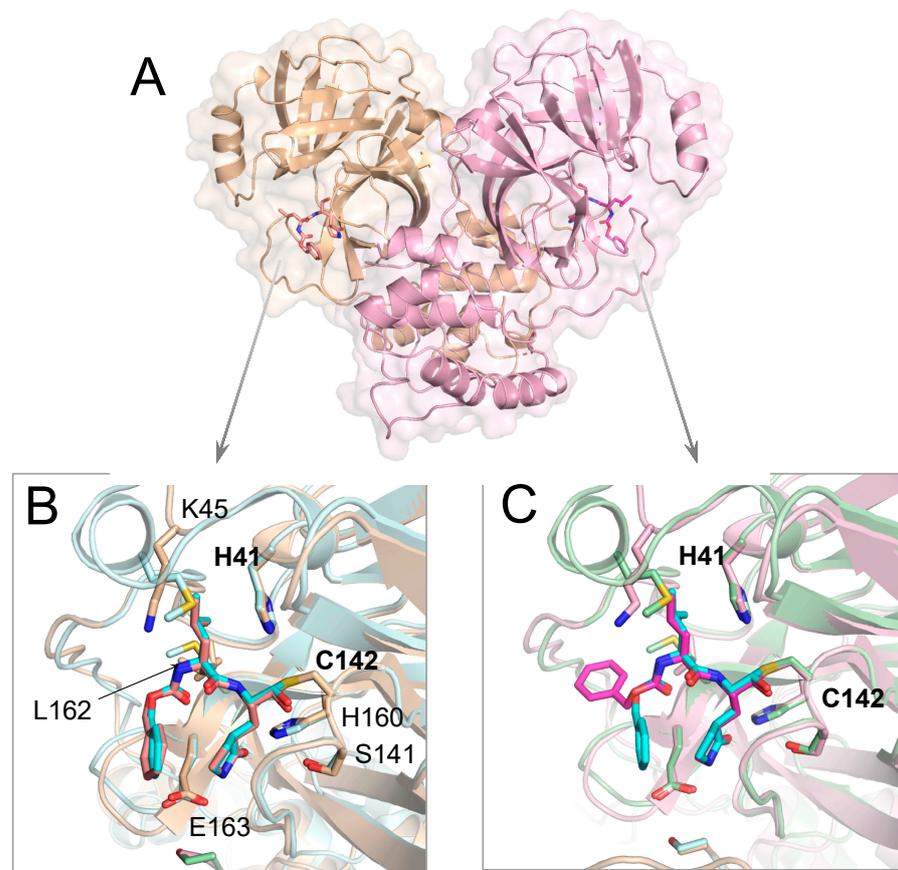
### 3. Results

The crystal structures of M<sup>Pro</sup> from gamma-CoV SW1 (beluga whale) were determined with a bound covalent inhibitor at the active site and in the apo form with no inhibitor. Both structures were solved in the P2<sub>1</sub> space group, with the apo form at 2.57 and the inhibitor-bound structure at sub-2 Å resolution. We also determined the inhibitor-bound structure of M<sup>Pro</sup> to 2.45 Å resolution from the porcine delta-CoV HKU15, and the same inhibitor (PF-00835231) bound to M<sup>Pro</sup> of SARS-CoV-2 for direct comparison. All four structures were solved with the dimer in the asymmetric unit. The cocrystal structures had full occupancy of inhibitor at both active sites, covalently bound to the catalytic cysteine. Crystallographic data collection and refinement statistics are presented in Table S1.

#### 3.1. Crystal Structure of Inhibitor-Bound Gamma-Coronavirus SW1 M<sup>Pro</sup>

The crystal structure of M<sup>Pro</sup> from beluga whale gamma-CoV SW1 shares the same overall fold of previously reported coronavirus M<sup>Pro</sup> structures, with the C-terminal helical Domain III facilitating dimerization and the active site between Domains I and II in both protomers accessible for ligand binding (Figure 1). The bisulfite prodrug GC376 converted to the aldehyde form (GC373) and attached covalently to the catalytic Cys142; both active sites were fully occupied by the covalent inhibitor. As previously seen in other M<sup>Pro</sup> cocrystal structures [9,29], the attachment of the aldehyde warhead can result in either (S) or (R)-configuration of the hemithioacetal to orient the hydroxyl group. The electron density was best fit with a 50/50 occupancy of (S)- and (R) configurations at both active sites (in the figures, only the (S) configuration is displayed for clarity).

The inhibitor GC376 was bound with the glutamine mimic g-lactam at the P1 position, making direct interactions with Glu163 and inter-ligand stacking with the P3 benzyl ring in one of the protomers. At the active site of the other protomer, the P3 group was flipped such that the benzyl ring interacted with the protease instead of making inter-ligand interactions. A similar flipped binding mode of the P3 group for GC376 has been observed before in a crystal structure of this inhibitor with SARS-CoV-2 M<sup>Pro</sup> (PDB ID: 6WTT; [12]), although in most structures, the benzyl ring stacks with the P1 group g-lactam (PDB IDs: 7JSU, 6WTJ; [11,13]). This dual binding mode of the P3 group suggests interactions at the S3 subsite could be optimized, ideally while maintaining inter-ligand interactions.

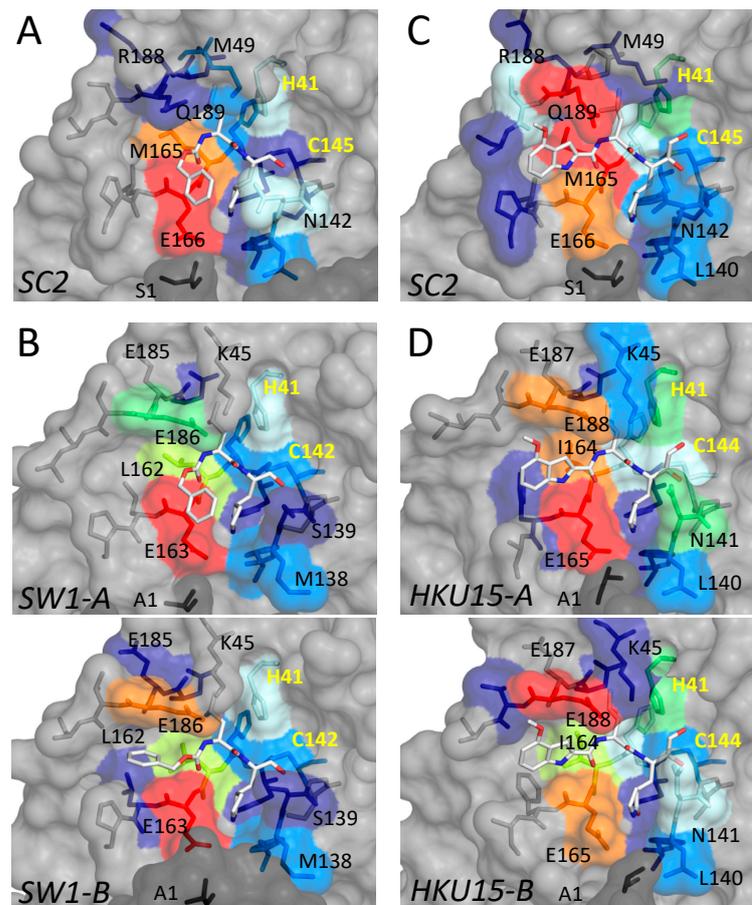


**Figure 1.** Crystal structure of  $M^{Pro}$  from gamma-CoV SW1 bound to inhibitor GC376. (A) The overall structure is displayed above in cartoon representation with a transparent surface and the two protomers colored pink and wheat. The inhibitor molecules covalently bound at the active sites are magenta and in stick representation. (B,C) The close-up views of the two active sites, where the SARS-CoV-2  $M^{Pro}$  structure (PDB 6WTJ) is superimposed and depicted in light blue and green cartoon representation. The key active site residues interacting with the inhibitor are labeled in (B). The terminal benzyl group of the inhibitor is flipped relative to the canonical binding mode (as seen in SARS-CoV-2  $M^{Pro}$ ; teal sticks) in one of the protomers (C).

The leucine side chain at the P2 position was tucked in the hydrophobic S2 subsite, consistent with the substrate amino acid preference at this position. The P2 group interacts mostly with Glu186, located in the 180s loop. Amino acid sequence alignment of  $M^{Pro}$  from various coronavirus species (Figure S2) shows that this position in the 180s loop is not conserved, with a Pro or Gln in alpha and beta-CoVs, respectively. Despite this variation, the binding conformation of the P2 moiety was mostly the same in this gamma-CoV  $M^{Pro}$  as in other reported structures, including SARS-CoV-2.

### 3.2. Inhibitor Binding to Delta and Gamma-Coronavirus $M^{Pro}$

In addition to the gamma-CoV SW1, we determined the crystal structure of  $M^{Pro}$  from the porcine delta-CoV HKU15 bound to PF-00835231. For both cocrystal structures, we quantified the contribution of active site residues to the packing with the inhibitor by calculating the van der Waals interactions between the inhibitor and protease (Figure 2). These interactions were compared with the packing of the same inhibitor at the active site of SARS-CoV-2  $M^{Pro}$ .



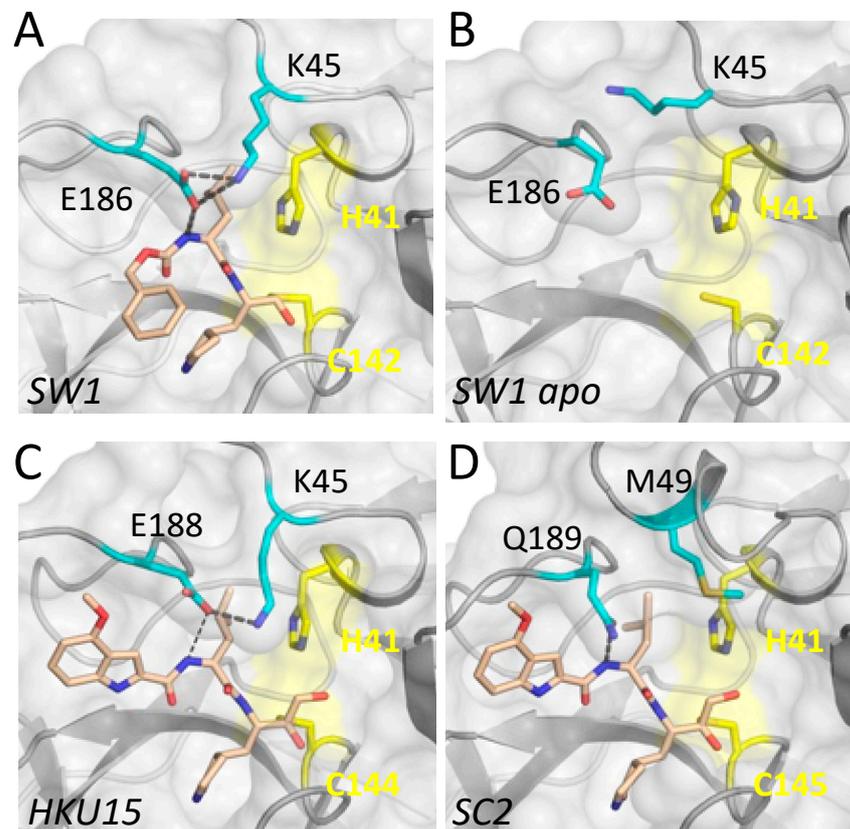
**Figure 2.** Packing of inhibitors with the active site residues in cocrystal structures of  $M^{Pro}$  in comparison with SARS-CoV-2 (SC2). (A,B) Inhibitor GC376 bound to  $M^{Pro}$  from SARS-CoV-2 (PDB ID: 6WTJ) and gamma-CoV SW1 from the beluga whale. (C,D) Inhibitor PF-00835231 (active form PF-07304814) bound to  $M^{Pro}$  from SARS-CoV-2 (PDB ID: 8DSU; presented in this work) and porcine delta-CoV HKU15. Interactions at the active sites of the two protomers are indicated by A and B for SW1 and HKU15. In all panels, the active site residues are colored blue to red for increasing interactions with the bound inhibitor. The catalytic residues are labeled in yellow font.

PF-00835231 bound at the active site of HKU15  $M^{Pro}$ , with the same binding mode observed with SARS-CoV-2. Due to the variation of amino acids at the active sites of the two variants, the degree of packing with the inhibitor varied despite the same binding mode. Interactions with the 180s loop were slightly lower, while those with the conserved Glu165/166 (HKU15/SARS-CoV-2 numbering, respectively) and N141/142 near the P1 group were higher in HKU15  $M^{Pro}$ . In addition to Glu165/166, residues I164/M165 and E188/Q189 had the most extensive interactions with PF-00835231 bound at the active site. There was some variation in the extent of van der Waals interactions of the inhibitor with the active site residues in the two protomers of HKU15  $M^{Pro}$  despite conserved binding mode due to slight rearrangements of the side chains (Figure 2D). Overall, the binding mode, as well as the interactions, were largely conserved between the two protomers and the two  $M^{Pro}$  variants.

For GC376, the interactions of the inhibitor with the active site residues of SW1  $M^{Pro}$  depended on the binding mode, as expected. In protomer A, where the P3 moiety bound in the same conformation as in SARS-CoV-2  $M^{Pro}$ , interactions with L162 were less extensive compared to the equivalent M165 (Figure 2A,B) while the leucine group of GC376 had considerably higher contacts with Glu186 of SW1  $M^{Pro}$ , compared to Gln189 at the equivalent location in SARS-CoV-2 (Figure 2A,B). These interactions were even more pronounced in protomer B where the P3 ring flipped to orient closer to Glu186. Thus,

for both inhibitors, PF-00835231 and GC376, the leucine-like P2 moiety had extensive interactions with the Glu186/188. The 180s loop containing this Glu/Gln amino acid varies in conformation, depending on the bound ligand (substrate peptide or inhibitor) in crystal structures reported [28]. Thus, the flexibility and variation in this loop might be relevant for accommodating different inhibitors.

Closer analysis of the determined cocrystal structures revealed that the E186/188 in the 180s loop establishes a salt bridge interaction with the side chain of K45 in inhibitor-bound structures of both SW1 and HKU15 M<sup>Pro</sup> (Figure 3). Amino acid sequence alignment shows that these two amino acids are not conserved in alpha or beta-CoV (Figure S2), and the 40s loop is three amino acids shorter in delta- and gamma-CoV. Accordingly, there are no salt bridges in SARS-CoV-2 M<sup>Pro</sup> structures involving this relatively variable loop. Interestingly, the side chain of E186 has rotated away from the 40s loop in the apo structure of SW1 M<sup>Pro</sup>; in one of the protomers, the K45 side chain had an interaction with this rotated E185, while in the other protomer, the salt bridge was completely missing. Thus inhibitor binding induced a flip in the E186 side chain conformer to establish a salt bridge with K45. This observation suggests that establishing the salt bridge may correlate with ligand binding and/or stabilization of the 180s loop.



**Figure 3.** Salt bridge involving the 180s loop in delta- and gamma-CoV M<sup>Pro</sup> crystal structures. (A) Inhibitor-bound M<sup>Pro</sup> from beluga whale gamma-CoV SW1. The side chains of residues K45 and E186 establish a salt bridge over the P2 moiety of the bound inhibitor. (B) The salt bridge is missing in the apo structure of SW1 M<sup>Pro</sup>. (C) Inhibitor PF-00835231 bound to M<sup>Pro</sup> from porcine delta-CoV HKU15. The salt bridge is between K45 and E188. (D) There are no residues that can form a similar salt bridge in alpha or beta-CoV M<sup>Pro</sup> variants, as shown for SARS-CoV-2 (SC2). In all panels, the protease is depicted in gray surface representation, and the catalytic residues are colored yellow.

#### 4. Discussion

The abundance of coronavirus species in nature, especially in wild bats and birds, urges the characterization of diverse coronavirus species that can adapt to infect domesticated

animals and humans to cause outbreaks. Here we characterized the main protease of two highly divergent delta- and gamma-CoVs that have already adapted to infect mammals (Supplementary Figure S3). The M<sup>Pro</sup> from these two genera had been severely lacking in structural characterization; the crystal structures we present here constitute the first M<sup>Pro</sup> structures from a mammalian-infecting gamma-CoV. We found that the covalent small molecule inhibitors GC-376 and PF-00835231 bind to SW1 and HKU15 M<sup>Pro</sup> in their canonical binding mode, with relatively minor rearrangements, in agreement with their broad potency against various viral species.

The main differences among M<sup>Pro</sup> variants from the beta, delta, and gamma-CoV we analyzed here were in the S2 subsite, where the amino acid sequences and structures are the most divergent (Figures 1–3 and S2). This subsite includes residues from the 180s loop, which is adaptable and changes conformation depending on the substrate peptide or inhibitor bound [28]. Residues in the 40s loop, which constitute the other part of the S2 subsite, are highly variable among CoV genera. These variations suggest the P2 position of inhibitors to be the most challenging to optimize to be able to effectively target all species. The shared leucine moiety at the P2 position of the inhibitors studied here (Figure S1) is likely critical for their broad activity, and modifications at this position to increase potency against a given species may abrogate binding to other M<sup>Pro</sup> variants. Thus, our results suggest that the design of potential pan-coronavirus inhibitors needs to consider the diversity of the S2 subsite among CoV species while optimizing the moiety at the P2 position. Our characterization of substrate peptide-bound cocrystal structures of SARS-CoV-2 M<sup>Pro</sup> to determine the substrate envelope indicated that many of the current inhibitors protrude from the substrate envelope at this P2 position [28]. Such protrusions not only make these inhibitors susceptible to resistance mutations [28,30] but will also likely prevent achieving pan-coronaviral activity.

The M<sup>Pro</sup> from the gamma-CoV SW1 has the same overall fold as other 3C and 3C-like proteases, as expected, despite low sequence identity to the main proteases of alpha- and beta-CoV, including SARS-CoV-2. Very few marine mammal-infecting gamma-CoVs have been identified to date: the whale-infecting SW1, whose M<sup>Pro</sup> we characterized here, and two species of dolphin-infecting coronaviruses [31,32]. Similarly, delta-CoVs are severely understudied. Both delta- and gamma-CoVs are thought to originate from birds and have already adapted to infect mammals [33]. The ability of coronaviruses to jump species barriers presents challenges for preventing and containing future outbreaks [33,34], signifying the need for pan-coronaviral treatment options. Virus crossover between species, including CoV infecting humans, is likely much more common than realized, and monitoring emerging viral species is important for pandemic preparedness. Considering the diversity and rapid evolution of the virus, especially of the spike protein, to evade antibody and vaccine protection, other viral proteins, such as M<sup>Pro</sup>, are more attractive and relatively well-conserved targets. Nevertheless, sequences from circulating SARS-CoV-2 isolates since the beginning of the COVID-19 pandemic display the high evolutionary potential of the virus. Moreover, there is vast diversity among other already known coronavirus species and characterization of their M<sup>Pro</sup> structures, as we have carried out here, should be useful in designing and developing effective pan-coronaviral inhibitors.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v15030781/s1>, Figure S1: Amino acid sequence alignment of M<sup>Pro</sup> from various coronavirus species, Figure S2: Phylogenetic tree of coronaviruses, Table S1: Crystallization and refinement statistics.

**Author Contributions:** Conceptualization, D.D., C.A.S. and N.K.Y.; methodology, A.M.S., Q.J.H., E.O. and M.M.; software, Q.J.H.; formal analysis, S.N.Z. and A.M.S.; investigation, S.N.Z., E.O., N.K.Y., E.O., M.M., S.M., M.K. and D.D.; resources, M.K., D.D. and C.A.S.; data curation, A.M.S.; writing—original draft preparation, N.K.Y.; writing—review and editing, N.K.Y., C.A.S. and D.D.; visualization, S.N.Z.; supervision, N.K.Y., D.D. and C.A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding. The research at the CAS laboratory was funded by Novartis Institutes for Biomedical Research.

**Data Availability Statement:** The data presented in this study are openly available in Protein Data Bank, accession codes 8DSU, 8FWX, 8E7N, and 8E7C.

**Acknowledgments:** We would like to thank Akbar Ali for assistance with Figure S1.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ruiz-Aravena, M.; McKee, C.; Gamble, A.; Lunn, T.; Morris, A.; Snedden, C.E.; Yinda, C.K.; Port, J.R.; Buchholz, D.W.; Yeo, Y.Y.; et al. Ecology, evolution and spillover of coronaviruses from bats. *Nat. Rev. Microbiol.* **2022**, *20*, 299–314. [[CrossRef](#)]
2. Cui, J.; Li, F.; Shi, Z.L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17*, 181–192. [[CrossRef](#)]
3. Drosten, C.; Kellam, P.; Memish, Z.A. Evidence for camel-to-human transmission of MERS coronavirus. *N. Engl. J. Med.* **2014**, *371*, 1359–1360. [[CrossRef](#)] [[PubMed](#)]
4. Guan, Y.; Zheng, B.J.; He, Y.Q.; Liu, X.L.; Zhuang, Z.X.; Cheung, C.L.; Luo, S.W.; Li, P.H.; Zhang, L.J.; Guan, Y.J.; et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **2003**, *302*, 276–278. [[CrossRef](#)]
5. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; et al. Structure of M. *Nature* **2020**, *582*, 289–293. [[CrossRef](#)] [[PubMed](#)]
6. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* **2020**, *368*, 409–412. [[CrossRef](#)] [[PubMed](#)]
7. Hoffman, R.L.; Kania, R.S.; Brothers, M.A.; Davies, J.F.; Ferre, R.A.; Gajiwala, K.S.; He, M.; Hogan, R.J.; Kozminski, K.; Li, L.Y.; et al. Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J. Med. Chem.* **2020**, *63*, 12725–12747. [[CrossRef](#)] [[PubMed](#)]
8. Owen, D.R.; Allerton, C.M.N.; Anderson, A.S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R.D.; Carlo, A.; Coffman, K.J.; et al. An oral SARS-CoV-2 M(pro) inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021**, *374*, 1586–1593. [[CrossRef](#)] [[PubMed](#)]
9. Kim, Y.; Lovell, S.; Tiew, K.C.; Mandadapu, S.R.; Alliston, K.R.; Battaile, K.P.; Groutas, W.C.; Chang, K.O. Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. *J. Virol.* **2012**, *86*, 11754–11762. [[CrossRef](#)] [[PubMed](#)]
10. Lu, J.; Chen, S.A.; Khan, M.B.; Brassard, R.; Arutyunova, E.; Lamer, T.; Vuong, W.; Fischer, C.; Young, H.S.; Vederas, J.C.; et al. Crystallization of Feline Coronavirus M(pro) With GC376 Reveals Mechanism of Inhibition. *Front. Chem.* **2022**, *10*, 852210. [[CrossRef](#)] [[PubMed](#)]
11. Iketani, S.; Forouhar, F.; Liu, H.; Hong, S.J.; Lin, F.Y.; Nair, M.S.; Zask, A.; Huang, Y.; Xing, L.; Stockwell, B.R.; et al. Lead compounds for the development of SARS-CoV-2 3CL protease inhibitors. *Nat. Commun.* **2021**, *12*, 2016. [[CrossRef](#)] [[PubMed](#)]
12. Ma, C.; Sacco, M.D.; Hurst, B.; Townsend, J.A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M.T.; Chen, Y.; et al. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* **2020**, *30*, 678–692. [[CrossRef](#)] [[PubMed](#)]
13. Vuong, W.; Khan, M.B.; Fischer, C.; Arutyunova, E.; Lamer, T.; Shields, J.; Saffran, H.A.; McKay, R.T.; van Belkum, M.J.; Joyce, M.A.; et al. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun.* **2020**, *11*, 4282. [[CrossRef](#)] [[PubMed](#)]
14. Arutyunova, E.; Khan, M.B.; Fischer, C.; Lu, J.; Lamer, T.; Vuong, W.; van Belkum, M.J.; McKay, R.T.; Tyrrell, D.L.; Vederas, J.C.; et al. N-Terminal Finger Stabilizes the S1 Pocket for the Reversible Feline Drug GC376 in the SARS-CoV-2 M(pro) Dimer. *J. Mol. Biol.* **2021**, *433*, 167003. [[CrossRef](#)] [[PubMed](#)]
15. Mihindukulasuriya, K.A.; Wu, G.; St Leger, J.; Nordhausen, R.W.; Wang, D. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J. Virol.* **2008**, *82*, 5084–5088. [[CrossRef](#)] [[PubMed](#)]
16. Wang, F.; Chen, C.; Wang, Z.; Han, X.; Shi, P.; Zhou, K.; Liu, X.; Xiao, Y.; Cai, Y.; Huang, J.; et al. The Structure of the Porcine Deltacoronavirus Main Protease Reveals a Conserved Target for the Design of Antivirals. *Viruses* **2022**, *14*, 486. [[CrossRef](#)]
17. Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.; Shen, W.; Li, J.; Zhou, Z.; Ding, Y.; Zhao, Q.; et al. Structures of two coronavirus main proteases: Implications for substrate binding and antiviral drug design. *J. Virol.* **2008**, *82*, 2515–2527. [[CrossRef](#)]
18. Adams, P.D.; Afonine, P.V.; Bunkoczi, G.; Chen, V.B.; Davis, I.W.; Echols, N.; Headd, J.J.; Hung, L.W.; Kapral, G.J.; Grosse-Kunstleve, R.W.; et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213–221. [[CrossRef](#)]
19. McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674. [[CrossRef](#)]

20. Lockbaum, G.J.; Reyes, A.C.; Lee, J.M.; Tilvawala, R.; Nalivaika, E.A.; Ali, A.; Kurt Yilmaz, N.; Thompson, P.R.; Schiffer, C.A. Crystal Structure of SARS-CoV-2 Main Protease in Complex with the Non-Covalent Inhibitor ML188. *Viruses* **2021**, *13*, 174. [[CrossRef](#)] [[PubMed](#)]
21. Moriarty, N.W.; Grosse-Kunstleve, R.W.; Adams, P.D. electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **2009**, *65*, 1074–1080. [[CrossRef](#)] [[PubMed](#)]
22. Emsley, P.; Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126–2132. [[CrossRef](#)] [[PubMed](#)]
23. Brunger, A.T. Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **1992**, *355*, 472–475. [[CrossRef](#)] [[PubMed](#)]
24. Davis, I.W.; Leaver-Fay, A.; Chen, V.B.; Block, J.N.; Kapral, G.J.; Wang, X.; Murray, L.W.; Arendall, W.B., 3rd; Snoeyink, J.; Richardson, J.S.; et al. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, *35*, W375–W383. [[CrossRef](#)]
25. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)]
26. Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10*, 980. [[CrossRef](#)]
27. DeLano, W.L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **2002**, *40*, 82–92.
28. Shaqra, A.M.; Zvornicanin, S.N.; Huang, Q.Y.J.; Lockbaum, G.J.; Knapp, M.; Tandeske, L.; Bakan, D.T.; Flynn, J.; Bolon, D.N.A.; Moquin, S.; et al. Defining the substrate envelope of SARS-CoV-2 main protease to predict and avoid drug resistance. *Nat. Commun.* **2022**, *13*, 3556. [[CrossRef](#)]
29. Ho, C.Y.; Yu, J.X.; Wang, Y.C.; Lin, Y.C.; Chiu, Y.F.; Gao, J.Y.; Lai, S.J.; Chen, M.J.; Huang, W.C.; Tien, N.; et al. A Structural Comparison of SARS-CoV-2 Main Protease and Animal Coronaviral Main Protease Reveals Species-Specific Ligand Binding and Dimerization Mechanism. *Int. J. Mol. Sci.* **2022**, *23*, 5669. [[CrossRef](#)]
30. Flynn, J.M.; Samant, N.; Schneider-Nachum, G.; Barkan, D.T.; Yilmaz, N.K.; Schiffer, C.A.; Moquin, S.A.; Dovala, D.; Bolon, D.N.A. Comprehensive fitness landscape of SARS-CoV-2 M(pro) reveals insights into viral resistance mechanisms. *Elife* **2022**, *11*, e77433. [[CrossRef](#)]
31. Wang, L.; Maddox, C.; Terio, K.; Lanka, S.; Fredrickson, R.; Novick, B.; Parry, C.; McClain, A.; Ross, K. Detection and Characterization of New Coronavirus in Bottlenose Dolphin, United States, 2019. *Emerg. Infect. Dis.* **2020**, *26*, 1610–1612. [[CrossRef](#)] [[PubMed](#)]
32. Woo, P.C.; Lau, S.K.; Lam, C.S.; Tsang, A.K.; Hui, S.W.; Fan, R.Y.; Martelli, P.; Yuen, K.Y. Discovery of a novel bottlenose dolphin coronavirus reveals a distinct species of marine mammal coronavirus in Gammacoronavirus. *J. Virol.* **2014**, *88*, 1318–1331. [[CrossRef](#)]
33. Woo, P.C.; Lau, S.K.; Lam, C.S.; Lau, C.C.; Tsang, A.K.; Lau, J.H.; Bai, R.; Teng, J.L.; Tsang, C.C.; Wang, M.; et al. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J. Virol.* **2012**, *86*, 3995–4008. [[CrossRef](#)] [[PubMed](#)]
34. Dhama, K.; Patel, S.K.; Sharun, K.; Pathak, M.; Tiwari, R.; Yatoo, M.I.; Malik, Y.S.; Sah, R.; Rabaan, A.A.; Panwar, P.K.; et al. SARS-CoV-2 jumping the species barrier: Zoonotic lessons from SARS, MERS and recent advances to combat this pandemic virus. *Travel Med. Infect. Dis.* **2020**, *37*, 101830. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.