



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

# Production of Authentic SARS-CoV M<sup>Pro</sup> with Enhanced Activity: Application as a Novel Tag-cleavage Endopeptidase for Protein Overproduction

Xiaoyu Xue<sup>1,2,†</sup>, Haitao Yang<sup>1,2,†</sup>, Wei Shen<sup>1</sup>, Qi Zhao<sup>1</sup>, Jun Li<sup>1</sup>  
Kailin Yang<sup>1</sup>, Cheng Chen<sup>1</sup>, Yinghua Jin<sup>1</sup>, Mark Bartlam<sup>1,2</sup>  
and Zihao Rao<sup>1,2,3\*</sup>

<sup>1</sup>*Tsinghua-Nankai-IBP Joint Research Group for Structural Biology, Tsinghua University Beijing 100084, China*

<sup>2</sup>*National Laboratory of Biomacromolecules, Institute of Biophysics (IBP), Chinese Academy of Sciences Beijing 100101, China*

<sup>3</sup>*Nankai University Tianjin 300071, China*

The viral proteases have proven to be the most selective and useful for removing the fusion tags in fusion protein expression systems. As a key enzyme in the viral life-cycle, the main protease (M<sup>Pro</sup>) is most attractive for drug design targeting the SARS coronavirus (SARS-CoV), the etiological agent responsible for the outbreak of severe acute respiratory syndrome (SARS) in 2003. In this study, SARS-CoV M<sup>Pro</sup> was used to specifically remove the GST tag in a new fusion protein expression system. We report a new method to produce wild-type (WT) SARS-CoV M<sup>Pro</sup> with authentic N and C termini, and compare the activity of WT protease with those of three different types of SARS-CoV M<sup>Pro</sup> with additional residues at the N or C terminus. Our results show that additional residues at the N terminus, but not at the C terminus, of M<sup>Pro</sup> are detrimental to enzyme activity. To explain this, the crystal structures of WT SARS-CoV M<sup>Pro</sup> and its complex with a Michael acceptor inhibitor were determined to 1.6 Å and 1.95 Å resolution respectively. These crystal structures reveal that the first residue of this protease is important for sustaining the substrate-binding pocket and inhibitor binding. This study suggests that SARS-CoV M<sup>Pro</sup> could serve as a new tag-cleavage endopeptidase for protein overproduction, and the WT SARS-CoV M<sup>Pro</sup> is more appropriate for mechanistic characterization and inhibitor design.

© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** SARS coronavirus; main protease; fusion protein; crystal structure; endopeptidase

\*Corresponding author

## Introduction

In the age of proteomics, the production of pure protein in a high-throughput manner is required

† X.X. and H.Y. made equal contributions to this work.

Abbreviations used: SARS, severe acute respiratory syndrome; SARS-CoV, SARS coronavirus; M<sup>Pro</sup>, main protease; GST, glutathione-S-transferase; WT, wild-type enzyme with authentic N and C termini; GS-WT, wild-type enzyme with two additional amino acid residues (GS) at the N terminus; GPLGS-WT, wild-type enzyme with five additional amino acid residues (GPLGS) at the N terminus; WT-GPH<sub>6</sub>, wild-type enzyme with eight additional amino acid residues (GPHHHHHH) at the C terminus.

E-mail address of the corresponding author:  
raozh@xtal.tsinghua.edu.cn

for both structural and functional studies. Especially for protein structure studies, one of the main bottlenecks is to produce adequate quantities of soluble and properly folded recombinant proteins.<sup>1</sup> Fusion protein expression systems have been widely used for this purpose in basic research and in industry. Fusion domains (or small “tags”) are expressed as partners of the passenger proteins and are generally removed after purification, as they might interfere with the function or other biochemical or biophysical characteristics of the protein.<sup>2</sup> Hence, some endopeptidases, such as bovine thrombin, bovine Factor Xa, human rhinovirus 3C protease and tobacco etch virus (TEV) protease, are routinely used to remove fusion domains. Of these, the viral proteases, e.g. human rhinovirus 3C protease<sup>3,4</sup> and TEV protease,<sup>5,6</sup> have proven to be the most selective and useful to date. Therefore, the development of novel viral

proteases with high specificity and activity will be helpful to proteomics research.

The SARS coronavirus (SARS-CoV) is the etiological agent responsible for the global outbreak of a life-threatening disease that caused approximately 800 deaths worldwide.<sup>7–11</sup> The coronavirus main protease (M<sup>Pro</sup>), which plays a key role in mediating viral replication and transcription, has been identified as the most attractive target for anti-SARS drug design.<sup>12–14</sup> In 2003, our group published the first crystal structure of SARS-CoV M<sup>Pro</sup>.<sup>14</sup> Thereafter, several other groups published the structures of M<sup>Pro</sup>,<sup>15–17</sup> and its complex with an aza-peptide epoxide inhibitor.<sup>18</sup> The M<sup>Pro</sup> can form a homodimer both in the crystal<sup>14,17,18</sup> and in solution.<sup>19,20</sup> Each protomer consists of three domains: domains I and II resemble chymotrypsin, whereas domain III has a globular cluster of five, mostly antiparallel,  $\alpha$ -helices. SARS-CoV M<sup>Pro</sup> has a catalytic dyad consisting of His41 and Cys145, and its substrate-binding pocket is located in the cleft between domains I and II. The N-terminal residues 1–7 of domain I (or N-finger) of M<sup>Pro</sup> are considered to have an important role in its proteolytic activity,<sup>14,21–23</sup> although their importance in dimerization has been reported with inconsistent results.<sup>21–23</sup> In particular, mutation of Arg4, which is involved in forming an ion pair, results in a fourfold decrease in activity.<sup>21</sup> Deletion of the N-terminal residues 1–7 results in an almost completely inactive SARS-CoV M<sup>Pro</sup>.<sup>23</sup> SARS-CoV M<sup>Pro</sup> has been characterized sufficiently and two particular aspects drew our attention. First, it is highly selective for substrate sequence.<sup>24</sup> Second, it displays a high level of proteolytic activity, although inconsistent kinetic parameters have been reported for SARS-CoV M<sup>Pro</sup> with  $k_{cat}/K_m$  ranging from 20–29,000 M<sup>-1</sup>s<sup>-1</sup>.<sup>19,20,25–27</sup> These properties indicate that SARS-CoV M<sup>Pro</sup> is a promising candidate as an endopeptidase to remove fusion tags.

In this study, SARS-CoV M<sup>Pro</sup> was utilized to remove the fusion domain in overproduction of calbindin D28k (a member of the calmodulin superfamily) from a novel glutathione-S-transferase (GST) fusion protein expression system.<sup>28–31</sup> To obtain the peptidase with high catalytic efficiency, we developed a new method to produce wild-type (WT) SARS-CoV M<sup>Pro</sup> with authentic N and C termini, and investigated the activity difference among several SARS-CoV M<sup>Pro</sup> constructs. The crystal structures of WT M<sup>Pro</sup> and its complex with an inhibitor were solved to explain the mechanism of enhanced activity.

## Results

### Construction of a new GST fusion protein expression vector (pGSTM)

In order to take advantage of SARS-CoV M<sup>Pro</sup> as an endopeptidase, a new GST fusion protein expression vector (designated pGSTM) was con-

structed from another fusion vector pGEX-6p-1 (GE Healthcare). The resulting plasmid retains the whole multiple cloning sites of pGEX-6p-1. However, the linker between the GST gene and the gene of interest was replaced by nucleotides encoding a unique N-terminal autocleavage site of SARS-CoV M<sup>Pro</sup> consisting of 11 amino acid residues (see Figure 1(a)). Sequencing confirmed that the cleavage site (TSAVLQSGFRK) was inserted correctly.

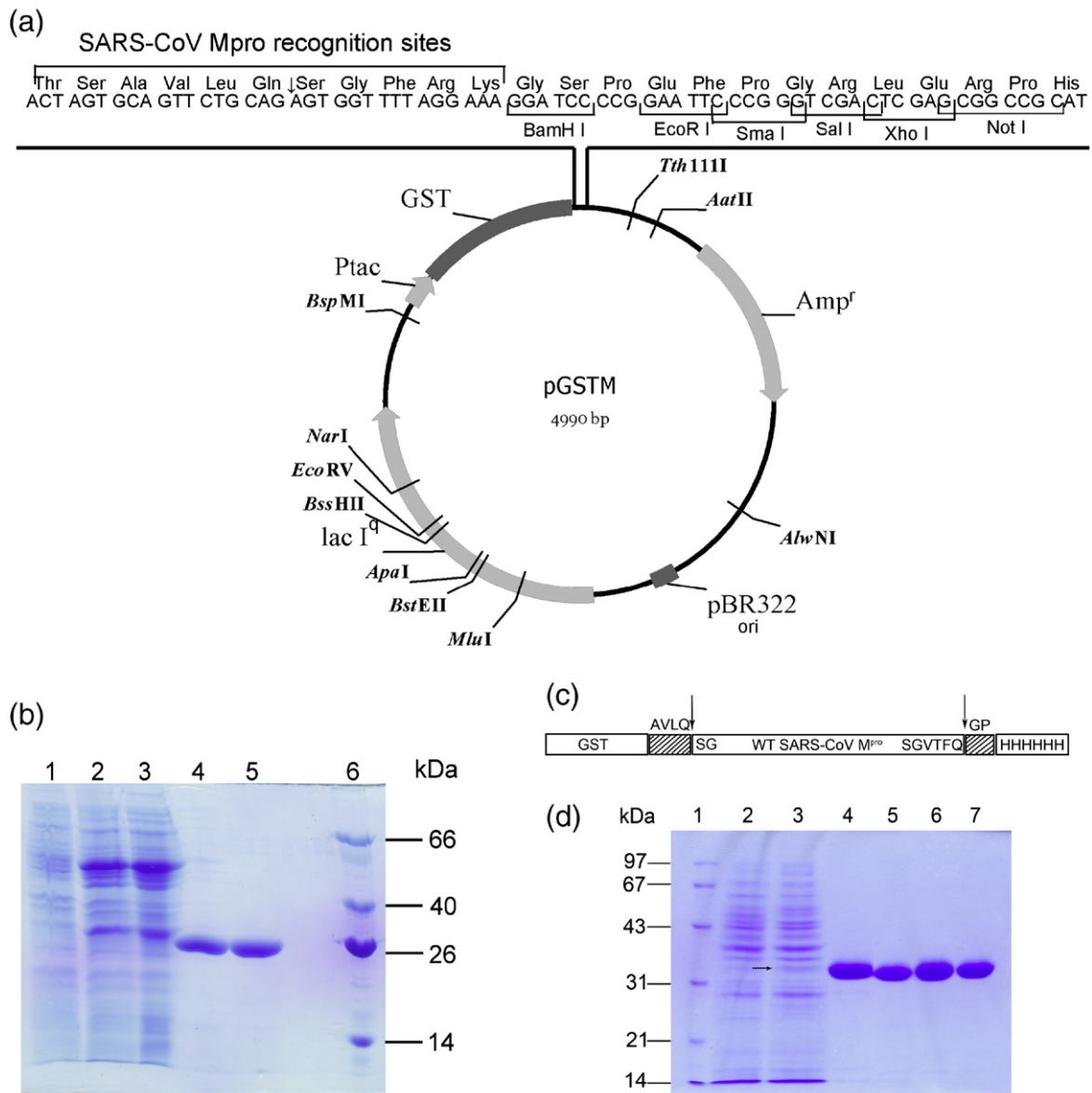
### Expression and purification of calbindin D28k using the pGSTM expression system

Calbindin D28k, a member of the calmodulin superfamily, has been proposed to function as an important intracellular Ca<sup>2+</sup>-buffering protein.<sup>31</sup> We took it as our protein of interest for structural studies. This protein was cloned, expressed and purified using our new protein expression system. After one-step purification, calbindin D28k protein, with five additional residues (SGFRK-) at the N terminus and an authentic C terminus, was analyzed by SDS-PAGE and the purity was shown to be >90% (Figure 1(b)). Around 70–80 mg of calbindin D28k was obtained from 1 l of bacterial culture. The crystallographic analysis of calbindin D28k is underway.

### Preparation of WT SARS-CoV M<sup>Pro</sup>

In an earlier study, wild-type SARS-CoV M<sup>Pro</sup> with two additional residues (GS) at the N terminus (GS-WT) was used for the determination of kinetic parameters.<sup>12</sup> To avoid any potential effects from the extraneous N-terminal residues on enzyme activity, we developed a new method to produce WT SARS-CoV M<sup>Pro</sup>. First, the four amino acids AVLQ, which correspond to the P4–P1 sites of the N-terminal autocleavage sequence of SARS-CoV M<sup>Pro</sup>, were introduced between the GST tag and the first residue of the protease. Thus, the authentic N terminus would become available by autocleavage during protein expression. Second, the eight amino acids GPHHHHHH (abbreviated as GPH<sub>6</sub>, where GP correspond to the P1' and P2' sites of rhinovirus 3C protease; see Table 1) were added after the last glutamine residue at the C terminus (Figure 1(c)). This strategy was intended to yield an authentic C terminus following cleavage by rhinovirus 3C protease, despite appearing unreasonable, since SARS-CoV and rhinovirus belong to different families (Coronaviridae and Picornaviridae, respectively). This strategy was used for the following reasons.

With an extra domain III, SARS-CoV M<sup>Pro</sup> differs from rhinovirus 3C protease in many aspects. However, they share several similarities in both cleavage sequence and the substrate-recognition pocket. From the native/complex structures of these two proteases published to date, the S4, S2, S1, S1' and S2' subsites are critical for substrate binding to the two proteases.<sup>12,14,18,32,33</sup> Thus, the corresponding P4, P2, P1, P1' and P2' sites were



**Figure 1.** A new GST fusion protein expression system. (a) The map of the pGSTM vector design. The cleavage site for SARS-CoV M<sup>Pro</sup> is labeled. (b) Expression and purification of calbindin D28k using the pGSTM expression system. Lane 1, total cell extract for calbindin D28k before induction; lane 2, total cell extract for calbindin D28k after induction overnight; lane 3, supernatant of the cell lysate; lanes 4 and 5, purified calbindin D28k; lane 6, protein molecular mass marker. (c) Schematic plot of WT SARS-CoV M<sup>Pro</sup> construct designed. (d) Expression and purification of WT M<sup>Pro</sup>. Lane 1, protein molecular mass marker; lane 2, total cell extract for WT M<sup>Pro</sup> before induction; lane 3, total cell extract for WT M<sup>Pro</sup> after induction overnight; lane 4, WT-GPH<sub>6</sub> after affinity chromatography; lane 5, WT after cleavage by rhinovirus 3C protease; lane 6, GPLGS-WT; lane 7, GS-WT.

taken into consideration when designing recognition sequences. Table 1 lists the C-terminal autocleavage site for SARS-CoV M<sup>Pro</sup>, and the substrate sequence for human rhinovirus 3C protease. It

**Table 1.** C-terminal autocleavage site for SARS-CoV M<sup>Pro</sup> and typical substrate sequence for human rhinovirus 3C protease

Protease	P6	P5	P4	P3	P2	P1	P1'	P2'
SARS-CoV M <sup>Pro</sup> (C terminus)	S	G	V	T	F	Q	G	K
Human rhinovirus 3C	L	E	V	L	F	Q	G	P

shows that the five sites are identical for the two proteases, with the exception of P2'. The substantial difference embodied in the two substrate sequences is limited to P2'. Hence, we reasoned that rhinovirus 3C protease could recognize the octapeptide substrate SGVTFQ↓GP, whose P1–P6 sites were derived from the C-terminal autoprocessing site of SARS-CoV M<sup>Pro</sup>. However, replacing lysine with proline at the P2' site would arrest autocleavage. Our group has recently solved the crystal structure of a mutant of SARS-CoV M<sup>Pro</sup> in complex with an 11 amino acid residue peptidyl substrate (unpublished results). The complex structure revealed that substitution of

P2' with Pro would result in steric hindrance between P2' (Pro) and the main chain of SARS-CoV, hindering substrate binding. On the basis of the above analysis, we hypothesized that rhinovirus 3C protease but not SARS-CoV M<sup>Pro</sup> could recognize the octapeptide substrate, SGVTFQ↓GP. SDS-PAGE analysis (see Figure 1(d)) showed that after Ni-affinity chromatography, the SARS-CoV M<sup>Pro</sup> is in the WT-GPH<sub>6</sub> form, implying that the N-terminal GST tag was removed efficiently by autocleavage during protein expression in *Escherichia coli*. The difference between lane 4 and lane 5 indicates that the GPH<sub>6</sub> tag was indeed removed by rhinovirus 3C protease cleavage, confirming that the modified substrate of SARS-CoV M<sup>Pro</sup> could be recognized by rhinovirus 3C protease. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis indicated that WT-GPH<sub>6</sub> and WT M<sup>Pro</sup>s were in accord with their predicted molecular mass (see Table 2).

### Enzyme activity assay of four types of SARS-CoV M<sup>Pro</sup>s

To ascertain whether the additional residues would interfere with the catalytic activity of SARS-CoV M<sup>Pro</sup>, we determined the kinetic parameters for GPLGS-WT (with five additional residues (GPLGS) at the N terminus of WT), GS-WT, WT-GPH<sub>6</sub> and WT SARS-CoV M<sup>Pro</sup>s (see Table 3). The catalytic efficiency of an enzyme is best defined by  $k_{cat}/K_m$ .<sup>34</sup> Table 3 shows that the WT protease has the highest cleavage efficiency ( $k_{cat}/K_m=26,500\text{ M}^{-1}\text{ s}^{-1}$ ) while GPLGS-WT has the lowest ( $k_{cat}/K_m=167\text{ M}^{-1}\text{ s}^{-1}$ ). The activity of the WT protease with authentic N and C termini was more than 150-fold greater than that of GPLGS-WT and 20-fold greater than that of GS-WT. Furthermore, our results suggest that increasing the number of additional residues at the N terminus would result in a greater decrease in activity. However, the activity of WT-GPH<sub>6</sub> was about one-third of that of the WT enzyme, which suggests that additional residues at the C terminus had less effect on activity.

### Crystal structure of WT SARS-CoV M<sup>Pro</sup>

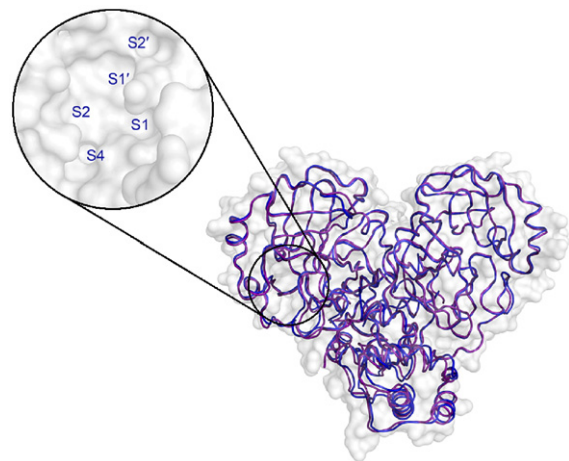
In order to shed light on the difference in activity, we determined the crystal structure of the WT SARS-CoV M<sup>Pro</sup> to 1.6 Å resolution. In our published crystal structure of SARS-CoV M<sup>Pro</sup> (in the GPLGS-WT form), the crystal belongs to the space group  $P2_1$  and the asymmetric unit contains a dimer (with the two protomers designated A and

**Table 3.** Comparison of enzyme activities of four types of SARS-CoV M<sup>Pro</sup>

SARS-CoV M <sup>Pro</sup>	$K_m$ ( $\mu\text{M}$ )	$k_{cat}$ ( $\text{s}^{-1}$ )	$k_{cat}/K_m$ ( $\text{M}^{-1}\text{ s}^{-1}$ )
GPLGS-WT	126±8	0.021±0.001	167
GS-WT <sup>12</sup>	129±7	0.14±0.01	1100
WT-GPH <sub>6</sub>	61.0±2.9	0.41±0.02	6800
WT	40.0±0.8	1.06±0.04	26,500

B). Within the dimer, protomer A was in the active form, while protomer B showed an inactive form, resulting from a partially collapsed S1 subsite.<sup>14</sup> In contrast, the crystal of WT SARS-CoV M<sup>Pro</sup> belongs to space group C2, and each asymmetric unit contains only one protomer of a typical dimer (with the two protomers designated A\* and B\*). The two protomers in the dimer are related by a crystallographic 2-fold symmetry axis, and each has a catalytically competent conformation. All residues of the protomer (residues 1–306) were identified from electron density maps. Apart from differences in the S1 subsite, the SARS-CoV M<sup>Pro</sup> protomer as seen in the new structure is very similar to each member of the dimer in the original structure (see Figure 2). Compared to the latter structure, the protomer in the new crystal forms display overall rmsd for C $\alpha$  atoms of 0.6 Å for protomer A and 0.7 Å for protomer B. In the following discussion, we consider the differences between the substrate-binding sites of the two structures in further detail.

The S1 site of SARS-CoV M<sup>Pro</sup>, which has absolute specificity for Gln in the P1 site, consists of the side-chains of His163 and Phe140, and the main-chain atoms of GluA166, Asn142, Gly143 and HisA172. In the S1 site of protomer A in our original structure (GPLGS-WT), the NH group of SerB1 was unable to donate hydrogen bonds simultaneously to the carboxylate group of GluA166 and the main-chain carbonyl group of PheA140, due to the presence of additional residues at the N terminus, although the



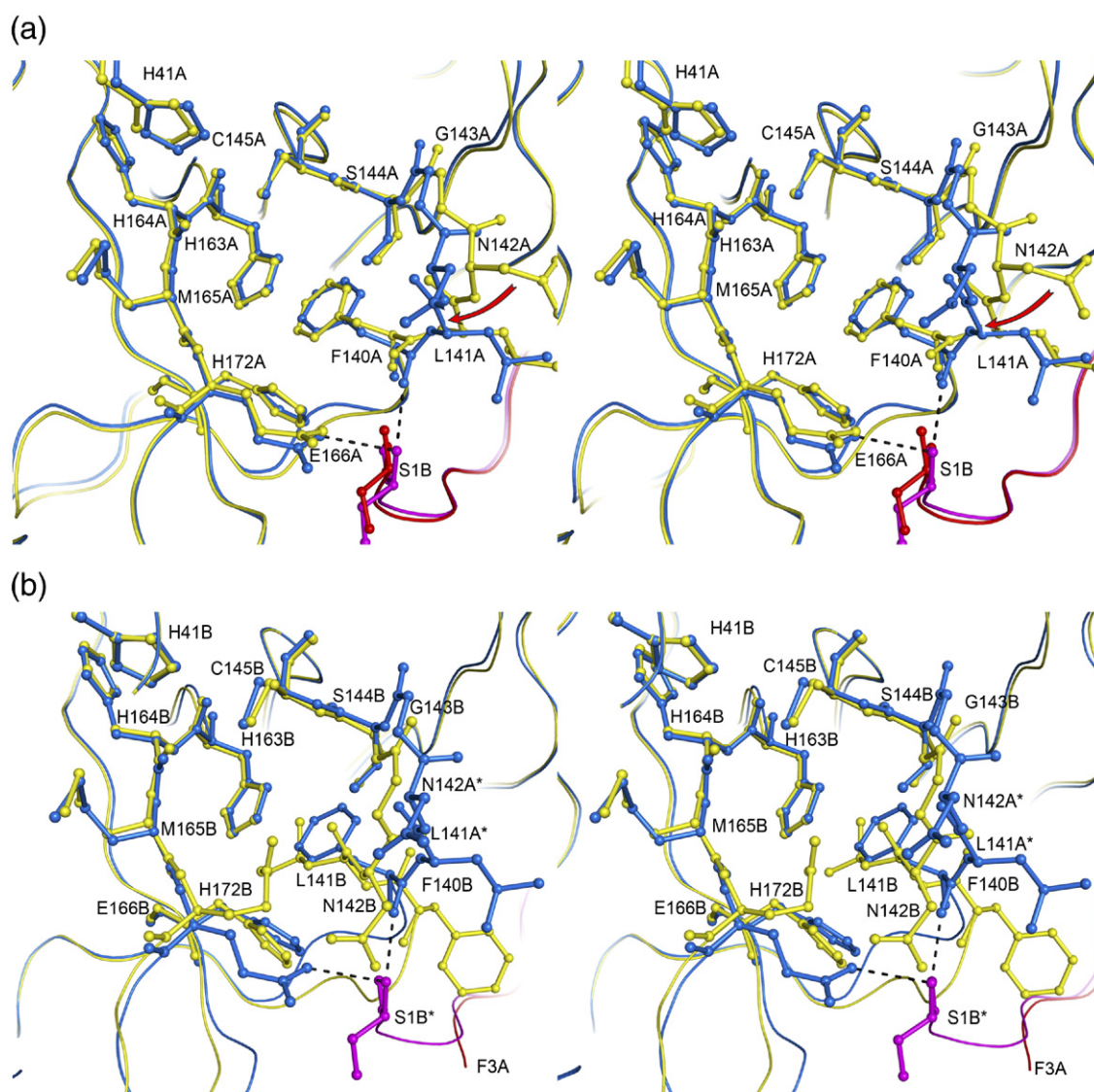
**Figure 2.** Superposition of WT and GPLGS-WT M<sup>Pro</sup>s. The substrate-binding pocket of one protomer is in surface representation. GPLGS-WT is in blue, WT is in magenta.

**Table 2.** Molecular mass of WT-GPH<sub>6</sub> and WT measured by MALDI-TOF MS

SARS-CoV M <sup>Pro</sup>	Predicted (Da)	Measured (Da)
WT-GPH <sub>6</sub>	34,822	34,819
WT	33,846	33,817

distance is suitable for hydrogen bond formation. However, in the newly solved WT structure, the amino group (NH<sub>2</sub>) of SerB\*1 in one protomer donates a 3.0 Å hydrogen bond to the carboxylate group of GluA\*166 (Figure 3(a)) and a 2.7 Å hydrogen bond to the main-chain carbonyl group of PheA\*140, thus stabilizing the S1 pocket. This induces a series of conformational changes. For instance, the NH of GlyA\*143, which participates directly in the formation of the oxyanion hole, moves by 0.8 Å towards the active site; the main chain of residues 142-143 moves towards the S1

subsite; and the side-chain of AsnA\*142 flips over with a 6 Å shift. Stabilized by SerB\*1, protomer A\* displays a more catalytically competent conformation in the S1 subsite than protomer A in our original structure. In our original structure, the S1 pocket of protomer B is partly collapsed compared with the WT protomer: no electron density was visible for residues A1 and A2; GluB166 reorientates to interact with the possibly protonated HisB163; PheB140 undergoes a dramatic conformational change, with the phenyl ring moving by as much as 10 Å; and GlyB143 moves about 3 Å towards the active site,



**Figure 3.** Superposition of the S1 pockets of GPLGS-WT and WT SARS-CoV M<sup>PRO</sup> (in stereo). (a) Superposition of the S1 pockets in protomer A of GPLGS-WT and that of protomer A\* of WT SARS-CoV M<sup>PRO</sup>. Protomer A\* of WT is in blue; protomer A of GPLGS-WT is in yellow; protomer B\* of WT is in magenta; protomer B of GPLGS-WT is in red. In the WT structure, the amino group (NH<sub>2</sub>) of Ser1 in protomer B\* donates a 3.0 Å hydrogen bond to the carboxylate group of Glu166 and a 2.7 Å hydrogen bond to the main-chain carbonyl group of Phe140 in protomer A\*, stabilizing the S1 pocket. The NH of Gly143 moves 0.8 Å toward the activity site; the main chain of residues 142-143 moves toward the S1 subsite; the side-chain of Asn-A\*142 flips over with a 6 Å shift compared with protomer A of GPLGS-WT. (b) Superposition of the S1 pockets in protomer B of GPLGS-WT and that of Protomer A\* of WT SARS-CoV M<sup>PRO</sup>. Protomer A\* of WT is in blue; protomer B of GPLGS-WT is in yellow; protomer B\* of WT is in magenta; protomer A of GPLGS-WT is in red. The S1 pocket of protomer B collapses partly with reorientation of Glu166 and residues 140-143. No electron density was visible for residues A1 and A2.

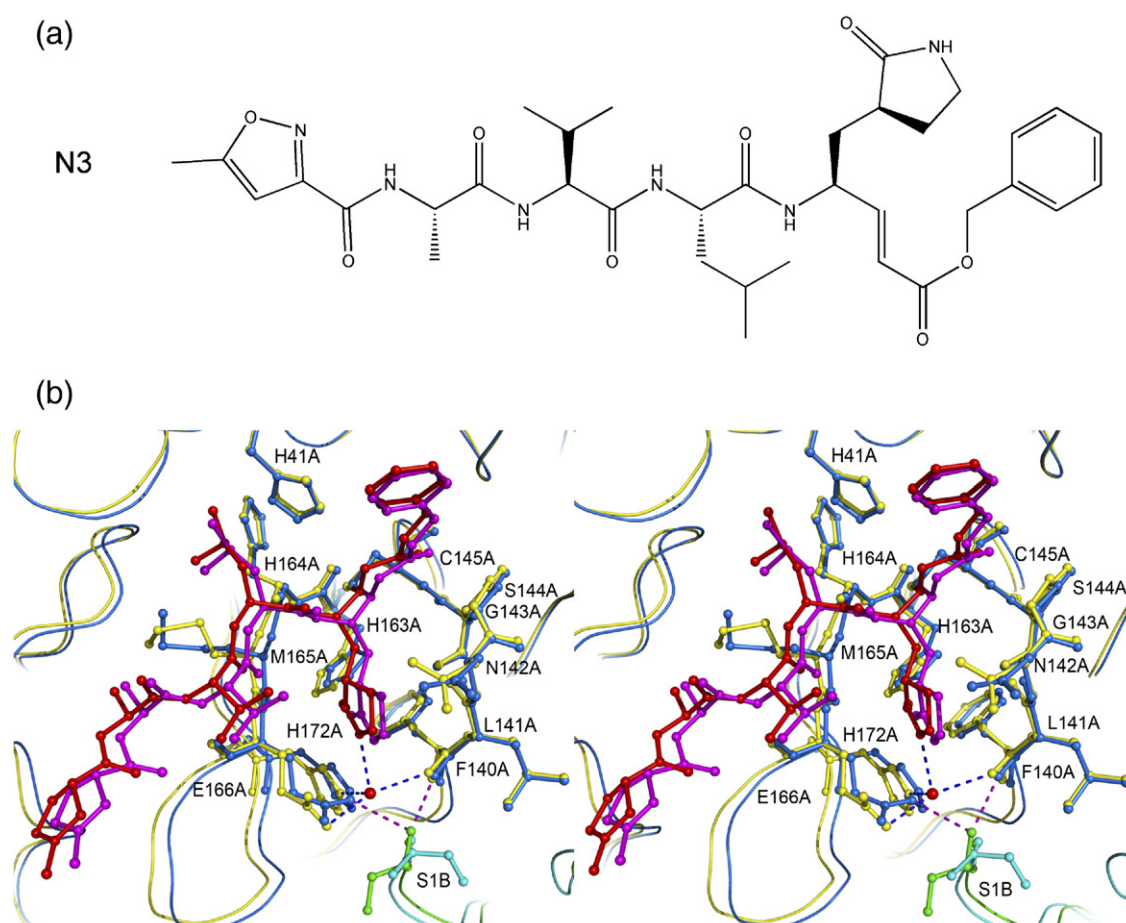
leaving no space to accommodate a tetrahedral reaction intermediate (Figure 3(b)). These structural variations account for the higher activity of the WT protease compared with the GPLGS-WT and GS-WT proteases. We observed also that the GPLGS-WT protease has lower activity than the GS-WT protease. This might result from the additional flexible residues at the N terminus, which are located close to the active site and would hinder substrate binding.

In contrast to the N terminus, the C terminus of the WT protomer (residues 301–306) is located far from the substrate-binding pocket of its partner protomer ( $\sim 10$  Å). Therefore, the additional residues (GPH<sub>6</sub>) at the C terminus are expected to have less effect on enzyme activity (see Table 3).

### Inhibition assay of SARS-CoV M<sup>Pro</sup>

In our previous study, we designed an irreversible anti-coronavirus inhibitor (designated as N3,

Figure 4(a)) consisting of an  $\alpha,\beta$ -unsaturated ester (one type of Michael acceptor) incorporated with a peptidyl portion.<sup>12</sup> The evaluation of this series of time-dependent inhibitors requires a pseudo second-order rate constant ( $k_3/K_i$ ).  $K_i$  and  $k_3$  represent the equilibrium binding constant and inactivation rate constant for covalent bond formation, respectively.<sup>12</sup> We assayed the inhibition of N3 against different types of M<sup>Pro</sup>s to determine whether or not the additional residues would affect inhibitor binding. In our preliminary inhibition assays, we observed that N3 could completely inactivate WT and WT-GPH<sub>6</sub> proteases after its preincubation with the proteases (fivefold molar excess of the enzyme) for 5 min, but not GPLGS-WT and GS-WT proteases (see Supplementary Data Figure S2). The strict kinetic parameters listed in Table 4 show that the second-order rate constant of N3 against WT protease ( $k_3/K_i=18,800$  M<sup>-1</sup>s<sup>-1</sup>) is approximately equal to that against WT-GPH<sub>6</sub> ( $k_3/K_i=15,400$  M<sup>-1</sup>s<sup>-1</sup>). However, the second-order rate constant



**Figure 4.** Differences between the complex structures of WT and GPLGS-WT. (a) Inhibitor N3. (b) Superposition of the substrate-binding pockets in protomer A of GPLGS-WT and that in protomer A\* of WT. In the WT-N3 complex structure, the NH<sub>2</sub> group of Ser1 in protomer B\* was still hydrogen-bonded to the carboxylate group of Glu166 and the carbonyl group of Phe140 in protomer A\*, stabilizing the S1 pocket. In the GPLGS-WT-N3 complex structure, however, the two hydrogen bonds described above were not found. Instead, an ordered water molecule was observed in the S1 pocket. Protomer A\* of WT is in blue; protomer A of GPLGS-WT is in yellow; inhibitor N3 (complexed with WT) is in magenta; inhibitor N3 (complexed with GPLGS-WT) is in red; protomer B\* of WT is in green; protomer B of GPLGS-WT is in cyan.

**Table 4.** Enzyme inhibition data of inhibitor N3 against four types of SARS-CoV M<sup>Pro</sup>

Protease	$K_i$ ( $\mu\text{M}$ )	$k_3$ ( $\text{s}^{-1}$ )	$k_3/K_i$ ( $\text{M}^{-1} \text{s}^{-1}$ )
GPLGS-WT	N/A	N/A	N/A
GS-WT	9.0±0.8	0.0031±0.0005	340±27
WT-GPH <sub>6</sub>	2.3±0.1	0.034±0.001	15,400±1,200
WT	1.9±0.1	0.035±0.002	18,800±1,800

of N3 ( $k_3/K_i=340 \text{ M}^{-1} \text{ s}^{-1}$ ) against GS-WT<sup>12</sup> is decreased by greater than 50-fold. This difference implies that the first residue of SARS-CoV M<sup>Pro</sup> also plays an important role in inhibitor binding.

### The crystal structure of WT SARS-CoV M<sup>Pro</sup> in complex with a Michael acceptor

We determined the crystal structure of WT SARS-CoV M<sup>Pro</sup> in complex with inhibitor N3 to 1.9 Å resolution. The substantial difference between the structures of the WT complex and the GPLGS-WT complex reported previously<sup>12</sup> still lies in the S1 subsite. In the GPLGS-WT complex structure, N3 binds to protomers A and B of SARS-CoV M<sup>Pro</sup> in an identical and normal manner,<sup>12</sup> thus we discuss only protomer A and A\*. In protomer A of GPLGS-WT M<sup>Pro</sup> complexed with N3, the NH group of SerB1 was 3.4 Å from the carboxylate group of GluA166 and 4.7 Å from the main-chain carbonyl group of PheA140, both of which are beyond the distance for formation of a hydrogen bond. However, an ordered water molecule is situated at the bottom of the S1 subsite, connecting N3 and the protease. This water donates two hydrogen bonds, to the carboxylate group of GluA166 and the main-chain carbonyl group of PheA140, and accepts two hydrogen bonds from the lactam and the side-chain of HisA172. Although the water molecule helps to stabilize the inhibitor binding in the S1 pocket, it occupies part of the space of the S1 subsite. Due to steric hindrance, the lactam was not able to insert further into the S1 subsite. In protomer A\* of the WT M<sup>Pro</sup> complexed with N3, no water molecule was found at the bottom of the S1 pocket. As a consequence, the carboxylate group of GluA\*166 moves 1.7 Å upwards to form a 2.8 Å hydrogen bond with the NH of the lactam. The competent conformation of the S1 subsite is still maintained *via* the interaction of NH<sub>2</sub> SerB\*1 with the carboxylate group of GluA\*166 (Figure 4(b)) and the main-chain carbonyl group of PheA\*140. These structural data demonstrate that the first Ser1 residue is important also for inhibitor binding, and accounts for the more

potent inhibition of N3 against the WT than the GPLGS-WT and GS-WT proteases.

## Discussion

Although SARS-CoV is notorious for causing a lethal disease in humans, some positive elements may result from this life-threatening virus. In this study, SARS-CoV M<sup>Pro</sup> was engineered to serve as a novel endopeptidase to remove fusion tags in recombinant protein overproduction. Table 5 shows the advantages of SARS-CoV M<sup>Pro</sup> compared with other routinely used proteases in methods for production, substrate specificity and cleavage efficiency. Human thrombin and bovine factor Xa are both extracted from plasma, which requires more complicated procedures, although they possess a high level of cleavage efficiency. In addition, they are not as selective for substrate as viral proteases. WT SARS-CoV M<sup>Pro</sup>, which can be suitably overexpressed in *E. coli*, is highly specific for substrate. Furthermore, it has superior activity ( $k_{\text{cat}}/K_m=26500 \text{ M}^{-1} \text{ s}^{-1}$ ) to rhinovirus and TEV proteases. Therefore, SARS-CoV M<sup>Pro</sup> is a suitable candidate for site-specific cleavage in fusion protein expression systems. The cleavage efficiency of WT-GPH<sub>6</sub> is still high, despite additional residues at the C terminus. Highly purified M<sup>Pro</sup> in this form could be obtained by a simple one-step purification, as shown by the overloaded SDS-PAGE gel (Figure 1(d)) For the sake of an additional six histidine residues at the C terminus, it can be readily separated from the protein of interest by affinity chromatography after removal of the fusion tags. These advantages suggest this form of SARS-CoV M<sup>Pro</sup> could have important industrial applications.

One disadvantage for recombinant protein expression is to commonly produce additional amino acid residues at the termini of the wild type protein. In SARS-CoV M<sup>Pro</sup> studies, several research groups have reported inconsistent results for the kinetic parameters of this protease, with  $k_{\text{cat}}/K_m$  ranging from 20–29,000  $\text{M}^{-1} \text{ s}^{-1}$ .<sup>19,20,25–27</sup> The slight differences in methods (HPLC and FRET-based methods) and substrates used would not account entirely for this phenomenon. The first published crystal structure of SARS-CoV M<sup>Pro</sup> (with additional residues at the N terminus) by our group provided some clues that the first residue might play an important role in substrate binding. In order to clarify this point, we designed a new strategy to produce the WT SARS-CoV M<sup>Pro</sup>. We created a GST fusion product with a tag that can be removed *via*

**Table 5.** Comparison of the different proteases

Proteases	Methods for production	Substrate specificity	Cleavage efficiency $k_{\text{cat}}/K_m$ ( $\text{M}^{-1} \text{ s}^{-1}$ )	References
Human thrombin	Extraction from plasma	-XLVPR↓GSX-	94,600	39–41
Bovine Factor Xa	Extraction from plasma	-XIEGR↓X-	39,000	42
Human rhinovirus 3C protease	Recombinant expression	-XLEVLFQ↓GPX-	920	3,4
TEV protease	Recombinant expression	-XENLYFQ↓G(S)X-	2620	5,6
SARS-CoV M <sup>Pro</sup>	Recombinant expression	-XTSAVLQ↓SGFRKX-	26,500	



the autocleavage mechanism of this enzyme for three reasons: (1) autoprocessing will not produce additional residues at the N terminus; (2) SARS-CoV M<sup>pro</sup> has been reported to have highly efficient expression in GST fusion systems; and (3) this method could be used to characterize the autocleavage efficiency of SARS-CoV *in vitro*. As for point (3), the affinity GST tag was a mimic of the transmembrane domain upstream of SARS-CoV M<sup>pro</sup> in polyprotein 1a and 1ab, representing an autocleavage model of the M<sup>pro</sup> *in vitro*. SDS-PAGE analysis shows that the GST tag was removed entirely from the N terminus through autoprocessing during expression of SARS-CoV M<sup>pro</sup> in *E. coli*, exhibiting its high efficiency for autocleavage *in vitro*. In previous studies, it was hypothesized that the autocleavage of M<sup>pro</sup> may occur in proximity to the membrane.<sup>13</sup> Our data suggest that completion of autocleavage might be achieved in the cytoplasm, although the precise whereabouts of the assembly of replicase complex components remains to be identified. According to separate reports by Hsu and Lin,<sup>35,36</sup> using autocleavage to remove tags such as thioredoxin can further support our hypothesis. Rhinoviruses and coronaviruses belong to the Picornaviridae and the Coronaviridae, respectively. No previous report has shown that any CoV M<sup>pro</sup> could efficiently process the substrate of picornavirus 3C proteases, or *vice versa*. In our study, it is interesting to observe that only substitution at the P2' site resulted in the conversion of the SARS-CoV main protease substrate into that of rhinovirus 3C protease.

In our study, we demonstrated that the first Ser1 residue at the N terminus of SARS-CoV M<sup>pro</sup> is important for its activity and inhibitor binding. The critical interactions involved in stabilizing the substrate-binding pocket are two hydrogen bonds formed by the free amino group of the first residue of one protomer with two residues, Phe140 and Glu166, comprising the S1 subsite of its partner protomer within the dimer structure. The stabilizing effect of the free amino group is obvious. In our published GPLGS-WT structure, the S1 subsite of protomer B is partially collapsed without stabilization by the first residue of protomer A, suggesting that the S1 pocket would be less stable due to micro-environmental changes. During the crystallization step of the structure determination procedure, a large quantity of WT SARS-CoV M<sup>pro</sup> crystals could be produced over one night, which could easily diffract to very high resolution under the conditions described previously. In contrast, it usually took five to ten days to produce a very limited amount of GPLGS-WT SARS-CoV M<sup>pro</sup> or GS-WT SARS-CoV M<sup>pro</sup> crystals suitable for diffraction, albeit with comparatively lower resolution. As a consequence, the WT SARS-CoV M<sup>pro</sup> reported here will be more suitable for activity assays, inhibitor screening and crystallization. It should accelerate development of anti-coronavirus inhibitors through a structure-assisted approach to drug design.

## Materials and Methods

### Construction of the pGSTM expression vector

The 654–965 region of the pGEX-6p-1 vector was cloned into the pMD18-T vector with two primers:

forward, 5'-**TTCGAA**GATCGTTTATGTCATAAA-3'  
reverse, 5'-**GGATCC**TTTCTAAACCCTCTGCAGA-  
ACTGCACTAGTATCCGATTTGGAGGATG-3'

The 33 nucleotides encoding the specific 11 amino acids TSAVLQSGFRK recognized by SARS-CoV M<sup>pro</sup> were introduced by the reverse primer. The recombinant pMD18-T plasmid was double-digested with BstBI and BamHI. The segment of interest was recombined into the pGEX-6p-1 vector, which had also been double-digested by BstBI and BamHI, to construct the new protein expression vector pGSTM.

### Cloning, expression and purification of calbindin D28k by the pGSTM protein expression system

#### Construction of recombinant plasmid

The gene encoding calbindin D28k was amplified by the polymerase chain reaction (PCR) by primers:

forward, 5'-**GGATCC**ATGGCAGAATCCACCTG-3'  
reverse, 5'-**CCGCTCGAG**CTAGTTATCCCCAGCACA-3'

The PCR products were ligated to pMD18-T vectors with bacteriophage T4 DNA ligase. After digestion by BamHI and XhoI, the gene of interest was inserted between the BamHI and XhoI sites of the pGSTM vector.

#### Expression and purification

The resulting recombinant plasmid was transformed into the *E. coli* strain BL21 (DE3). The cells were cultured in LB medium containing 0.1 mg ml<sup>-1</sup> ampicillin. When the absorbance at 600 nm (*A*<sub>600</sub>) reached 0.6, IPTG was added to 0.5 mM and the cell culture was incubated at 16 °C for 10 h. After harvesting by centrifugation at 4600g (Beckman JLA-10-5), the pellet was resuspended in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.3) and sonicated on ice. The lysate was centrifuged at 27,000g (Beckman JA-25-50) for 30 minutes and the precipitate was discarded. The supernatant was loaded onto 2 ml GST-glutathione affinity columns (Pharmacia) equilibrated with PBS and washed with 30 column volumes of PBS. After that, 0.1 mg of SARS-CoV M<sup>pro</sup> (GS-WT) was added to the column at 4 °C for 12 h to remove the GST tag. The protein of interest was collected and analyzed by SDS-PAGE.

### Expression and purification of SARS-CoV M<sup>pro</sup>

Expression and purification of SARS-CoV M<sup>pro</sup> with different numbers of additional residues (5, 2, 0) at the N terminus is described below.

#### SARS-CoV M<sup>pro</sup> (with five additional residues, GPLGS, at the N terminus)

Expression and purification of GPLGS-WT SARS-CoV M<sup>pro</sup> has been reported.<sup>14</sup> Briefly, the coding sequence of

the SARS-CoV M<sup>Pro</sup> was inserted into the BamHI and XhoI sites of pGEX-6p-1 plasmid DNA (GE Healthcare). The resulting plasmid was used to transform *E. coli* BL21 (DE3) cells. The GST fusion protein, GST-SARS-CoV M<sup>Pro</sup>, was purified by GST-glutathione affinity chromatography, cleaved with GST rhinovirus 3C protease, and the recombinant SARS-CoV M<sup>Pro</sup> was further purified by anion-exchange chromatography.

#### SARS-CoV M<sup>Pro</sup> (with two additional residues, GS, at the N terminus)

Expression and purification of GS-WT SARS-CoV M<sup>Pro</sup> has been reported.<sup>12</sup> Briefly, the coding sequence was inserted into the BamHI and XhoI sites of the pGEX-4T-1 vector (GE Healthcare). The following procedure is similar to that used for the expression and purification of SARS-CoV M<sup>Pro</sup> in pGEX-6p-1 plasmid, except that the GST fusion protein was cleaved by thrombin.

#### WT SARS-CoV M<sup>Pro</sup> (without additional residues at the termini)

The coding sequence for SARS-CoV M<sup>Pro</sup> was amplified by polymerase chain reaction (PCR) using the PCR primers:

forward, 5'-CGGGATCCGCGGTACTGCAGAGTGGTTT-CAGGAAAATGGCA-3'  
reverse 5'-CCGCTCGAGTTAGTGGTGGTGGTGGTGGTGGTGGTCTTGGGAAGGTAATCC-3'

The 12 nucleotides coding for the four amino acids AVLQ (corresponding to the P1–P4 autocleavage sites at the N terminus of SARS-CoV M<sup>Pro</sup>; nomenclature for the substrate amino acid residues is P<sub>n</sub>,..., P<sub>2</sub>, P<sub>1</sub>, P<sub>1</sub>', P<sub>2</sub>',..., P<sub>n</sub>', where P<sub>1</sub>–P<sub>1</sub>' denotes the hydrolyzed bond while S<sub>n</sub>,..., S<sub>2</sub>, S<sub>1</sub>, S<sub>1</sub>', S<sub>2</sub>',..., S<sub>n</sub>' denote the corresponding enzyme binding sites<sup>37</sup>) were added before the first Ser1 residue. The 24 nucleotides coding for the eight amino acids GPH<sub>6</sub> were added at the C terminus by the reverse primer. The PCR products were inserted into the BamHI and XhoI sites of the pGEX-6p-1 plasmid (GE Healthcare). The resulting plasmid was then used to transform *E. coli* BL21 (DE3) cells. The sequence of the insert was verified by dideoxynucleotide sequencing. Positive clones harboring the recombinant plasmid were grown to an A<sub>600</sub> of 0.6 at 37 °C by shaking in LB medium containing 0.1 mg ml<sup>-1</sup> ampicillin. The GST fusion protein was expressed by introducing IPTG to 0.5 mM with incubation continued at 16 °C for 10 h. Cells were then harvested by centrifugation at 4600g (Beckman JLA-10-5), resuspended in lysis buffer (20 mM Tris–HCl (pH 8.0), 300 mM NaCl) and sonicated on ice. The lysate was centrifuged at 27,000g (Beckman JA-25-50) for 30 min and the supernatant was collected. The His tag fused protein was purified by Ni-NTA affinity chromatography and concentrated in PreScission Cleavage Buffer (50 mM Tris–HCl (pH 7.0), 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol). 50 µl of 3 mg ml<sup>-1</sup> human rhinovirus 3C protease was added to 1 ml of the above 10 mg ml<sup>-1</sup> WT-GPH<sub>6</sub> solution to cleave the C-terminal His tag, producing an SARS-CoV M<sup>Pro</sup> with an authentic C terminus. The WT SARS-CoV M<sup>Pro</sup> was further purified using anion-exchange chromatography. The protein samples in each step were prepared for SDS-PAGE analysis. WT-GPH<sub>6</sub> and WT were analyzed by MALDI-TOF MS. The purified and concentrated WT SARS-CoV M<sup>Pro</sup> (10 mg ml<sup>-1</sup>) was

stored in 50 mM Tris–HCl (pH 7.3), 1 mM EDTA at –80 °C for enzyme activity assays and crystallization.

#### Enzymatic activity and inhibition assays

Enzyme activity assays of GS-WT SARS-CoV M<sup>Pro</sup> have been described.<sup>12</sup> Activity assays of GPLGS-WT, WT-GPH<sub>6</sub> and WT SARS-CoV M<sup>Pro</sup> followed a similar protocol. Briefly, the substrate of the N and C-terminal authentic SARS-CoV M<sup>Pro</sup> was the fluorogenic compound MCA-AVLQSGFR-Lys(Dnp)-Lys-NH<sub>2</sub> (greater than 95% purity, GL Biochem Shanghai Ltd, Shanghai, China). The excitation and emission wavelengths of the fluorogenic substrate were 320 nm and 405 nm, respectively. A buffer consisting of 50 mM Tris–HCl (pH 7.3), 1 mM EDTA was used for enzyme activity assays at a temperature of 30 °C. The reaction was initiated by adding protease (final concentration of 0.2 µM for WT and WT-GPH<sub>6</sub>, 2 µM for GPLGS-WT) to a solution containing different final concentrations of the substrate (3.2–40 µM for WT and WT-GPH<sub>6</sub>, and 6.4–80 µM for GPLGS-WT). The kinetic constants K<sub>m</sub> and k<sub>cat</sub> were obtained from a double-reciprocal plot (Supplementary Data Figure S1). Strict

**Table 6.** Data collection and refinement statistics

	WT SARS-CoV M <sup>Pro</sup>	WT SARS-CoV M <sup>Pro</sup> : N3
<b>A. Data collection statistics</b>		
Wavelength (Å)	1.0000	1.5418
Resolution limit (Å)	50.0–1.50 (1.55–1.50)	50.0–1.95(2.02–1.95)
Space group	C2	C2
Cell parameters		
<i>a</i> (Å)	108.4	108.6
<i>b</i> (Å)	81.8	81.2
<i>c</i> (Å)	53.6	53.3
β (deg.)	104.7	104.5
Total reflections	219,294	148,615
Unique reflections	63,241	32,068
Completeness, %	87.7 (40.6)	97.9 (94.6)
Redundancy	3.5 (2.5)	4.7 (3.7)
R <sub>merge</sub> <sup>a</sup>	0.038 (0.286)	0.042 (0.277)
σ cutoff	0	0
I/σ (I)	30.4 (2.0)	32.4 (4.2)
<b>B. Refinement statistics</b>		
Resolution range (Å)	50.0–1.6	50.0–1.95
R <sub>work</sub> <sup>b</sup> (%)	20.1	20.2
R <sub>free</sub> (%)	21.4	22.1
rmsd from ideal		
geometry		
Bond lengths (Å)	0.009	0.015
Bond angles (deg.)	1.59	1.83
Average B factor, Å <sup>2</sup>		
Chain A	26.1	37.2
Solvent	41.5	49.5
Ramachandran plot <sup>c</sup>		
Favored (%)	90.9	91.7
Allowed (%)	7.9	7.2
Generously allowed (%)	0.8	0.8
Disallowed (%)	0.4	0.4

<sup>a</sup>  $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i$ , where  $I_i$  is the intensity of an individual reflection  $i$ , and  $\langle I \rangle$  is the average intensity of that reflection.

<sup>b</sup>  $R_{\text{work}} = \sum |F_p| - |F_c| / \sum |F_p|$ , where  $F_c$  is the calculated and  $F_p$  is the observed structure factor amplitude.

<sup>c</sup> Ramachandran plots were generated with the program PROCHECK.

kinetic parameters were determined for the inhibition assay.<sup>12</sup>

### Crystallization, data collection and structure determination

Crystallization of WT SARS-CoV M<sup>pro</sup> was carried out as described.<sup>14</sup> The preparation of the co-crystals of SARS-CoV M<sup>pro</sup> in complex with the inhibitor N3 has been reported.<sup>12</sup> A set of WT SARS-CoV M<sup>pro</sup> data was collected from a single crystal on beamline BL19-ID of the Advanced Photon Source (APS), Argonne National Lab at a wavelength of 1.00 Å. Data for the SARS-CoV M<sup>pro</sup> complex were collected at 100 K in-house on a Rigaku CuK $\alpha$  rotating-anode X-ray generator (MM007) at 40 kV and 20 mA (1.5418 Å) with a Rigaku R-Axis IV++ image-plate detector. Data were processed, integrated, scaled and merged using HKL2000.<sup>38</sup> The methods used for structure determination were as described.<sup>12</sup> Briefly, the structures were determined by molecular replacement from our native structure of SARS-CoV M<sup>pro</sup> (pH 7.6) (PDB ID 1UK3). Data collection and structure refinement statistics are summarized in Table 6.

### Data Bank with accession numbers

Coordinates and structure factors for WT SARS-CoV M<sup>pro</sup> have been deposited in the Protein Data Bank with accession number 2H2Z. Coordinates and structure factors for WT SARS-CoV M<sup>pro</sup> in complex with the inhibitor N3 have been deposited in the Protein Data Bank with accession number 2HOB.

### Acknowledgements

We thank Xuemei Li and Sheng Ye for technical assistance; Huanming Yang, Jian Wang, and Jun Yu for providing cDNA of SARS-CoV M<sup>pro</sup>; Rongguang Zhang and Andrzej Joachimiak for data collection of WT SARS-CoV M<sup>pro</sup>. This work was supported by Project 973 of the Ministry of Science and Technology of China (grant number 2004BA519A30), the NSFC (grant number 30221003), 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 PhD student innovation fund.

### Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2006.11.073](https://doi.org/10.1016/j.jmb.2006.11.073)

### References

1. Edwards, A. M., Arrowsmith, C. H., Christendat, D., Dharamsi, A., Friesen, J. D., Greenblatt, J. F. & Vedadi, M. (2000). Protein production: feeding the crystallographers and NMR spectroscopists. *Nature Struct. Biol.* **7**, (Suppl.), 970–972.
2. Bucher, M. H., Evdokimov, A. G. & Waugh, D. S. (2002). Differential effects of short affinity tags on the crystallization of *Pyrococcus furiosus* maltodextrin-binding protein. *Acta Crystallog. sect. D*, **58**, 392–397.
3. Walker, P. A., Leong, L. E., Ng, P. W., Tan, S. H., Waller, S., Murphy, D. & Porter, A. G. (1994). Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Biotechnology (NY)*, **12**, 601–605.
4. Wang, Q. M., Johnson, R. B., Cox, G. A., Villarreal, E. C. & Loncharich, R. J. (1997). A continuous colorimetric assay for rhinovirus-14 3C protease using peptide p-nitroanilides as substrates. *Anal. Biochem.* **252**, 238–245.
5. Stols, L., Gu, M., Dieckman, L., Raffin, R., Collart, F. R. & Donnelly, M. I. (2002). A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr. Purif.* **25**, 8–15.
6. Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D. & Waugh, D. S. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14**, 993–1000.
7. Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S. *et al.* (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976.
8. Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S. *et al.* (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1953–1966.
9. Kuiken, T., Fouchier, R. A., Schutten, M., Rimmelzwaan, G. F., van Amerongen, G., van Riel, D. *et al.* (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet*, **362**, 263–270.
10. Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W. *et al.* (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet*, **361**, 1319–1325.
11. Fouchier, R. A., Kuiken, T., Schutten, M., van Amerongen, G., van Doornum, G. J., van den Hoogen, B. G. *et al.* (2003). Aetiology: Koch's postulates fulfilled for SARS virus. *Nature*, **423**, 240.
12. Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W. *et al.* (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **3**, e324.
13. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*, **300**, 1763–1767.
14. Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z. *et al.* (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl Acad. Sci USA*, **100**, 13190–13195.
15. Xu, T., Ooi, A., Lee, H., Wilmouth, R., Liu, D. & Lescar, J. (2005). Structure of the SARS coronavirus main proteinase as an active C2 crystallographic dimer. *Acta Crystallog. sect. F*, **61**, 964–966.
16. Tan, J., Verschuere, K. H., Anand, K., Shen, J., Yang, M., Xu, Y. *et al.* (2005). pH-dependent conformational flexibility of the SARS-CoV main proteinase (M<sup>pro</sup>) dimer: molecular dynamics simulations and multiple X-ray structure analyses. *J. Mol. Biol.* **354**, 25–40.

17. Hsu, M.-F., Kuo, C.-J., Fang, J.-M., Shie, J.-J., Chang, K.-T., Chang, H.-C. *et al.* (2005). Understanding the maturation process and inhibitor design of SARS-CoV 3CL<sup>pro</sup> from the crystal structure of C145A in a product-bound form. *J. Biol. Chem.*, M502577200.
18. Lee, T. W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C. *et al.* (2005). Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. *J. Mol. Biol.* **353**, 1137–1151.
19. Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A. & Freire, E. (2004). Identification of novel inhibitors of the SARS coronavirus main protease 3CL<sup>pro</sup>. *Biochemistry*, **43**, 4906–4912.
20. Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L. *et al.* (2004). Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* **279**, 1637–1642.
21. Chou, C. Y., Chang, H. C., Hsu, W. C., Lin, T. Z., Lin, C. H. & Chang, G. G. (2004). Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry*, **43**, 14958–14970.
22. Hsu, W. C., Chang, H. C., Chou, C. Y., Tsai, P. J., Lin, P. I. & Chang, G. G. (2005). Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease. *J. Biol. Chem.* **280**, 22741–22748.
23. Chen, S., Chen, L., Tan, J., Chen, J., Du, L., Sun, T. *et al.* (2005). Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization. Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. *J. Biol. Chem.* **280**, 164–173.
24. Ziebuhr, J. (2005). The coronavirus replicase. *Curr. Top. Microbiol. Immunol.* **287**, 57–94.
25. Huang, C., Wei, P., Fan, K., Liu, Y. & Lai, L. (2004). 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry*, **43**, 4568–4574.
26. Kao, R. Y., To, A. P., Ng, L. W., Tsui, W. H., Lee, T. S., Tsoi, H. W. & Yuen, K. Y. (2004). Characterization of SARS-CoV main protease and identification of biologically active small molecule inhibitors using a continuous fluorescence-based assay. *FEBS Letters*, **576**, 325–330.
27. Blanchard, J. E., Elowe, N. H., Huitema, C., Fortin, P. D., Cechetto, J. D., Eltis, L. D. & Brown, E. D. (2004). High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase. *Chem. Biol.* **11**, 1445–1453.
28. Mikkonen, M., Soinenen, H. & Pitkanen, A. (1997). Distribution of parvalbumin-, calretinin-, and calbindin-D28k-immunoreactive neurons and fibers in the human entorhinal cortex. *J. Comp. Neurol.* **388**, 64–88.
29. Fortin, M. & Parent, A. (1997). Distribution of calretinin, calbindin-D28k and parvalbumin in the hypothalamus of the squirrel monkey. *J. Chem. Neuroanat.* **14**, 51–61.
30. Winsky, L. & Kuznicki, J. (1995). Distribution of calretinin, calbindin D28k, and parvalbumin in subcellular fractions of rat cerebellum: effects of calcium. *J. Neurochem.* **65**, 381–388.
31. Berggard, T., Szczepankiewicz, O., Thulin, E. & Linse, S. (2002). Myo-inositol monophosphatase is an activated target of calbindin D28k. *J. Biol. Chem.* **277**, 41954–41959.
32. Matthews, D. A., Dragovich, P. S., Webber, S. E., Fuhrman, S. A., Patick, A. K., Zalman, L. S. *et al.* (1999). Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. *Proc. Natl Acad. Sci. USA*, **96**, 11000–11007.
33. Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W. *et al.* (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell*, **77**, 761–771.
34. Copeland, R. A. (2000). Enzymes. In *A Practical Introduction to Structure, Mechanism, and Data Analysis* (Copeland, R. A., eds), 2nd edit., pp. 120–122, Wiley VCH, New York.
35. Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T. P. *et al.* (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* **280**, 31257–31266.
36. Lin, C. W., Tsai, C. H., Tsai, F. J., Chen, P. J., Lai, C. C., Wan, L. *et al.* (2004). Characterization of trans- and cis-cleavage activity of the SARS coronavirus 3CL<sup>pro</sup> protease: basis for the in vitro screening of anti-SARS drugs. *FEBS Letters*, **574**, 131–137.
37. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. *Papain. Biochem. Biophys. Res. Commun.* **27**, 157–162.
38. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collection in oscillation mode. *Methods Enzymol.* **276**, 307–326.
39. Pozsgay, M., Szabo, G., Bajusz, S., Simonsson, R., Gaspar, R. & Elodi, P. (1981). Study of the specificity of thrombin with tripeptidyl-p-nitroanilide substrates. *Eur. J. Biochem.* **115**, 491–495.
40. Smith, D. B. & Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*, **67**, 31–40.
41. Chang, J. Y. (1985). Thrombin specificity. Requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate. *Eur. J. Biochem.* **151**, 217–224.
42. Cho, K., Tanaka, T., Cook, R. R., Kisiel, W., Fujikawa, K., Kurachi, K. & Powers, J. C. (1984). Active-site mapping of bovine and human blood coagulation serine proteases using synthetic peptide 4-nitroanilide and thio ester substrates. *Biochemistry*, **23**, 644–650.

Edited by R. Huber

(Received 13 September 2006; received in revised form 24 November 2006; accepted 28 November 2006)  
Available online 1 December 2006