

RESEARCH ARTICLE

# Liberation of SARS-CoV main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode

Shuai Chen<sup>1</sup>, Felix Jonas<sup>1</sup>, Can Shen<sup>1</sup>, Rolf Hilgenfeld<sup>1,2</sup> (✉)

<sup>1</sup> Institute of Biochemistry, Center for Structural and Cell Biology in Medicine, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

<sup>2</sup> Laboratory for Structural Biology of Infection and Inflammation, c/o DESY, Building 22a, Notkestr. 85, 22603 Hamburg, Germany

✉ Correspondence: hilgenfeld@biochem.uni-luebeck.de

Received October 27, 2009; accepted November 17, 2009

## ABSTRACT

The main protease ( $M^{\text{pro}}$ ) plays a vital role in proteolytic processing of the polyproteins in the replicative cycle of SARS coronavirus (SARS-CoV). Dimerization of this enzyme has been shown to be indispensable for *trans*-cleavage activity. However, the auto-processing mechanism of  $M^{\text{pro}}$ , i.e. its own release from the polyproteins through autocleavage, remains unclear. This study elucidates the relationship between the N-terminal autocleavage activity and the dimerization of “immature”  $M^{\text{pro}}$ . Three residues (Arg4, Glu290, and Arg298), which contribute to the active dimer conformation of mature  $M^{\text{pro}}$ , are selected for mutational analyses. Surprisingly, all three mutants still perform N-terminal autocleavage, while the dimerization of mature protease and *trans*-cleavage activity following auto-processing are completely inhibited by the E290R and R298E mutations and partially so by the R4E mutation. Furthermore, the mature E290R mutant can resume N-terminal autocleavage activity when mixed with the “immature” C145A/E290R double mutant whereas its *trans*-cleavage activity remains absent. Therefore, the N-terminal auto-processing of  $M^{\text{pro}}$  appears to require only two “immature” monomers approaching one another to form an “intermediate” dimer structure and does not strictly depend on the active dimer conformation existing in mature protease. In conclusion, an auto-release model of  $M^{\text{pro}}$  from the polyproteins is proposed, which will help understand

the auto-processing mechanism and the difference between the autocleavage and *trans*-cleavage proteolytic activities of SARS-CoV  $M^{\text{pro}}$ .

**KEYWORDS** SARS-CoV  $M^{\text{pro}}$ , N-terminal autocleavage, autocleavage activity, *trans*-cleavage activity, dimerization

## INTRODUCTION

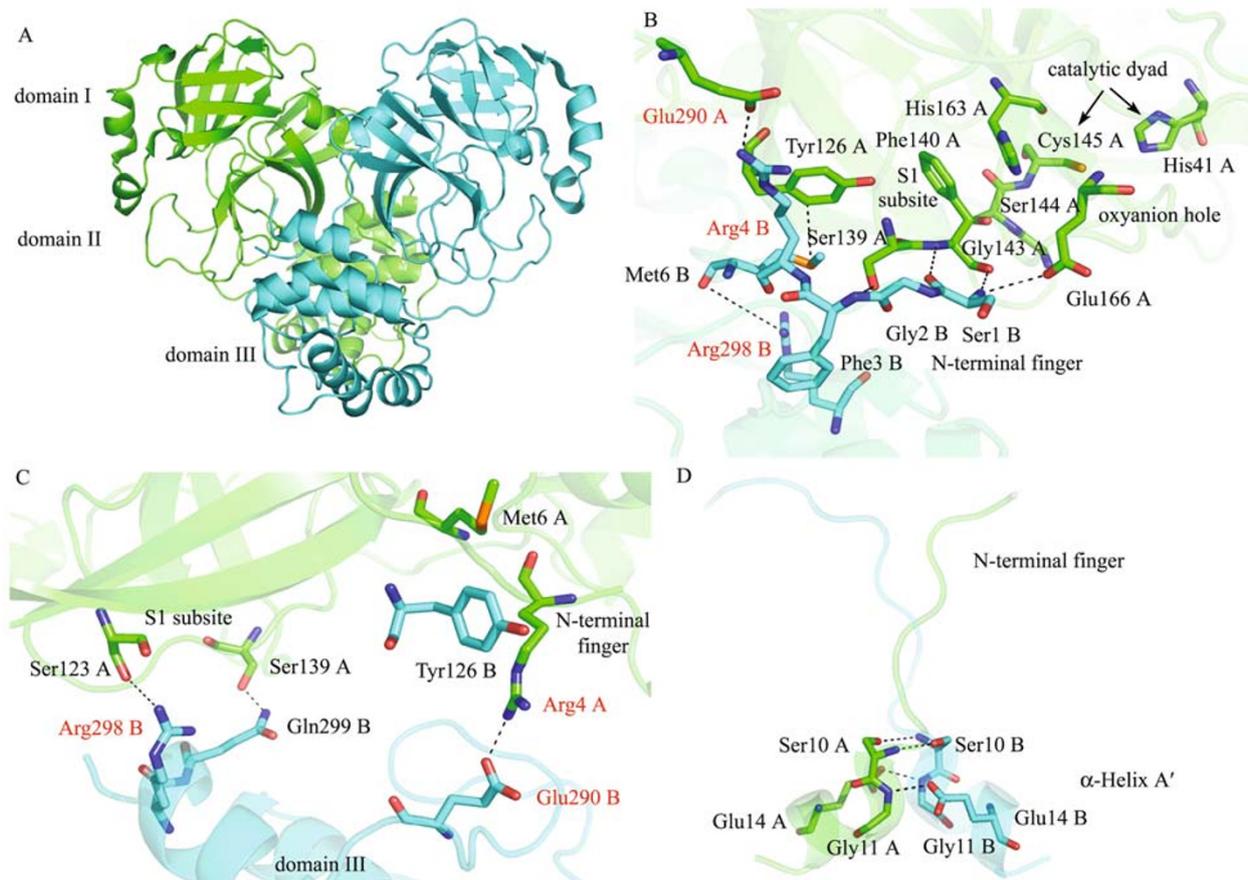
The global outbreak of the highly infectious Severe Acute Respiratory Syndrome (SARS) between November, 2002, and June, 2003, was caused by a new human coronavirus, SARS coronavirus (SARS-CoV) (Drosten et al., 2003; Peiris et al., 2003). SARS-CoV is an enveloped, positive-sense RNA virus and involves the largest viral RNA genome known to date, encoding several structural and auxiliary proteins as well as two large overlapping polyproteins, pp1a (replicase 1a, around 450 kDa) and pp1ab (replicase 1ab, around 750 kDa) necessary for viral RNA synthesis (Marra et al., 2003; Thiel et al., 2003). These two polyproteins are cleaved extensively by the main protease ( $M^{\text{pro}}$ , also often called 3C-like protease) and the papain-like cysteine protease ( $PL2^{\text{pro}}$ ), both of which are encoded by the viral genome, to yield a multi-subunit protein complex termed “viral replicase-transcriptase” (Thiel et al., 2003; Ziebuhr, 2004; Groneberg et al., 2005). Because of its functional indispensability in viral replication, the main protease ( $M^{\text{pro}}$ ) has become an attractive

target in developing inhibitors directed at SARS-CoV and other coronaviruses (Anand et al., 2005).

SARS-CoV M<sup>Pro</sup> exists as a dimer in all crystal structures of the wild-type enzyme determined so far (Yang et al., 2003; Hsu et al., 2005a; Lee et al., 2005; Tan et al., 2005) and dimerization is also observed in solution, in a concentration-dependent manner (Chou et al., 2004; Fan et al., 2004; Shi et al., 2004; Chen et al., 2005; Graziano et al., 2006b). Structurally, two monomers orient perpendicular to one another to form a dimer. Each monomer contains three domains (Fig. 1A) (Yang et al., 2003): domains I (residues 8–101) and II (residues 102–184) form a chymotrypsin fold and domain III (residues 201–306) is an antiparallel globular cluster of five  $\alpha$ -helices connected to domain II by a long loop (residues 185–200). The substrate-binding pocket is located in a cleft between domains I and II and the active site consists of a Cys145–His41 catalytic dyad (Huang et al., 2004). In the active conformation of the dimer, the pocket is accessible for interaction with the respective amino-acid residues of the

substrate and the oxyanion loop (residues 138–145) has the correct shape to donate two hydrogen bonds from main-chain amides to stabilize the tetrahedral transition-state of the proteolysis reaction. To date, numerous experimental results have indicated that only the dimer is the biological functional form of SARS-CoV M<sup>Pro</sup> and the monomer is considered enzymatically inactive (Chou et al., 2004; Hsu et al., 2005a, b; Barrila et al., 2006; Graziano et al., 2006a; Chen et al., 2008b). It has also been revealed that the activity loss of the dissociated monomer is mainly due to the collapse of the oxyanion hole and the S1 substrate-binding subsite (Tan et al., 2005).

Since the dissociated monomer of M<sup>Pro</sup> is inactive, the dimer interface has been suggested as another potential target for inhibitor design (Shi et al., 2004; Hsu et al., 2005a). As revealed by the crystal structure, the dimer interface of the protease mainly involves the following interactions (Fig. 1B–D): (i) between the N-terminal finger (residues 1–7) of one monomer and residues near the S1 subsite of the other



**Figure 1.** The overall dimeric structure and detailed dimer interface of mature SARS-CoV M<sup>Pro</sup>. (A) A ribbon diagram for the dimer structure of mature SARS-CoV M<sup>Pro</sup> (PDB code: 1UK4). Monomers A and B are represented in green and cyan, respectively, and the three domains are labeled. The dimer interface mainly involves the interactions: (B) between the S1 subsite of monomer A (green) and the N-terminal finger (cyan) of monomer B, (C) between the N-terminal finger and the S1 subsite of monomer A and the helical domain III of monomer B, (D) between the two A'  $\alpha$ -helices of each monomer. The residues selected for site-directed mutations in this study are marked in red.

monomer, in particular the oxyanion loop, (ii) between the N-terminal finger and the helical domain III from each monomer, (iii) between the two A'  $\alpha$ -helices (residues 10–15), one from each monomer, that immediately follow the N-terminal finger. The contributions of several individual residues on the dimer interface to dimerization and enzymatic activity of M<sup>Pro</sup> have been identified by several groups (Chou et al., 2004; Hsu et al., 2005b; Barrila et al., 2006; Shi and Song, 2006; Wei et al., 2006; Chen et al., 2008a, b; Lin et al., 2008). Firstly, residues on the N-terminal finger were considered to play an important role in both dimerization and activity (Chen et al., 2005; Hsu et al., 2005b; Chen et al., 2008b). In particular, Arg4 of the N-terminal finger can form a salt-bridge with residue Glu290 of the other monomer, which is vital for stabilizing the dimer structure to maintain the correct conformation of the active site (Chou et al., 2004). Furthermore, residues of domain III have also been found to extensively mediate dimerization and to be responsible for positioning the N-terminal finger to interact with the active site of the neighboring monomer (Shi and Song, 2006; Lin et al., 2008). Residue Arg298 in the C-terminal helix (residues 293–301) has been identified as a key component for maintaining the dimer conformation and its mutation was found to trigger a structural switch from dimer to monomer (Shi et al., 2008). In addition, the A'  $\alpha$ -helix (residues 10–15) is another critical part of the dimer interface since mutation of Gly11 can also result in complete dimer dissociation, as shown by X-ray crystallography (Chen et al., 2008a).

Before proteolytic processing of the viral polyproteins pp1a and pp1ab into a total of 15 or 16 non-structural proteins (Nsp) occurs, SARS-CoV M<sup>Pro</sup> itself is embedded in these polyproteins as the Nsp5 domain. On both sides, it is flanked by putative transmembrane (TM) domains (Nsp4 and Nsp6) that are anchored to the double-membrane vesicles where viral replication takes place in the infected host cell (Snijder et al., 2003, 2006; Knoop et al., 2008). Therefore, the M<sup>Pro</sup> has to first liberate itself from the polyproteins through autocleavage, and then the self-released mature M<sup>Pro</sup> would form a dimer and *trans*-cleave pp1a and pp1ab at other sites. Despite being the first and essential step for viral maturation, the autocleavage mechanism of M<sup>Pro</sup> has only been rarely addressed (Lin et al., 2004; Shan et al., 2004; Hsu et al., 2005a) and remains poorly characterized. On the other hand, the relationship between dimerization and enzymatic activity of the mature M<sup>Pro</sup> is well documented (Chou et al., 2004; Shi et al., 2004; Hsu et al., 2005b; Graziano et al., 2006a, b; Chen et al., 2008a; Lin et al., 2008). Since most of these studies use a protease preparation that resembles the sequence of mature M<sup>Pro</sup> after auto-processing and formation of the final dimer structure, all these results are only relevant for the *trans*-cleavage activity of the mature enzyme.

In this paper, we focus on the correlation between the N-terminal autocleavage activity and dimerization of “immature” SARS-CoV M<sup>Pro</sup>, which has not been reported yet. A total of

three residues involved in maintaining the dimer conformation of mature M<sup>Pro</sup> are selected for mutational analyses, i.e., Arg4, Glu290, and Arg298 (Fig. 1B and 1C). In the crystal structure of the mature dimer, the side-chain guanidyl of Arg4 forms a salt bridge with the side-chain of Glu290 of the neighboring monomer in the dimer. Mutation of the conserved Glu290 has been reported to cause dimer dissociation of M<sup>Pro</sup>, leading to an inactive monomer in solution, while mutation of the not absolutely conserved Arg4 shifts the dimer-monomer equilibrium and induces a significant decrease of *trans*-cleavage activity (Chou et al., 2004). Furthermore, mutation of Arg298 has also been shown to trigger dimer dissociation in both solution and crystal as well as a complete loss of activity (Shi et al., 2008). As the construction strategy of all the plasmids used in the study presented here, a SARS-CoV M<sup>Pro</sup> autocleavage site (AVLQ↓S) was introduced between the N-terminal tag and the N-terminal residue (Ser1) of the protease. Thus, the N-terminal autocleavage activity of wild type and mutated M<sup>Pro</sup>s can be evaluated by the extent to which the N-terminal tag is removed. Our results surprisingly reveal that all three mutants can still perform autocleavage during protein production and purification. In the following, we also investigate the effects of these mutations on dimerization and *trans*-cleavage activity of mature SARS-CoV M<sup>Pro</sup>. Consistent with the published results, dimer formation of mature protease in solution is completely abolished by the E290R and R298E mutations and partially so by the R4E mutation, resulting in an entire or dramatic loss of *trans*-cleavage activity. Furthermore, the reconstructed cleavage assay indicates that the mature E290R mutant can resume N-terminal autocleavage activity when mixed with the “immature” C145A/E290R double mutant whereas its *trans*-cleavage activity remains absent. These results indicate that N-terminal autocleavage of SARS-CoV M<sup>Pro</sup> from the polyproteins only requires two “immature” proteases approaching one another to form an “intermediate” dimer structure and does not depend on the active dimer conformation existing in the mature protease. The present study is expected to help us better understand the maturation mechanism and the difference between autocleavage and *trans*-cleavage proteolytic processing of SARS-CoV M<sup>Pro</sup>.

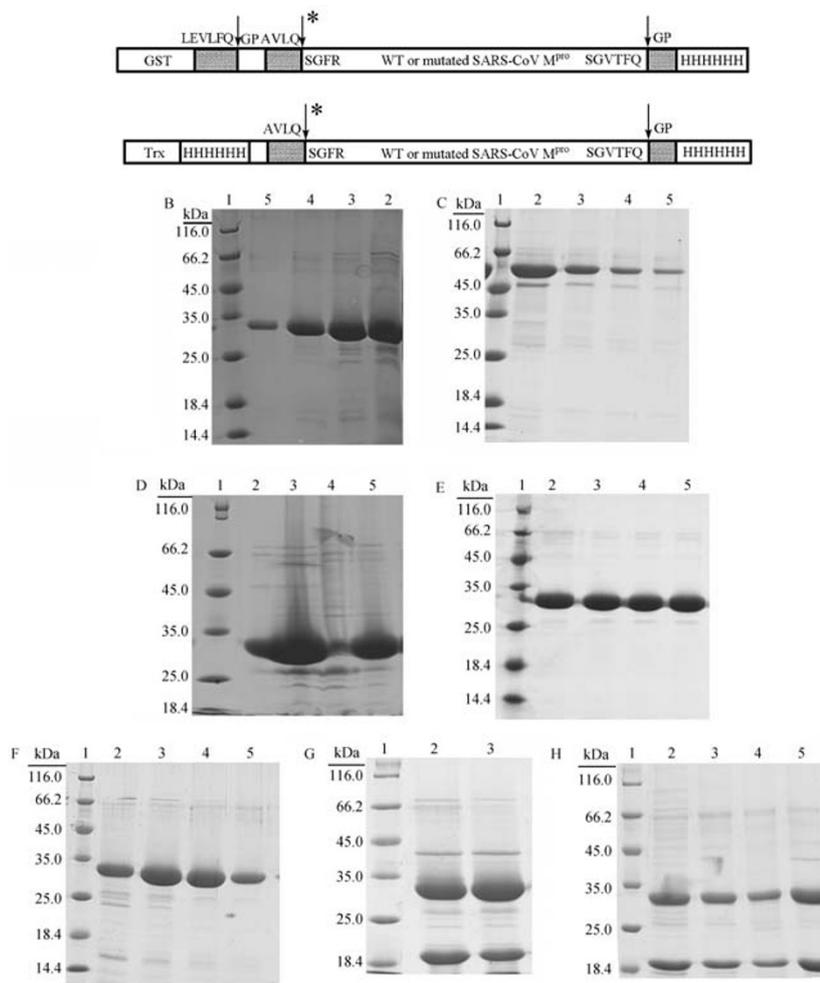
## RESULTS AND DISCUSSION

### Autocleavage of the N-terminally tagged wild-type and mutated SARS-CoV M<sup>Pro</sup> during lysate preparation

The goal of this study was to elucidate the relationship between the N-terminal autocleavage activity and dimerization of “immature” SARS-CoV M<sup>Pro</sup>. Three residues (Arg4, Glu290, and Arg298), which are involved in maintaining the integrity of the dimer interface of mature SARS-CoV M<sup>Pro</sup> (Fig. 1B and 1C), were selected for mutational analysis. The Arg4...Glu290 salt bridge has been demonstrated to be a key

element in dimerization of mature M<sup>Pro</sup> and residues on domain III (e.g., Arg298, Gln299 et al.) have also been shown to be vital regulators of dimerization and *trans*-cleavage activity (Chou et al., 2004; Shi and Song, 2006; Lin et al., 2008). Structurally, domain III of monomer B in the dimer not only directly interacts with the opposite S1 subsite (Gln299B...Ser139A, Arg298B...Ser123A; Fig. 1C), but also helps its own N-terminal finger to properly insert into the counterpart monomer A by making both inter-monomer (Arg4B...Glu290A) and intra-monomer (Arg298B...Met6B) interactions (Fig. 1B). Since these residues contribute to the active dimer conformation of mature M<sup>Pro</sup> mainly by electrostatic interactions, we mutated them into oppositely

charged residues and evaluated the influence of these mutations on the N-terminal autocleavage activity of M<sup>Pro</sup>. An M<sup>Pro</sup> autocleavage site (AVLQ↓S), corresponding to the C-terminus of Nsp4 which precedes the M<sup>Pro</sup> (Nsp5) in the viral polyproteins pp1a and pp1ab, was inserted between the N-terminal GST tag and the first residue (Ser1) of the protease (Fig. 2A) to resemble the “immature” M<sup>Pro</sup> before autocleavage. In addition, 24 nucleotides coding for eight extra residues (GPH<sub>6</sub>) were added at the C-terminus of M<sup>Pro</sup> for purification convenience. Using this construct, the N-terminal autocleavage activity of M<sup>Pro</sup> was evaluated in terms of the extent of removal of the GST tag during gene expression and protein purification. As shown by SDS-PAGE analysis



**Figure 2. SDS-PAGE analyses of the autocleavage activity of N-terminally tagged SARS-CoV M<sup>Pro</sup>.** (A) Schematic plots of N-terminal GST- or Trx-tagged SARS-CoV M<sup>Pro</sup> constructs designed in this study. The arrow with “\*” represents the autocleavage site of M<sup>Pro</sup> and the arrow alone indicates the cleavage site of PreScission protease. (B–F) Purification of WT, C145A, R4E, E290R and R298E mutants of GST-tagged M<sup>Pro</sup>. Lane 1, protein molecular-mass marker; lanes 2–5 represent the proteases eluted by 500 mM imidazole from the Ni-NTA column. (G–H) Purification of WT and R298E mutant of Trx-tagged M<sup>Pro</sup>. Lanes 2–3 and lanes 2–5 represent the 500 mM imidazole eluants from Ni-NTA column respectively.

(Fig. 2B), only one purified protein band with an  $M_r$  of approximately 33.0 kDa was obtained in the eluants of wild-type (WT)  $M^{\text{Pro}}$  from the Ni-NTA affinity column, well in agreement with the molar mass calculated from the amino-acid sequence (33.8 kDa) of mature  $M^{\text{Pro}}$ . This indicated that the N-terminal GST fusion tag (~26 kDa) had been removed by WT protease through autocleavage. As a negative control (Fig. 2C), we mutated the catalytic residue Cys145 to Ala. The C145A mutant exhibited a band with an  $M_r$  of around 59 kDa, suggesting that it still existed as a GST-fusion protein and excluding the possibility that the autocleavage of WT protease was caused by non-specific side reactions in the *E. coli* expression system used to produce the proteins. Subsequently, autocleavage assays were carried out for all three mutants using the same procedure. For the R4E mutant (Fig. 2D), the autocleavage activity was not affected since the GST tag was cleaved off completely. It has been reported (Chou et al., 2004) that mutation of the highly, but not absolutely conserved Arg4 to Ala can result in unstable dimers while the mutant still maintains a *trans*-cleavage activity comparable to that of WT protease. Thus, it is possible that the "immature" R4E mutant can still perform efficiently in N-terminal autocleavage during expression, as a small amount of unstable dimers might form. On the other hand, mutation of the totally conserved Glu290 to Ala was reported to induce complete dimer dissociation of  $M^{\text{Pro}}$  in solution (Shi and Song, 2006) and the Arg298Ala mutation also produced a monomeric structure in the crystal (Shi et al., 2008), thereby inducing complete loss of *trans*-cleavage activity. These results clearly indicated that Glu290 and Arg298 are key factors in maintaining the dimeric form of mature  $M^{\text{Pro}}$ . Surprisingly, as shown in Fig. 2E and 2F, both the E290R and R298E mutants displayed no obvious difference of autocleavage behavior compared to WT protease. Since mutation of either of these two residues were reported to completely abolish the dimer of mature  $M^{\text{Pro}}$ , our finding raises the intriguing question of how the E290R and R298E mutants can auto-process their N-terminal GST tags when they are unable to form the active dimer structure. In order to better explain these results and elucidate the autocleavage mechanism of  $M^{\text{Pro}}$ , the contributions of these residues to dimerization and *trans*-cleavage activity of mature  $M^{\text{Pro}}$  needed to be further investigated.

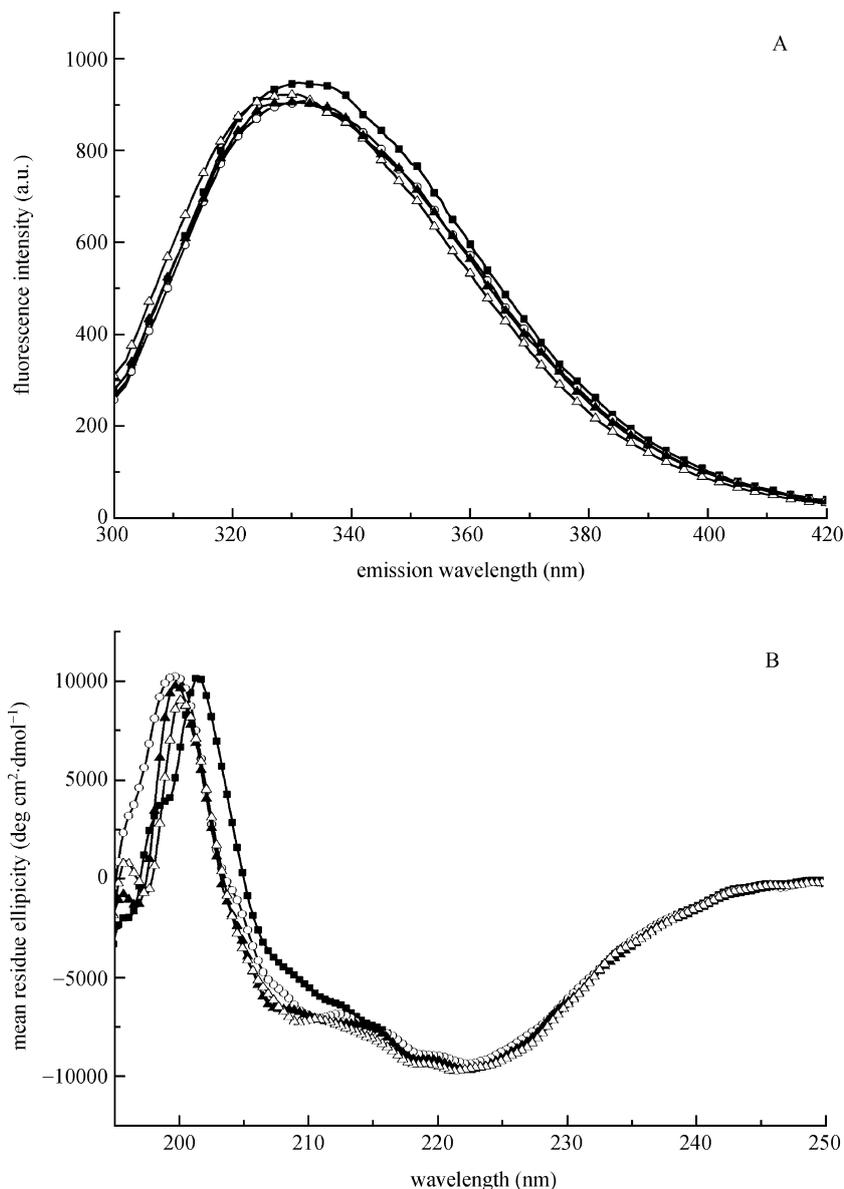
The GST protein alone also exists as a homodimer in the crystal (Nishida et al., 1998) and the N-terminal autocleavage of the E290R or R298E mutant might possibly be due to dimerization of the GST-fused  $M^{\text{Pro}}$  through the N-terminal GST tag. For excluding this possibility, WT  $M^{\text{Pro}}$  and all the mutants were also subcloned into the vector pET-32a (Fig. 2A). This construct encodes an N-terminal thioredoxin (Trx) tag, which does not form dimers by itself (Jeng et al., 1994). As shown by SDS-PAGE (Fig. 2G and 2H), two purified protein bands corresponding to cleaved  $M^{\text{Pro}}$  (~33 kDa) and Trx (~19 kDa including an additional in-frame sequence from

the vector) were observed for both WT and R298E mutant, suggesting that the Trx-tagged R298E mutant maintains autocleavage activity comparable to WT protease. These findings therefore clearly indicate that the N-terminal fusion tag has no impact on autocleavage of  $M^{\text{Pro}}$  and further validate the reliability of our autocleavage assay.

### Folding behavior and dimerization features of mature WT and mutated SARS-CoV $M^{\text{Pro}}$

As described above, all three mutants maintain efficient N-terminal autocleavage activity despite the residues we selected for mutation, especially E290 and R298, have been reported to be indispensable for dimerization of mature  $M^{\text{Pro}}$  (Shi and Song, 2006; Shi et al., 2008). In view of this apparent inconsistency, we further investigated the folding behavior and dimerization features of all mature mutants after autocleavage. Fig. 3A shows the fluorescence emission spectra of mature WT and the mutants of  $M^{\text{Pro}}$ . The emission  $\lambda_{\text{max}}$  of WT  $M^{\text{Pro}}$  is around 330 nm. Similar to WT protease, all three mutants show only minor differences in the emission  $\lambda_{\text{max}}$  (varying from 329 to 331 nm), demonstrating that mutation of a single residue on the dimer interface into one carrying the opposite charge has not changed the fold of the protease. In addition, the far-UV CD spectra of mature WT and mutated  $M^{\text{Pro}}$  are also similar to one another (Fig. 3B). All spectra feature a positive peak around 200 nm and dual negative peaks at 209 and 222 nm, typical of proteins containing significant amounts of  $\alpha$ -helix and  $\beta$ -sheet. These data suggest that all three mutants have well-defined secondary structure and exclude the possibility of misfolding caused by the mutation of individual residues.

The dimerization of mature  $M^{\text{Pro}}$  has been successfully characterized by various biochemical and biophysical methods (Shi et al., 2004; Chen et al., 2005; Hsu et al., 2005b; Graziano et al., 2006b). According to a published method (Prakash et al., 2002), we performed a chemical cross-linking analysis of mature WT  $M^{\text{Pro}}$ . When incubated with 0.01% glutaraldehyde, the protease at a concentration of 0.1 mg/mL predominantly displayed the monomeric form near 33.0 kDa, with a minor band around 66.0 kDa corresponding to the dimer (Fig. 4A, lane 5). With increasing protein concentration, the dimeric form increased and was almost equivalent to the amount of the monomeric form at 1 mg/mL (Fig. 4A, lanes 6–8). When using a higher concentration of glutaraldehyde (0.1%), a similar cross-linking pattern of the protease was observed with slightly higher efficiency, as the dimer had become the predominant species at the same protein concentrations (Fig. 4A, lanes 1–4). These results further demonstrate that mature WT  $M^{\text{Pro}}$  possesses a dimer-monomer equilibrium in solution and its dimerization is concentration-dependent, in good agreement with the literatures (Chou et al., 2004; Shi et al., 2004; Chen et al., 2005; Graziano et al., 2006b). Nevertheless, the possibility of minor

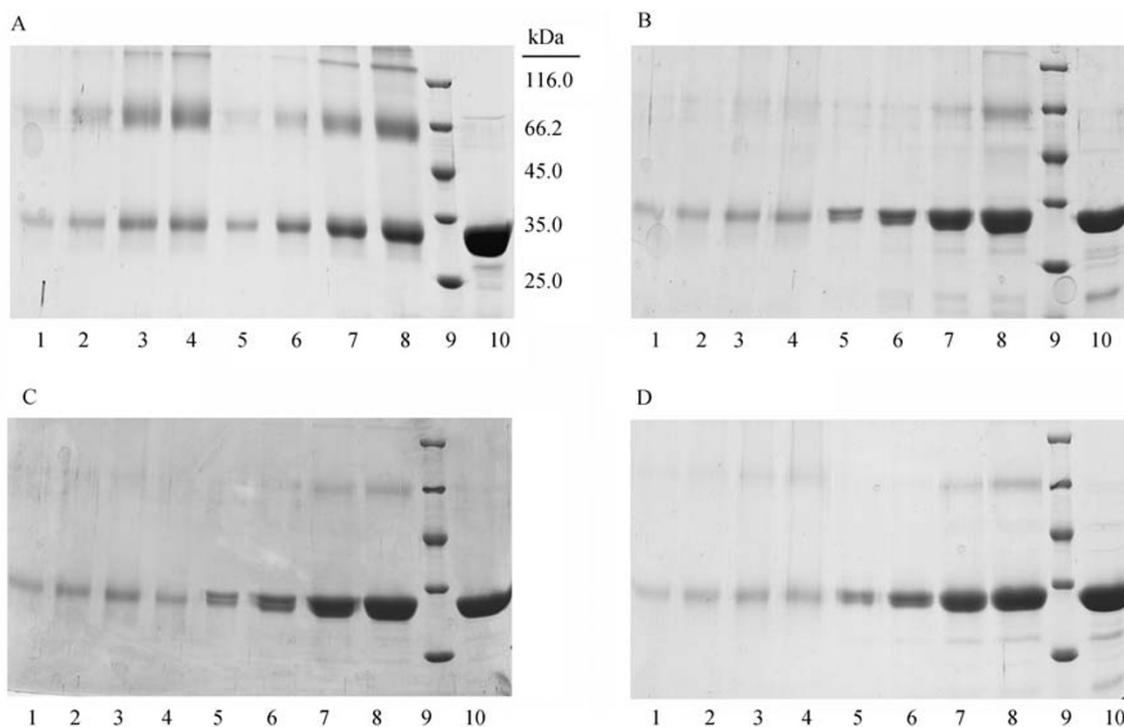


**Figure 3. Fluorescence emission and Far-UV CD spectra of mature WT and mutated SARS-CoV M<sup>Pro</sup>.** (A) The fluorescence spectra were recorded at 25°C after excitation at 280 nm. SARS-CoV M<sup>Pro</sup>: (■) WT; (○) R4E; (▲) E290R; (△) R298E. (B) The CD spectra of M<sup>Pro</sup> at 25°C are shown as: (■) WT; (○) R4E; (▲) E290R; (△) R298E.

artificial cross-linking effects might still exist due to the appearance of high-order multimers in SDS-PAGE (Fig. 4A). However, since the cross-linking analyses of all three mutants can be performed under exactly the same experimental conditions, this method is still very useful for examining the effects of these mutations on dimerization of mature M<sup>Pro</sup>. As indicated in Fig. 4B, the R4E mutant showed an obvious tendency to monomer and the amount of the dimer decreased significantly at lower protein concentrations compared to WT protease, which is consistent with the published result that mutation of Arg4 would result in weakened dimerization of M<sup>Pro</sup> (Chou et al., 2004). For the E290R and R298E mutants

(Fig. 4C and 4D), dimerization seemed to be disrupted even more severely than for the R4E mutant since almost no dimer can be observed at low protein concentrations, further demonstrating the indispensability of Glu290 and Arg298 for maintaining the dimeric structure. However, a very small amount of dimeric form of the E290R and R298E mutants were visible in SDS-PAGE at relatively high protein concentrations (Fig. 4C and 4D, lanes 7, 8). This might be the real dimer of M<sup>Pro</sup> or just a dimeric product caused by minor artificial cross-linking effects as mentioned above.

Subsequently, size-exclusion chromatography (SEC) analyses were carried out to evaluate the dimer-monomer



**Figure 4.** SDS-PAGE profiles of chemically cross-linked mature WT and mutated SARS-CoV M<sup>pro</sup>. Cross-linking analyses of (A) WT SARS-CoV M<sup>pro</sup>, (B) R4E, (C) E290R, and (D) R298E mutants, respectively. Lanes 1–4 and 5–8 represent M<sup>pro</sup> cross-linked by 0.1% or 0.01% glutaraldehyde with a concentration series of 0.1, 0.2, 0.5, 1.0 mg/mL; lane 9, protein molecular-mass marker; lane 10, untreated M<sup>pro</sup> (1.0 mg/mL).

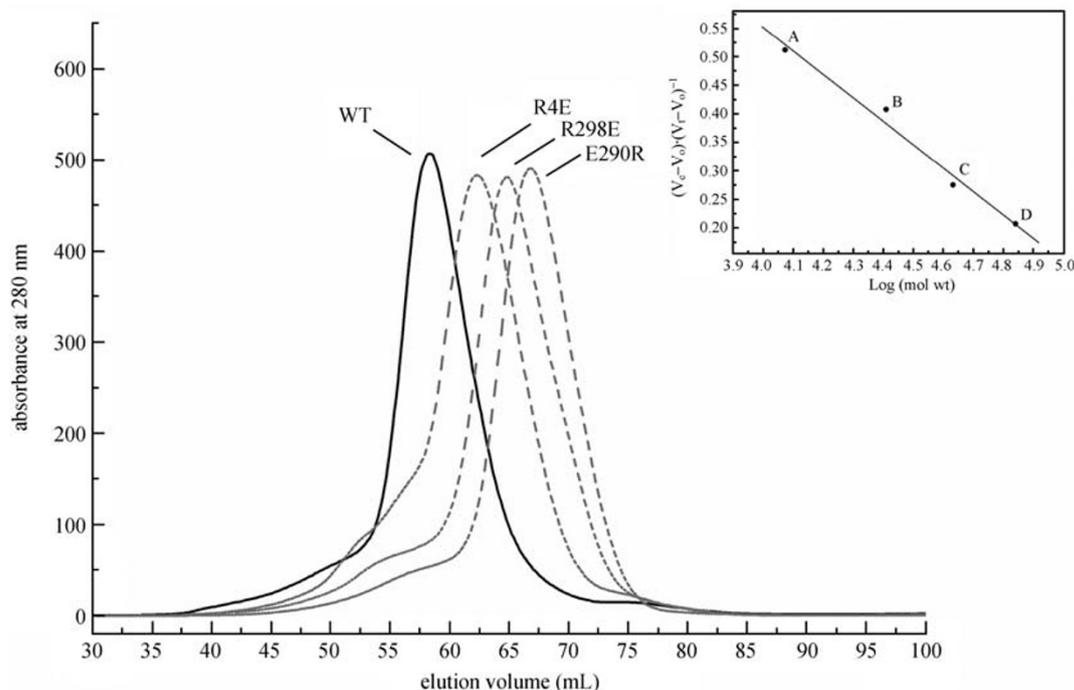
equilibria of the three mutants more precisely. In view of the appearance of a dimeric form of the mutants at high concentration in the cross-linking analyses, we used a protein concentration of 5 mg/mL for each SEC run, which is much higher than the highest concentration used for cross-linking. The physical state corresponding to the native monomer or dimer of M<sup>pro</sup> in the gel-filtration column was then monitored. As shown in Fig. 5, WT M<sup>pro</sup> eluted as a single peak with a retention volume of 57.01 mL. The elution profiles of four molecular-mass marker proteins confirmed that the peak represented the dimeric species of M<sup>pro</sup> (Table 1, estimated  $M_r$ : ~66.7 kDa), implying that in solution, WT protease exists almost exclusively as a dimer at relatively high protein concentration. In comparison, the R4E mutant eluted at a higher volume (61.86 mL), consistent with it being a mixture of dimer and monomer, with the monomer being the predominant species (Table 1, estimated  $M_r$ : ~47.6 kDa). This further supports the cross-linking result and the literature report that Arg4 affects dimerization only to a moderate extent (Chou et al., 2004). For E290R and R298E mutants, the elution volumes dramatically shifted to 66.47 and 64.79 mL, corresponding to estimated  $M_r$  values of 34.6 kDa and 38.9 kDa, respectively, clearly indicating a monomeric state in solution. The data therefore demonstrate that individual mutation of Glu290 or Arg298 is sufficient to prevent dimerization of

mature M<sup>pro</sup> even at a high protein concentration, which convincingly supports the published results (Shi and Song, 2006; Shi et al., 2008) and also suggests that the very minor amounts of E290R and R298E dimers that appeared in SDS-PAGE after cross-linking (Fig. 4C and 4D) is likely due to a small extent of nonspecific cross-linking.

**Table 1** Estimation of protein multimeric states of mature WT and mutated SARS-CoV M<sup>pro</sup> based upon gel filtration data

	WT	R4E	E290R	R298E
$V_e$ (mL)	57.01	61.86	66.47	64.79
$K_{av}$	0.21	0.27	0.33	0.31
estimated $M_r$	66696	47610	34585	38859
monomer $M_r$	34823	34796	34851	34796
estimated $M_r$ / monomer $M_r$	1.92	1.37	0.99	1.12
solution state	dimer	mixture of dimer and monomer	monomer	monomer

Note:  $V_e$  is the elution volume and  $K_{av}$  is the gel-phase distribution coefficient. Calibration of the column determined the relationship between  $K_{av}$  and  $M_r$  as:  $K_{av} = -0.4109 \log M_r + 2.1949$ . Full values were used in all calculations, but were rounded to the nearest two decimals for tabulation.



**Figure 5. Dimer-monomer equilibria of mature WT and mutated SARS-CoV M<sup>Pro</sup> analyzed by SEC.** Elution profiles of WT and mutated M<sup>Pro</sup>s at pH 7.5 with a concentration of 5 mg/mL were measured. Each protein sample was loaded onto a HiLoad 16/60 Superdex 75 prep grade column and then eluted at a flow rate of 1.5 mL/min with detection of absorbance at 280 nm. Inset shows calibrated gel filtration data for four marker proteins, as described in *Materials and Methods*.

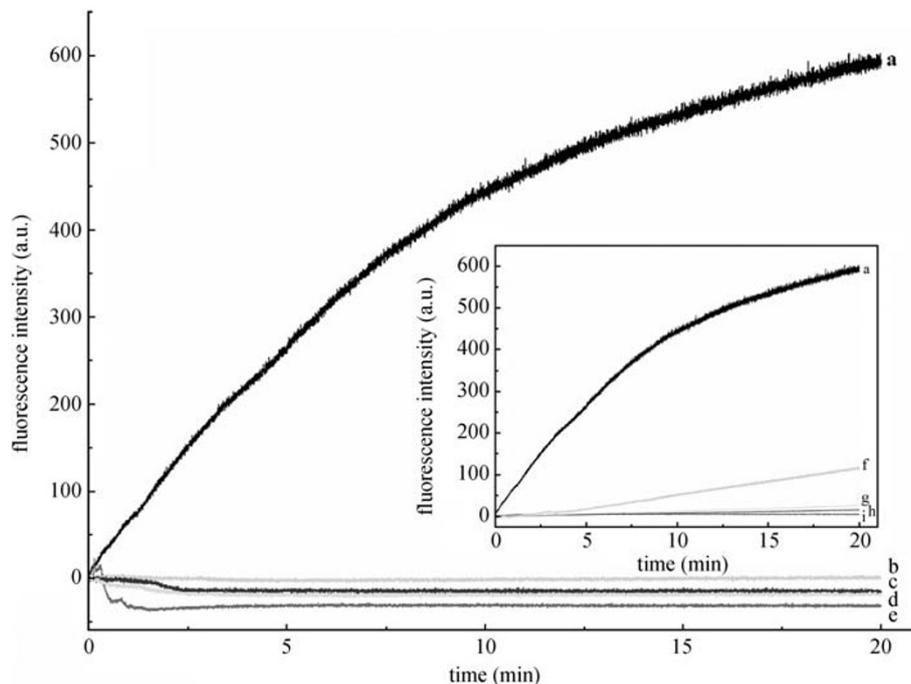
### **Trans-cleavage activity of mature WT and mutated SARS-CoV M<sup>Pro</sup>**

Numerous published data have proposed that only the dimer is the biologically functional form of mature M<sup>Pro</sup> and the dissociated monomer has no *trans*-cleavage activity. Since mutation of a single residue on the dimer interface could result in partial or complete dimer dissociation of mature M<sup>Pro</sup> as demonstrated above, the three mutants were expected to show decreased or no *trans*-cleavage activity. To verify this prediction, we performed *trans*-cleavage assays of mature WT protease and the mutants using a fluorogenic peptide substrate (Verschueren et al., 2008). As shown in Fig. 6, the fluorescence increase following hydrolysis of the substrate by WT M<sup>Pro</sup> was significant and time-dependent, implying that the protease *trans*-cleaved the substrate efficiently, whilst the C145A mutant showed negligible activity against the substrate, thus proving the validity of the *trans*-cleavage assay. Similar to the C145A mutant, the *trans*-cleavage activities of the three “dimerization mutants” at a concentration of 0.1  $\mu$ M were all below the detection limit of the assay during the entire measuring process (Fig. 6), even when the reaction time was prolonged to overnight (data not shown). Next, we increased the concentration of the mutants up to 1  $\mu$ M, and the data showed that the R4E mutant displayed a slight increase of activity but still no measurable *trans*-cleavage reactivity was detected for the E290R and R298E mutants (see inset of

Fig. 6). These results clearly reveal that mutation of Glu290 or Arg298 induces a complete loss of *trans*-cleavage activity of mature M<sup>Pro</sup>, correlating well with the complete dimer dissociation of the two mutants determined by SEC analyses. On the other hand, the R4E mutant exhibits a weak activity when increasing protein concentration, well in agreement with its unstable dimerization in solution.

### **Reconstruction of N-terminal autocleavage of M<sup>Pro</sup> with mature E290R and “immature” C145A/E290R mutants**

According to all the results described above, mutation of residues E290 or R298 seemed to have no obvious influence on the N-terminal autocleavage of “immature” M<sup>Pro</sup>, although it did cause complete dimer dissociation and disruption of *trans*-cleavage activity of the mature form of the mutated M<sup>Pro</sup>. This surprising observation led to the interesting conclusion that the N-terminal autocleavage of SARS-CoV M<sup>Pro</sup> is not dependent on dimerization, at least it does not require the final dimer conformation existing in the mature protease. On the other hand, this final dimer conformation is indispensable for *trans*-cleavage activity of mature M<sup>Pro</sup>. In order to prove this hypothesis, an N-terminal autocleavage assay was constructed. The inactive C145A/E290R double mutant, which cannot perform autocleavage and still contains the N-terminal GST tag, was prepared as an “immature” substrate for mature WT M<sup>Pro</sup> and the E290R mutant. The cleavage



**Figure 6. Fluorescence profiles of hydrolysis of the fluorogenic substrate by mature WT and mutated SARS-CoV M<sup>Pro</sup>.** The fluorogenic substrate at a concentration of 3  $\mu$ M was incubated with WT or mutated SARS-CoV M<sup>Pro</sup> at 25°C. Increase of emission fluorescence intensity at 490 nm was recorded continuously,  $\lambda_{EX} = 340$  nm. The emission spectra were recorded for 20 min as follows: (a) WT, 0.1  $\mu$ M; (b) R4E, 0.1  $\mu$ M; (c) E290R, 0.1  $\mu$ M; (d) R298E, 0.1  $\mu$ M; (e) C145A, 0.1  $\mu$ M; Inset: (f) R4E, 1  $\mu$ M; (g) R298E, 1  $\mu$ M; (h) E290R, 1  $\mu$ M; (i) C145A, 1  $\mu$ M.

process can be easily followed from the appearance of free GST in SDS-PAGE. As shown in Fig. 7A, GST-C145A/E290R was cleaved efficiently by mature WT M<sup>Pro</sup>. The band (~26 kDa) corresponding to the cleaved GST tag appeared after 5 min of incubation time (lane 2) but not in the substrate-only lane (lane 1), and the cleavage was totally completed within 2 hours (lanes 3–6). Since mature WT M<sup>Pro</sup> can form the active dimer even at a low concentration in solution, the free GST tag is the product of *trans*-cleavage of the “immature” substrate by the WT M<sup>Pro</sup> dimer (Fig. 7C). As a control, an unrelated protein (the effector domain of influenza A virus nonstructural protein 1, H1N1 NS1 ED, Fig. 7A, lane 8) containing an N-terminal His tag and an M<sup>Pro</sup> cleavage site (AVLQ↓S), was also prepared as a *trans*-cleavage substrate for WT M<sup>Pro</sup>. Compared to GST-C145A/E290R, a similar cleavage pattern was observed for His-H1N1 NS1 ED and the cleaved substrate (~17 kDa) was clearly visible in SDS-PAGE (lanes 9–13), further demonstrating the high *trans*-cleavage activity of the WT M<sup>Pro</sup> dimer. With the mature E290R mutant, quite a different cleavage behavior was observed between His-H1N1 NS1 ED and GST-C145A/E290R (Fig. 7B). The mature E290R mutant showed no detectable cleavage activity with His-H1N1 NS1 ED during the entire measuring procedure (lanes 9–13). As His-H1N1 NS1 ED is a *trans*-cleavage substrate for M<sup>Pro</sup> (see above; Fig. 7D), this result is well in agreement with the conclusion that the mature E290R mutant exists as a monomer in solution as revealed by SEC

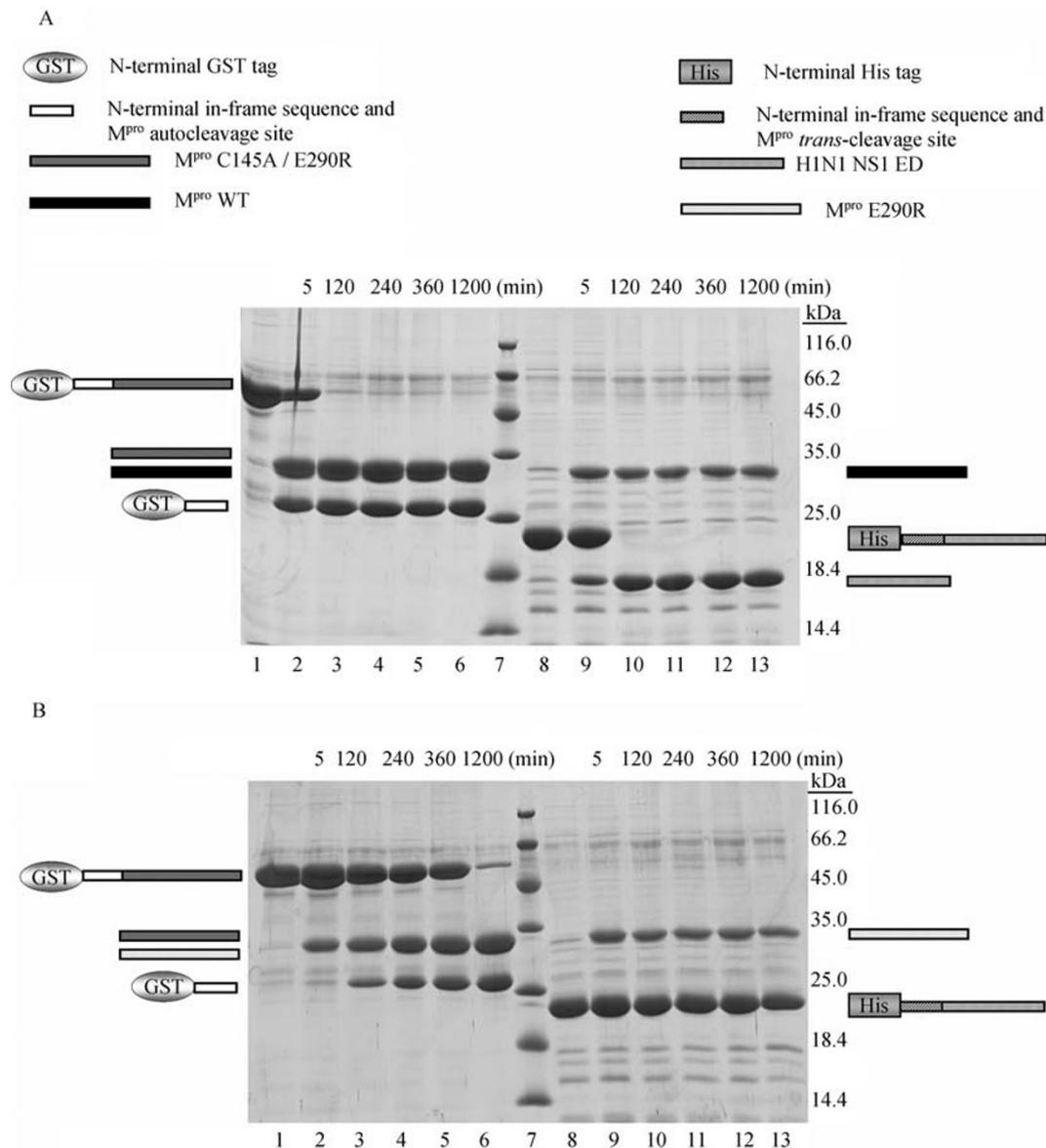
analysis (Fig. 5), without any *trans*-cleavage activity (Fig. 6). On the other hand, when mixing the E290R mutant with GST-C145A/E290R, the cleaved GST tag appeared after 2 h (lane 3) and cleavage was almost complete after overnight incubation (lanes 4–6), indicating that the mature E290R mutant can still cleave “immature” GST-C145A/E290R, although the efficiency is much lower than for WT protease. Since the mature E290R monomer cannot *trans*-cleave His-H1N1 NS1 ED (Fig. 7B) nor the M<sup>Pro</sup> peptide substrate (Fig. 6) at all, the cleavage of GST-C145A/E290R by the mature E290R monomer might be due to the reconstruction of only the N-terminal autocleavage by these two M<sup>Pro</sup> monomers (Fig. 7D). This strongly supports the idea that autocleavage of coronavirus M<sup>Pro</sup> occurs in *trans* (Anand et al., 2003). Furthermore, as the E290R mutation completely abolishes formation of the mature dimer in solution (Fig. 5), these data strongly support the hypothesis that the active dimer conformation existing in mature WT protease is unnecessary for N-terminal autocleavage of M<sup>Pro</sup>.

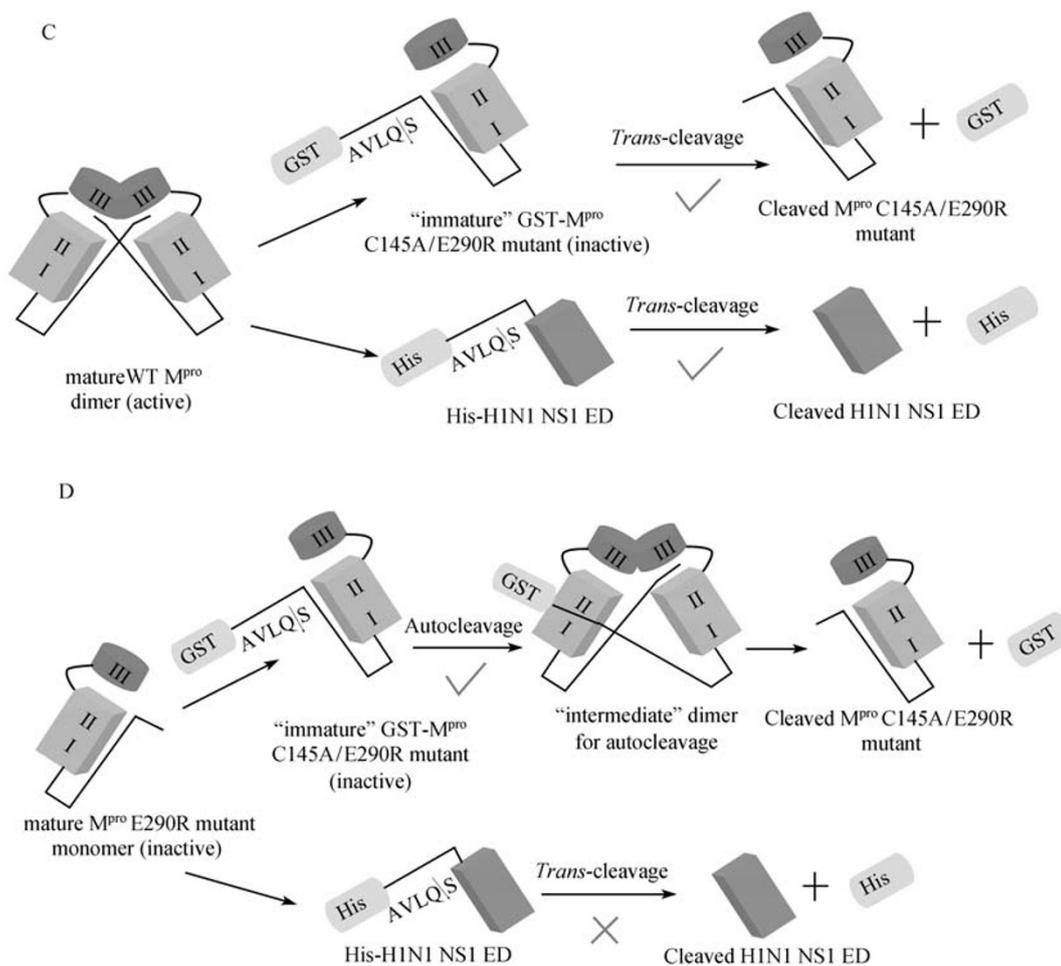
#### How does “immature” M<sup>Pro</sup> release itself from the precursor polyproteins without forming the mature dimer structure?

What we find puzzling is why the N-terminal autocleavage activity of M<sup>Pro</sup> is independent from the dimer structure that is vital to *trans*-cleavage activity of the protease? What is the

difference between the proteolytic autocleavage and *trans*-cleavage mechanisms of SARS-CoV M<sup>Pro</sup>? It has been reported (Shi et al., 2004) that domain III of M<sup>Pro</sup> itself possesses an intrinsic tendency to form a dimer even at a very low concentration. Recently, the crystal structure of domain III alone has been determined, revealing a novel 3D domain-swapped dimer structure (Zhong et al., 2009). Furthermore, it was reported (Zhong et al., 2008) that the N-terminal finger deletion mutant of M<sup>Pro</sup> also forms a novel stable dimer through domain III-domain III interactions, which is different from the dimer interface observed in the WT M<sup>Pro</sup> dimer. These results indicate that in addition to the dimer conformation in the crystal structure of WT protease, M<sup>Pro</sup> might have other possible dimerization modes. M<sup>Pro</sup> is still a part of pp1a and pp1ab when the polyproteins are produced in coronavirus-infected cells. Through the hydrophobic domains of Nsp4 and Nsp6 that flank the M<sup>Pro</sup> (Nsp5)

domain, the unprocessed polyproteins are anchored to the double-membrane vesicles found in infected host cells. Subsequently, when these precursor proteins accumulate to high concentration, the "immature" M<sup>Pro</sup> can release itself from the polyproteins by inter-molecular cleavage, and then the self-released mature M<sup>Pro</sup> triggers the *trans*-cleavage processing of the polyproteins (Anand et al., 2003). In the dimer structure of mature M<sup>Pro</sup>, the N-terminal finger of one monomer can form intensive interactions with the S1 subsite and Domain III (the C-terminal helix) of the other monomer (Fig. 1B and 1C), which are key residues for maintaining the active dimer conformation. It is very unlikely that the "immature" M<sup>Pro</sup> would form such dimer because of the two TM domains flanking it at both N- and C-termini before self-releasing (Oostra et al., 2008). In our proposed model (Fig. 8), the N-terminal autocleavage might only need two "immature" M<sup>Pro</sup> domains in monomeric polyproteins approach one





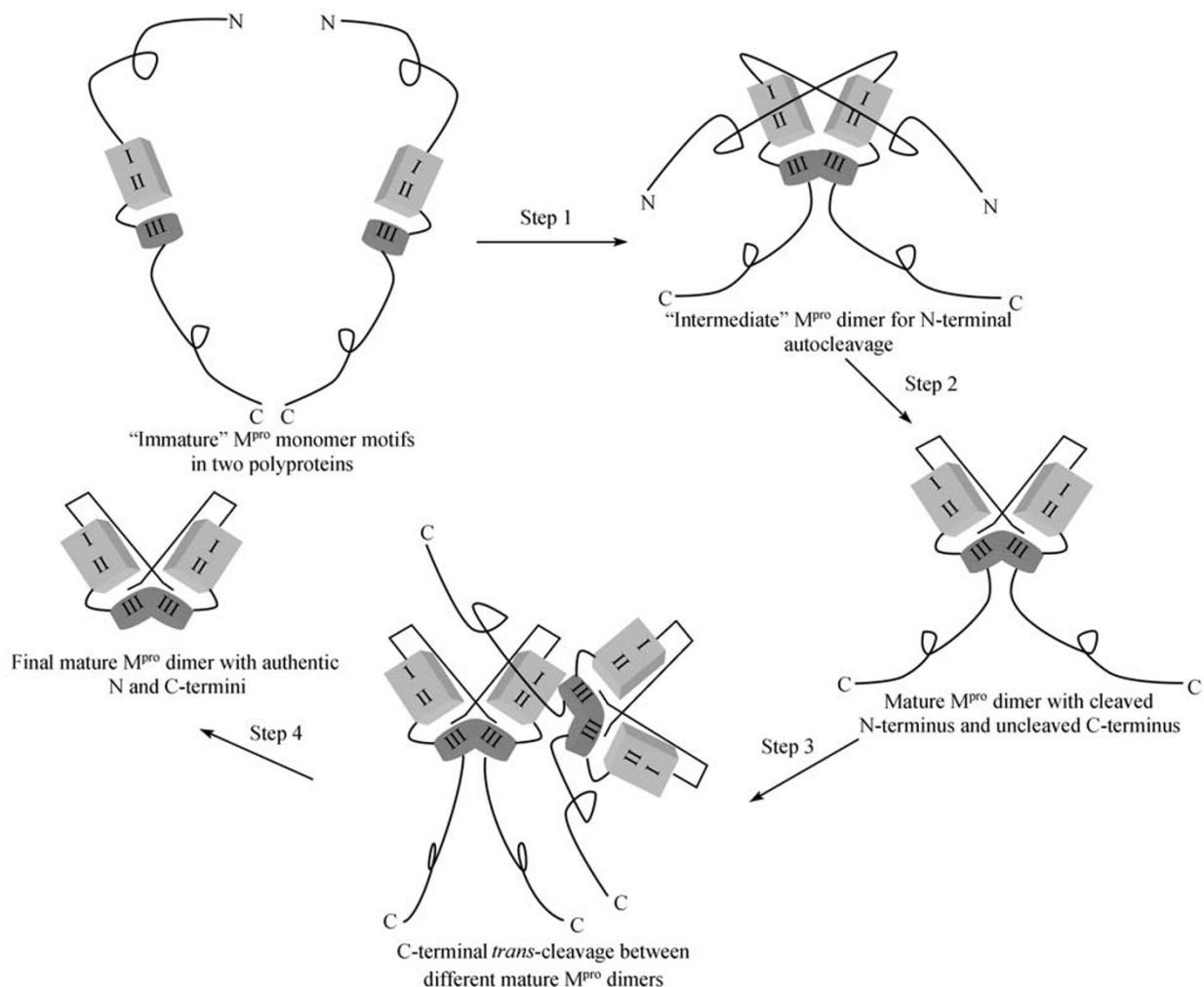
**Figure 7. Reconstruction of N-terminal autocleavage of  $M^{\text{pro}}$  by mature E290R and “immature” C145A/E290R mutants.** SDS-PAGE analyses of cleavage by mature WT  $M^{\text{pro}}$  (A) and E290R mutant (B). The inactive C145A/E290R double mutant (to prevent autocleavage) containing an N-terminal GST tag and autocleavage site was prepared as an “immature” substrate for mature WT  $M^{\text{pro}}$  and E290R mutant. The substrate protein at a concentration of 20  $\mu\text{M}$  (lane 1) was mixed with 5  $\mu\text{M}$  of WT  $M^{\text{pro}}$  or E290R mutant at 25°C and the products were monitored after specified periods of incubation time shown on the top of the figure (lanes 2–6). Meanwhile, the effector domain of influenza A virus nonstructural protein 1 (H1N1 NS1 ED, lane 8) containing an N-terminal His tag and  $M^{\text{pro}}$  cleavage site (AVLQ↓S) was also prepared as a *trans*-cleavage substrate for WT  $M^{\text{pro}}$  and E290R mutant, and the cleavage assay was performed under the same experimental conditions (lanes 9–13). Lane 7, protein molecular-mass marker. (C) The proposed *trans*-cleavage scheme of mature WT  $M^{\text{pro}}$ . (D) The proposed scheme of autocleavage reconstruction of mature E290R with “immature” C145A/E290R mutant.

another to form an “intermediate” dimer, possibly through Domain III-III dimerization, which also agrees well with the plausible dimer formation pathway proposed for the  $M^{\text{pro}}$  previously (Chen et al., 2008a; Hu et al., 2009). The formation of the “intermediate” dimer could trigger the rotation of their domains I/II relative to domains III and thereby insert their “uncleaved” N-termini into the substrate-binding pockets of the opposite monomer, which might induce the active conformation of the S1 subsites through an induced-fit catalytic mechanism. Once the autocleavage is finished, the “cleaved” N-terminal fingers should slip away from the active sites and switch to their final spatial positions, which is observed in the mature dimer of WT  $M^{\text{pro}}$  (Fig. 1B), thereby locking the dimer in a catalytic competitive state. The conformational shift of the N-terminal finger from the

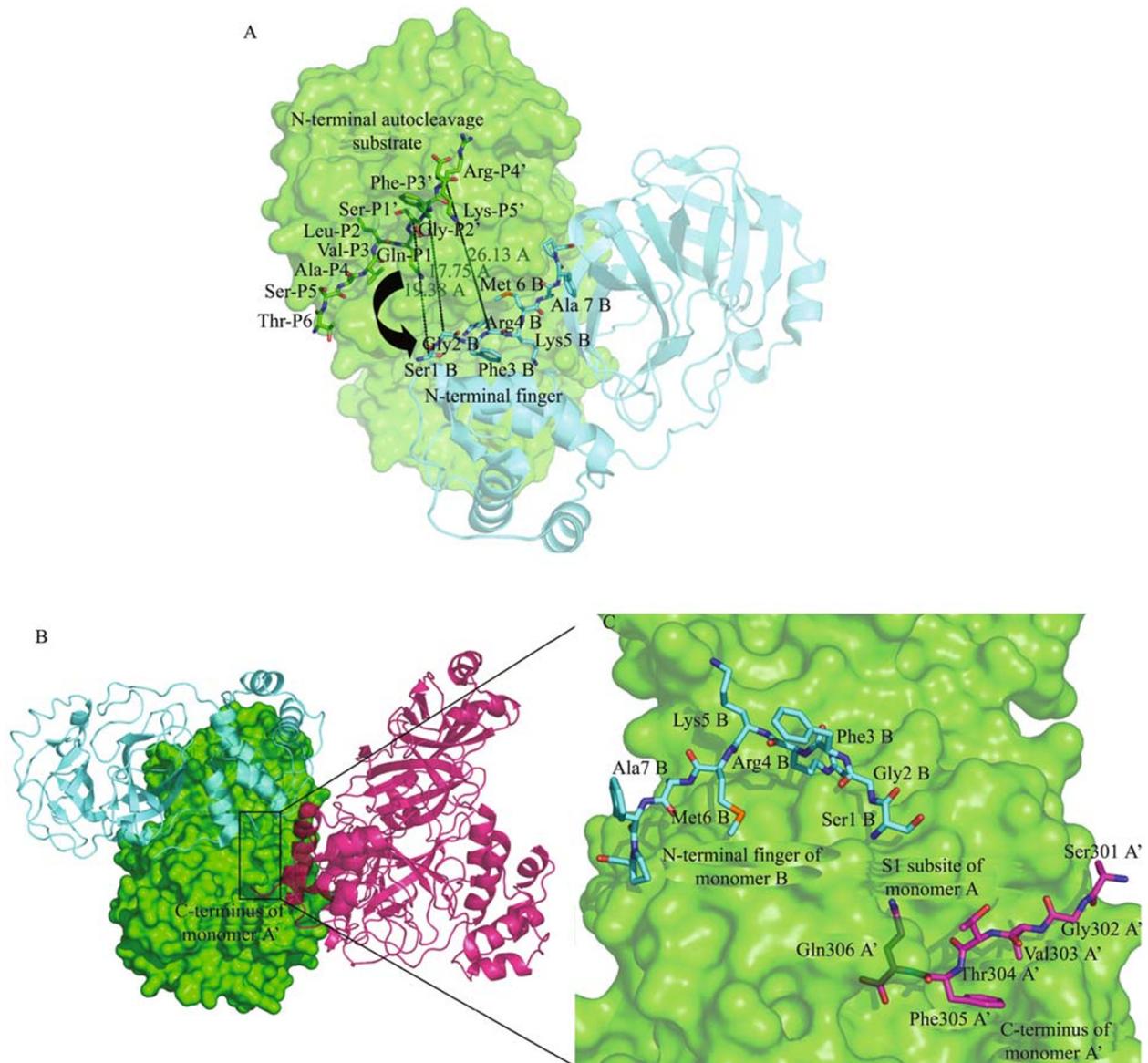
auto-processed state to the final “cleaved” state is also supported by the crystal structure of the  $M^{\text{pro}}$  H41A mutant in complex with an N-terminal auto-cleavage substrate (Xue et al., 2008), indicating an estimated distance of  $\sim 20$  Å between residues P1'–P5' (SGFRK) of the bound substrate and the free N-terminal finger of the mature dimer (Fig. 9A). Meanwhile, Domains III should also glide to their “final” positions to produce the mature active dimer since the last helices of domains III are tethered with the N-terminal fingers in the mature dimer structure (Fig. 1C). The dimer conformation existing in mature  $M^{\text{pro}}$  is not the cause but more likely the product of N-terminal auto-processing of “immature” protease from polyproteins. Thus mutation of residues that are indispensable for dimerization of mature  $M^{\text{pro}}$  has little effect on N-terminal autocleavage since the conformation of the

“intermediate” dimer is different from that of mature dimer. On the other hand, the *trans*-cleavage activity of the mature dimer is much higher than the autocleavage activity of the “intermediate” dimer, as indicated in Fig. 7A and 7B. However, the “intermediate” dimer might be the only feasible structure for “immature” M<sup>Pro</sup> to form within pp1a and pp1ab. In order to perform the first step of autocleavage from the polyproteins, “immature” M<sup>Pro</sup> possibly tolerates a less-than-ideal dimer. Once the mature dimer is formed after N-terminal self-releasing, the unprocessed C-terminus of one mature dimer can insert into an active site of another mature dimer for

*trans*-cleavage (Fig. 8), as revealed by a product-bound crystal structure of the M<sup>Pro</sup> C145A mutant (Hsu et al., 2005a) (Fig. 9B and 9C). After the formation of the final mature M<sup>Pro</sup> dimer with authentic N- and C-termini, *trans*-cleavage of the polyproteins at other sites could subsequently be triggered at a high rate. In conclusion, our current work is expected to provide more insights into the auto-processing mechanism of SARS-CoV M<sup>Pro</sup>. In addition, the proposed auto-releasing model should be further verified by X-ray crystallographic determination of the “intermediate” dimer structure.



**Figure 8. A proposed mode of SARS-CoV M<sup>Pro</sup> auto-release from the precursor polyproteins.** The chymotrypsin fold (domains I and II) and domain III of M<sup>Pro</sup> monomers in two polyproteins are shown as *boxes* and *cylinders*, respectively. The N- and C-termini are also labeled. The auto-release mode involves four steps. *Step 1*, two “immature” M<sup>Pro</sup> monomers approach one another and their domains III form an “intermediate” dimer structure, which triggers the rotations of their chymotrypsin folds and insert their “uncleaved” N-termini into the substrate-binding pockets of the opposite monomers. *Step 2*, with an induced-fit catalytic mechanism, the active site of M<sup>Pro</sup> could be activated and catalyze the N-terminal autocleavage. Afterwards, the “cleaved” N-terminal fingers slip away from the active sites and switch to their final spatial positions, as well as domains III, locking the mature dimer with “uncleaved” C-termini at the active state. *Step 3*, the “uncleaved” C-terminus of one mature dimer can insert into an active site of another mature dimer, as supported by a product-bound crystal structure of the M<sup>Pro</sup> C145A mutant (Hsu et al., 2005a). *Step 4*, once the C-terminus is processed by *trans*-cleavage, the final mature dimer with authentic N- and C-termini is formed, which is observed in the crystal structure of WT M<sup>Pro</sup> (Xue et al., 2007).



**Figure 9.** The crystal structures of SARS-CoV M<sup>pro</sup> H41A mutant in complex with an N-terminal autocleavage substrate (PDB code: 2Q6G) and the C145A mutant (PDB code: 1Z1J). (A) In the dimer of the H41A mutant, monomer A is shown as surface diagram in green and monomer B is represented as ribbon in cyan. The 11-peptide substrate (TS AVLQSGFRK) and the N-terminal finger (SGFRKMA) of monomer B are shown as sticks. Dashes represent the distance (in Å) between residues P1'–P5' of the substrate and the N-terminal finger; (B) The dimer of the C145A mutant is composed of monomer A (shown with surface in green) and monomer B (shown with ribbon in cyan). The C-terminus of monomer A' from the neighboring dimer (shown as ribbon in magenta) in another asymmetric unit is inserted into the active site of monomer A. (C) An enlarged view of (B) near the active site, indicating the C-terminus of monomer A' (residues 301–306, shown as sticks) as well as the N-terminal finger of monomer B in the neighborhood of the active site of monomer A.

## MATERIALS AND METHODS

### Expression and purification

The plasmid of WT SARS-CoV M<sup>pro</sup> was a gift from Prof. Z. Rao (Xue et al., 2007). Briefly, 12 nucleotides coding for the four amino acids AVLQ (corresponding to the P4–P1 residues of the N-terminal

autocleavage site of SARS-CoV M<sup>pro</sup>) were added to precede the N-terminal Ser1 residue of the protease. 24 nucleotides coding for eight extra residues (GPH<sub>6</sub>) were added at the C-terminus of M<sup>pro</sup>. The resulting coding sequence was inserted into BamHI and XhoI sites of the pGEX-6P-1 plasmid. The plasmid was then transformed into *E. coli* BL21 (DE3) cells, and the N-terminal GST fusion protein with a C-terminal His tag was produced by introducing IPTG to 0.5 mM at 25°C for 6 h.

Because of the plasmid construction strategy, the N-terminal GST tag can be cleaved off by autocleavage of active M<sup>Pro</sup>. The purification of the C-terminally GPH<sub>6</sub>-tagged product was performed according to the published method (Xue et al., 2007). The protein was first subjected to Ni-NTA affinity chromatography and concentrated in PreScission protease cleavage buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT). The PreScission protease was then added to cleave off the C-terminal GPH<sub>6</sub> tag, producing an authentic mature SARS-CoV M<sup>Pro</sup>. For the inactive M<sup>Pro</sup> C145A mutant, which cannot perform autocleavage, the product was first purified by Ni-NTA column, followed by adding the PreScission protease to remove both the N-terminal GST and the C-terminal GPH<sub>6</sub> tags. All the cleaved proteins were further purified using anion-exchange chromatography. Finally, the purified and concentrated protease was dialyzed against 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT, 1 mM EDTA and stored at -20°C.

#### Site-directed mutagenesis of SARS-CoV M<sup>Pro</sup>

Site-directed mutagenesis of SARS-CoV M<sup>Pro</sup> was processed by a modified recombinant PCR method. The mutated proteases (Cys145Ala, Arg4Glu, Glu290Arg, Arg298Glu single mutants, and Cys145Ala/Glu290Arg double mutant) were prepared with the QuikChange II site-directed mutagenesis kit (Stratagene) using pGEX-6p-1-SARS-CoV M<sup>Pro</sup> as a template. The nucleotide sequences of the primers used for mutation are given in Table 2. The plasmids encoding mutated forms of M<sup>Pro</sup> were verified by sequencing, and then *E. coli* BL21 (DE3) cells were transformed by the resulting plasmids. The mutated proteins were produced and purified using a procedure similar to that described above for WT protease. The purity and structural integrity of the mutated proteases were analyzed by SDS-PAGE.

All the gene fragments coding for WT and mutated proteases were also cleaved off from the plasmids of pGEX-6p-1-SARS-CoV M<sup>Pro</sup> using BamHI and XhoI, and subcloned into the vector pET-32a with N-terminal thioredoxin (Trx) tag. The resulting plasmids were verified by sequencing and transformed into *E. coli* BL21 (DE3) cells. The proteins were produced and purified using a procedure similar to that for the GST fusion protease mentioned above.

#### N-terminal autocleavage assay

According to the construction strategy of all the plasmids, a SARS-CoV M<sup>Pro</sup> autocleavage site (AVLQ↓S) was located between the

N-terminal GST or Trx tag and the first Ser1 residue of the protease (Fig. 2A). The N-terminal GST or Trx tag can be removed by autocleavage of active M<sup>Pro</sup>. Thus, the N-terminal autocleavage activity of WT and mutated M<sup>Pro</sup>s can be evaluated by the removal extent of GST or Trx tag during the process of protein production and purification. The tag-cleaved protease was then checked using SDS-PAGE after Ni-NTA affinity chromatography purification.

#### Fluorescence spectroscopy

The fluorescence measurements were performed on a Cary Eclipse fluorescence spectrophotometer. The protease samples were prepared in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT, and 1 mM EDTA, with a protein concentration of 5 μM. The fluorescence emission spectra from 300 to 420 nm were recorded after excitation at 280 nm, with a spectral slit width of 5 nm for excitation and emission. Fluorescence spectra of mature WT and mutated SARS-CoV M<sup>Pro</sup>s were measured in a 1-mL quartz cuvette at 25°C. All final spectra were corrected for the buffer contribution, and were the average of three parallel measurements.

#### Circular dichroism (CD) spectroscopy

Far-UV CD spectra from 195 to 250 nm were recorded on a JASCO-715 spectropolarimeter. The protein samples were prepared in 50 mM Tris-HCl pH 7.5, 1 mM EDTA at 25°C, with a concentration of 10 μM. The spectra were collected with 1 nm band width using a 0.1-cm path length cuvette, and normalized by subtracting the baseline recorded for the buffer. Each measurement was repeated five times and the final result was the average of five independent scans. The CD spectra of mature mutated proteases were compared to that of WT SARS-CoV M<sup>Pro</sup> to exclude the possibility of misfolding caused by mutation.

#### Glutaraldehyde cross-linking

Chemical cross-linking was performed for mature WT and mutated SARS-CoV M<sup>Pro</sup>s (final concentrations from 0.1 to 1.0 mg/mL in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT, and 1 mM EDTA) with glutaraldehyde. An aliquot of 25% (v/v) glutaraldehyde was added to the protein samples to give a final concentration of 0.1% or 0.01%, and the samples were incubated at 25°C for 20 min followed by quenching the reaction with the addition of 1.0 M Tris-HCl pH 8.0 (0.5%, v/v). Orthophosphoric acid was thereafter added to the

**Table 2** Nucleotide sequences of the primers used for site-directed mutagenesis of SARS-CoV M<sup>Pro</sup>

oligonucleotide sequence (5'→3')	polarity	mutation introduced
CTTTCCTTAATGGATCAGCTGGTAGTGTGGTTTTAAC	forward	SARS-CoV M <sup>Pro</sup> Cys145Ala
GTAAAACCAACTACCAGCTGATCCATTAAGGAAAG	reverse	SARS-CoV M <sup>Pro</sup> Cys145Ala
GTGTTGCAGAGTGGTTTCG <b>A</b> GAAAAATGGCATTCCCG	forward	SARS-CoV M <sup>Pro</sup> Arg4Glu
CGGAATGCCATTTTC <b>T</b> CGAAACCACTCTGCAACAC	reverse	SARS-CoV M <sup>Pro</sup> Arg4Glu
GGTAGCACTATTTTGAAGAT <b>A</b> GGTTTACACCATTTGATGTTG	forward	SARS-CoV M <sup>Pro</sup> Glu290Arg
CAACATCAAATGGTGTAAAC <b>C</b> TATCTTCTAAATAGTGCTACC	reverse	SARS-CoV M <sup>Pro</sup> Glu290Arg
CCATTTGATGTTGTTGAACAATGCTCTGGTGTG	forward	SARS-CoV M <sup>Pro</sup> Arg298Glu
CACACCAGAGCATTGTT <b>T</b> CAACAACATCAAATGG	reverse	SARS-CoV M <sup>Pro</sup> Arg298Glu

Note: The mutant codons in the oligonucleotide sequences are highlighted in boldface. SARS-CoV M<sup>Pro</sup> amino acids are numbered continuously from the N-terminal residue, Ser1, to the C-terminal residue, Gln306.

reaction mixture to result in precipitation of the cross-linked proteins. After centrifugation (12,000 rpm, 4°C), the precipitate was redissolved in loading buffer and heated at 70°C for 10 min. SDS-PAGE was run with 12.5% acrylamide gels.

#### Size-exclusion Chromatography (SEC) analysis

The dimer-monomer equilibria of mature WT and mutated SARS-CoV M<sup>Pro</sup>s were measured using size-exclusion chromatography on a HiLoad 16/60 Superdex 75 prep grade column through an ÄKTA FPLC system. Buffer used was 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT, and 1 mM EDTA. Protein samples with a concentration of 5 mg/mL were loaded onto the column and then eluted with the buffer by detection of absorbance at 280 nm. The column was calibrated with four molecular-mass marker proteins (labeled in the inset of Fig. 5): (A) cytochrome c (11.8 kDa), (B) chymotrypsinogen A (25.7 kDa), (C) ovalbumin (42.9 kDa), and (D) albumin (69.3 kDa). A linear relationship between the gel-phase distribution coefficient ( $K_{av}$ ) and the average molecular weight ( $M_r$ ) was fit by  $K_{av} = -0.4109 \log M_r + 2.1949$ .

#### Trans-cleavage assay

A FRET-based assay was used to assess the *trans*-cleavage activity of mature WT and mutated SARS-CoV M<sup>Pro</sup>s (Verschuere et al., 2008). The assay made use of a 14-amino acid fluorogenic substrate, Dabcyl-KTSAVLQ↓SGFRKME-EDANS (95% purity, Biosynton GmbH, Berlin, Germany), which contains an optimal M<sup>Pro</sup> cleavage site (indicated by the arrow). The enhanced emission fluorescence due to cleavage of this substrate as catalyzed by the protease was monitored as a function of time at 490 nm with excitation at 340 nm, using a Cary Eclipse fluorescence spectrophotometer. The experiments were performed in a buffer consisting of 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT, and 1 mM EDTA at 25°C. The reaction was initiated by adding M<sup>Pro</sup> (final concentration 0.1–1 μM) in a 96-well black microplate containing a final fluorogenic substrate concentration of 3 μM. The incubation of the substrate in the assay buffer without the protease was performed as a control and the *trans*-cleavage activity was the average of three parallel assays.

#### ACKNOWLEDGEMENTS

This work was supported, in part, by the Sino-European Project on SARS Diagnostics and Antivirals (SEPSDA, contract NO. SP22-CT-2004-003831) and by VIZIER (contract no. LSHG-CT-2004-511960), both funded by the European Commission. We acknowledge support from the Sino-German Center for the Promotion of Research, Beijing (grant no. 233 (202/6)), the Schleswig-Holstein Innovation Fund, and the DFG (Hi 611/4 and Cluster of Excellence "Inflammation at Interfaces"). We thank Professor Z. Rao (Tsinghua University, Beijing) for providing the plasmid of WT SARS-CoV M<sup>Pro</sup>, Dr. Jinzhi Tan for discussion, and Silke Schmidtke for expert technical assistance.

#### ABBREVIATIONS

CD, circular dichroism; Dabcyl, 4-[[4-(dimethylamino)phenyl]azo] benzoic acid; EDANS, 5-[[2'-aminoethyl]-amino] naphthelene sulfonic acid; FRET, fluorescence resonance energy transfer; GST, glutathion S-transferase; M<sup>Pro</sup>, main protease; Nsp, non-structural protein; SARS-CoV, severe acute respiratory syndrome coronavirus;

SEC, size-exclusion chromatography; TM, transmembrane; Trx, thioredoxin; WT, wild type

#### REFERENCES

- Anand, K., Yang, H., Bartlam, M., Rao, Z., and Hilgenfeld, R. (2005). Coronavirus main proteinase: target for antiviral drug therapy. In *Coronaviruses with special emphasis on first insights concerning SARS*. A. Schmidt, M.H. Wolf and O. Weber, ed. (Basel, Birkhäuser) pp. 173–199.
- Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R., and Hilgenfeld, R. (2003). Coronavirus main proteinase (3CL<sup>Pro</sup>) structure: basis for design of anti-SARS drugs. *Science* 300, 1763–1767.
- Barrila, J., Bacha, U., and Freire, E. (2006). Long-range cooperative interactions modulate dimerization in SARS 3CL<sup>Pro</sup>. *Biochemistry* 45, 14908–14916.
- Chen, S., Chen, L., Tan, J., Chen, J., Du, L., Sun, T., Shen, J., Chen, K., Jiang, H., and Shen, X. (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.
- Chen, S., Hu, T., Zhang, J., Chen, J., Chen, K., Ding, J., Jiang, H., and Shen, X. (2008a). Mutation of Gly-11 on the dimer interface results in the complete crystallographic dimer dissociation of severe acute respiratory syndrome coronavirus 3C-like protease: crystal structure with molecular dynamics simulations. *J Biol Chem* 283, 554–564.
- Chen, S., Zhang, J., Hu, T., Chen, K., Jiang, H., and Shen, X. (2008b). Residues on the dimer interface of SARS coronavirus 3C-like protease: dimer stability characterization and enzyme catalytic activity analysis. *J Biochem* 143, 525–536.
- Chou, C.Y., Chang, H.C., Hsu, W.C., Lin, T.Z., Lin, C.H., and Chang, G.G. (2004). Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry* 43, 14958–14970.
- Drosten, C., Preiser, W., Gunther, S., Schmitz, H., and Doerr, H.W. (2003). Severe acute respiratory syndrome: identification of the etiological agent. *Trends Mol Med* 9, 325–327.
- Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J., et al. (2004). Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J Biol Chem* 279, 1637–1642.
- Graziano, V., McGrath, W.J., DeGruccio, A.M., Dunn, J.J., and Mangel, W.F. (2006a). Enzymatic activity of the SARS coronavirus main proteinase dimer. *FEBS Lett* 580, 2577–2583.
- Graziano, V., McGrath, W.J., Yang, L., and Mangel, W.F. (2006b). SARS-CoV main proteinase: The monomer-dimer equilibrium dissociation constant. *Biochemistry* 45, 14632–14641.
- Groneberg, D.A., Hilgenfeld, R., and Zabel, P. (2005). Molecular mechanisms of severe acute respiratory syndrome (SARS). *Respir Res* 6, 8–23.
- Hsu, M.F., Kuo, C.J., Chang, K.T., Chang, H.C., Chou, C.C., Ko, T.P., Shr, H.L., Chang, G.G., Wang, A.H., and Liang, P.H. (2005a). Mechanism of the maturation process of SARS-CoV 3CL protease. *J Biol Chem* 280, 31257–31266.
- Hsu, W.C., Chang, H.C., Chou, C.Y., Tsai, P.J., Lin, P.I., and Chang, G.G. (2005b). 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.
- Hu, T., Zhang, Y., Li, L., Wang, K., Chen, S., Chen, J., Ding, J., Jiang, H., and Shen, X. (2009). Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure. *Virology* 388, 324–334.
- Huang, C., Wei, P., Fan, K., Liu, Y., and Lai, L. (2004). 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry* 43, 4568–4574.
- Jeng, M.F., Campbell, A.P., Begley, T., Holmgren, A., Case, D.A., Wright, P.E., and Dyson, H.J. (1994). High-resolution solution structures of oxidized and reduced *Escherichia coli* thioredoxin. *Structure* 2, 853–868.
- Knoops, K., Kikkert, M., van den Worm, S.H., Zevenhoven-Dobbe, J. C., van der Meer, Y., Koster, A.J., Mommaas, A.M., and Snijder, E. J. (2008). SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol* 6, e226.
- Lee, T.W., Cherney, M.M., Huitema, C., Liu, J., James, K.E., Powers, J.C., Eltis, L.D., and James, M.N. (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.
- Lin, C.W., Tsai, C.H., Tsai, F.J., Chen, P.J., Lai, C.C., Wan, L., Chiu, H. H., and Lin, K.H. (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 Lett* 574, 131–137.
- Lin, P.Y., Chou, C.Y., Chang, H.C., Hsu, W.C., and Chang, G.G. (2008). Correlation between dissociation and catalysis of SARS-CoV main protease. *Arch Biochem Biophys* 472, 34–42.
- Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khattri, J., Asano, J.K., Barber, S.A., Chan, S.Y., *et al.* (2003). The genome sequence of the SARS-associated coronavirus. *Science* 300, 1399–1404.
- Nishida, M., Harada, S., Noguchi, S., Satow, Y., Inoue, H., and Takahashi, K. (1998). Three-dimensional structure of *Escherichia coli* glutathione S-transferase complexed with glutathione sulfonate: catalytic roles of Cys10 and His106. *J Mol Biol* 281, 135–147.
- Oostra, M., Hagemeyer, M.C., van Gent, M., Bekker, C.P., te Lintelo, E.G., Rottier, P.J., and de Haan, C.A. (2008). Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J Virol* 82, 12392–12405.
- Peiris, J.S., Lai, S.T., Poon, L.L., Guan, Y., Yam, L.Y., Lim, W., Nicholls, J., Yee, W.K., Yan, W.W., Cheung, M.T., *et al.* (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 361, 1319–1325.
- Prakash, K., Prajapati, S., Ahmad, A., Jain, S.K., and Bhakuni, V. (2002). Unique oligomeric intermediates of bovine liver catalase. *Protein Sci* 11, 46–57.
- Shan, Y.F., Li, S.F., and Xu, G.J. (2004). A novel auto-cleavage assay for studying mutational effects on the active site of severe acute respiratory syndrome coronavirus 3C-like protease. *Biochem Biophys Res Commun* 324, 579–583.
- Shi, J., and Song, J. (2006). The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *Febs J* 273, 1035–1045.
- Shi, J., Sivaraman, J., and Song, J. (2008). Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. *J Virol* 82, 4620–4629.
- Shi, J., Wei, Z., and Song, J. (2004). Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors. *J Biol Chem* 279, 24765–24773.
- Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., and Gorbalenya, A.E. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* 331, 991–1004.
- Snijder, E.J., van der Meer, Y., Zevenhoven-Dobbe, J., Onderwater, J. J., van der Meulen, J., Koerten, H.K., and Mommaas, A.M. (2006). Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. *J Virol* 80, 5927–5940.
- Tan, J.Z., Verschuere, K.H.G., Anand, K., Shen, J.H., Yang, M.J., Xu, Y.C., Rao, Z.H., Bigalke, J., Heisen, B., Mesters, J.R., *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.
- Thiel, V., Ivanov, K.A., Putics, A., Hertzog, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E.J., Rabenau, H., Doerr, H.W., *et al.* (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol* 84, 2305–2315.
- Verschuere, K.H., Pumpor, K., Anemuller, S., Chen, S., Mesters, J. R., and Hilgenfeld, R. (2008). A structural view of the inactivation of the SARS coronavirus main proteinase by benzotriazole esters. *Chem Biol* 15, 597–606.
- Wei, P., Fan, K., Chen, H., Ma, L., Huang, C., Tan, L., Xi, D., Li, C., Liu, Y., Cao, A., *et al.* (2006). The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochem Biophys Res Commun* 339, 865–872.
- Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K., Chen, C., Jin, Y., Bartlam, M., and Rao, Z. (2007). Production of authentic SARS-CoV M<sup>pro</sup> with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J Mol Biol* 366, 965–975.
- Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q., *et al.* (2008). Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J Virol* 82, 2515–2527.
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., *et al.* (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A* 100, 13190–13195.
- Zhong, N., Zhang, S., Xue, F., Kang, X., Zou, P., Chen, J., Liang, C., Rao, Z., Jin, C., Lou, Z., *et al.* (2009). C-terminal domain of SARS-CoV main protease can form a 3D domain-swapped dimer. *Protein Sci* 18, 839–844.
- Zhong, N., Zhang, S., Zou, P., Chen, J., Kang, X., Li, Z., Liang, C., Jin, C., and Xia, B. (2008). Without its N-finger, the main protease of severe acute respiratory syndrome coronavirus can form a novel dimer through its C-terminal domain. *J Virol* 82, 4227–4234.
- Ziebuhr, J. (2004). Molecular biology of severe acute respiratory syndrome coronavirus. *Curr Opin Microbiol* 7, 412–419.