Structure of the main protease from a global infectious human coronavirus, HCoV-HKU1

Qi Zhao¹, Shuang Li¹, Fei Xue¹, Yilong Zou¹, Cheng Chen¹, Mark Bartlam²

& Zihe Rao¹,²,³*

¹ Tsinghua-Nankai-IBP Joint Research Group for Structural Biology, Tsinghua University, Beijing 100084, China.
² College of Life Sciences and Tianjin State Laboratory of Protein Sciences, Nankai University, Tianjin 300071, China.
³ National Laboratory of Biomacromolecules, Institute of Biophysics (IBP), Chinese Academy of Sciences, Beijing 100101, China.

* Correspondence should be addressed to:
Zihe Rao, Ph.D.
Laboratory of Structural Biology,
Life Sciences Building,
Tsinghua University,
Beijing 100084, China
Tel: +86 10 62771493; Fax: +86 10 62773145; E-mail: raozh@xtal.tsinghua.edu.cn

Running title: Human coronavirus HKU1 M⁰ structure

Word count for abstract: 151

Word count for manuscript: 3,689
Abstract:
The newly emergent human coronavirus HKU1 (HCoV-HKU1) was first identified in Hong Kong in 2005. Infection by HCoV-HKU1 occurs worldwide and causes syndromes such as the common cold, bronchitis and pneumonia. The coronavirus main protease (M\textsuperscript{pro}), which is a key enzyme in viral replication via proteolytic processing of the replicase polyproteins, has been recognized as an attractive target for rational drug design. In this study, we report the structure of HCoV-HKU1 M\textsuperscript{pro} in complex with a Michael acceptor inhibitor N3. The structure of HCoV-HKU1 provides a high quality model for group 2A coronaviruses, which are distinct from group 2B coronaviruses such as SARS-CoV. The structure, together with activity assays, supports the relative conservation at the P1 position discovered by sequencing of the HCoV-HKU1 genome. Combined with structural data from other coronavirus M\textsuperscript{pro}, the HCoV-HKU1 M\textsuperscript{pro} structure reported here provides insights into both substrate preference and the design of anti-virals targeting coronaviruses.

Keywords: coronavirus, HCoV-HKU1, main protease, drug discovery, Michael acceptor inhibitor
Introduction

Coronaviruses (CoVs) are positive-strand RNA viruses that have been identified as the main etiologic agents responsible for a vast number of enteric, gastric and respiratory syndromes among both humans and/or animals (14-17, 19, 21, 23, 26, 30, 31, 34, 45). Coronaviruses can be divided into 3 groups: group 1 (including human coronavirus 229E and transmissible gastric enteric virus), group 2 (including human coronavirus OC43, murine hepatitis virus and bovine coronavirus) and group 3 (including avian infectious bronchitis virus).

Shortly after the emergence of SARS-CoV in 2003, group 2 CoVs were further divided into two subgroups, termed 2A and 2B (46). The classical group 2 viruses constitute subgroup 2A, while the newly-emergent SARS-CoV and its animal counterparts (37) form subgroup 2B. Group 1 and group 2 coronaviruses have more impact on human health than group 3, since group 3 coronaviruses (such as avian infectious bronchitis virus) can only infect avian species. Following the outbreak of SARS, group 2 coronaviruses have continued to attract greater attention for two reasons. First, they consist of human viruses (SARS-CoV, human coronavirus OC43), as well as several important animal viruses (mouse hepatitis virus and bovine coronavirus) that serve as useful models for coronavirus-host interactions. Second, group 2 coronaviruses are reported to have crossed the animal-to-human species barrier in two instances: one bat-to-human transmission in group 2B (27, 37) and one transmission event in group 2A coronaviruses when bovine
coronavirus led to the emergence of human coronavirus OC43 (36).

Group 2A human coronaviruses (HCoVs) were less widely studied prior to
the global SARS epidemic in 2003. However, they are closely associated with a
wide range of acute or chronic respiratory syndromes (4, 5, 8-12, 15, 20, 22, 35,
39, 40, 47). In the wake of the SARS outbreak, several novel HCoVs have been
discovered, one of which is the human coronavirus HKU1 (HCoV-HKU1) (10,
39). HCoV-HKU1 has achieved global distribution since it was first identified in
2005: infections were first characterized in Hong Kong (26), followed by the
identification of several strains of the virus in Korea (10), Europe (6, 17),
Australia (31), and North America (14). In contrast to the lethal SARS-CoV,
infection by HCoV-HKU1 usually leads to self-limiting syndromes affecting the
lower respiratory tract. Nevertheless, the consequences could be more severe
in patients with a compromised or immature immune system, such as asthma
sufferers or newborn infants (24). Genome sequencing has confirmed that the
HCoV-HKU1 virus belongs to coronavirus group 2A and shares high sequence
homology with MHV and BCoV (39).

The functional components of the coronavirus replication machinery are
released via post-translational cleavage by two or three proteases. These
proteases were first designated as the papain like protease (PLP) and 3C-like
protease (3CL) for their respective sequence homology to the papain and
rhinovirus 3C proteases. The 3CL protease is also commonly known as the
main protease (M^pro^) because of the major role it plays in the proteolytic
pathway, which makes it the most attractive pharmacological target for anti-coronavirus drug design. Coronavirus M\textsuperscript{pro} have been intensively studied and crystal structures have been determined of the M\textsuperscript{pro} from the following coronaviruses: human coronavirus strain 229E (HCoV-229E) (3), porcine transmissible gastric enteritis virus (TGEV) (2), avian infectious bronchitis virus (IBV) (41) and SARS-CoV (44). These structures are representative of group 1 (HCoV-229E, TGEV), group 2B (SARS CoV) and group 3 (IBV) coronaviruses. However, no structure of the M\textsuperscript{pro} from a group 2A coronaviruses (MHV, HCoV-HKU1, HCoV-OC43) has been determined to date. The absence of structural data presents a major obstacle for structure-aided drug optimization targeting group 2A coronaviruses.

The M\textsuperscript{pro} from different coronavirus groups are homologous in both sequence and main-chain architecture. They share a similar substrate binding sequence with a requirement for glutamine at the P1 position and a strong preference for leucine/methionine at P2. Based on this information, broad-spectrum lead compounds (43) with micromolar Ki values have been designed targeting CoV M\textsuperscript{pro}. However, structural data for the M\textsuperscript{pro} from classical group 2A CoVs are still not available, posing a problem for further optimization.

Although CoV M\textsuperscript{pro} exhibit absolute specificity for glutamine in the P1 position, recent research (38) has shown that the M\textsuperscript{pro} from HCoV-HKU1 may possess an unusual substrate preference at P1 site quite different from other
coronavirus M\textsuperscript{pro}. Here we report the structure of HCoV-HKU1 M\textsuperscript{pro}, which serves as a model for group 2A coronaviruses, in complex with a synthetic peptidomimetic inhibitor N3. The structure and subsequent enzyme activity assays help to resolve the issue of relative conservation at the P1 position based on genome sequencing. Moreover, this complex structure provides further structural data for rational drug design against human coronaviruses.

**Methods**

**cDNA and plasmid**

The cDNA encoding HCoV-HKU1 M\textsuperscript{pro} was kindly provided by Prof. KY Yuen from Department of Microbiology, Hong Kong University, HKSAR, China. BamH1 and Xho1 restriction sites were attached to the 5’ and 3’ ends separately by PCR, and the PCR product was first inserted into the pMD-18T vector (TAKARA). DNA of interest was then cleaved from the T vector and sub-cloned into a GST tagged expression vector pGEX-4T-1. The validity of the whole procedure was confirmed by DNA sequencing.

**Protein expression and purification**

The plasmid was first transformed into a commercial *E. coli* strain BL21(DE3) Rosetta (Invitrogen). After incubation at 37 °C overnight on an Amp+ algae LB plate, fresh transformants were inoculated into 5 ml LB media in the presence of 100 µg/ml ampicillin. After growth for 12 hours, the
incubation system was scaled up to 1L LB media with the same concentration of antibiotics in a 2L flask, and was shaken vigorously at 37 °C until the OD$_{600}$ reached 0.6. Cells were induced by 0.5 mM IPTG (Sigma) at 16 °C overnight.

Cell pellets were harvested by centrifugation, resuspended in 40 ml PBS buffer with 2 mM DTT and 7 mM β-ME, and sonicated on ice for 25 min. The supernatant was collected after centrifugation of the sonicant at 15,000 rpm for 40 min.

Affinity purification was achieved by letting the supernatant flow through 2 ml GST affinity media twice. On-column digestion lasted for 16 hours at 4 °C with thrombin (New England BioLabs), and the protein of interest was harvested and concentrated to 30 mg/ml. The N3 inhibitor was then added to a final molar ratio of 1:1 and incubated at 4 °C overnight. The HCoV-HKU1 M$^{\text{pro}}$-inhibitor complex was finally purified by gel filtration using a Superdex200 (10/30) column (GE Healthcare). Protein concentration was adjusted to 20 mg/ml for crystallization trials.

**Crystallization and structure determination**

Crystals of HCoV-HKU1 M$^{\text{pro}}$ were grown in 0.1 M imidazole, pH 6.0 and 0.6 M sodium acetate by the hanging-drop vapor diffusion method. Synchrotron X-ray diffraction data were collected on beamline BL-5A of the Photon Factory (Tsukuba, Japan) and processed to 2.5 Å resolution, using HKL2000 (29) for data indexing and scaling. Molecular replacement using the SARS M$^{\text{pro}}$
structure (PDB entry: 2AMQ, 48% identity) as a template was performed with PHASER (32). Manual rebuilding of the structure was performed using Coot (13) and the structure was refined using REFMAC in the CCP4 suite (1). Final modification was carried out using CNS (7). The volume of the S1 cavity was calculated using VOIDOO (25).

**Enzyme activity assays**

Substrate and analogs were designed through 3 rounds of affinity optimization (42) by substrate mimicry and from a library of substrate analogs. The substrate and analogs were synthesized by Prof. Dawei Ma from the Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai, China.

The strategy employed for enzyme activity assays of HCoV-HKU1 M\textsuperscript{pro} has been described previously (43). Activity assays for HCoV-HKU1 M\textsuperscript{pro} against the CoV consensus substrate and the HCoV-HKU1 specific substrate followed a similar protocol, described briefly below. The consensus substrate and HCoV-HKU1-specific substrates were fluorescent compounds with the sequences MCA-AVLQSGFR-Lys(Dnp)-Lys-NH\textsubscript{2} and (MCA)-PRLHCTTN-Lys(Dnp)-Lys-NH\textsubscript{2} respectively (greater than 95% purity, GL Biotech Shanghai Ltd, Shanghai, China). A P1-single mutant substrate was also synthesized with sequence MCA-AVLHSGFR-Lys (Dnp)-Lys-NH\textsubscript{2}.

The excitation and emission wavelengths of the fluorescent substrates
were 320 and 405 nm, respectively. A buffer consisting of 50 mM Tris-HCl (pH 7.3) and 1 mM EDTA was used for enzyme activity assays at 30 °C. The reaction was initiated by adding protease (final concentration of 2 µM) to a solution containing different final concentrations of the substrate (3.2-40 µM).

Strict kinetic parameters for the inhibition assay were determined according to the previously reported protocol (43). All results from enzyme activity assays were calculated using data based on at least 3 independent parallel experiments.

Results

Structural overview

Four protein molecules (denoted A, B, C and D) occupy one asymmetric unit, with one N3 molecule per protomer. Two of the protomers form a typical homodimer, while the remaining two protomers dimerize with their adjacent symmetry-related counterparts (see Figure 1a). Each protomer exhibits a three-domain (I−III) architecture common to other coronavirus M\textsuperscript{pro} structures (2, 3, 42, 44): domains I and II have chymotrypsin-like folds, and domain III displays a globular α-helical cluster unique to coronavirus M\textsuperscript{pro}. The catalytic site, including the Cys-His dyad, and the relatively shallow substrate binding pocket of HCoV-HKU1 M\textsuperscript{pro} are located in the cleft between domains I and II. The substrate binding pocket features 2 deeply buried sites (P1 and P2) and several sites with different levels of solvent exposure (P3, P4, and P5) (see
Figure 1b). X-ray data collection and refinement statistics are summarized in Table 2.

Michael acceptor and catalytic dyad

Clear and continuous electron density was observed between the reactive backbone carbon atom of the N3 substrate and the $S_\gamma$ atom of Cys145 in the inhibitor-bound HCoV-HKU1 M$^{pro}$ structure. We conclude that this reaction can be categorized as an electrophilic addition mediated by a Michael acceptor, obeying the $K_i - k_3$ kinetics (see Scheme 1), where $K_i$ is the dissociation constant and $k_3$ is the turnover number. As the covalently bound inhibitor is a mimic of the real peptide substrate, it is possible to model the transition state by treating the enzyme-inhibitor complex structure as a snapshot of the catalytic dyad, and hence to predict parameters of the $K_m - k_{cat}$ kinetics (see Scheme 2).

This catalytic dyad involves residues His41 and Cys145, and the intermediate state might be stabilized by the oxyanion hole (28) formed by the backbone amides of the oxyanion loop from Phe140 to Cys145 (see Figure 2a). The oxyanion hole is crucial to the stabilization of the intermediate state, and so formation of the oxyanion hole has a significant influence on $k_{cat}$. As discussed in the inhibitor design targeting human rhinovirus 3C proteases (28), correct organization of this oxyanion loop is also essential to the $k_3$ step for mechanism based suicide inhibitors.

Surrounding the $S_\gamma$ atom of Cys145, we observe well-defined amides from
the loop from residues 142-145. Similar to human rhinovirus 3C proteases (28), these amide dipoles construct a tetrahedral oxyanion hole. From native and complex structures of SARS-CoV M\textsuperscript{pro}, the correct orientation of these backbone amides is triggered and maintained by substrate-binding, in particular by the binding of the P1 residue and interaction between the “N-finger” and substrate (3, 44) where the backbone carbonyl of Leu141 is hydrogen bonded to the sidechain oxygen of Ser144. The correct position of Leu141 is maintained by a hydrogen bond between the carbonyl group of Phe140 and the amide group of the P1 side-chain, and by hydrophobic stacking between His163 and Phe140. Although analysis of the HCoV-HKU1 M\textsuperscript{pro} structure in complex with N3 (see Figure 2a) shows that the P1 sidechain exerts no direct influence on the residues forming the oxyanion hole, its side chain oxygen atom forms a strong hydrogen bond (2.6 Å) with His163 and helps to strengthen the stacking interaction with Phe140. Furthermore, the nitrogen atom of the P1 sidechain also forms a hydrogen bond (3.1 Å) with the backbone of Phe140, thus helping to maintain the oxyanion loop (Phe140-Cys145) in its proper conformation. For the above reasons, we conclude that the P1 sidechain is important for the network of interaction stabilizing the oxyanion hole.

The S1 pocket has a smaller size to accommodate P1 histidine.

Given its crucial role in the catalytic process, glutamine outperforms other
residues as the signature of the $\text{M}^{\text{pro}}$ substrate at the P1 position. In addition to this advantage, the sidechain of glutamine in the P1 position suitably fits with residues forming the S1 subsite via Van der Waals interactions (see Figure 2b).

From the HCoV-HKU1 genome sequence, 11 out of 12 $\text{M}^{\text{pro}}$ recognition sites have Gln at the P1 position. In our structure, the N3 molecule has a lactam ring as an analog to the glutamine residue (the cross-linking between the $\text{C}_\gamma$ and $\text{N}$ atoms helps to select the stretching conformation from the ensemble of rotamers, and better occupy the binding cleft). In HCoV-HKU1 Mpro structures, the lactum ring protrudes into the S1 pocket via a hydrogen bond to the imidazole ring NH of His162 at a distance of 2.6 Å. However, unlike the SARS-CoV $\text{M}^{\text{pro}}$ structure in complex with N3, the NH of the HKU1-N3 lactam ring fails to recruit a water molecule to satisfy a second S1 hydrogen bond. Instead, the N-terminal $\text{O}_\gamma$ atom might provide a weak electronegative interaction to stabilize the NH atom; the interaction is likely to be stronger due to the presence of redundant residues as a cloning artifact, hindering the N-terminal Ser from coming any closer to the NH of P1 side chain.

Nevertheless, compared with the main proteases from other coronavirus groups, the structure of the HCoV-HKU1 $\text{M}^{\text{pro}}$ has a S1 pocket with a relatively smaller volume of $\sim$18.1 Å$^3$. In contrast, the volume of the S1 pocket of TGEV $\text{M}^{\text{pro}}$ is $\sim$19.1 Å$^3$, of IBV $\text{M}^{\text{pro}}$ is $\sim$21.7 Å$^3$, and of SARS-CoV $\text{M}^{\text{pro}}$ is $\sim$19.5 Å$^3$. The reduced size of the S1 pocket might be caused by the position of the loop Leu167-Cys171, which is bent upward by about 90°. As a result, the smaller S1
pocket might tolerate mutation to short chain residues at the P1 position, in which case a weakened oxyanion hole is to be expected. Novel substrate specificity has already been found in the HCoV-HKU1 genome, where the Mpro recognition site between the helicase and exonuclease utilizes histidine instead of glutamine at the P1 position. Mimicking proteolysis in the cell, enzyme activity assays using a synthetic fluorogenic substrate confirm the existence of such a cleavage event in vitro and exhibit novel enzymatic properties compared with the consensus substrate (see Table 1).

Enzyme activity assays indicate that the affinity for a substrate containing a single mutation at the P1 position decreases to 30 percent of the affinity for the native consensus substrate, which can be attributed to the loss of a hydrogen bond resulting from the mutation of glutamine to histidine. The scissile velocity decreases to 3 percent, which is to be expected since the histidine residue lacks an oxygen atom required to form a strong hydrogen bond and support the intermediate oxyanion hole. However, when determining the influence of P2-P5 variance in the HCoV-HKU1 specific substrate, we observed an unusual 4-fold elevation in $K_m$ compared with the consensus substrate and a minor rescue of the scissile velocity (a 3-fold elevation of the single mutant substrate). This could be explained by the contribution of the non-P1 residues in the HCoV-HKU1 specific substrate, since activity assays for the single mutant substrate imply that mutation at the P1 position has a detrimental effect not only on the substrate binding affinity, but also on the substrate scissile velocity.
The S2 pocket presents group-specific features but no group-specific substrate preferences.

The P2 side-chain of the ligand protrudes into the S2 pocket via interactions with the hydrophobic side-chains of Met25, Pro52 and Tyr54 (see Figure 2B).

The lid of the pocket is covered by a short 3\_10 helical region from Ser45-Asn51.

To compare the diversity in the S2 pocket among all 3 coronavirus groups, the backbones of M\textsuperscript{pro} complex structures from all groups were superimposed (see Figure 3): group 1: TGEV M\textsuperscript{pro} in complex with the inhibitor N1, an ancestor of N3; group 3: IBV M\textsuperscript{pro} in complex with N3; group 2B: SARS-CoV in complex with inhibitor N3. We observed three modes of secondary structure: 3\_10 helix (HCoV-HKU1, SARS-CoV), a loose loop (IBV) and a tight loop (TGEV).

Interestingly, the clustering of secondary structure correlates with the temporary classification of coronaviruses. We then explored the natural recognition sequences to examine whether the group specific features could result in different substrate specificities at the P2 site (see table 3). After summarizing the P2 residue type in the protease recognition sites of the HKU1 PP1ab genome, we observe that M\textsuperscript{pro} prefer a hydrophobic residue at this position, which is also the case for SARS-CoV, IBV and TGEV. Although there are a few exceptions, such as asparagine or valine residues, leucine/methionine are the most abundant. This is consistent with the observation from our structure that the hydrophobic P2 side chain extends into
the deep S2 site without clashing with the Van der Waals surface of the pocket (Figure 2b). Therefore, on one hand, considering the flexibility of S2 pocket as well as the residual space after occupation by the P2 residue, the optimal choice for leucine or methionine might be related to the size of the S2 pocket. On the other hand, the similar preference on S2 sites among group 1, 2a, 2b, and 3 CoV M^{pro} does challenge the efficacy of designing group specific inhibitors by only altering P2 moieties.

The P3 position

Two out of four protomers in one asymmetric unit exhibit a solvent exposed P3 side-chain, which may interact weakly with the edge of the substrate-binding cleft via Van der Waals forces (see Figure 2b). To assess whether P3 side-chain variation can influence the potency of inhibition, we synthesized a small library of 6 inhibitors (see supplementary Table 1) to assay their affinity by the second order reaction coefficients. The results are summarized in supplementary Figure 1. It appears unlikely that the preference can be attributed exclusively to weak interactions. When investigating other possible P3 related interactions found in our HCoV-HKU1 M^{pro} structure, we scrutinized and evaluated the mainly hydrophobic structural interaction (see Figure 4a) between molecules B and D (see Figure 4b). From the crystal packing, we observed a crystallographic contact close to the P3 position (see supplementary material Figure 3) that we suspect might affect this property of the P3 residue. To
examine the physiological relevance, we conducted dynamic light scattering (DLS) experiments to check for higher molecular weight (MW) states that should be expected in solution if this interaction is related to one of the stable physiological states. However, DLS experiments did not identify a tetramer or higher MW complex in aqueous solution (see supplementary material Figure 2). Thus, the substrate-selectivity at the P3 position may be attributed to other weak factors such as solvent-sidechain interactions and Van der Waals interactions with the substrate binding cleft, rather than to a strong and direct interaction which is more likely to be influenced by crystal contacts.

Discussion

The HKU1 structure is a suitable model for group 2A CoV M<sup>pro</sup>

The crystal structure of M<sup>pro</sup> from HCoV-HKU1 is the first to be determined from a group 2A M<sup>pro</sup>. Since members of the group 2A coronaviruses share particularly high sequence identity (5) (see Figure 5a), non-conservative changes occur mainly in flexible regions of the HCoV-HKU1 M<sup>pro</sup> structure, including domain linkages and the molecular surface. Notable variations from residues 46 to 71 in group 2A sequences are located in or nearby the S2 pocket, which might infer properties relating to enzyme activity. However, since even greater differences between the different groups of CoVs exhibit no particular enzyme-specific preferences in the S2 pocket, the relatively small variations here may be unlikely to challenge the consensus substrate preference among
coronaviruses at the P2 position. The residue ranges 20-40, 140-160, 187-189 and residue 166 are highly conserved and are involved in formation of the S-1 and S1 pockets, together with the walls of the binding pocket of the P3 side-chain. Thus, it is reasonable to conclude that the HCoV-HKU1 M<sup>pro</sup> structure is a suitable model for the study of group 2A coronaviruses, both in terms of enzyme activity and inhibitor design.

**Michael acceptor inhibitors interact with CoV M<sup>pro</sup> in a similar manner**

As Michael acceptor suicide inhibitors, N3 and its derivatives co-crystallize with the CoV M<sup>pro</sup> in a similar manner (see Figure 6). The backbones of the peptidomimetic compounds align anti-parallel to the β-strands constituting the binding cleft. The P1 and P2 residues fit into the S1 and S2 pockets respectively and have a major contribution towards substrate preference: glutamine at P1, and leucine/methionine at P2. In our future optimization of M<sup>pro</sup> inhibitors, we consider the glutamine (or its analog) might be worth keeping in the P1 position, while it would reasonable to conduct a comparison of leucine or methionine for the evaluation of P2 residue. Aside from these deeply buried side-chains, the solvent exposed P3 provides no straightforward information for the substrate-enzyme interaction, though the variation at this position shows an obvious impact on inhibition. Alternatively, we might employ random screening for further optimization at the P3 position.
Conclusions

Structural data are now available for coronavirus M\(^{\text{pro}}\)-inhibitor complexes from all coronavirus groups, including the two subgroups of the group 2 coronaviruses. Moreover, these structures provide further confirmation for the efficacy of wide spectrum inhibitors at atomic resolution. From enzyme activity assays, we succeeded in identifying the atypical substrate specificity of HCoV-HKU1 M\(^{\text{pro}}\) with higher affinity (\(K_m\)) and lower reactivity (\(k_{\text{cat}}\)) than the consensus coronavirus M\(^{\text{pro}}\) substrate. We attributed these properties to the contribution of non-P1 residues and the distortion of the oxyanion hole. Although the S2 pockets from different groups share group-specific features, investigation of the natural recognition sequences does not observe different residue-type specificity at P2 site.

Considering the high identity shared by group 2A coronaviruses, these structural features of HCoV-HKU1 M\(^{\text{pro}}\), together with corresponding enzyme activity assays, will help to profile HCoV-HKU1 and other newly emerging etiologic agents from this group of coronaviruses.

Coordinates

Coordinates and structure factors for the HKU1 M\(^{\text{pro}}\) in complex with inhibitor N3 have been deposited in the Protein Data Bank with entry ID 3D23.
Acknowledgements

We thank K.Y. Yuen for providing cDNA of HCoV-HKU1 M\textsuperscript{pro}; Zhiyong Lou and Xiaohang Tong for technical assistance and data collection of HCoV-HKU1 M\textsuperscript{pro} in complex with N3. This work was supported by Project 973 of the Ministry of Science and Technology of China (Grant numbers 2006CB806503, 2007CB914301), the National Natural Science Foundation of China (Grant numbers 30221003, 30730022), the Sino-German Center (Grant number GZ236(202/9)), the Sino-European Project on SARS Diagnostics and Antivirals (SEPSDA) of the European Commission (Grant number 003831), and the Tsinghua University Ph.D student innovation fund.

Abbreviations used in this paper: SARS, severe acute respiratory syndrome; SARS-CoV, SARS coronavirus; M\textsuperscript{pro}, main protease; HCoV-HKU1, human coronavirus strain HKU1.
References


Figure Legends

Figure 1

a: Structural overview of 4 protomers (A: green, B: cyan, C: magenta, D: yellow) in one asymmetric unit, represented as cartoons. N3 inhibitors are shown as blue sticks.

b: Structural overview of the enzyme-inhibitor complex of one monomer unit. The main-chain of the enzyme is represented as blue cartoons and the synthetic inhibitor is shown as yellow sticks. The three domains are labeled.

Figure 2

a: Details of the interaction between the P1 side-chain and the defined oxyanion loop, shown in stereo representation. Side-chains are shown as sticks, and the crucial hydrogen bond between His163 and the substrate side-chain is shown by a cyan dash line.

b: Details of the substrate binding pocket. The inhibitor was shown in color scheme(C:white O:red N:blue), crucial residues of the enzyme was shown in color scheme(C: cyan O:red N: blue S:yellow). Hydrogen bonds were shown as red dash line.

Figure 3
S1 and S2 binding sites of HCoV-HKU1 M\textsuperscript{pro}

Main-chains of four M\textsuperscript{pro} structures are superimposed and displayed in the neighborhood of substrate binding site. The S1 and S2 binding sites are highlighted by light green shadows. The main-chains are represented in worm from. Different colors are used to represent the strain of coronavirus. Lemon: synthetic compound; magenta: HCoV-HKU1; light green: SARS CoV; light blue: TGEV; yellow: AIBV.

**Figure 4**

a: Overview of the P3 pocket. The inhibitor resides between the interface of molecule B and D. A cyan surface model is shown covering protomer B and D. The inhibitor is shown in magenta.

b: P3 interaction site of substrate in detail. Neighboring residues within 4 Å of the S3 site are colored green. An Fo-Fc map contoured at 1.5σ around the inhibitor is displayed in cyan. For the inhibitor, C atoms are colored yellow; N atoms are colored blue; O atoms are colored red. The protein carbon atoms are colored grey. Neighboring main-chains are displayed as white ribbons.

**Figure 5**

a: Sequence alignment of typical M\textsuperscript{pro} from coronavirus group 2A exhibits high homology.

The alignment was performed with ClustalW (33) and the final figure was
White letters with red background refer to identical residues; red letters with white background refer to conservative variation, black letters with white ground refer to non-conservative mutations.

b:

3D representation of non-conserved mutations in group 2A coronavirus $M^{\text{pro}}$ mapped onto the HCoV-HKU1 $M^{\text{pro}}$ structure. Identical residues are colored white, conserved mutations are colored yellow, non-conserved mutations are colored red. The inhibitor is colored blue.

**Figure 6:**

Superposition of representative coronavirus $M^{\text{pro}}$ in complex with Michael acceptor based inhibitors: group 1 (TGEV, blue), group 2A (HCoV-HKU1, green), group 2B (SARS CoV, magenta), group 3 (IBV, red).

The color of each inhibitor is consistent with its host.
### Scheme 1

\[ E + S \xrightleftharpoons[K_m]{k_{cat}} ES \xrightarrow{k_{cat}} E - S \]

### Scheme 2

\[ E + I \xrightleftharpoons[K_i]{k_i} EI \xrightarrow{k_i} E - I \]

### Table 1

<table>
<thead>
<tr>
<th>Substrate</th>
<th>( K_m ) (( \mu M ))</th>
<th>( k_{cat} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate C</td>
<td>83.2 ± 13.3</td>
<td>1.1 ± 0.12</td>
</tr>
<tr>
<td>Substrate H</td>
<td>22.3 ± 5.2</td>
<td>0.09 ± 0.013</td>
</tr>
<tr>
<td>Substrate SMC</td>
<td>265.3 ± 21.5</td>
<td>0.03 ± 0.001</td>
</tr>
</tbody>
</table>

Substrate C refers to the consensus substrate of coronavirus M\(^{\text{pro}}\) with the sequence MCA-AVLQSGFR-Lys(Dnp)-Lys-NH\(_2\).

Substrate H refers to the specific substrate of HCoV-HKU1 strain, including a mutation at the P1 site from glutamine to histidine, with the sequence: MCA-PRLHCTTN-Lys(Dnp)-Lys-NH\(_2\).

Substrate SMC refers to the consensus substrate, including a mutation at the P1 site from glutamine to histidine, with the sequence: MCA-AVLHSGFR-Lys(Dnp)-Lys-NH\(_2\).
### Table 2

X-ray data processing and refinement statistics

<table>
<thead>
<tr>
<th>Data collection statistics</th>
<th>HCoV-HKU1 M&lt;sup&gt;pro&lt;/sup&gt;-N3 complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.0</td>
</tr>
<tr>
<td>Resolution limit (Å)</td>
<td>50.0-2.48 (2.62 – 2.48)</td>
</tr>
<tr>
<td>Space group</td>
<td>P4₁</td>
</tr>
<tr>
<td>Cell parameters</td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>91.770</td>
</tr>
<tr>
<td>b (Å)</td>
<td>91.770</td>
</tr>
<tr>
<td>c (Å)</td>
<td>187.914</td>
</tr>
<tr>
<td>β (°)</td>
<td>90</td>
</tr>
<tr>
<td>Total reflections</td>
<td>510408</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>108501</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.1 (98.64)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.0 (4.9)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;*&lt;/</td>
<td>0.11</td>
</tr>
<tr>
<td>Sigma cutoff</td>
<td>0</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>16 (5)</td>
</tr>
<tr>
<td>Refinement statistics</td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.0-2.5</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;† (%)</td>
<td>22.9</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>28.5</td>
</tr>
<tr>
<td>rms deviation from ideal geometry</td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.012</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>1.59</td>
</tr>
<tr>
<td>Average B factor (Å&lt;sup&gt;2&lt;/sup&gt;)</td>
<td></td>
</tr>
<tr>
<td>protein</td>
<td>38.7</td>
</tr>
<tr>
<td>Small molecule</td>
<td>43.7</td>
</tr>
<tr>
<td>Ramachandran plot‡</td>
<td></td>
</tr>
<tr>
<td>Favored (%)</td>
<td>86.8</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>12.2</td>
</tr>
<tr>
<td>Generously allowed (%)</td>
<td>0.9</td>
</tr>
<tr>
<td>Disallowed (%)</td>
<td>0.1</td>
</tr>
</tbody>
</table>

*R<sub>merge</sub> = Σ||I<sub>i</sub>−⟨I⟩||/Σ|I|, where I<sub>i</sub> is the intensity of an individual reflection and ⟨I⟩ is the average intensity of that reflection.

†R<sub>work</sub> = Σ|F<sub>p</sub>|−|F<sub>c</sub>|/Σ|F<sub>p</sub>|, where F<sub>c</sub> is the calculated and F<sub>p</sub> is the observed structure factor amplitude.

‡Ramachandran plots were generated by using the program PROCHECK.

* Numbers in parentheses correspond to the highest resolution shell.
**Table 3**

P2 residues in different coronavirus genomes

<table>
<thead>
<tr>
<th>M(^{\text{pro}}) cleavage site</th>
<th>SARS-CoV</th>
<th>HCoV-HKU1</th>
<th>IBV</th>
<th>TGEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>L</td>
<td>L</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>V</td>
<td>I</td>
<td>M</td>
<td>V</td>
</tr>
<tr>
<td>4</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>5</td>
<td>L</td>
<td>M</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>6</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>V</td>
<td>V</td>
<td>M</td>
</tr>
<tr>
<td>8</td>
<td>L</td>
<td>M</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>9</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>10</td>
<td>L</td>
<td>L</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>11</td>
<td>L</td>
<td>M</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>