

Coronavirus main proteinase: target for antiviral drug therapy

Kanchan Anand¹, Haitao Yang², Mark Bartlam², Zihe Rao^{2,3} and Rolf Hilgenfeld¹

¹ *Institute of Biochemistry, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany*

² *Laboratory of Structural Biology, Tsinghua University, Beijing, China*

³ *Institute of Biophysics, Chinese Academy of Science, Beijing, China*

Introduction

During the first outbreak of severe acute respiratory syndrome (SARS) in early 2003, almost 8500 people worldwide were diagnosed with the disease, and approximately 800 died of it. The epidemic was contained by July, 2003, but since then, the disease has re-emerged in the Guangdong province in Southern China around the turn of the year 2003, although the number of cases was very limited this time. In addition, there were two cases of SARS caused by improper sample handling in research laboratories in Singapore and Taiwan. A “mini-outbreak” originated from several independent laboratory infections at a research laboratory in Beijing in March and April, 2004, and led to infection chains encompassing three generations. This time, China escaped another epidemic since fortunately, the number of diseased people was limited to nine, although there was one fatality. In any case, these isolated mini-outbreaks remind us of the global threat that SARS still represents.

In late March, 2003, a new coronavirus, SARS-CoV, was isolated from affected tissue of SARS patients and identified as the cause of the disease [1, 2]. *Coronaviridae* are RNA viruses that have been united with the families *Arteriviridae* and *Roniviridae* in the order *Nidovirales*. The phylogenetic relationship of these morphologically unrelated virus families is based on their similar polycistronic genome organization, common transcriptional and post-translational strategies, and a conserved array of homologous domains in the viral polyproteins [3–5]. Coronaviruses are responsible for a number of diseases of man and animals, some of which have major economical impacts, although the human coronaviruses have clearly been neglected before SARS emerged. Transmissible gastroenteritis virus (TGEV) infection is associated with severe and often fatal diarrhoea in young pigs [6, 7]. The human coronaviruses (HCoV) 229E and OC43 are major causes

of upper respiratory tract illness in humans, in particular, common cold-like conditions [8]. To date, only the 229E strain of HCoV has been characterized in detail because until the discovery of SARS-CoV, it used to be the only isolate to grow efficiently in cell culture. In contrast to the relatively harmless disease caused by these human coronaviruses, SARS is characterized by high fever, malaise, rigor, headache, non-productive cough, or dyspnea and may progress to generalized, interstitial infiltrates in the lung, requiring intubation and mechanical ventilation [9]. The infection, however, is not restricted to the lungs. Other affected organs include the kidneys and the liver, as well as the gastrointestinal tract; hence SARS can be considered a systemic disease. Epidemiological evidence suggests that the transmission of this newly emerging pathogen occurs mainly by face-to-face contact, although other routes of transmission cannot be fully excluded.

Coronavirus main proteinase: function and overall structure

Coronaviruses are positive-stranded RNA viruses featuring a large (27–31 kB), single-stranded RNA genome. The replicase gene, encompassing more than 20,000 nucleotides, encodes two overlapping polyproteins, pp1a (replicase 1a, ~450 kDa) and pp1ab (replicase 1ab, ~750 kDa) [10] that mediate all the functions required for viral replication and transcription [11]. Expression of the 3'-proximal portion of the replicase gene requires (-1) ribosomal frameshifting [10]. The functional non-structural proteins are released from the viral polyproteins by extensive proteolytic processing. The vast majority of cleavages are mediated by the viral main proteinase, M^{Pro} [12], a cysteine proteinase which is frequently also called 3C-like proteinase (3CL^{Pro}) to indicate a similarity of its substrate specificity to that observed for picornavirus 3C proteinases (3C^{Pro}) (see Tab. 1), although the structural similarities between the two families of proteinases are limited [13]. The M^{Pro} (3CL^{Pro}) cleaves the polyprotein at no less than 11 conserved sites involving Leu-Gln↓(Ser,Ala,Gly) sequences, a process initiated by the enzyme's own autolytic cleavage from pp1a and pp1ab [14, 15]. Importantly, this cleavage pattern appears to be conserved in the M^{Pro} from SARS coronavirus (SARS-CoV), as deduced from the genomic sequence [16, 17] and shown experimentally for one cleavage site [18]. The SARS-CoV polyproteins have three non-canonical M^{Pro} cleavage sites with Phe, Met or Val in the P2 position, but the same cleavage sites are unusual in other coronaviruses as well. The functional importance of M^{Pro} in the viral life cycle makes this proteinase an attractive target for the development of drugs directed against SARS and other coronavirus infections.

Our laboratories have determined crystal structures for HCoV strain 229E M^{Pro} at 2.54 Å resolution [18], porcine TGEV M^{Pro} at 1.96 Å [13], and SARS-CoV M^{Pro} (1.9 Å resolution) [19]. In addition, we have also elucidated the structures of substrate-analogous hexapeptidyl chloromethyl

Table 1: Autocleavage sites of coronavirus M^{Pro} and picornavirus 3C^{Pro}

	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'
HCoV M ^{Pro}										
N-terminal	Y	G	S	T	L	Q	A	G	L	R
C-terminal	F	G	V	N	L	Q	S	G	K	T
TGEV M ^{Pro}										
N-terminal	V	N	S	T	L	Q	S	G	L	R
C-terminal	Y	G	V	N	L	Q	A	G	K	V
MHV M ^{Pro}										
N-terminal	T	T	S	F	L	Q	S	G	I	V
C-terminal	A	G	V	K	L	Q	S	K	R	V
SARS-CoV M ^{Pro}										
N-terminal	T	S	A	V	L	Q	S	G	F	R
C-terminal	S	G	V	T	F	Q	G	K	F	I
HRV-2 3C ^{Pro}										
N-terminal	L	P	V	V	V	Q	G	P	N	T
C-terminal	Y	F	T	D	V	Q	G	Q	I	T
HAV 3C ^{Pro}										
N-terminal	K	G	L	F	S	Q	A	K	I	S
C-terminal	M	E	L	R	T	Q	S	F	S	N
PV 3C ^{Pro}										
N-terminal	E	T	A	K	V	Q	G	P	G	F
C-terminal	G	A	V	T	E	Q	G	Y	L	N
Consensus										
Coronavirus M ^{Pro}	-	-	small	-	L	Q	small	-	-	-
Picornavirus 3C ^{Pro}	-	-	φ	-	small	Q	(G)	-	-	-

Cleavage site specificities of coronavirus M^{Pro}s and picornavirus 3C^{Pro}s are similar but different in detail. N- and C-terminal autocleavage sites of M^{Pro}s and 3C^{Pro}s are shown. HCoV, human coronavirus 229E; TGEV, porcine transmissible gastroenteritis virus; MHV, mouse hepatitis virus; SARS-CoV, SARS coronavirus; HRV 2, human rhinovirus serotype 2; HAV, hepatitis A virus; PV, poliovirus type 1 (strain Mahoney); φ, hydrophobic

ketone inhibitor complexes of the TGEV and SARS-CoV enzymes, at around 2.4 Å resolution [18, 19]. The structures of all three enzymes show that the coronavirus main proteinase consists of three domains, the first two of which together distantly resemble chymotrypsin while the third has a unique α -helical fold (Fig. 1). Located between domains I and II, the active site comprises a Cys-His catalytic dyad, rather than the usual triad. Domains I and II (residues 8-99 and 100-183, respectively) are six-stranded antiparallel β -barrels. A long loop (residues 184 to 199) connects domain II to the C-terminal domain (domain III, residues 200-300), which is a globular cluster of five, mostly antiparallel helices.

The interior of the β -barrel of domain I consists entirely of hydrophobic residues. A short α -helix (helix A; residues 53-58) closes the barrel like a

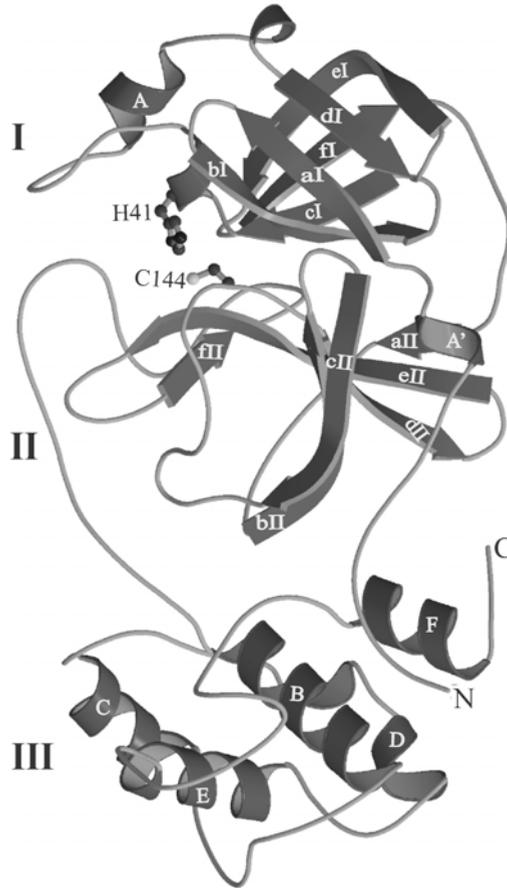


Figure 1. Domain structure of the coronavirus main proteinase monomer (shown here is TGEV M^{pro}).

lid. Domain II is smaller than domain I and also smaller than the homologous domain II of chymotrypsin and the 3C proteinase (3C^{pro}) of hepatitis A virus (HAV) [20–22]. Several secondary structure elements of this domain in HAV 3C^{pro} (strands b2II and cII and the intervening loop) are missing in the coronavirus M^{pro}. The domain II barrel of the M^{pro} is far from perfect (Fig. 1). The segment from Gly135 to Ser146 forms a part of the barrel, even though it consists mostly of consecutive loops and turns. In fact, in contrast to domain I, a structural alignment of domain II onto the picornaviral 3C proteinases has proven difficult. The superposition of domains I and II of the TGEV M^{pro} onto those of the HAV 3C^{pro} yields a root mean square deviation (r.m.s.d.) of 1.85 Å for 114 equivalent (out of 184 compared) C_α pairs, while domain II alone displays an r.m.s.d. of 3.25 Å for 57 (out of 85) C_α pairs.

The three M^{pro} structures determined so far are quite similar to one another. The overall r.m.s. deviations for C α atoms between the structures is below 2 Å for all 300 C α positions of the molecules that could be located in the electron density maps but the isolated domains exhibit significantly smaller deviations. The most variable domain appears to be the helical domain III, which shows a better spatial overlap between TGEV M^{pro} and HCoV M^{pro} than between each of these and SARS-CoV M^{pro}. To a certain extent, these numbers reflect the degree of sequence identity between the different M^{pro}s. HCoV 229E and TGEV, both being group I coronaviruses, share 61% sequence identity between their main proteinases, whereas SARS-CoV as a distant outlier of group II has 40% and 44%, respectively, of the sequence of its M^{pro} in common with HCoV 229E and TGEV [18]. Interestingly, domains I and II show a higher degree of sequence conservation (42–48% identity) than domain III (36–40%) between SARS-CoV M^{pro} and the coronavirus group I enzymes.

Dimer formation

The crystal structure of the TGEV M^{pro} provided the first indication that coronavirus main proteinases form tight dimers [13]. This has since been confirmed by the X-ray structures of the HCoV 229E and SARS-CoV M^{pro}s (Fig. 2) [18, 19]. Dimerisation also occurs in solution: we could show by dynamic light scattering that both HCoV 229E and TGEV M^{pro} exist as a mixture of monomers (~65%) and dimers (~35%) in diluted solutions (1–2 mg proteinase/ml). Lai and coworkers have demonstrated that the enzymatic activity of the SARS-CoV M^{pro} increases linearly with enzyme concentration, suggesting that the enzymatically active species is the dimer and not the monomer [23]. Crystal structures of the SARS-CoV enzyme at different pH values provided an explanation for this observation [19] (see below).

It should be noted that dimer formation as a prerequisite for enzymatic activity makes coronavirus main proteinases unique amongst the proteinases of RNA viruses, and present a clear distinction from the 3C proteinases of picornaviruses. The TGEV M^{pro} X-ray structure showed, and the other coronavirus M^{pro} structures confirmed, that the monomer-monomer contact interface comprises two regions, i.e. i) the N-terminal eight residues of molecule A and domain II of molecule B, and ii) the domains III of each protomer. Although the latter part of the interface is significantly smaller than the former, it is apparently sufficient to hold the isolated domains III together in solution, as has recently been shown for SARS-CoV M^{pro} [24]. The crystal structures show that the two molecules are oriented perpendicular to one another (Fig. 2) [13, 18, 19]. Since the architecture of the dimers including most details of intermolecular interaction are the same in TGEV M^{pro} – three independent dimers per asymmetric unit – and HCoV 229E as

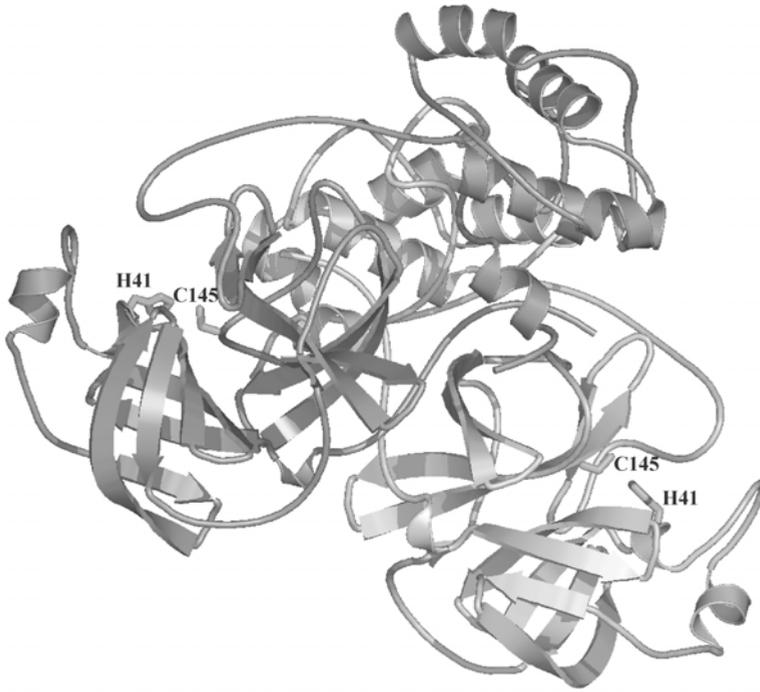


Figure 2. The SARS-CoV main protease dimer.

well as SARS-CoV M^{pro}s – both with one dimer per asymmetric unit –, i.e., in completely different crystalline environments, there can be little doubt that dimer formation is of biological relevance in these enzymes.

In the M^{pro} dimer, the N-terminal amino acid residues are squeezed in between domains II and III of the parent monomer – designated “A” in this paragraph – and domain II of the other monomer (“B”), where they make a number of very specific interactions that appear tailor-made to bind this segment with high affinity. These interactions include *intermolecular* salt bridges between the positively charged amino terminus (residue A1) and GluB165 (all residue numbers given in this section are for TGEV M^{pro}) of domain II of the neighboring molecule in the dimer, between ArgA4 and GluB286, as well as an *intramolecular* salt bridge between LysA5 and GluA291 of domain III of the parent molecule. Also, there is an important hydrophobic contact between the semiconserved LeuA3 and a hydrophobic patch on domain III which includes PheA206, AlaA209, PheA287, ValA292, the C^β atom of GlnA295, and MetA296. All sequenced members of the coronavirus proteinase family have a hydrophobic residue in position 3 (SARS-CoV has Phe), while glycine is absolutely conserved in position 2.

The latter residue adopts the α_L conformation, which is easily accessible only to glycine. This conformation ensures that the N-terminal segment fits into its narrow binding canyon in the dimerization interface.

The existence of a specific binding site for the N-terminal segment of the M^{pro} could be important in the activation of the main proteinase by catalyzing its release from the viral polyprotein. The free N-terminus is created as a consequence of the first cleavage reaction performed by the proteinase, at its N-terminal autocleavage site. In the mature M^{pro} monomer, the first eight residues from the N-terminus are not a part of the globular protein structure. After the autocleavage reaction, they could easily remain bound to the substrate-binding site of the proteinase if an alternative high-affinity binding site was not provided by dimer formation. The existence of such high-affinity binding site ensures that the catalytic site is immediately liberated from the product and can act on other cleavage sites in the polyprotein. However, the exact placement of the amino terminus also has a very important structural role for the *mature* M^{pro} , and is in fact critical for its catalytic activity, as we will see below.

Catalytic site

In the active site of the M^{pro} , Cys144(145) and His41 (SARS-CoV numbering in brackets where different) form a catalytic dyad. In contrast to serine proteinases and other cysteine proteinases, which have a catalytic triad, there is no third catalytic residue present. HCoV 229E and TGEV M^{pro} s have Val84 in the corresponding sequence position (Cys in SARS-CoV M^{pro}), with its side chain pointing away from the active site. A buried water molecule is found in the place that would normally be occupied by the third member of the triad; this water is hydrogen-bonded to His41 N δ 1, Gln163 N ϵ 2, and Asp186 O δ 1 in HCoV 229E M^{pro} (His, His, and Asp in both SARS-CoV and TGEV M^{pro}). Asp186(187) does not act on the active-site His41 through the mentioned water molecule, but rather interacts with conserved Arg40 by forming a strong inter-domain salt bridge.

There has been a long-standing debate on the exact nature of the catalytic centre in cysteine proteinases. Polgár has shown that in papain-like cysteine proteinases, the active species is a thiolate-imidazolium ion pair, with the imidazolium cation held in the correct position by an asparagine residue, the third residue of the catalytic triad [25]. However, such a mechanism does not seem to be predominant in the coronavirus main proteinases; Lai and coworkers have recently shown that if the active-site cysteine residue of SARS-CoV M^{pro} is replaced by serine, some residual enzymatic activity – decreased by a factor of 40 – can still be detected. This can be seen as an indication for the existence of a classical general base mechanism in the wild-type proteinase, involving the non-ionized Cys145 hydrogen bonded to the uncharged imidazole of His41 in the ground state of the free

enzyme [26]. We note that in viral cysteine proteinases, both picornaviral 2A and 3C proteinases as well as coronaviral main proteinases, the S_{γ} of the active-site cysteine residue is exactly in the plane of the histidine imidazole ring, at a distance of between 3.5 and 4.0 Å, while in the papain-like cysteine proteinases, this is not the case. For hydrogen bonding to occur between the two side chains, but not for ion-pair formation, the sulfur has to be more or less in the plane of the imidazole. While this preliminary observation has to be investigated further, it seems that X-ray crystallography can contribute to this important mechanistic question by providing exact geometric details of the ground-states of the free enzymes.

Substrate binding

Comparison of the structures shows that the substrate-binding sites are well conserved among coronavirus main proteinases. This is supported by our experimental finding that recombinant SARS-CoV M^{pro} cleaves a pentadecapeptide corresponding to the N-terminal autocleavage site of the TGEV M^{pro} [18].

To allow structure-based design of drugs directed at coronavirus M^{pro}s, we sought to determine the exact binding mode of M^{pro} substrates. To this end, we synthesized the substrate analog hexapeptidyl chloromethyl ketone inhibitor, Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK (subsequently named CMK), and soaked it into crystals of TGEV and SARS-CoV M^{pro} [18, 19]. The sequence of the inhibitor was derived from the P6–P1 residues of the N-terminal autoprocessing site of TGEV M^{pro} (SARS-CoV M^{pro} and HCoV 229E M^{pro} have Thr-Ser-Ala-Val-Leu-Gln and Tyr-Gly-Ser-Thr-Leu-Gln, respectively, at the corresponding positions). At the outset, we will present here the complex between the TGEV M^{pro} and the inhibitor [18]. In the case of the SARS-CoV enzyme, some unusual binding modes were observed [19], which we will discuss later.

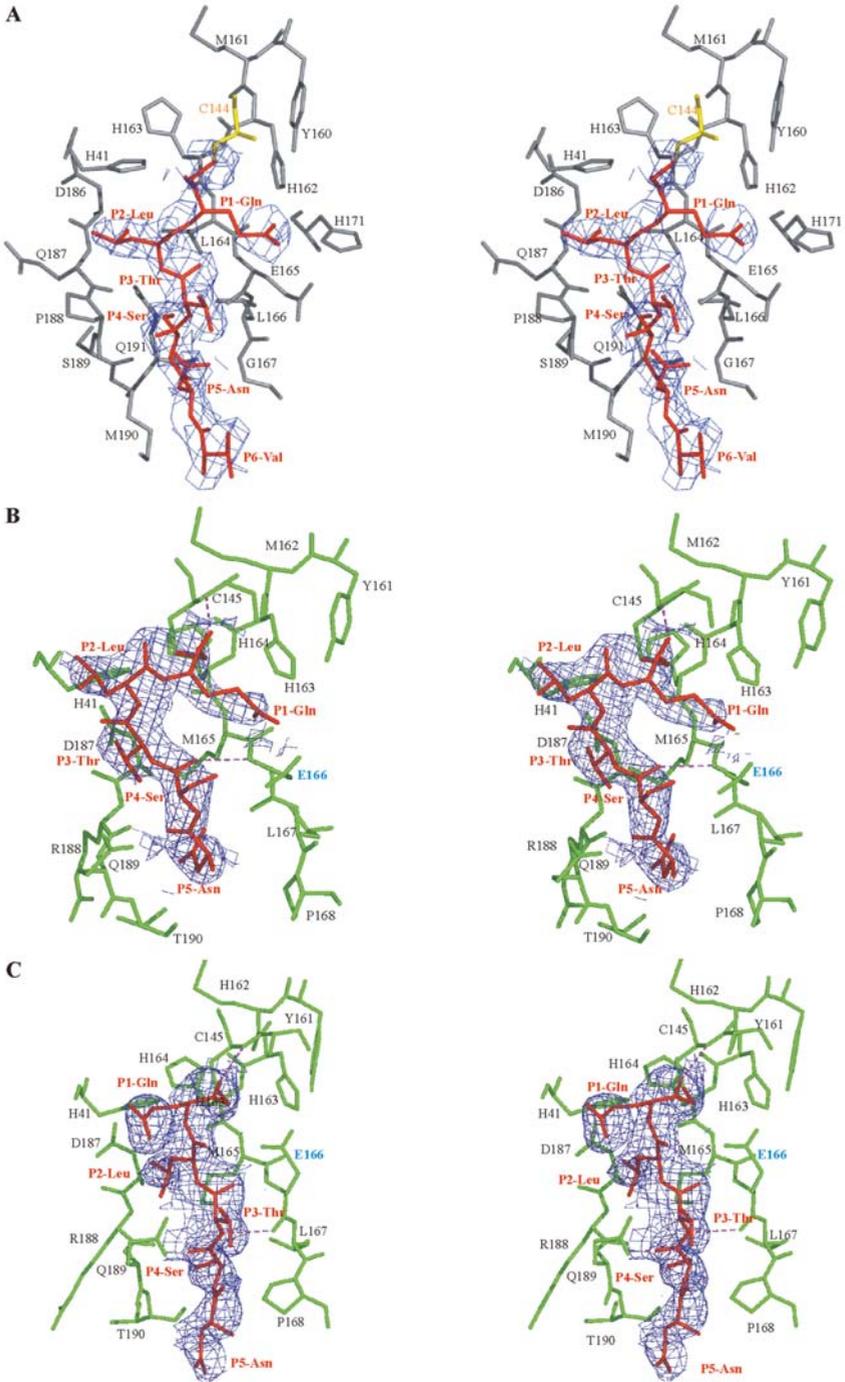
X-ray crystallographic analysis of the TGEV M^{pro} complex at 2.37 Å resolution revealed difference density for all residues (except the benzoyloxycarbonyl (Cbz) protective group) of the inhibitor, in two (B and F) out of the six TGEV M^{pro} monomers in the asymmetric unit [18]. In these monomers, there is a covalent bond between the S_{γ} atom of the active-site nucleophile Cys144 and the methylene group of the chloromethyl ketone. There are no significant differences between the structures of the enzyme in the free and in the complexed state. The substrate-analogue inhibitor binds in the shallow substrate-binding site at the surface of the proteinase, between domains I and II (Fig. 3A). The residues Val-Asn-Ser-Thr-Leu-Gln occupy, and thereby define, the subsites S6 to S1 of the proteinase. Residues P5 to P3 form an antiparallel β -sheet with segment 164–167 of the long strand eII on one side, and they also interact with segment 189–191 of the loop linking domains II and III on the other. In the discussion that follows,

residue numbers for TGEV M^{pro} are followed by those for the SARS-CoV enzyme in brackets.

S1 specificity site

In coronavirus M^{pro} polyprotein cleavage sites, the P1 position is invariably occupied by Gln. At the very bottom of the M^{pro} S1 subsite, the imidazole of His162(163) is suitably positioned to donate a hydrogen bond from N ϵ 2 to the side-chain carbonyl of the P1 glutamine (Fig. 3A). In order to specifically bind glutamine and not glutamate in the S1 pocket, this histidine residue is required to be uncharged over a broad pH range. This appears to be maintained by two important interactions made by the imidazole ring: i) stacking onto the phenyl ring of Phe139(140), and ii) accepting a hydrogen bond (to N δ 1) from the hydroxyl group of the buried Tyr160(161). In agreement with this structural interpretation, any replacement of His162 completely abolishes the proteolytic activity of HCoV 229E and feline infectious peritonitis virus (FIPV) M^{pro} [27, 28]. Furthermore, FIPV M^{pro} Tyr160 mutants have their proteolytic activity reduced by a factor of >30 [27]. All of these residues are conserved in SARS-CoV M^{pro} and, in fact, in all coronavirus main proteinases. Other elements involved in the S1 pocket of the M^{pro} are the *main-chain* atoms of Ile51, Leu164(Met165), Glu165(166), and His171(172). The latter two residues form a salt bridge at the outer wall of the S1 specificity pocket in the TGEV and HCoV 229E M^{pro}s, as well as in one conformation of the SARS-CoV M^{pro}. However, the latter displays another conformation which will be discussed below.

In all structures of free coronavirus main proteinases that we determined so far, we find electron density at the bottom of the S1 pocket. We interpreted this as a water molecule or, in the case of TGEV M^{pro} where the density was more dumb-bell-shaped, as two water molecules. However, especially in the latter case, we realized that apart from a possible hydrogen bond between one of these waters and the His162 imidazole, no interactions with the proteins were made by these water molecules. Also, the positions of the two water molecules were in fact too close to one another (about 2.1–2.3 Å) [13]. When we subsequently determined the structure of the complex between the TGEV M^{pro} and the CMK inhibitor, we noted that the oxygen and the nitrogen of the side-chain amide occupied a position exactly matching that of the two water molecules in the free enzyme. It is therefore possible that the free enzyme in fact also contains some residual peptide cleavage product (perhaps from an unidentified *E. coli* protein). Having realized this possibility, we checked for residual difference electron density in other subsites but could find nothing, with the exception of a patch of relatively strong electron density at the P2 position which we had interpreted and successfully refined as a 2-methyl-2,4-pentandiol (MPD) molecule from the crystallization medium [13]. The MPD does not



fully protrude into the S2 pocket but rather occupies a position corresponding to the main chain of the peptide substrate. At this point, we still think the assignment of this density was correct, but studies on other proteinases – e.g. thermolysin – that revealed the presence of cleavage product after years of investigation taught us to be cautious.

S2 specificity site

Apart from a few exceptions, coronavirus M^{pro} cleavage sites have a Leu residue in the P2 position [28]. The hydrophobic S2 subsite of the TGEV enzyme is formed by the side chains of Leu164, Ile51, Thr47, His41, and Tyr53. In addition, residues 186–188 line the S2 subsite with some of their main-chain atoms. The P2 Leu side chain of the inhibitor is well accommodated in this pocket (Fig. 3A).

In the SARS-CoV M^{pro} [19], the S2 pocket is similar to what has been seen in HCoV M^{pro} [18] and TGEV M^{pro} [13]. One wall of the subsite is formed by methionine residues 49 and 165, which adopt positions spatially similar to TGEV and HCoV 229E M^{pro} residues Ile51 and Leu/Ile164, respectively. The longer methionine side chains restrict the S2 pocket in one direction and probably require a slight reorientation of the P2 leucine residue of the substrate, compared to TGEV M^{pro}. On the other side, the S2 pocket appears to be somewhat enlarged in SARS M^{pro}, due to the insertion of Ala46 which is not present in TGEV or HCoV 229E M^{pro}, as well as to many differences in amino acid sequence. In SARS-CoV M^{pro}, the stretch 45–50 forms a short 3_{10} -helix, whereas this region has no regular secondary structure in the other coronavirus main proteinases. The net effect of these modifications appears to be a widening of the S2 pocket on this side. This may be required for the efficient binding of a P2 phenylalanine, which is present in the C-terminal autocleavage site of the M^{pro}.

Specificity sites beyond S1 and S2

There is no specificity for any particular side chain at the P3 position of coronavirus M^{pro} cleavage sites. This agrees with the P3 side chain of our

Figure 3. X-ray crystallographic analysis of M^{pro} complexes with CMK inhibitor (stereo figures) (A) The substrate analogue inhibitor binds in the shallow substrate-binding site at the surface of the TGEV proteinase, between domains I and II. (B) In the active monomer A of the SARS-CoV M^{pro} crystallized at pH 6.0, the P1 glutamine of the inhibitor protrudes into the S1 pocket, but P2–P5 adopt an unusual binding mode. (C) In the inactive monomer B, the S1 pocket is not accessible, and the P1 glutamine points towards bulk solvent.

A: After [18], with kind permission of the American Association for the Advancement of Science (AAAS); B, C: After [19], with kind permission of The National Academy of Sciences of the USA.

substrate analog being oriented towards bulk solvent in the TGEV M^{pro} complex (Fig. 3A). At the P4 position, there has to be a small amino acid residue such as Ser, Thr, Val, or Pro because of the congested cavity formed by the side chains of Leu164, Leu166, and Gln191 as well as the main-chain atoms of Ser189. These residues are conserved or conservatively substituted (Leu164⇒Met165, Ser189⇒Thr190) in SARS-CoV M^{pro}. In the TGEV M^{pro} complex, the P5 Asn side chain interacts with the main chain at Gly167, Ser189, and Gln191 (Pro168, Thr190, Gln192 in the SARS-CoV enzyme), thus involving the loop linking domains II and III, whereas the P6 Val residue is not in contact with the protein (Fig. 3A).

Although our CMK inhibitor does not include a P1' residue, it is easily seen that the common small P1' residues (Ser, Ala, or Gly) can be easily accommodated in the S1' subsite of TGEV M^{pro} formed by Leu27, His41, and Thr47, with the latter two residues also being involved in the S2 subsite (Leu, His, and Asp in SARS-CoV M^{pro}).

Superimposition of the structures of the TGEV M^{pro}-CMK complex and the free HCoV 229E M^{pro} shows that the two substrate-binding sites are basically the same. All residues along the S4-S1 side of the cleft are identical, with the exception of the conservative M190L replacement (Ala191 in SARS-CoV M^{pro}). In the SARS pathogen, M^{pro} residues 168 and 188 - 190 show some substitutions but since these residues contribute to substrate binding with their main-chain atoms only, the identity of the side chains is less important. Indeed, the substrate-binding site of the SARS-CoV M^{pro} matches those of its TGEV and HCoV 229E counterparts quite well (Fig. 3B). And yet, the CMK inhibitor bound to the SARS-CoV M^{pro} in a remarkably different mode (Figs. 3B,C). But before we discuss substrate or inhibitor binding to the SARS-CoV main proteinase, we have to analyze the dynamical nature of the structure of the free enzyme.

pH-dependent conformational changes of the SARS-CoV M^{pro}

The SARS-CoV M^{pro} also forms a dimer in the crystal (and in solutions at concentrations of >1 mg/ml) with the two protomers oriented almost at right angles to each other [19], in an arrangement that is similar to the HCoV and TGEV M^{pro} structures [13, 18]. However, when we first determined the structure of the SARS-CoV enzyme by X-ray analysis of crystals grown at pH 6.0 [19], we were surprised to see that the substrate-binding site of one monomer (denoted "A", Fig. 4A) was in the normal, active conformation described above for the TGEV and HCoV 229E M^{pro}s, whereas the other ("B") had adopted a catalytically incompetent form (Fig. 4B). There were two major conformational changes that led to this inactivation: i) the S1 binding pocket for the glutamine residue at the cleavage site of the substrate had collapsed, and ii) so had the oxyanion hole (Fig. 4B). We believe that these changes originate from protonation of HisB163, the con-

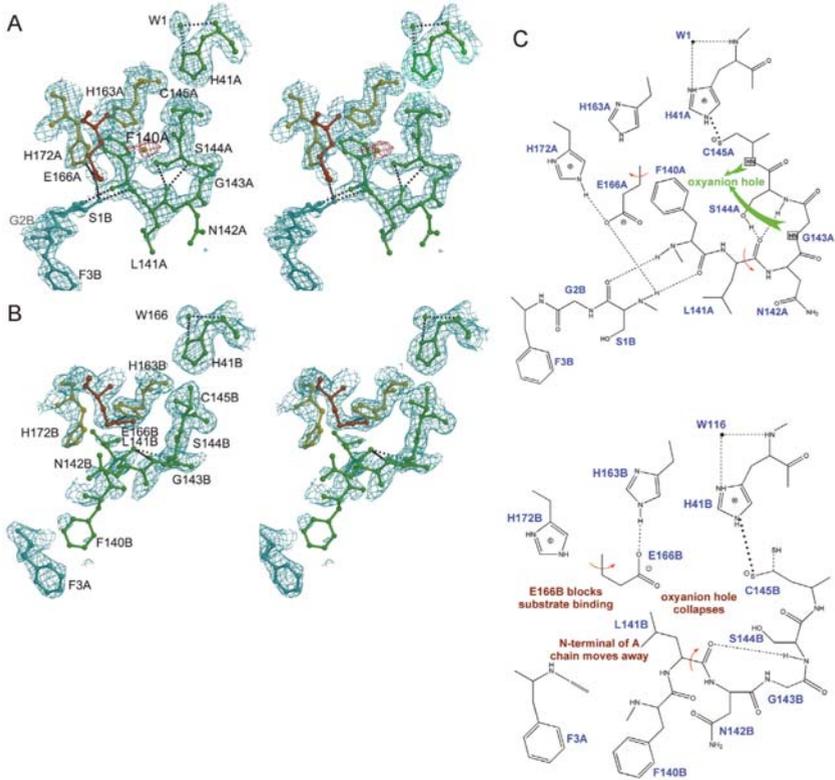


Figure 4. Conformational changes of SARS-CoV M^{Pro}. (A) Substrate-binding site of one monomer in the normal, active conformation; (B) catalytically incompetent form; (C) Schematic drawing showing the S1 pocket in the active conformation (top) and the inactive conformation (bottom). After [19], with kind permission of The National Academy of Sciences of the USA.

served residue found at the bottom of the S1 pocket, in monomer B. The pH of crystallization of 6.0 is close to the pK of histidine, explaining why the residue appears to be protonated in monomer “B” but not monomer “A” of the SARS-CoV M^{Pro} dimer. As a consequence of the protonation, the side chain of PheB140, which stacks onto HisB163 as long as the latter is unprotonated, moves away and reorients towards the A3 of the other monomer. In order to neutralize the positive charge of HisB163 in its relatively hydrophobic environment, GluB166 moves into the S1 pocket and forms a very strong salt bridge of 2.43 Å with HisB163. This leads to breakage of the ion pair between GluB166 and HisB172 as well as of the amino terminus, residue A1, of the other monomer in the dimer. This very amino terminus also interacts with the main chain of residue B140, the phenylalanine that moves away in response to the protonation of HisB163. This interaction,

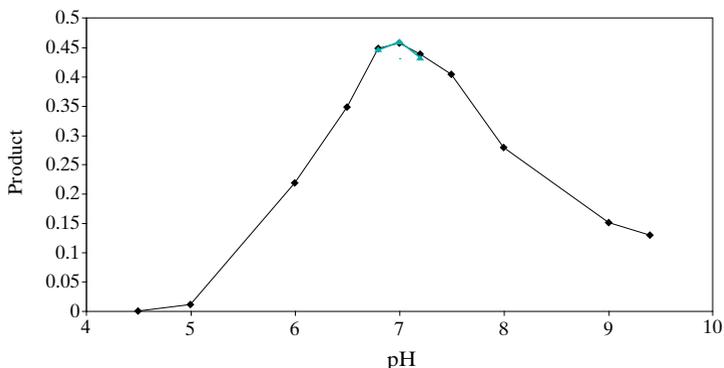


Figure 5. Activity of the SARS-CoV M^{Pro} as a function of pH.

consisting of two hydrogen bonds, between the hydrogen bond with the amino group of A1 and the CO of PheB140 as well as between the CO of A1 and the NH of PheB140, is also lost as part of the conformational rearrangements. As a result, the segment around PheB140 (residues B138 - B143) is no longer fixed. It adopts a different conformation with positional differences from the A chain of up to 5.9 Å and becomes poorly ordered, although we do see continuous electron density for the main chain and for all side chains except PheB140 (Fig. 4B). In fact, the entire loop translates towards the active site. Importantly, residue 140 is part of the oxyanion loop, which stabilizes the transition state of the proteolytic reaction by donating hydrogen bonds from the main-chain amides of Gly143 and Cys145 to the negatively charged oxygen of the tetrahedral intermediate. It is essential that this loop be in the right conformation for the enzyme to exhibit full catalytic activity. However, in the inactive protomer, Gly143 moves by >3.2 Å towards the active site and leaves no space for the oxyanion to bind.

As mentioned above, the S1 binding site is partly occupied by GluB166 interacting with HisB163. In addition, CysB145 adopts two conformations, each about half occupied, one of which is filling the remainder of the collapsed S1 pocket. The other one is interacting with the catalytic HisB41 as in the active protomer. In summary, protomer “B” in the dimer is inactive because it cannot stabilize the oxyanion transition state and because it is unable to bind the P1 glutamine residue of the substrate. This agrees with the measured activity of the enzyme as dependent on pH, which displays a bell-shaped curve, with a maximum at pH 7.0 and less than 50% activity both at pH 6.0 and at pH 8.5 (Fig. 5) [29].

In a follow-up study, we equilibrated at pH 7.6 and 8.0 crystals of the SARS-CoV M^{Pro} that had been grown at pH 6.0, and determined the structures by X-ray analysis [19]. At pH 7.6 (Fig. 6A), the S1 binding pocket is

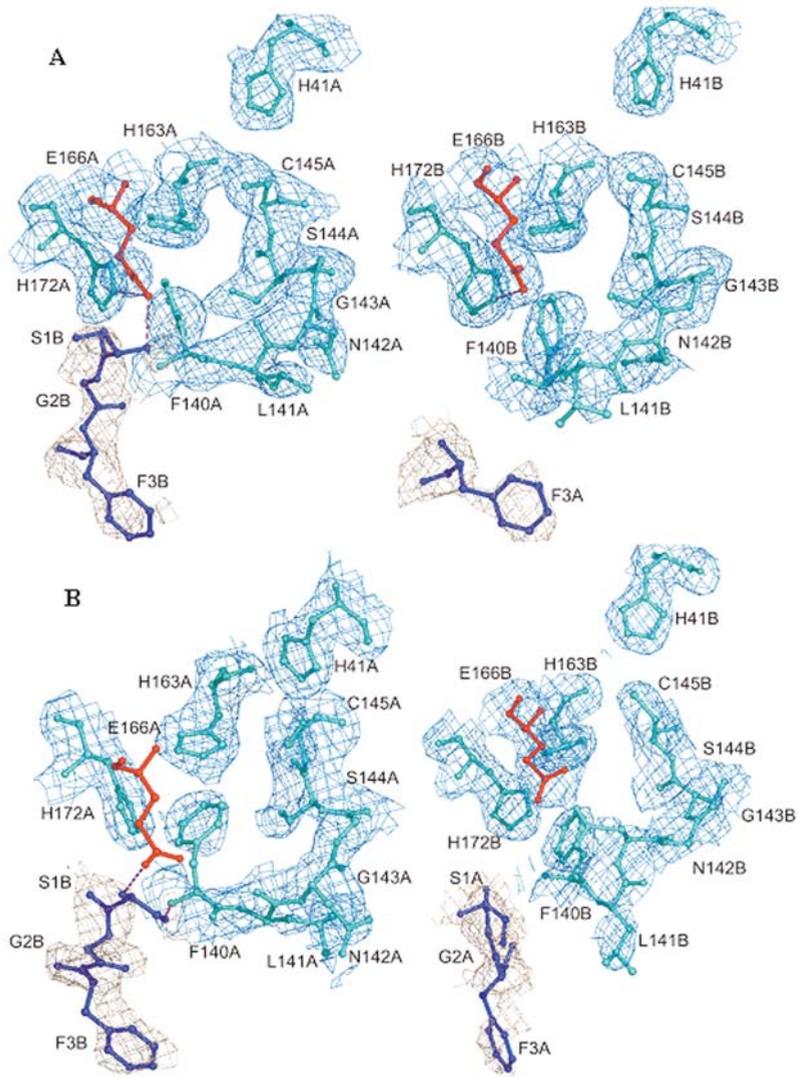


Figure 6. The S1 pocket of SARS-CoV M^{Pro} equilibrated at pH 7.6 (A) and pH 8.0 (B). Left panels: molecule A, right panels: molecule B.

fully formed in both monomers and Phe140 is interacting with the now unprotonated His163. The oxyanion is correctly shaped. Glu166 forms ion pairs with His172 and the N-terminus of the other monomer in the dimer. At pH 8.0 (Fig. 6B), another interesting change takes place. The salt bridge between Glu166 and His172 breaks – but not the one between Glu166 and the amino terminus of the other protomer in the dimer –, resulting in

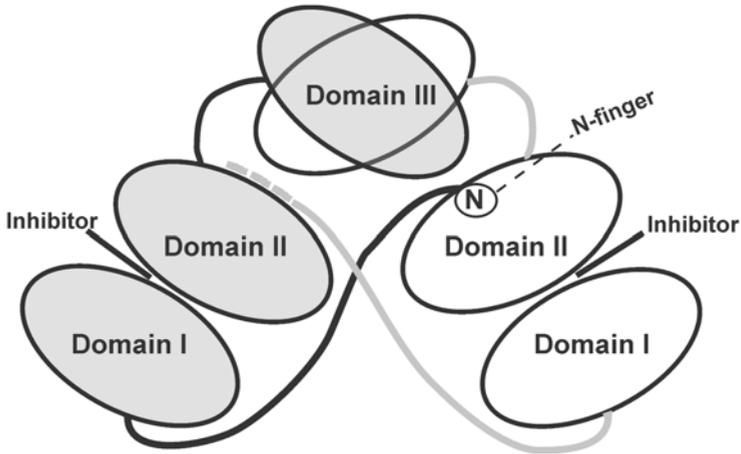


Figure 7. Schematic drawing of the SARS-CoV M^{Pro} dimer as seen in the crystal structure at pH 6.0. Note the important role of the “N-finger”.

Glu166 becoming more mobile and partly filling the S1 pocket, thereby decreasing the activity of the enzyme. This is presumably due to the deprotonation of His172 at the high pH value. The pK value of histidine residues involved in salt bridges has been shown to be around 8.0 to 8.5, fully consistent with our interpretation of the phenomenon. Thus, we propose that the low-pH and high-pH flanks of the bell-shaped pH-activity curve (Fig. 5) are determined by the protonation and deprotonation of His163 and His172, respectively. In this respect, our interpretation differs from that of Lai and coworkers who proposed that deactivation of the enzyme at high pH is due to deprotonation of the active-site cysteine [26]. In our opinion, the latter would only play a role at even higher pH values of >9.

The different protonation states of His163 and His172 and their influence on the conformation of the substrate-binding site were further explored by molecular-dynamics calculations over 10 ns each (Tan et al., in preparation). The results are in full agreement with the interpretation of the crystal structures put forward here. Furthermore, molecular dynamics calculations have the advantage that they can simulate states of the system under investigation that are not easily accessible to experiment. A dynamics run at pH 5.0, with His163 protonated in both monomers of the dimer, revealed a tendency of the two S1 pockets to collapse, with Glu166 switching between the protonated His163 and His172 residues over the time of the simulation.

It is important to note that none of the conformational differences seen between monomers “A” and “B” of the SARS-CoV M^{Pro} when crystallized at pH 6.0 is due to crystal packing artefacts. Symmetry-related contacts between dimers mainly involve domain III of both molecules; the only

residues in domains I and II to display strong crystal contacts are His A134, Asn B53, Asp B56, Ile B59, Arg B60, His B134. Furthermore, the fact that the conformation of monomer B flips back to the normal, active one within the same crystal lattice after equilibration of the crystals at higher pH values, proves that crystal packing does not play a role here.

The amino terminus of the SARS-CoV M^{pro} plays an important role in our structural model, by interacting with essential residues of the substrate-binding site of the other monomer in the dimer. Like a finger, the N-terminus docks into its specific binding site (Fig. 7) and flicks on the enzymatic activity of the other monomer. Hydrogen bonding from the “N-finger” to the main chain of Phe140 helps shape the oxyanion loop, and ion-pair formation with Glu166 prevents the latter from blocking the S1 specificity pocket. If a free N-terminus is not available at position 1 of the M^{pro}, such as in the maltose-binding protein (MBP) fusion protein or in the presence of an N-terminal histidine tag, the enzymatic activity should be reduced and this has indeed been observed, at least with the MBP fusion protein [13]. In order to further demonstrate the importance of the N terminus, we deleted residues 1 to 5 in the TGEV M^{pro} and indeed found the enzyme to be almost totally inactive [13]. This observation also solves the mystery of the importance of the “extra” helical domain III of the main proteinase. Its task is to provide defined interaction points for the N-terminal segment and thereby locating the tip of the N-finger to its specific interaction site with Phe140 and Glu166 of the other monomer. Accordingly, deletion of domain III also results in an inactive enzyme [13]. The enzyme is dimeric because of the essential role played by the N-finger, but the N-terminal segment is not essential for dimerization, as shown recently by Shi et al. [24] and also indicated by steered molecular dynamics calculations of a SARS-CoV M^{pro} dimer devoid of the N-terminal residues (Tan et al., personal communication).

A transition between a catalytically competent and an incompetent conformation of the oxyanion-binding site has also been seen in the crystal structure of equine arterivirus main proteinase (nsp4), where the loop structurally equivalent to the M^{pro} 139–143 loop undergoes a peptide flip, resulting in misorientation of the amide corresponding to Gly143 [30]. Also, we have little doubt that the same pH-induced conformational changes that we observed in the SARS-CoV M^{pro} does exist in all other coronavirus main proteinases. We simply failed to detect it earlier because we had crystallized the M^{pro}s of TGEV and HCoV 229E at pH > 8.0, i.e. with both monomers adopting an active conformation [13, 18].

In conclusion, SARS-CoV M^{pro} appears to possess a pH-triggered activation switch, which may regulate proteolytic activity in different cellular compartments. Such a pH switch would provide the virus with a mechanism to synchronize the polyprotein processing with the microenvironment. Further mutagenesis and biochemical studies on the details of the switch are clearly desirable.

Self-activation of the main proteinases

The M^{pro} is part of the viral polyprotein itself and has to be activated by autocleavage from this inactive precursor before it can cleave the polyprotein at other cleavage sites. So far, it has been completely unclear how this autoactivation might occur but the crystallographic work on SARS-CoV M^{pro} now provides a clue. The pH-dependent conformational changes seen in the X-ray structures [19] could well correlate with the structural arrangements occurring during self-activation of the enzyme.

Moreover, a comparison with the transition between chymotrypsin and its proenzyme, chymotrypsinogen, reveals a remarkable similarity in the mechanisms used. In mature chymotrypsin, the charged amino terminus, created after cleavage of the proenzyme at the Arg15-Ile16 bond, forms an ion pair with Asp194, a residue involved in forming the wall of the S1 specificity pocket. In chymotrypsinogen, on the other hand, the N-terminal extension of the polypeptide chain by 15 residues causes disruption of this ion pair [31]. The Asp194 side chain changes conformation and so does, along with it, Gly193, whose amide forms part of the oxyanion hole. As a consequence, interaction of Gly193, which corresponds to Gly143 of SARS-CoV M^{pro}, with the tetrahedral intermediate is no longer possible. In SARS-CoV and other coronaviruses, the polyprotein upstream to the N-terminal autocleavage site of the M^{pro} corresponds to the propeptide in chymotrypsinogen. The role of Asp194 of the latter is taken by Glu166 in SARS-CoV M^{pro} and its proenzyme, and the oxyanion-binding region changes in a similar way. The major difference between the two enzymes is that the event is *intramolecular* in chymotrypsin but *intermolecular* in SARS-CoV M^{pro}.

It is a matter of debate whether the autocleavage of the M^{pro}s from the polyprotein precursors occur in *cis* or in *trans*. In the case of mouse hepatitis (corona)virus (MHV) M^{pro}, autocleavage has exclusively been observed in *trans* [28]. On the other hand, it has been shown for several picornavirus 3C proteinases, e.g. hepatitis A virus 3C^{pro} [32], that the first step of autocleavage of the enzyme, i.e. processing at the N-terminal cleavage site, occurs in *cis*, followed by other cleavage reactions in *trans*. Molecular modeling studies based on the TGEV M^{pro} crystal structure suggested that N-terminal precursor cleavage in *cis* should be possible, because by introducing a bend around residues 11 to 13, the extended N-terminal chain can be quite easily folded back into the active site of its own molecule, even though the distance between the amino terminus (A1) and CysA145 S_γ is about 34 Å [13]. On the other hand, bringing the N terminus from the position seen in the crystal structure into the catalytic site of the other molecule in the dimer proves quite difficult, even though the distance is only about 11 Å. Thus, the fact that the M^{pro}s are dimers does not necessarily support cleavage in *trans*. The SARS-CoV M^{pro} structure rather suggests a different role for the dimerization, the mutual activation of the subunits by transition from the proenzyme

conformation to that of the active enzyme, triggered by binding of N-fingers to the activation switch. Clearly, asymmetric dimers consisting of one molecule in the mature conformation and one in the proenzyme conformation, as seen in the SARS-CoV M^{PRO} crystal structure at pH 6.0, would be important intermediates in this process. Polyprotein processing in SARS-CoV-infected cells could thus be initiated by the occasional dimerization of two membrane-anchored M^{PRO} domains within polyproteins and mutual activation through the N-finger tips, or by the rare event of cis-autocleavage of a single M^{PRO} domain which then binds to another one within polyprotein, again activating it through the N-finger. The results are highly active M^{PRO} dimers which can then cleave other substrates in *trans*. Such a mechanism could lead to self-acceleration of M^{PRO} autoprocessing, after a slow onset triggered by a rare initial event. This mechanism is consistent with the fact that in cells infected with HCoV 229E or MHV, long-lived cleavage intermediates are usually not observed, but polyprotein is processed with increasing speed, once the first cleavage has occurred [28, 33].

Complex of the SARS-CoV M^{PRO} with a chloromethyl ketone inhibitor

After dissecting the details of conformational dynamics of the SARS-CoV main proteinase, we are now finally in the position to competently discuss the interactions of the enzyme with substrate. To visualize these, the substrate-analogue hexapeptidyl CMK inhibitor, Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK, was soaked into the crystals of the SARS-CoV enzyme grown at pH 6.0, and the structure of the complex was determined at 2.5 Å resolution [19]. We expected to see one of two possible cases, i.e. either i) the inhibitor would only bind to the protomer in the active conformation (monomer “A”), or ii) it would bind to the inactive protomer (monomer “B”) as well, thereby imposing the active conformation onto the latter by an induced-fit mechanism. Quite unexpectedly, neither assumption turned out to be true. Instead, whereas the inhibitor does bind to both subunits and covalent bonds are being formed between the S_γ of Cys145 and the methylene group of the CMK, the inhibitor exhibits two different modes of binding, each of which is unusual and unexpected (Fig. 3B, C). Protomer “B” remains in the inactive conformation, and the inhibitor does not open the occluded S1 pocket to allow the P1 Gln to enter the specificity site (Fig. 3C). Instead, the P1 Gln points out of the substrate-binding cleft and towards bulk solvent. On the other hand, P2 Leu and P4 Ser bind to the appropriate specificity pockets that have been observed in the structure of TGEV M^{PRO} [13], whereas P3 Thr is oriented toward bulk solvent. This result can be interpreted as resulting from the high chemical reactivity of the CMK moiety with obvious concurrent inability of the inhibitor to overcome the protonation of HisB163 at the bottom of the S1 pocket at pH 6.0.

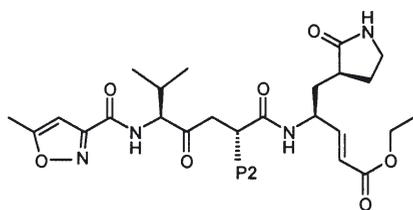
Binding of the CMK inhibitor to protomer A, which exhibits the active conformation, is even less conventional (Fig. 3B). The P1 glutamine resides in the S1 pocket and its side-chain carbonyl accepts a 2.8-Å hydrogen bond from the Nε2 atom of HisA163, whereas its side-chain amide nitrogen donates a 2.8-Å hydrogen bond to the carboxylate of GluA166, ensuring the specificity for glutamine in the S1 subsite. However, the P2 residue fails to penetrate the S2 subsite and becomes partially solvent-accessible. This non-canonical binding results in a frameshift in subsite interaction: P3 Thr occupies the S2 pocket instead and P5 Asn binds at the S4 subsite. We note that the P2 residue specificity of SARS-CoV M^{Pro} is less stringent than that of other CoV M^{Pro}s, which, with few exceptions, are restricted to leucine [28]. The unexpected binding mode of the substrate-like inhibitor may well be a consequence of the more relaxed specificity requirements of the SARS-CoV M^{Pro}, compared to other coronaviral main proteinases. However, before further conclusions are made, it has to be excluded that the binding mode observed is due to an artefact, perhaps because the sequence of the peptidyl CMK inhibitor had been derived from the TGEV M^{Pro} and not the SARS-CoV M^{Pro} N-terminal autocleavage site. Therefore, we are presently determining the crystal structures of the complexes of SARS-CoV M^{Pro} with other inhibitors, both peptidic and non-peptidic.

Structure-based design of anti-SARS inhibitors

When the genome sequence of the SARS coronavirus was published on the internet on April 13, 2003 [16], we immediately located the segment in orf1a coding for the main proteinase and aligned the amino acid sequence of the enzyme with the M^{Pro}s of other coronaviruses. Next, we constructed a homology model of the SARS-CoV M^{Pro}, based on this alignment and our crystal structures of the TGEV and HCoV 229E main proteinases [18]. This theoretical model revealed that the substrate-binding site of the SARS-CoV enzyme is very similar to those of the two other viral enzymes. At that time, less than one month after the discovery of the new SARS virus and at the peak of the epidemic in China, this structural model was the best we could get, and it proved in fact very useful (see below). Of course, a theoretical homology model will always be inferior to an experimental crystal structure, but in the absence of the latter, it can provide a reasonably good basis for structure-assisted drug discovery [34, 35]. When we had determined the crystal structure of the SARS-CoV M^{Pro} another 10 weeks later [19], the overall r.m.s. deviation for all Cα atoms of the protein between theoretical model and experimental structure was found to be as high as 3.8 Å. However, this unreasonably high value is only caused by large deviations of the last four residues at the flexible C-terminus, which deviate by more than 12 Å, and by slight differences of the mutual orientation of the

three domains. If the models for the individual domains are compared with the crystal structure, and only this is a fair comparison in view of the large flexibility of the enzyme, the r.m.s. deviation is down to little more than 1 Å.

Whatever the view taken on this issue, the homology model served primarily one purpose, and that was to facilitate the design of anti-SARS inhibitors directed at the main proteinase. With the template of the peptidyl CMK inhibitor – as bound to the TGEV M^{pro} – at hand, we compared its binding mode to those seen for all other inhibitors of cysteine proteinases that occupy the N-terminal half of the substrate binding site (subsites Sn-S1), as long as atomic coordinates were available in the Protein Data Bank (PDB). We found that AG7088, a compound carrying a vinylogous ethyl ester instead of a CMK group and designed to block human rhinovirus (HRV) 3C proteinase [36], interacted in much the same way with its target enzyme than did our CMK inhibitor (Fig. 8) [18].



P2 = *p*-fluoro-benzyl; AG7088 (CAS number: 223537-30-2, Pfizer Inc.)

As mentioned before, the picornavirus 3C proteinases (3C^{pro}) have a chymotrypsin-related structure, reminiscent of domains I and II of the coronavirus M^{pro}, although several of the secondary-structure elements are arranged differently, making structural alignment difficult with sequence identities < 10%. Also, they completely lack a counterpart to domain III of coronavirus M^{pro}s. Nevertheless, the substrate specificity of picornavirus 3C^{pro}s [32, 37] for the P1', P1 and P4 sites is similar to that of the coronavirus M^{pro}s (hence the original designation “3C-like” for the latter; see Tab. 1). Only parts of the two structures, TGEV M^{pro} in complex with the CMK inhibitor and HRV-2 in complex with AG7088, can be spatially superimposed (Fig. 8; r.m.s. deviation of 2.10 Å for 134 pairs of C α positions out of the ~ 180 residues in domains I and II). Both inhibitors, the hexapeptidyl CMK and AG7088, bind to their respective target proteinases through formation of an antiparallel β -sheet with strand eII. However, completely different segments of the polypeptide chain interact with the substrate analogues on the opposite site: residues 188 to 191 of the loop connecting domains II and III in M^{pro}, as opposed to the short β -strand 126 to 128 in HRV 3C^{pro}. As a result, the architectures of the S2 subsites are entirely different between the two enzymes; hence, the different specificities for the P2 residues of the substrates (Leu *versus* Phe). The inhibitor

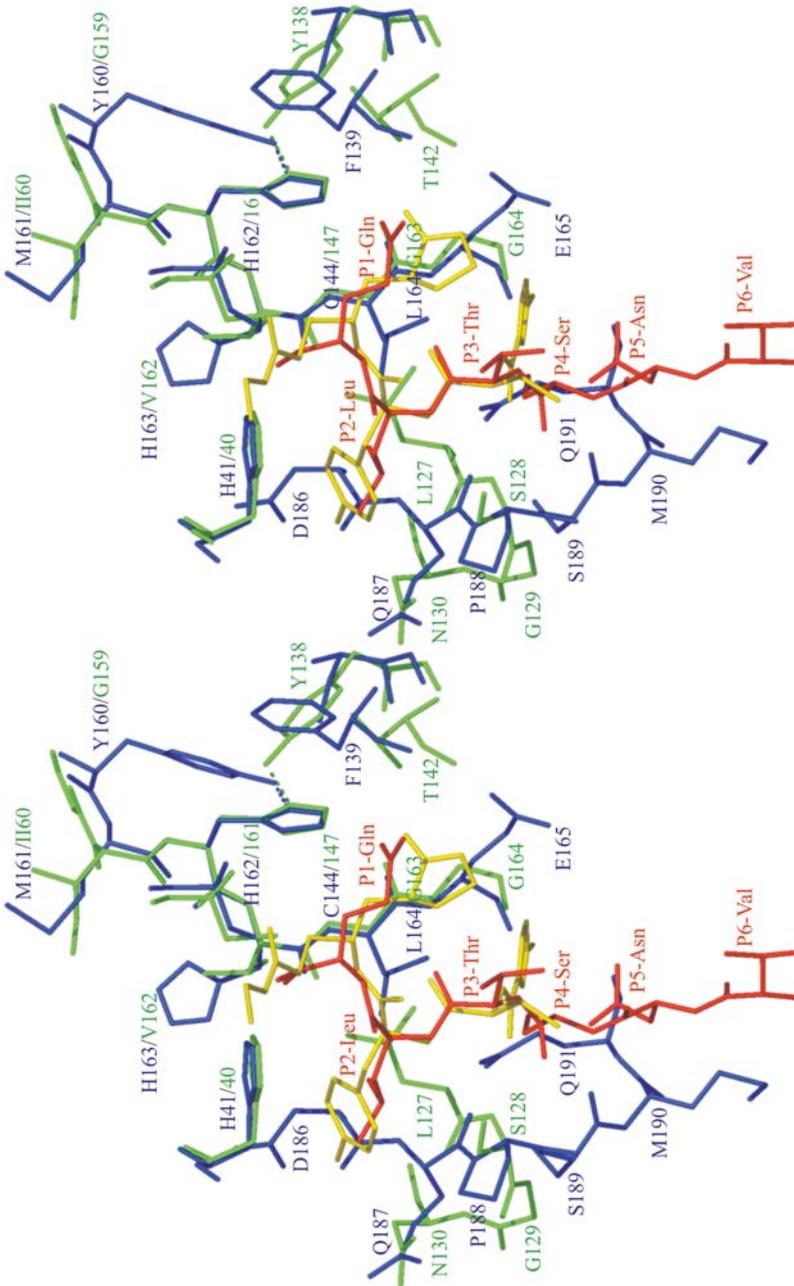


Figure 8. Superimposed structures of TGEV M^{pro} (blue) in complex with CMK inhibitor (red) and HRV-2 (green) in complex with AG7088 (yellow). Only the substrate-binding sites are shown. Stereo illustration. From [18], with permission of the American Association for the Advancement of Science (AAAS).

AG7088 has a *p*-fluorophenylalanine side chain (*p*-fluorobenzyl) in this position. Based on molecular modeling, we proposed that this side chain might be too long to fit into the S2 pocket of coronavirus M^{Pro}, but an unmodified benzyl group would probably fit, as evidenced by Phe occurring in the P2 position of the C-terminal autocleavage site of the SARS coronavirus enzyme (see Tab. 1). Apart from this difference, the superimposition of the two complexes (Fig. 8) suggested that the side chains of AG7088 binding to subsites S1 (lactone derivative of glutamine) and S4 (5-methyl-isoxazole-3-carbonyl) can be easily accommodated by the coronavirus M^{Pro}. Thus, we proposed that AG7088 could serve as a starting point for modifications which should quickly lead to an efficient and bioavailable inhibitor for coronavirus main proteinases. In fact, such modifications have been applied (work mostly unpublished so far; a published proposal for modification is found in [38]) and we will see a number of good inhibitors designed on this basis appear in the literature in the next few months.

Another possible approach to inhibiting the SARS-CoV main proteinase could make use of the insight into the importance of the N-finger in activating the enzyme. One could screen for competitive inhibitors that prevent M^{Pro} dimerization through binding to the N-finger docking site, but without activating the Glu166/Phe140 switch. Such an approach would open a second front in addition to active site-directed inhibitors, in the urgent search for substances efficiently blocking the SARS coronavirus.

Several groups have proposed to use HIV-1 proteinase inhibitors as starting points for designing inhibitors directed at the SARS-CoV M^{Pro}. However, many of these studies remained at the modelling level and failed to prove the usefulness of the HIV-1 proteinase inhibitors experimentally [39, 40]. Some of the reports did not even seem to acknowledge the fact that the enzymatic mechanisms of HIV-1 proteinase, an aspartic proteinase, and SARS-CoV main proteinase, a cysteine proteinase, are very different. One of the few studies reporting a clinical usefulness of the application of HIV-1 proteinase inhibitors was published by Peiris et al. [41]. According to their results, a combination of the HIV-1 proteinase inhibitor kaletra and ribavirin was more effective than ribavirin alone, which, however, was basically inactive, as shown later [42]. Based on the present structural knowledge, it is not obvious how HIV-1 proteinase inhibitors should block SARS-CoV.

In conclusion, prospects for discovering and developing inhibitors of the SARS-CoV main proteinase are good, given the structural knowledge that is available for the target today. It should also be emphasized that the crystal structures of the various M^{Pro}s allow very efficient virtual screening [43] for existing organic compounds that might bind and inhibit the target. This approach is currently yielding promising inhibitors, which also exhibit anti-SARS activity on virus-infected Vero cells (Jiang, personal communication).

Conclusions

The three-dimensional structures presented here for coronavirus main proteinases provide a solid basis for the design of anti-coronaviral drugs. The binding modes of peptidic inhibitors and substrates have been revealed by the crystal structure of TGEV M^{pro} in complex with a hexapeptidyl chloromethyl ketone, and, within limitations, of the complex with the SARS-CoV M^{pro} and the same inhibitor. The structure of the SARS-CoV enzyme was determined at three different pH values and revealed a pH-dependent activation switch for the M^{pro}. The pH-activity profile of the main proteinase was shown to be governed by the protonation and deprotonation of two histidine residues in the substrate-binding site. The conformational changes observed also suggested a mechanism for the self-activation of the M^{pro} by autocleavage from the viral polyprotein. On the basis of structural comparisons between the TGEV M^{pro} in complex with a hexapeptidyl CMK inhibitor and the HRV 3C proteinase in complex with the vinyllogous ethyl ester AG7088, it has been proposed that whereas the latter would not be a good inhibitor of the SARS-CoV M^{pro}, it should be a good starting point for designing such inhibitors. Such design studies are being performed at the moment and will start to yield potent anti-SARS compounds in the next few months and years.

Acknowledgements

The work described here was, in part, supported by grants from the “Deutsche Forschungsgemeinschaft”, the “Sino-German Center for the Promotion of Science”, the “Sino-European Project on SARS Diagnostics and Antivirals” (SEPSDA) of the European Commission, the Chinese Ministry of Science and Technology, and the “Fonds der Chemischen Industrie”.

References

- 1 Drostén C, Günther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RAM et al (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 348: 1967–1976
- 2 Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W et al (2003) A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348: 1953–1966
- 3 Cavanagh D (1997) *Nidovirales*: a new order comprising *Coronaviridae* and *Arteriviridae*. *Arch Virol* 142: 629–633
- 4 Cowley JA, Dimmock CM, Spann KM, Walker PJ (2000) Gill-associated virus

- of *Penaeus monodon* prawns: an invertebrate virus with ORF1a and ORF1b genes related to arteri- and coronaviruses. *J Gen Virol* 81: 1473–1484
- 5 den Boon JA, Snijder EJ, Chirnside ED, de Vries AA, Horzinek MC, Spaan WJ (1991) Equine arteritis virus is not a togavirus but belongs to the coronavirus-like superfamily. *J Virol* 65: 2910–2920
- 6 Enjuanes L, van der Zeijst BAM (1995) Molecular basis of transmissible gastroenteritis virus epidemiology. In: SG Siddell (ed): *The Coronaviridae*. Plenum Press, New York, NY, 337–376
- 7 Saif LJ, Wesley R (1999) Transmissible gastroenteritis virus. In Straw BES, Allaire WL, Mengeling WL, Taylor DJ (eds): *Diseases of Swine*, 8th edn. Iowa State University Press, Ames, Iowa, 295–325
- 8 Myint SH (1995) Human coronavirus infections. In: Siddell SG (ed): *The Coronaviridae*. Plenum Press, New York, 389–401
- 9 Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt GM, Ahuja A, Yung MY, Leung CB, To KF et al (2003) A major outbreak of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* 348: 1986–1994
- 10 Herold J, Raabe T, Schelle-Prinz B, Siddell SG (1993) Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology* 195: 680–691
- 11 Thiel V, Herold J, Schelle B, Siddell SG (2001) Viral replicase gene products suffice for coronavirus discontinuous transcription. *J Virol* 75: 6676–6681
- 12 Ziebuhr J, Herold J, Siddell SG (1995) Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *J Virol* 69: 4331–4338
- 13 Anand K, Palm J, Mesters J, Siddell SG, Ziebuhr J, Hilgenfeld R (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain. *EMBO J* 21: 3213–3224
- 14 Ziebuhr J, Snijder EJ, Gorbalenya AE (2000) Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *J Gen Virol* 81: 853–879
- 15 Hegyi A, Ziebuhr J (2002) Conservation of substrate specificities among coronavirus main proteases. *J Gen Virol* 83: 595–599
- 16 Marra M, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattra J, Asano JK, Barber SA, Chan SY et al (2003) The genome sequence of the SARS-associated coronavirus. *Science* 300: 1399–1404
- 17 Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH et al (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300: 1394–1399
- 18 Anand K, Ziebuhr J, Wadhwani P, Mesters J, Hilgenfeld R (2003) Coronavirus main proteinase (3CL^{Pro}) structure: basis for design of anti-SARS drugs. *Science* 300: 1763–1767
- 19 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 USA* 100: 13190–13195
- 20 Tsukada H, Blow DM (1985) Structure of α -chymotrypsin refined at 1.68 Å resolution. *J Mol Biol* 184: 703–711

- 21 Allaire M, Chernaia MM, Malcolm BA, James MN (1994) Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* 369: 72–76
- 22 Bergmann EM, Mosimann SC, Chernaia MM, Malcolm BA, James MN (1997) The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *J Virol* 71: 2436–2448
- 23 Fan K, Wei P, Feng Q, Chen S, Huang C, Ma L, Lai B, Pei J, Liu Y, Chen J, Lai L (2004) Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J Biol Chem* 279: 1637–1642
- 24 Shi J, Wei Z, Song J (2004) Dissection study on the severe acute respiratory syndrome virus 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme. *J Biol Chem* 279: 24765–24773
- 25 Polgár, L (1974) Mercaptide-imidazolium ion-pair: the reactive nucleophile in papain catalysis. *FEBS Lett* 47: 15–18
- 26 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.
- 27 Hegyi A, Friebe A, Gorbalenya AE, Ziebuhr J (2002) Mutational analysis of the active centre of coronavirus 3C-like proteases. *J Gen Virol* 83: 581–593
- 28 Ziebuhr J, Snijder EJ, Gorbalenya AE (2000) Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *J Gen Virol* 81: 853–879
- 29 Heisen B (2004) Expression, Reinigung und Aktivitätsbestimmung der Hauptproteinase des SARS-Coronavirus. Bachelor Thesis, University of Lübeck, 2004
- 30 Barrette-Ng IH, Ng KK, Mark BL, Van Aken D, Cherney MM, Garen C, Kolodenko Y, Gorbalenya AE, Snijder EJ, James MN (2002) Structure of arterivirus nsp4. The smallest chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. *J Biol Chem* 277: 39960–39966
- 31 Bode W, Schwager P, Huber R (1978) The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine trypsinogen-pancreatic trypsin inhibitor complex and of its ternary complex with Ile-Val at 1.9 Å resolution. *J Mol Biol* 118: 99–112
- 32 Ryan MD, Flint M (1997) Virus-encoded proteinases of the picornavirus supergroup. *J Gen Virol* 78: 699–723
- 33 Ziebuhr J, Heusipp G, Siddell SG (1997) Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *J Virol* 71: 3992–3997
- 34 Hillisch A, Pineda F, Hilgenfeld R (2004) The use of homology models in the drug discovery process. *Drug Discov Today* 9: 659–669
- 35 Hillisch A, Hilgenfeld R (eds) *Modern Methods of Drug Discovery*. Birkhäuser, Basel, 2002.
- 36 Matthews DA, Dragovitch PS, Webber SE, Fuhrman SA, Patick AK, Zalman LS, Hendrickson TF, Love RA, Prins TJ, Marakovits JT (1999) Structure-assist-

- ed 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
- 37 Kräusslich HG, Wimmer E (1988) Viral proteinases. *Annu Rev Biochem* 57: 701–754
 - 38 Chou K, Wei D, Zhong W (2003) Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. *Biochem Biophys Res Commun* 308: 148–151
 - 39 Jenwitheesuk E, Samudrala R (2003) Identifying inhibitors of the SARS coronavirus proteinase. *BMC Lett* 13: 3989–3992
 - 40 Zhang XW, Yap YL (2004) Old drugs as lead compounds for a new disease? Binding analysis of SARS coronavirus main proteinase with HIV, psychotic, and parasite drugs. *Bioorg Med Chem* 12: 2517–2521
 - 41 Peiris JSM, Chu CM, Cheng VCC, Chan KS, Hung IFN, Poon LLM, Law KI, Tang BSF, Hon TYW, Chan CS et al (2003) Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* 361: 1767–1762
 - 42 Cyranoski D (2003) Critics slam treatment for SARS as ineffective and perhaps dangerous. *Nature* 423: 4
 - 43 Shen J, Xu X, Cheng F, Liu H, Luo X, Shen J, Chen K, Zhao W, Shen X, Jiang H (2003) Virtual screening on natural products for discovering active compounds and target information. *Curr Med Chem* 10: 1241–1253