

STRUCTURE AND DYNAMICS OF SARS CORONAVIRUS MAIN PROTEINASE (M^{pro})

Rolf Hilgenfeld, Kanchan Anand, Jeroen R. Mesters, Zihe Rao, Xu Shen, Hualiang Jiang, Jinzhi Tan, and Koen H. G. Verschueren*

1. INTRODUCTION: A WORD ON NOMENCLATURE

All protein functions required for SARS coronavirus replication are encoded by the replicase gene.^{1,2} This gene encodes two overlapping polyproteins (pp1a and pp1ab), from which the functional proteins are released by extensive proteolytic processing. This is primarily achieved by the 34-kDa main proteinase (M^{pro}), which is frequently also called 3C-like proteinase (3CL^{pro}) to indicate a similarity in substrate specificity with the 3C proteinase of picornaviruses.³ While useful at the time of initial description of the coronaviral enzyme, there are in fact large differences between the structures and mechanisms of these enzymes, making the designation of the coronavirus main proteinase as 3CL^{pro} rather misleading. We will therefore use the term M^{pro} exclusively.

2. OVERALL STRUCTURE OF CORONAVIRUS MAIN PROTEINASE

The functional importance of the SARS-CoV M^{pro} in the viral life cycle makes it a preferred target for discovering anti-SARS drugs.⁴⁻⁷ However, in order to apply rational drug design or virtual screening, information on the structure of the target enzyme is required. Initially, this came from homology models of the SARS-CoV M^{pro} that were constructed on the basis of crystal structures of human CoV (HCoV) 229E M^{pro} and of porcine transmissible gastroenteritis virus (TGEV) M^{pro} that we had previously determined.^{4,8} The SARS virus enzyme shares about 40% sequence identity with these proteinases of group I coronaviruses. More recently, the crystal structure of the SARS-CoV M^{pro} has been determined.⁹⁻¹¹ As with other CoV M^{pro}s, the molecule comprises three domains (Figure 1). Domains I (residues 8–101) and II (residues 102–184) are β -

*Rolf Hilgenfeld, Kanchan Anand, Jeroen R. Mesters, Jinzhi Tan, Koen H. G. Verschueren, University of Lübeck, Lübeck, Germany, 23538. Zihe Rao, Tsinghua University, Beijing, China, 100084, and Chinese Academy of Sciences, Beijing, China, 100101. Xu Shen, Hualiang Jiang, Jinzhi Tan, Chinese Academy of Sciences, Shanghai, China, 201203.

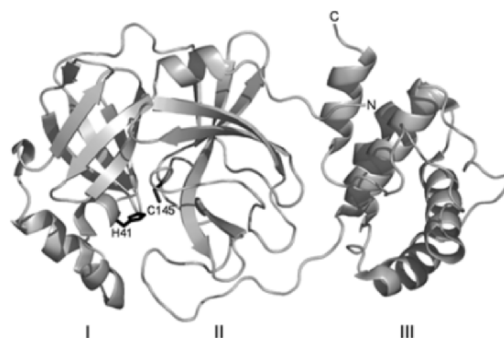


Figure 1. The three domains of the SARS-CoV M^{pro} monomer. The catalytic dyad (His41...Cys145, shown in black) is located in the interface between domains I and II.

barrels and together resemble the structure of chymotrypsin, whereas domain III (residues 201–306) consists mainly of α -helices. The active site, containing a Cys...His catalytic dyad, is located in a cleft between domains I and II. Domains II and III are connected by a long loop (residues 185–200). *In vitro* experiments demonstrated that deletion of domain III abolished almost completely the proteolytic activities of the main proteinases of TGEV and SARS-CoV.^{8,12} This domain is essential for the dimerization of the M^{pro},¹³ which in turn assures proper orientation of the N-terminal residues of monomer B that play an important role for the catalytic activity of monomer A (and *vice versa*; see below).^{9,10} In all known crystal structures of coronavirus main proteinases, the enzyme exists as a dimer,^{4,8-11} and dimerization is also observed in solution at slightly elevated concentrations.^{8,12-14} The dimer is the enzymatically active species because the specific activity increases linearly with increasing enzyme concentration.¹⁴

3. PLASTICITY OF THE SUBSTRATE-BINDING SITE

A special feature first discovered for the SARS-CoV M^{pro} (but most probably present in all coronavirus main proteinases) is that in the monoclinic crystals grown at pH 6.0, the two monomers have different conformations around the S1 substrate-binding site, because the loop 138–145, in particular Phe140, as well as Glu166 undergo dramatic conformational rearrangements. As a result, one protomer exists in an active and the other in an inactive conformation.⁹ In the latter, the S1 substrate-binding pocket has virtually collapsed as a consequence of the reorientation of Glu166, and the oxyanion hole no longer exists due to the conformational change of residues 138–145. When the crystals are equilibrated at pH 7.6 and 8.0, both monomers are in an active conformation.⁹ We have proposed^{9,10} that these conformational changes are controlled by the protonation state of His163, an absolutely conserved residue at the bottom of the S1 substrate-specificity pocket (Figure 2). This subsite is designed to accommodate the P1-glutamine residue of M^{pro} substrates, with high specificity. No other amino-acid side chain must be accepted in this position, in particular not glutamate (as opposed to glutamine). This is achieved by ensuring that over a broad pH range, His163 is uncharged. Two important interactions made by the imidazole ring are responsible for keeping it in the neutral state: (i), stacking or edge-on-face interaction with the phenyl ring of Phe140 (Figure 2, left

panel), and (ii), acceptance by its N δ 1 atom of a strong hydrogen bond from the hydroxyl group of the buried Tyr161. This means that only the N ϵ 2 atom of His163 can normally carry a proton, and it is this nitrogen that will donate a hydrogen bond to the side chain oxygen of the P1 glutamine residue of the substrate (Figure 2, right panel).⁸ In agreement with this structural interpretation, any replacement of the conserved histidine residue (His162 in this case) abolishes the proteolytic activity of HCoV 229E and feline infectious peritonitis virus (FIPV) M^{pro}.^{15,16} Furthermore, FIPV M^{pro} Tyr160 (corresponding to SARS-CoV M^{pro} Tyr161) mutants have their proteolytic activity reduced by a factor of > 30.¹⁵ These observations and the absolute conservation of these residues in the coronavirus main proteinases underline the importance of the uncharged state of His163 in binding the substrate with the required specificity.

However, when the SARS-CoV M^{pro} crystals are grown at a pH near or below the pK_a value of this histidine residue, the latter can be protonated, leading to drastic structural consequences. In order to compensate for a positive charge on His163, which is in a largely hydrophobic environment, Glu166, which forms part of the wall of the S1 pocket, will move inwards and form a salt-bridge with the protonated His163 (Figure 2, middle). Through this conformational rearrangement, Glu166 will fill the S1 pocket, thereby preventing the binding of substrate. But the consequences are even more far-reaching. When His163 is protonated, its hydrophobic interaction with the phenyl ring of Phe140 is no longer possible, and the latter undergoes a major displacement with an amplitude of > 5.5 Å (compare the middle panel with the two other images in Figure 2). Along with this, the oxyanion loop (residues 138–145) changes conformation and is no longer able to stabilize the tetrahedral intermediate of peptide-bond cleavage through donation of hydrogen bonds from the amide groups of Gly143 and Cys145.

In addition to His163, there is a second histidine residue involved in formation of the S1 pocket. In the crystal structure obtained at pH 7.6, where both subunits are found in the active conformation,⁹ His172 forms part of the wall of the subsite, being engaged in a salt-bridge with Glu166. At this pH, the histidine is likely to be positively charged, because pK_a values of histidine residues involved in salt-bridges tend to be 2 units higher than those of isolated histidines.¹⁷ The His172...Glu166 ion pair also exists in the active subunit of the dimer at pH 6.0, whereas in the inactive one, His172 loses its partner which moves into the S1 pocket, in order to compensate for the positive charge on the protonated His163 at the bottom of the subsite. Furthermore, at pH 8.0, His172 is likely to be uncharged. As a result, Glu166 is no longer fixed at its position in the wall of the S1 pocket, but tends to be flexible and partly blocks the entry to the S1 pocket.¹⁰ These observations nicely agree with the enzymatic activity of SARS-CoV M^{pro},^{10,12} which has its maximum at pH 7.0 and about 50% activity each at pH 6.0 (S1 pocket and oxyanion hole collapsed in one subunit of the dimer) and pH 8.0 (wall of S1 pocket no longer stable due to interruption of His172...Glu166 salt-bridge, Glu166 partly blocking the entry to the S1 site). Therefore, we propose that the bell-shaped pH-activity curve of the SARS-CoV M^{pro} is governed by the protonation of His163 on its low-pH side, and by deprotonation of His172 on its high-pH side.¹⁰ In nice agreement with this, Chou *et al.*¹² reported that the apparent pK_a values characteristic of this curve are 5.7 ± 0.4 and 8.7 ± 0.4. However, they discuss that these may originate from the Glu290...Arg4 ion pair, or from the His41...Cys145 catalytic dyad, both less likely options from our point of view.

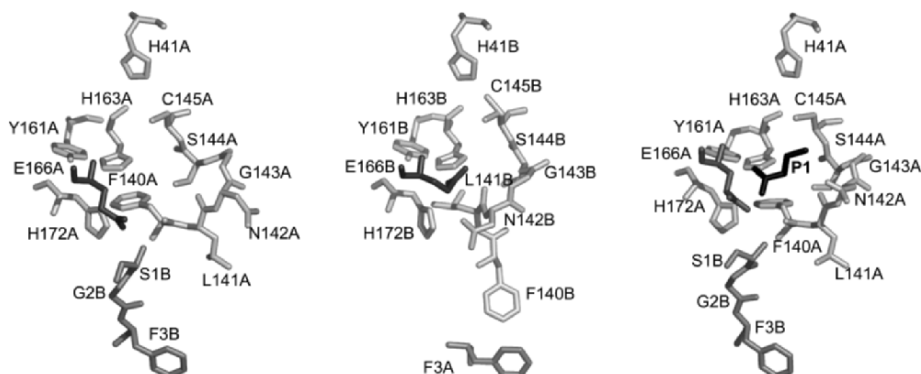


Figure 2. View of the S1 substrate binding pocket and oxyanion hole (loop 138–145, first two residues are not shown) in the active monomer (left), inactive monomer (middle) and as seen in the active monomer of a substrate-analogous inhibitor-bound SARS-CoV M^{pro} (right panel, only the P1 Gln residue is shown in black, occupying the S1 specificity pocket). The N-terminal residues of the other monomer are indicated in dark gray. The alternating position of the Glu166 side chain in the active (left and right panel) and inactive conformation (middle panel) is highlighted in darker gray.

4. MOLECULAR DYNAMICS SIMULATIONS

We have confirmed by molecular dynamics (MD) simulations that large conformational changes of the observed type can indeed be triggered by protonation of His163 and deprotonation of His172.¹⁰ In three different 10-ns simulations at pH 6.0, 7.6, and 8.0, we found the same type of rearrangements as seen in the crystals. This is reassuring, because it should not be forgotten that unless the crystals diffract to better than 1.0 Å resolution, X-ray diffraction cannot normally determine hydrogen positions because of their low scattering power. The force fields used in MD simulations, on the other hand, fully take into account the hydrogens bound to non-carbon atoms. Thus, because these simulations yield the same conformational rearrangements as those seen by X-ray diffraction, the interpretation of the crystallographic results is likely to be correct. Also, our preliminary NMR data with ¹⁵N-labeled SARS-CoV M^{pro} (J. George *et al.*, unpublished) appear to support these conclusions.

In addition to the MD simulations, we also investigated the dynamic behavior of the system with His163 in both subunits of the dimer protonated, i.e., at a presumed pH < 6.0. Starting from the (energy-minimized) crystal structure at pH 6.0, this simulation led to a dimer that had both monomers in the inactive conformation, with their S1 pockets and oxyanion holes collapsed. Thus, the MD simulations were apparently able to transform one conformation of the substrate-binding site of SARS-CoV M^{pro} into the other.¹⁰

5. NEW CRYSTAL FORMS OF SARS-CoV M^{pro}

Experimental support for this theoretical prediction was provided very recently by the analysis of two new crystal forms of the SARS-CoV M^{pro}. In addition to the monoclinic crystal form of the enzyme originally described in 2003,⁹ we managed to

obtain tetragonal (space group P4₃2₁2) and orthorhombic (P2₁2₁2) crystals of the enzyme, and determined these crystal structures at 2.0 and 2.8 Å resolution.¹⁰ Both of these crystal forms contain one SARS-CoV M^{PRO} monomer per asymmetric unit; the dimer is generated through the crystallographic twofold axis. Thus, by necessity, the two monomers of the dimer are identical in these crystals. The question, however, was whether they would be in the active or in the inactive conformation. Because the monoclinic form had one monomer in the active and the other in the inactive form when crystallized at pH 6.0, but both of them in the active conformation after equilibration of the crystals at pH 7.6, it was unclear what to expect for the new crystal forms which were crystallized at pH 5.9 (tetragonal form) and 6.6 (orthorhombic). Interestingly, it turned out that both of them are in the inactive form, *i.e.* the S1 pocket and the oxyanion hole have collapsed.

We have also attempted to equilibrate tetragonal crystals at higher pH values. Structures were determined at pH 7.0 (1.65 Å resolution), 7.4 (1.57 Å), and 8.0 (1.65 Å). Although these are the highest-resolution structures reported so far for the SARS-CoV M^{PRO}, a reliable interpretation of the electron density in the substrate-binding region was difficult because of dual (or even multiple) conformations. In any case, our preliminary analysis shows that the same type of conformational rearrangements occur in this crystal form as was observed in the monoclinic crystals, but an ensemble of inactive and active conformations appear to co-exist at all pH values. In fact, the tetragonal crystals of SARS-CoV M^{PRO}, containing < 30% solvent, may be less suited for studying the conformational transition because the molecules are tightly packed and seem to have a contracted substrate-binding site. This is supported by the observation that the tetragonal crystals crack when equilibrated at pH 8.0 for longer than 4 hours.

Interestingly, the orthorhombic crystal form, which has the (generated) dimer in the inactive conformation when crystallized at pH 6.6 in the presence of malonate, can also be obtained in an active form at about the same pH (6.5) when ammonium sulfate is used as a precipitant (J. Lescar, personal communication). At present, it is unclear why the change of precipitant should induce such changes; more likely, it is subtle differences in the final pH of the crystallization medium that are determining the resulting conformation when working near the pK_a value of His163.

Apart from the differences in detail in the substrate-binding site, the SARS-CoV M^{PRO} dimers as seen in the new structures are very similar to the dimer in the original monoclinic crystals. From the monoclinic crystal structure obtained at pH 6.0, the monomers in the new crystal forms display overall r.m.s. deviations for C α atoms (monomers A and B, respectively) of 0.95/0.76 Å (tetragonal form) and 1.10/0.78 Å (orthorhombic form). It is reassuring that the monomers in the new crystal forms, which are in the inactive conformation, are more similar to the inactive monomer B of the dimer in the monoclinic form (second number), than to the active monomer A (first number).

6. THE ROLE OF THE “N-FINGER”

One important intermolecular interaction in the coronavirus M^{PRO} dimer was not mentioned so far. In the active conformation of the SARS-CoV M^{PRO} dimer, the N-terminus of monomer B (which is the inactive subunit at pH 6.0) was shown to interact with the main chain amide and carbonyl of Phe140 and with the carboxylate of Glu166, both of monomer A.⁹ This interaction appears to help shape the substrate-binding site of

monomer A. On the other hand, the collapsed binding site of monomer B lacks these intermolecular interactions, and as a result, residues 1 and 2 of monomer A are disordered and not seen in the electron density. The same observation was made in our X-ray structures derived from the tetragonal and orthorhombic crystal forms¹⁰: here, both monomers are in the inactive conformation, and accordingly, both N-termini are disordered to the extent that no electron density is seen for residue Ser1. When we determined the first structure of a coronaviral M^{pro}, that of the TGEV enzyme,⁸ we saw that residues 1–7 were squeezed in between domain III of its own monomer and domain II of the other monomer in the dimer. The same interactions of this segment, which we later called the “N-finger,”⁹ were seen in HCoV 229E M^{pro} and SARS-CoV M^{pro}.^{4,9} When we deleted residues 1–5 in the TGEV proteinase, the enzyme was almost totally inactive with a synthetic pentadecapeptide as substrate.⁸ In agreement with this, Chen *et al.*¹⁸ have recently shown for SARS-CoV M^{pro} that the enzyme lacking residues 1–7 is proteolytically inactive, but the rather surprising finding was that it still forms a dimer.¹⁸ This result has to be seen in light of the finding by Shi *et al.* that even isolated domain III of SARS-CoV M^{pro} will dimerize.¹³ On the other hand, when Hsu *et al.* removed only residues 1–3 from the enzyme, it retained 76% of its proteolytic activity.¹⁹ Further, in apparent contradiction to the findings by Chen *et al.*,¹⁸ they found that the $\Delta(1-4)$ SARS-CoV M^{pro} was predominantly monomeric at concentrations of 0.1 mg/ml. Whatever the reason for the discrepancy might be, it could well be that while the presence and correct placement of the “N-finger” is important, the tip of the finger may not be as essential as thought hitherto. In fact, our MD simulations as well as the X-ray structure obtained from the monoclinic form at pH 7.6 also suggested that the active conformation of monomer A can also be retained without direct interaction with residue Ser1 of monomer B, provided residues 3–7 are in the correct position.

7. CONCLUSIONS AND OUTLOOK

In summary, from the various crystallographic studies on the coronavirus M^{pro} and from our MD simulations, we conclude that the enzyme is a very flexible protein, the conformational state of which appears to depend on the pH value of the medium. This is probably of biological significance, because the viral polyproteins (of which the M^{pro} is a domain before self-activation by autocleavage) assemble on the late endosome where local pH tends to be acidic. For designing inhibitors of the M^{pro}, knowledge of the dynamics of the target will be essential. This has been clearly demonstrated in the case of HIV-1 proteinase over the years, where understanding the flexibility of the enzyme did not only turn out to be a prerequisite for designing potent inhibitors but also for explaining many of the observed drug-resistance mutations. Today, HIV proteinase is perhaps the one enzyme best understood in terms of structure and dynamics, and this knowledge is mainly based on several hundred crystal structures, most of them complexes with various inhibitors. In terms of number of crystal structures, HIV-1 proteinase is probably followed by trypsin and lysozyme, but extrapolating from the current research activities, coronavirus main proteinases are likely to catch up.

The work described here was supported, in part, by the Sino-European Project on SARS Diagnostics and Antivirals (SEPSDA, contract no. SP22-CT-2004-003831; www.sepsda.info) of the European Commission, by the Deutsche Forschungsgemeinschaft

(Hi 611/4-1) and by the Sino-German Center for the Promotion of Science (GZ 233-202/6).

8. REFERENCES

1. D. A. Groneberg, R. Hilgenfeld, and P. Zabel, Molecular mechanisms of severe acute respiratory syndrome (SARS), *Respir. Res.* **6**, 8-23 (2005).
2. V. Thiel, K. A. Ivanov, A. Putics, et al., Mechanisms and enzymes involved in SARS coronavirus genome expression, *J. Gen. Virol.* **84**, 2305-2315 (2003).
3. A. E. Gorbalenya, E. V. Koonin, A. P. Donchenko, and V. M. Blinov, Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid analysis, *Nucl. Acids Res.* **17**, 4847-4861 (1989).
4. K. Anand, J. Ziebuhr, P. Wadhvani, J. R. Mesters, and R. Hilgenfeld, Coronavirus main proteinase (3CL^{pro}) structure: basis for design of anti-SARS drugs, *Science* **300**, 1763-1767 (2003).
5. B. Xiong, C. S. Gui, X. Y. Xu, et al., A 3D model of SARS-CoV 3CL proteinase and its inhibitors design by virtual screening, *Acta Pharmacol. Sin.* **24**, 497-504 (2003).
6. K. Anand, H. Yang, M. Bartlam, Z. Rao, and R. Hilgenfeld, in: *Coronaviruses with Special Emphasis on First Insights Concerning SARS*, edited by A. Schmidt, M. H. Wolff, and O. Weber (Birkhäuser, Basel, 2005), pp. 173-199.
7. H. Yang, W. Xie, X. Xue, et al., Design of wide-spectrum inhibitors targeting coronavirus main proteinases, *PLoS Biology* **3**, e324 (2005).
8. K. Anand, G. J. Palm, J. R. Mesters, et al., Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain, *EMBO J.* **21**, 3213-3224 (2002).
9. H. Yang, M. Yang, Y. Ding, et al., 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 (2003).
10. J. Tan, K. H. G. Verschuere, K. Anand, et al., pH-Dependent conformational flexibility of the SARS-CoV main proteinase (M^{pro}) dimer: Molecular dynamics simulations and multiple X-ray structure analyses, *J. Mol. Biol.* **354**, 25-40 (2005).
11. M.-F. Hsu, C.-J. Kuo, K.-T. Chang, et al., Mechanism of the maturation process of SARS-CoV 3CL protease, *J. Biol. Chem.* **280**, 31257-31266 (2005).
12. C.-Y. Chou, H.-C. Chang, W.-C. Hsu, et al., Quarternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease, *Biochemistry* **43**, 14958-14970 (2004).
13. J. Shi, Z. Wei, and J. Song, 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 (2004).
14. K. Fan, P. Wei, Q. Feng, et al., Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase, *J. Biol. Chem.* **279**, 1637-1642 (2004).
15. A. Hegyi, A. Friebe, A. E. Gorbalenya, and J. Ziebuhr, Mutational analysis of the active centre of coronavirus 3C-like proteases, *J. Gen. Virol.* **83**, 581-593 (2002).
16. J. Ziebuhr, E. J. Snijder, and A. E. Gorbalenya, Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* **81**, 853-879 (2000).
17. J. Yang, M. Yu, Y. N. Jan, and L. Y. Jan, Stabilization of ion selectivity filter by pore loop ion pairs in an inwardly rectifying potassium channel, *Proc. Natl. Acad. Sci. USA* **94**, 1568-1572 (1997).
18. S. Chen, L. Chen, J. Tan, et al., 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 (2005).
19. W.-C. Hsu, H.-C. Chang, C.-Y. Chou, P.-J. Tsai, P.-I. Lin, and G.-G. Chang, 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 (2005).