



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

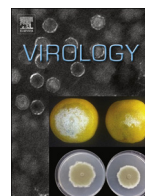
Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted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



ELSEVIER

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Structural basis for the dimerization and substrate recognition specificity of porcine epidemic diarrhea virus 3C-like protease

Gang Ye^{a,b}, Feng Deng^{a,b}, Zhou Shen^{a,b}, Rui Luo^{a,b}, Ling Zhao^{a,b}, Shaobo Xiao^{a,b}, Zhen F. Fu^{a,b,c}, Guiqing Peng^{a,b,d,*}

^a State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan, Hubei, China

^b College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, Hubei, China

^c Department of Pathology, College of Veterinary Medicine, University of Georgia, Athens, GA 30602, USA

^d The Cooperative Innovation Center for Sustainable Pig Production, Huazhong Agricultural University, Wuhan, Hubei, China

ARTICLE INFO

Article history:

Received 19 October 2015

Returned to author for revisions

14 April 2016

Accepted 15 April 2016

Available online 26 April 2016

Keywords:

Corona virus

PEDV 3CL^{PRO}

Crystal structure

Dimerization

Substrate specificity

ABSTRACT

Porcine epidemic diarrhea virus (PEDV), a member of the genus *Alphacoronavirus*, has caused significant damage to the Asian and American pork industries. Coronavirus 3C-like protease (3CL^{PRO}), which is involved in the processing of viral polyproteins for viral replication, is an appealing antiviral drug target. Here, we present the crystal structures of PEDV 3CL^{PRO} and a molecular complex between an inactive PEDV 3CL^{PRO} variant C144A bound to a peptide substrate. Structural characterization, mutagenesis and biochemical analysis reveal the substrate-binding pockets and the residues that comprise the active site of PEDV 3CL^{PRO}. The dimerization of PEDV 3CL^{PRO} is similar to that of other *Alphacoronavirus* 3CL^{PRO}s but has several differences from that of SARS-CoV 3CL^{PRO} from the genus *Betacoronavirus*. Furthermore, the non-conserved motifs in the pockets cause different cleavage of substrate between PEDV and SARS-CoV 3CL^{PRO}s, which may provide new insights into the recognition of substrates by 3CL^{PRO}s in various coronavirus genera.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Porcine epidemic diarrhea virus (PEDV), which belongs to the genus *Alphacoronavirus* in the family *Coronaviridae*, causes severe diarrhea, vomiting, dehydration and high mortality in neonatal piglets (Song and Park, 2012; Wood, 1977). PEDV was first observed in Europe in 1971 (Oldham, 1972). PEDV was isolated in Belgium in 1976 (Pensaert and de Bouck, 1978). Since PEDV was first identified, outbreaks have been reported in many swine-producing countries, notably in Europe and Asia. PEDV has recently re-emerged, with outbreaks in Asia and North America resulting in enormous economic losses (Pasick et al., 2014; Vlasova et al., 2014; Wang et al., 2014b). Continuous vaccine efforts have been made and some advances have been achieved since the outbreak (Collin et al., 2015; Song et al., 2015).

There are four genera within *Coronaviridae*: *Alphacoronavirus*, *Betacoronavirus* (A, B, C and D) (de Groot et al., 2012), *Gamma-coronavirus*, and *Deltacoronavirus* (Adams and Carstens, 2012). Transmissible gastroenteritis virus (TGEV), another

Alphacoronavirus, causes severe and often fatal diarrhea in young pigs (Garwes, 1988). The severe acute respiratory syndrome (SARS) coronavirus (lineage B *Betacoronavirus*) causes a life-threatening disease with a mortality rate of 11% (Weiss and Navas-Martin, 2005). Another novel lineage C *Betacoronavirus*, the Middle East respiratory syndrome coronavirus (MERS-CoV), has recently emerged in the Middle East and spread to Europe and other areas, causing a SARS-like infection in humans with a high mortality rate of approximately 40% (Chan et al., 2013).

The coronavirus replicase gene consists of two large open reading frames (ORF1a and ORF1b) located at the 5' end of the genome. ORF1a encodes the polyprotein pp1a, whereas ORF1a and ORF1b together encode pp1ab (Hegyi et al., 2002; Ziebuhr et al., 2000; Ziebuhr, 2005). The coronavirus 3C-like protease (3CL^{PRO}; non-structural protein 5, Nsp5) is part of the polyproteins pp1a and pp1ab, and this enzyme is encoded by ORF1a. With few exceptions, coronaviruses encode two papain-like proteases (called PL1^{PRO} and PL2^{PRO}, respectively), which are responsible for the cleavage of the N-proximal regions of the polyprotein. The central and C-proximal regions of the polyprotein are cleaved by 3CL^{PRO} at 11 conserved sites, and the products are essential for viral replication (Thiel et al., 2001; Thiel et al., 2003; Ziebuhr et al., 2000). Thus, 3CL^{PRO} is an appealing target for the design of anti-coronavirus therapies.

* Correspondence to: State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, The Cooperative Innovation Center for Sustainable Pig Production, Huazhong Agricultural University, Wuhan 430070, China.

E-mail address: penggq@mail.hzau.edu.cn (G. Peng).

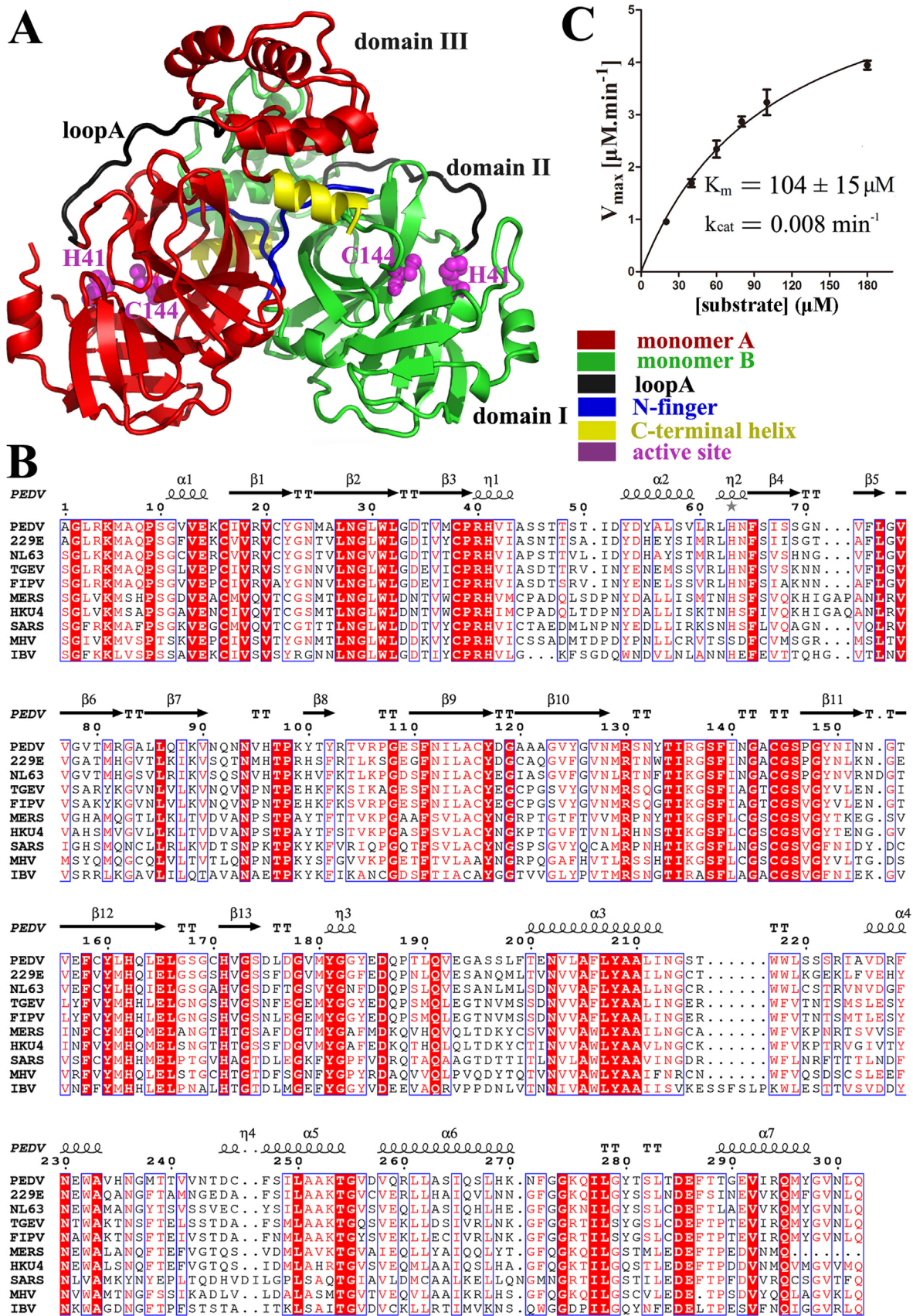


Fig. 1. Crystal structure and sequence alignment of PEDV 3CL^{pro}. **A.** The overall structure of the PEDV 3CL^{pro} homodimer. Loop A, the N-finger, the C-terminal helix and the active site are shown in black, blue, yellow and magenta, respectively. Loop A, which connects domain II to domain III, is labeled. Catalytic residues H41 and C144 are represented as spheres. **B.** The structure-based sequence alignment of several coronavirus 3CL^{pro}s. The sequence alignment was conducted with Clustal W and the figure was prepared with ESPript 3 (Robert and Gout, 2014). Residues conserved in all 3CL^{pro}s are shown in white on a red background. Residues conserved in most of the sequences are shown in red and boxed with a white background. **C.** Kinetic parameters of PEDV 3CL^{pro}. The fluorogenic substrate (Dabcyl-YNSTLQ₁AGLRKM-E-Edans) was used in the assays. The initial rates of the protease under different substrate concentrations were used to calculate the kinetic parameters by fitting with the Michaelis-Menten equation using GraphPad Prism5.

Coronavirus 3CL^{pro} employs conserved cysteine and histidine residues, which serve as the principal nucleophile and general acid-base catalyst, respectively, at its catalytic site (Anand et al., 2002; Hegyi et al., 2002; Lu et al., 1995; Thiel et al., 2003; Ziebuhr et al., 2000). An asparagine is the third member of the catalytic triad in proteases of the papain family. Chymotrypsin and other members of this serine protease family possess a catalytic triad (Anand et al., 2002). However, in coronavirus 3CL^{pro}, there is no third residue involved in catalysis (Hegyi et al., 2002; Liu and Brown, 1995; Lu and Denison, 1997). Notably, a buried water molecular hydrogen-bonded to three surrounding residues takes the place that is normally occupied by the side chain of the third member of the catalytic triad (Anand et al., 2002; Anand et al., 2003; Yang et al., 2003). Coronavirus 3CL^{pro} recognizes a conserved site containing a hydrophobic residue (preferably L) at the P2 position, a Q at the P1 position, and a small aliphatic amino acid residue (S, G, A) at the P1' position (Hegyi and Ziebuhr, 2002; Hsu et al., 2005; Ziebuhr et al., 2000). The previously reported nomenclature system was used to describe the residue sites of the peptide substrates and the substrate binding sites of the protease (Schechter and Berger, 1967).

Dimerization is a commonly used strategy for regulating viral protease activity. The formation of a dimer can be the mechanism for enzyme activation. Conversely, dimerization can inhibit an active monomeric enzyme (Marianayagam et al., 2004). In terms of coronavirus 3CL^{pro}s, only the dimeric form is functional (Chen et al., 2008; Li et al., 2010; Shi et al., 2008). In the structure of TGEV 3CL^{pro}, the N-terminus of one monomer helps shape the S1 pocket and the oxyanion hole of the opposite monomer; thus, dimerization is essential for its catalytic activity (Anand et al., 2002). The dimerization and trans-cleavage activity of SARS-CoV 3CL^{pro} are completely inhibited by the E290R and R298E (located at the C-terminal helix) variants and partly inhibited by the R4E (located at the N-terminal loop) variant (Hilgenfeld, 2014).

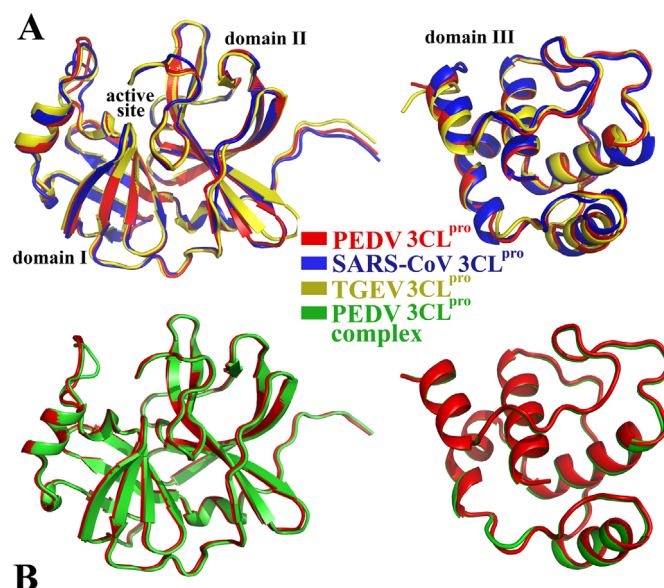
To investigate the properties of PEDV 3CL^{pro}, we determined the crystal structure of PEDV 3CL^{pro}, as well as that of an inactive PEDV 3CL^{pro} variant (C144A) bound to a peptide substrate. In addition, we performed biochemical analyses and structural comparisons to provide further insights into the dimerization and substrate specificity of coronavirus 3CL^{pro}.

2. Results and discussion

2.1. Overall structure of PEDV 3CL^{pro}

The PEDV 3CL^{pro} crystal structure was determined at a high resolution (1.65 Å) in the space group, *P*₂₁ (Fig. 1). Two molecules are observed in each asymmetric unit, with a root mean square deviation (RMSD) of 0.24 Å for 186 superimposed C α atoms of domain I and domain II, 0.89 Å for 97 superimposed C α atoms of domain III (computed through the PDBeFold service on the European Bioinformatics Institute website); the two protomers are oriented at approximately right angles to each other. Each monomer features three domains: domain I (residues 1–97), domain II (residues 98–186), and domain III (residues 202–298). Each of the first two domains exhibits an antiparallel β -barrel structure, which is similar to the 3CL^{pro}s from other coronaviruses (Anand et al., 2002; Needle et al., 2015; Yang et al., 2003). Domain III consists of five α helices and is connected to domain II by a long loop (residues 187–197, loop A in Fig. 1(A)). As observed in all of the reported 3CL^{pro} structures, the substrate-binding site near the catalytic dyad is located in a cleft between domains I and II.

Structural alignment of the domains of several coronavirus 3CL^{pro}s gives an overall view of the comparison (Fig. 2). The structures show high similarity, especially in the area around the



B

	PEDV 3CL ^{pro}		
	domain I	domain II	domain III
TGEV 3CL ^{pro}	0.8	0.7	0.7
HCoV 229E 3CL ^{pro}	0.7	0.5	0.7
SARS-CoV 3CL ^{pro}	0.9	0.5	1.2
PEDV 3CL ^{pro} complex	0.4	0.2	0.2

RMSDs of individual domains of PEDV 3CL^{pro} compared with that of another four enzymes

Fig. 2. Structure alignment of individual domains of several coronavirus 3CL^{pro}s. A. Schematic overview of superimposed domains of one monomer from several coronavirus 3CL^{pro}s (PEDV 3CL^{pro}, PDB Entry: 4XFQ; SARS-CoV 3CL^{pro}, PDB Entry: 2H2Z; TGEV 3CL^{pro}, PDB Entry: 1LVO; PEDV 3CL^{pro} complex, PDB Entry: 4ZUH; HCoV 229E 3CL^{pro}, PDB Entry: 1P9S). B. RMSD values of individual domains of the PEDV 3CL^{pro} compared with that of four additional enzymes (computed using the PDBeFold service on the European Bioinformatics Institute website).

active site, and several small shifts are found mainly in domain III (Fig. 2(A)). The RMSD value (1.2 Å) of domain III between SARS-CoV 3CL^{pro} and PEDV 3CL^{pro} is higher than that (0.7 Å) between the two *Alphacoronavirus* 3CL^{pro}s and PEDV 3CL^{pro}. The RMSD values of the first two domains between different coronavirus 3CL^{pro}s are relative small (all around 0.7 Å) (Fig. 2(B)). These results reveal that SARS-CoV 3CL^{pro} (*Betacoronavirus*) demonstrates a relatively high number of structural differences compared with the *Alphacoronavirus* 3CL^{pro}s, especially in domain III; the first two domains (chymotrypsin-like cores) show higher degrees of conservation compared with that of domain III among different genera of coronavirus.

The imidazole ring of H41 of the catalytic dyad adopts a different orientation from that observed in most current coronavirus 3CL^{pro} structures. The imidazole ring of H41 is not in contact with the corresponding nucleophile C144 in PEDV 3CL^{pro} (Fig. S1). However, this configuration may be an artifact caused by the low pH (4.2) of crystallization. At this pH, the histidine residue may be protonated and become unable to interact with the sulfhydryl of C144. Such a pH-effect has been previously described for the 3C protease of enterovirus 68 (Tan et al., 2013).

2.2. The dimerization of PEDV 3CL^{pro}

Tablesimeric form has consistently been shown to be the catalytically active form for all coronavirus 3CL^{pro}s studied to date. At the dimer interface of PEDV 3CL^{pro}, the N-terminal finger (8 residues of the N-terminus) assembles into the dimerization surface, interacting with domain II of the opposite monomer and the C-terminal helix of the same monomer (Fig. 1(A)). A total surface

Table 1
Data and model statistics.

	PEDV 3CL ^{PRO}	Complex of PEDV 3CL ^{PRO} variant (C144A) bound to a peptide
Data collection		
Space group	P 1 2 ₁ 1	P 1 2 ₁ 1
Cell parameter(a, b, c(Å))	56.64, 91.06, 57.98	56.68, 91.88, 58.01
α, β, γ	90.00°, 100.25°, 90.00°	90.00°, 100.24°, 90.00°
Wavelength	0.97917	0.97917
Resolution range (Å)	27.87–1.65	31.78–2.40
% Completeness	99.7 (99.9)	99.8(99.9)
R _{merge} (last shell)	0.055(0.480)	0.103(0.487)
I/σ (last shell)	23.25(3.42)	18.10(4.55)
Redundancy (last shell)	3.7(3.7)	3.2(3.1)
Refinement		
Resolution (Å)	27.45–1.65	31.78–2.40
R _{work} /R _{free}	17.9/20.3	16.5/21.8
No. reflections	69,111	22,868
No of protein atoms	4514	4512
No. of solvent atoms	465	267
No. of ions/ligands	0	1
r.m.s.d.		
Bond length (Å)	0.007	0.008
Bond angle (Å)	1.126	1.100
B factor (Å²)		
Protein	25.0	29.37
Water	23.54	28.91
Ligand	33.49	33.60
		45.09
Ramachandran plot:	97.78%, 2.22%,	96.45%, 3.21%,
core,		
allow, disallow	0.00%	0.34%

Highest resolution values are written in parenthesis.

$R_{\text{merge}} = \frac{\sum \sum |I_i - \langle I \rangle|}{\sum \sum I_i}$; where I_i is the intensity measurement of reflection h and $\langle I \rangle$ is the average intensity from multiple observations.

$R_{\text{work}} = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$; where F_o and F_c are the observed and calculated structure factors respectively.

R_{free} is equivalent to R_{work} but where 5% of the measured reflections have been excluded from refinement and set aside for cross-validation.

area of 2085 Å² is buried upon dimerization (computed using PDBePISA tool, <http://pdbe.org/pisa/>). In this molecule, 16 residues from monomer A are found to interact (hydrogen bonds or salt bridges) with 16 residues from monomer B at the dimer interface (Table 2 and Fig. 3). All the residues are conserved among the three Alphacoronavirus 3CL^{PRO}s, except three substitutions for the A1,

Table 2
The comparison of residue-residue interactions at the dimer interface from PEDV 3CL^{PRO} with these from other 3CL^{PRO}s.

PEDV 3CL ^{PRO}		TGEV 3CL ^{PRO}		HCoV 229E 3CL ^{PRO}		SARS-CoV 3CL ^{PRO}	
Monomer A	Monomer B	Monomer A	Monomer B	Monomer A	Monomer B	Monomer A	Monomer B
1A	139F,165E	1S	139F, 163E	1A	139F,165E	1S	166E
2G	138S	2G	138S	2G	138S	2G	139S, 140F
4R	126G, 286E	4R	126G, 286E	4R	126G, 286E	4R	127Q, 139S, 290E
7A	124V	7A	124V	7A	124V	7A	125V
10S	10S	10S	10S	10S	10S, 14E	10S	10S
11G	14E	11G	14E	11G	14E	11G	14E
14E	11G	14E	11G	14E	10S, 11G	14E	11G
124V	7A	124V	7A	124V	7A	125V	7A
126G, 286E	4R	126G, 286E	4R	126G, 286E	4R	127Q, 290E	4R
138S	2G,295Q	138S	2G, 295Q	138S	2G,295Q	139S	2G, 299Q
139F, 165E	1A	139F, 165E	1S	139F, 165E	1A	166E	1S
279G	281T	279S	281G	281S	282S		
281T	276Q	281G	279S	282S	281S		
295Q	138S	295Q	138S	295Q	138S	299Q	139S
						298R	123S

The PDB identifiers for the four structures are 4XFQ, 1LVO, 2ZU2 and 1UK4. The hydrogen-bond and hydrophobic interactions are all analyzed in PyMOL (Schrodinger, 2006). For the three Alphacoronavirus 3CL^{PRO}s, the residue-residue interactions are conserved except the two interactions (in purple bold) between domain III of each monomer. However, these two interactions are not observed in SARS-CoV 3CL^{PRO}, and the interaction between 298R and 123S is not found in the Alphacoronavirus 3CL^{PRO}s.

G279 and T281 residues (Table 2). Meanwhile, extensively hydrophobic interactions appear at the dimer interface. The interactions are mainly observed between the N-terminal residues of one monomer and domain II and domain III of the opposite monomer. The intraprotomer interactions between the two N-terminal loops and between the two domain III are also involved in the dimerization (Fig. 3(B)).

To explore the relationship between oligomerization and enzymatic activity, four variants (R4A, S138A, R294A, and Q295A) and an N-finger deletion variant at the PEDV 3CL^{PRO} dimer interface were constructed. The elution peaks of the four variants show similar retention volumes to that of the wild type (WT) at 9.95 ml (approximately 54 kDa), which represents a dimer form similar to that of WT PEDV 3CL^{PRO} (Fig. 4(A), (B) and (D)). However, the elution peak of the N-finger deletion variant (Δ 8aa) shows a retention volume of 11.31 ml (approximately 30.4 kDa), which represents the monomer. The oligomerization of the proteins was further confirmed through sedimentation velocity experiments (Fig. 4(C) and (D)). The results were consistent with that of the gel filtration chromatography analysis. The majority (88.90%) of wild type PEDV 3CL^{PRO} exists as dimers, and few (5.24%) exists as monomers. The four variants still primarily exist as dimers, however the relative populations of the variants monomers, except S138A, increased slightly. The majority (90.92%) of the N-finger deletion variant exists as monomers. These data indicate that wild type PEDV 3CL^{PRO} primarily exists as dimers in solution and the engineered variants (R4A, S138A, R294A, and Q295A) show little effect on the monomer-dimer equilibrium. The N-finger deletion absolutely disrupts the dimerization.

FRET (Matayoshi et al., 1990) results demonstrate that the mutations reduced the catalytic activity to varying degrees (Fig. 4 (E) and (F)). Only the N-finger deletion variant exhibited a complete loss of activity. The large buried surface area is responsible for dimer association of PEDV 3CL^{PRO} and one residue substitution may not completely destroy the dimerization. The N-finger residues of coronavirus 3CL^{PRO} are key components of the buried interface and are essential for dimerization. Thus, the N-finger deletion mutation destroyed the dimer form and inactivated PEDV 3CL^{PRO}.

In SARS-CoV 3CL^{PRO}, a hydrophobic interaction between the side chains of M6 and Y126 of the opposite promoter greatly contributes to the stabilization of the dimer conformation (Wei et al., 2006). At the same time, the backbone oxygen of M6 and the

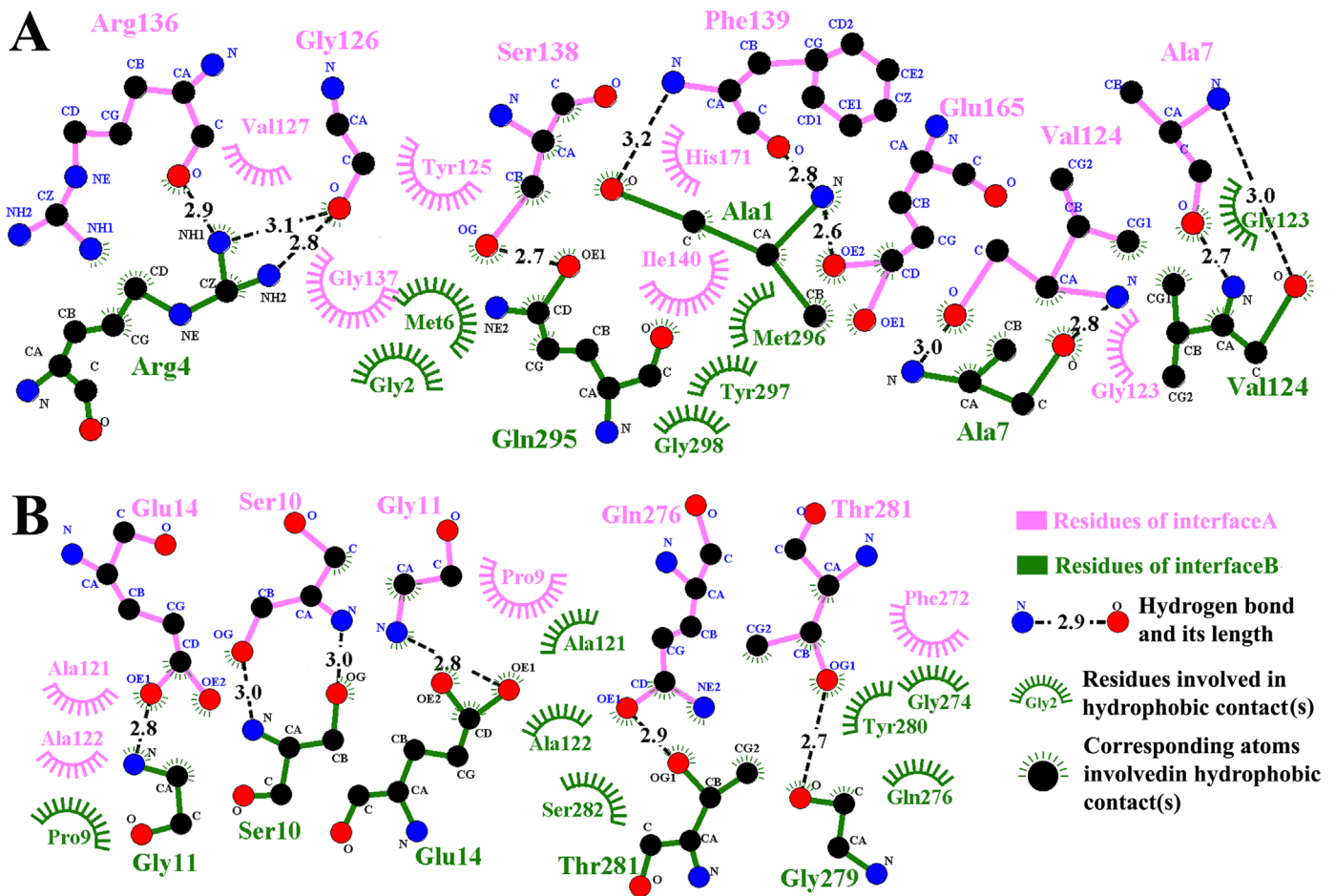


Fig. 3. A structural diagram of the distribution of hydrophobic and hydrophilic interactions at the dimer interface. A. The figure was generated using the LigPlot⁺ program. The residues of interface A and interface B are colored in pink and olive green, respectively. The carbon, nitrogen, and oxygen atoms are shown as black, blue and red circles, respectively. Hydrogen bonds are shown as black dashed lines labeled with the distance between the donor and corresponding acceptor atom. Hydrophobic interactions are demonstrated by arcs with spokes radiating toward the atoms (with spokes around) or residues (shown as arcs with spokes) they contact. B. Another interacting patch at the interface that is important for dimerization.

nitrogen atom of the side chain NH₂ of R298 in the same monomer form a hydrogen bond (2.94 Å), which can be further connected to the catalytic machinery of the opposite protomer through a two-step relay (Shi et al., 2008). Thus, R298 is very important for the catalytic activity of SARS-CoV 3CL^{pro}. As a result, the R298A mutation in the SARS-CoV 3CL^{pro} caused a dimer-monomer switch and inactivated the enzyme (Shi et al., 2008). However, the homologous variant (R294A) in PEDV 3CL^{pro} remained a dimer and an active enzyme (Fig. 4(E) and (F)). In the PEDV 3CL^{pro} structure, R294 shows no interactions with any residues from another monomer. The result shows that R294 is not a key component involved in the dimerization of PEDV 3CL^{pro}.

Despite high sequence and structure similarities, differences in dimerization exist among 3CL^{pro}s from different coronavirus genera. As in MERS-CoV 3CL^{pro}, non-conserved residues far from the dimer interface might be involved in dimer formation of a weakly associated dimer (Tomar et al., 2015). These structural differences may lead to differences in the dimerization and enzymatic properties of 3CL^{pro}s from different coronavirus groups.

2.3. Overall structure of the PEDV 3CL^{pro} C144A variant in complex with the peptide substrate

In order to investigate further the substrate recognition mechanism of PEDV 3CL^{pro}, we synthesized a peptide (TSAVLQ₁SGFRK) (Nanjing GenScript Company) for co-crystallization and determined

the crystal structure of the C144A variant complex at 2.2 Å resolution in the space group, *P*₂₁. Two 3CL^{pro} molecules are observed in the asymmetric unit, and the two monomers have the similar overall structures (RMSD of 0.25 Å for 186 superimposed Cα atoms of domain I and domain II, 0.92 Å for 99 superimposed Cα atoms of domain III). However, the first two domains show low RMSD values, it seems that they are more stable than the C-terminal helix domain. An eight amino acid peptide (SAVLQ₁SGF) is observed in only one monomer according to the density map (Fig. 5), but the substrate-binding sites of the two monomers are almost identical. The first two domains show almost no differences and domain III shows slight structure deviation. However, the substrate lies in the cleft between the first two domains. Thus, the absence of the peptide in another monomer may be caused by the poor density map or the crystal package. The four residues P4–P1 form an antiparallel β sheet with residues 162–166, and residues P5–P4 form an antiparallel β sheet on the other side with residues 188–190 of the long loop A, that links domain II and III (Fig. 5(A)). The P1' to P3' strand extends into the solvent, and P2'–G forms a short antiparallel strand interaction with A26.

2.4. Details of the substrate-binding sites of PEDV 3CL^{pro}

Coronavirus 3CL^{pro} recognizes a conserved residue (glutamine) at the P1 position for efficient hydrolysis (Hegyi et al., 2002; Ziebuhr and Siddell, 1999; Ziebuhr et al., 2000). The side chain of the

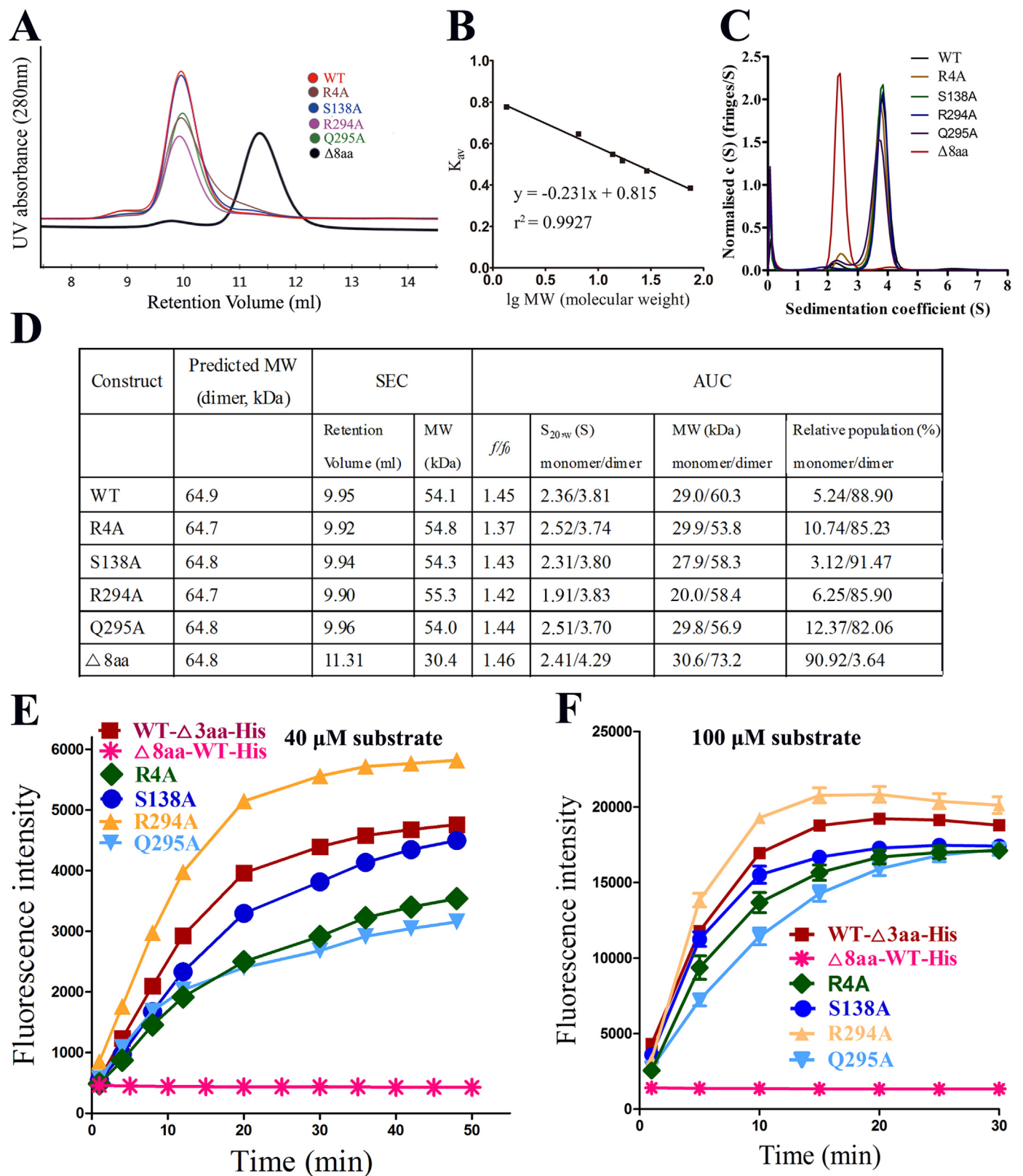


Fig. 4. Mutational studies of several residues at the dimer interface. **A.** Size exclusion chromatography (SEC) analyses of the four site variants and the N-terminal-deletion variant ($\Delta 8aa$). WT, $\Delta 8aa$, Q295A, R294A, S138A and R4A are shown in red, black, orange, blue, yellow and green, respectively. **B.** Calibration curve. Conalbumin, 75 kDa; carbonic anhydrase, 29 kDa; equine myoglobin, 17 kDa; ribonuclease A, 13.7 kDa; aprotinin, 6.5 kDa; vitamin B12, 1.35 kDa (Bio-Rad and GE Healthcare) were used to calibrate the column. The values of the Y-axis were calculated using the equation $K_{av} = (V_e - V_0)/(V_t - V_0)$. **C.** Sedimentation velocity analysis of the WT and variants. The major peak for each sample represents the major state of the protease. **D.** Calculated molecular masses based on the results of SEC and AUC assays, respectively. $\Delta 8aa$ primarily exists as monomers. **E.** FRET-based mutational studies of the residues that may be involved in the dimerization of PEDV 3CL^{pro}. The S138A, R4A, and Q295A variants exhibited reduced catalytic activity to various degrees (1, 0.74, 0.94, 1.22, 0.66, and 0.09, respectively; the activity of wild type PEDV 3CL^{pro} was taken as 1). None of the three variants caused a great or complete loss of activity, and the catalytic activity of the R294A variant exhibited no reduction. Only the N-finger deletion variant exhibited a complete loss of activity. **F.** Same as that in panel E but with different substrate concentration.

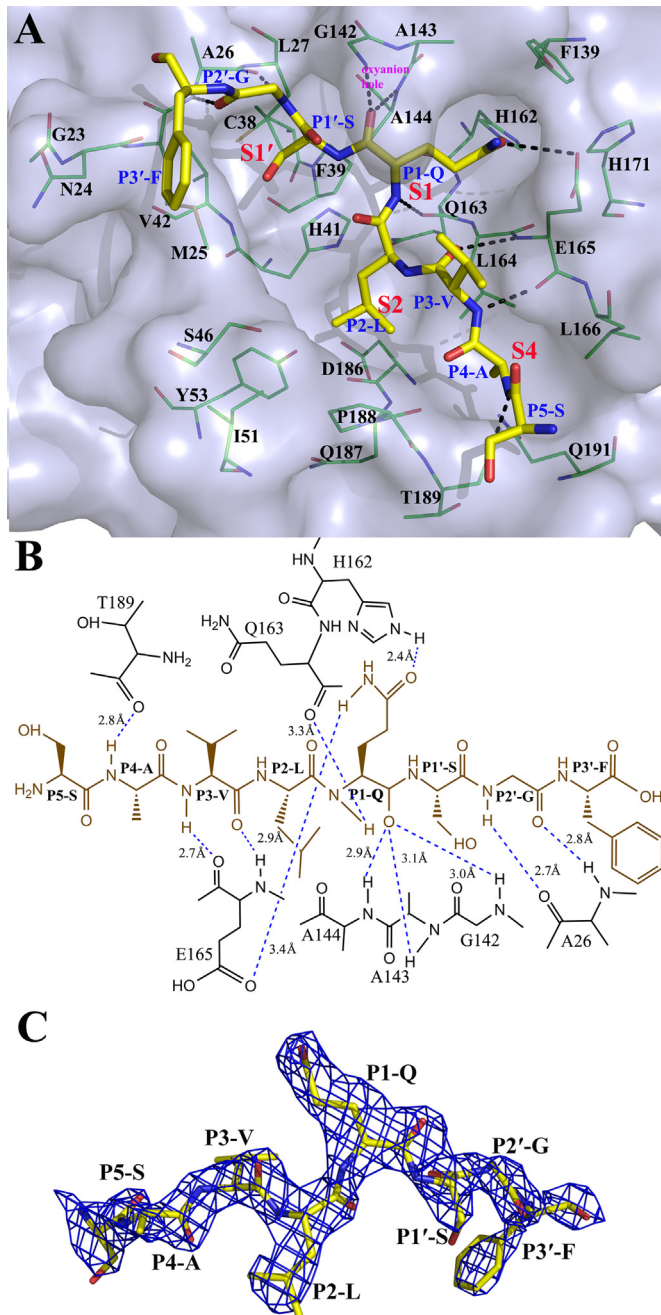


Fig. 5. Structure of a PEDV 3CL^{PRO} variant (C144A) bound to a peptide substrate. A. Three-dimensional structure of the substrate-binding pockets with a peptide. The residues of the pockets are represented in green as a stick diagram, and the substrate is shown in yellow as a stick diagram. The S1, S2, S4 and S1' pockets are labeled. The residues are all labeled. The nitrogen atoms are shown in blue, and oxygen atoms are shown in red. Hydrogen bond interactions are shown as black dashed lines. B. Diagram of the detailed molecular interactions between the substrate and the protease. The peptide substrate is shown in brown. Hydrogen bonds are shown as blue dashed lines, and the hydrogens were connected to their acceptors by the dashed lines. The distances labeled reflect the distances between the donor and the acceptor for all hydrogen bonds. Ceratina electron density map of the peptide substrate (2Fo-Fc, contoured at 1.0 σ).

conserved P1-Q fits comfortably in the pocket (Fig. 5(A)), stabilized by two hydrogen bonds: one between the N ϵ 2 atom of H162 and the O ϵ 1 atom of P1-Q and another between the side chain oxygen of E165 and the N ϵ 2 atom of P1-Q (Fig. 5(B)). The main chain amides of G142 and A144 form an oxyanion hole that stabilizes the carbonyl oxygen of P1-Q. The main chain nitrogen atom of P1-Q forms a hydrogen bond with the main chain oxygen of Q163. The

S2 sub-site of PEDV 3CL^{PRO} consists of P188, L164, I51, S46 and H41, similar to that of TGEV 3CL^{PRO}, which is lined with the side chains of P188, L164, I51, T47 and H41 (Anand et al., 2002). The side chains of these residues form a hydrophobic pocket that can accommodate a P2 residue with a large side chain, such as leucine or methionine (Fig. 5(A)). The main chain of P3-V forms two hydrogen bonds with the main chain of E165, and the side chain of P3-V is exposed to the solvent (Fig. 5(A)). Since the side chain of P3 Val is pointing to the surface and therefore to the solvent, the charged character of the P3 residue has no difference or preference. The P4 pocket consists of a series of hydrophobic amino acids and may prefer residues with high hydrophobicity. A hydrogen bond is found between the main chain NH of P4-A and the carbonyl oxygen of T189. The PEDV 3CL^{PRO} S1' sub-site is composed of M25, A26, L27, C38, P39, H41 and V42. Small residues such as serine, glycine and alanine are preferred in the small S1' pocket. The main chain of P2'-G forms two hydrogen bonds with the main chain atoms of A26 (Fig. 5(B)). Due to the poor density of P3'-F, its carbonyl group and the side chain phenyl group could easily be interchanged during structure refinement. The conservation in the substrate binding pockets provides a basic model for the design of potential antiviral agents (Yang et al., 2005). In the PEDV 3CL^{PRO} structure, the imidazole ring of H41 forms an abnormal conformation due to the low crystallization pH. However, in the complex structure of the active-site variant (C144A), the H41 imidazole ring is recovered (Fig. S1). The entire imidazole ring lies parallel with the side chain of P2-L and interacts through hydrophobic effects, and the N ϵ 2 of H41 becomes closer (3.3 Å) to the peptide bond between P1-Gln and P1'-Ser.

Part of the detailed interaction network is shown in Fig. 6(A). A water molecule is hydrogen-bonded to three residues (H41, Q163 and D186). Notably, the buried water molecule takes the place that is normally occupied by the side chain of D102 in the catalytic triad of the chymotrypsin and other members of this serine protease family. The oxyanion hole that stabilizes the main chain oxygen of P1-Q is made up of the main chain amides of G142 and A144. To determine whether the residues around the catalytic dyad are essential for hydrolytic activity, we performed proteolytic activity analyses of the variants (H41A, C144A, C144S, H162A and D186A). The results demonstrate that the fluorescence values of WT PEDV 3CL^{PRO} increase in a time-dependent manner, whereas none of the variants exhibit enzymatic activity (Fig. 6(B)). This finding indicates that the catalytic dyad of PEDV 3CL^{PRO} is essential for its proteolytic activity. Additionally, H162 and D186 in the substrate binding site are both essential, which are conserved among coronavirus 3CL^{PRO}s (Fig. 1(B)).

A His-tag was added to the C-terminus during construction. Unfortunately, we could not see the spatial orientation of the tag due to the poor density map. The WT- Δ 3aa variant without a His-tag was also constructed, and the protease performed similar enzymatic activity to that of the WT- Δ 3aa-His variant (data not shown). Therefore, the tag may not cause major effect on both the activity and on the conformation of the enzyme.

2.5. None-conserved motifs cause differences in substrate specificity

In general, Coronavirus 3CL^{PRO}s recognize a conserved cleavage site. To determine whether the two proteases from the first two coronavirus genera have differences in substrate specificities, two fluorogenic peptide substrates (Dabcyl-KTSAVLQ↓SGFRKME-Edans and Dabcyl-YNSTLQ↓AGLRKME-Edans contain the N-terminal auto-cleavage sites of PEDV and SARS-CoV 3CL^{PRO}s, respectively) were introduced in the FRET assays. In addition to processing viral polyproteins, coronavirus 3CL^{PRO} has the potential to cleave host proteins during virus-host interactions. PEDV 3CL^{PRO} regulates its interferon antagonism by cleaving nuclear transcription factor-

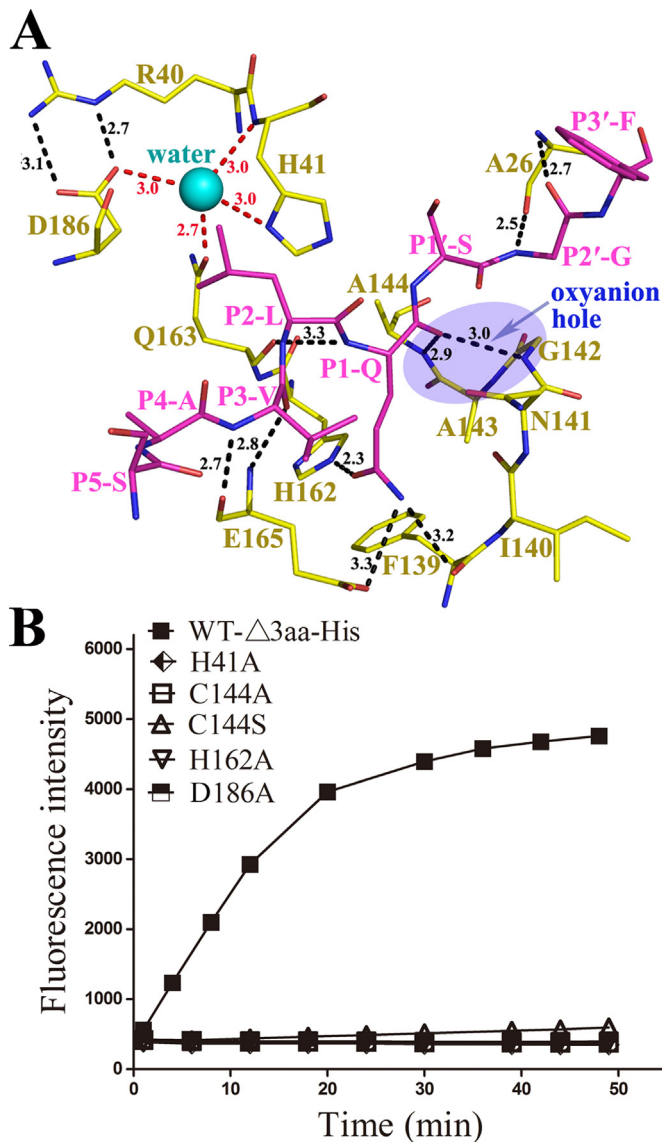


Fig. 6. Key residues for enzyme catalytic activity. A. A structural diagram of several key residues involved in enzyme catalytic activity. The residues of the protease are colored yellow, and the substrate is colored magenta. The residues of the substrate are labeled and shown as stick diagrams. Hydrogen bond interactions are shown as dashed lines, along with the bond distance. A water molecule buried by three residues (H41, Q163 and D186) is shown as a blue sphere. The oxyanion hole region is shown as a subtransparent blue oval. B. Fluorescence profiles of the hydrolysis of the fluorogenic substrate by WT- Δ 3aa-His and the variants of PEDV 3CL^{PRO}. The enhanced fluorescence intensity was monitored for 60 min. All five site-directed variants showed nearly no activity compared with WT- Δ 3aa-His.

kappaB (NF- κ B) essential modulator (NEMO) (Wang et al., 2016). The third fluorogenic peptide substrate (DabcyI-KLAQLQ_↓VAYHQE-Edans), which contains the cleavage site derived from NEMO was also applied in the cleavage assays. Both the proteases are able to cleave the two 3CL^{PRO} substrates and prefer the SARS-CoV 3CL^{PRO}-derived substrate; however PEDV 3CL^{PRO} demonstrates better hydrolysis efficiency than SARS-CoV 3CL^{PRO} (Fig. 7(A)). Surprisingly, SARS-CoV 3CL^{PRO} is unable to cleave the NEMO-derived substrate, whereas PEDV 3CL^{PRO} shows efficient hydrolysis (Fig. 7 (B)). Similar results were obtained when the concentration of the proteases increased by 10-fold (data not shown).

To understand the inability of SARS-CoV 3CL^{PRO} to cleave the NEMO-derived substrate, we analyzed its substrate preference at each sub-site according to the previous study. Each residue of “KLAQLQ_↓VAYHQE” is preferred at the corresponding sub-sites of

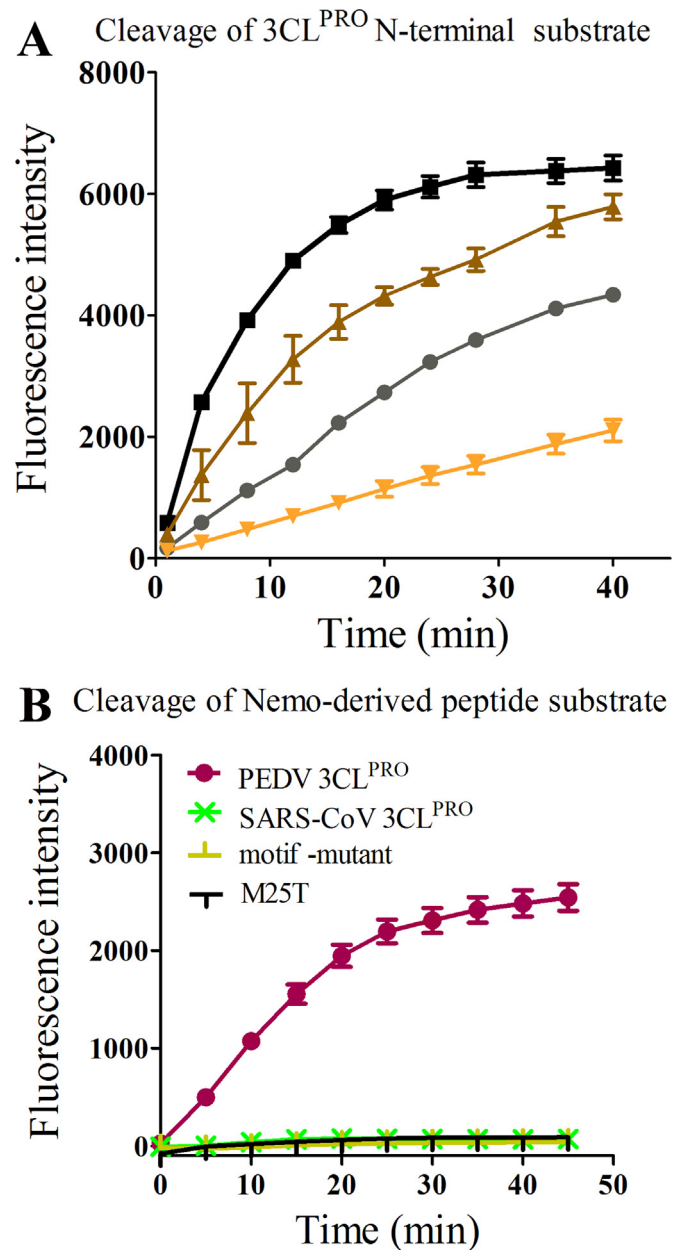


Fig. 7. Peptide substrate cleavage assays. A. Cleavage of coronavirus 3CL^{PRO} N-terminal substrate. Two fluorogenic peptide substrates (DabcyI-KTSAVLQ_↓SGFRKME-Edans and DabcyI-YNSTLQ_↓AGLRKME-Edans contain the N-terminal auto-cleavage site of PEDV and SARS-CoV 3CL^{PRO}, respectively) were introduced in the FRET assays. (□: SARS-CoV 3CL^{PRO}-derived substrate cleaved by PEDV 3CL^{PRO}; ○: PEDV 3CL^{PRO}-derived substrate cleaved by PEDV 3CL^{PRO}; △: SARS-CoV 3CL^{PRO}-derived substrate cleaved by SARS-CoV 3CL^{PRO}; ▽: PEDV 3CL^{PRO}-derived substrate cleaved by SARS-CoV 3CL^{PRO}). B. Cleavage of NEMO-derived substrate. Similar results were obtained when the concentration of the proteases increased by 10-folds (data not shown). Four proteases used in each assay were labeled in the figure. Motif-mutant and M25T represent two PEDV 3CL^{PRO} variants described in the text. Same concentration of the NEMO-derived substrate was used for the four proteases.

SARS-CoV 3CL^{PRO}, with the exception of the P1'-V, which prefers small residues with side chain volumes of less than 50 Å³ (Ser; Ala; Cys) at the S1' sub-site (Chuck et al., 2010; Chuck et al., 2011). The PEDV 3CL^{PRO} S1' sub-site is composed of M25, A26, L27, C38, P39, H41 and V42, and those of SARS-CoV 3CL^{PRO} are T25, T26, L27, C38, P39, H41 and V42. The hydrophobic side chain of M25 may form hydrophobic interactions with the side chain of P1'-V, which may lead to the tolerance of P1'-V of PEDV 3CL^{PRO}. To verify the hypothesis, we replaced the whole motif with that of SARS-CoV

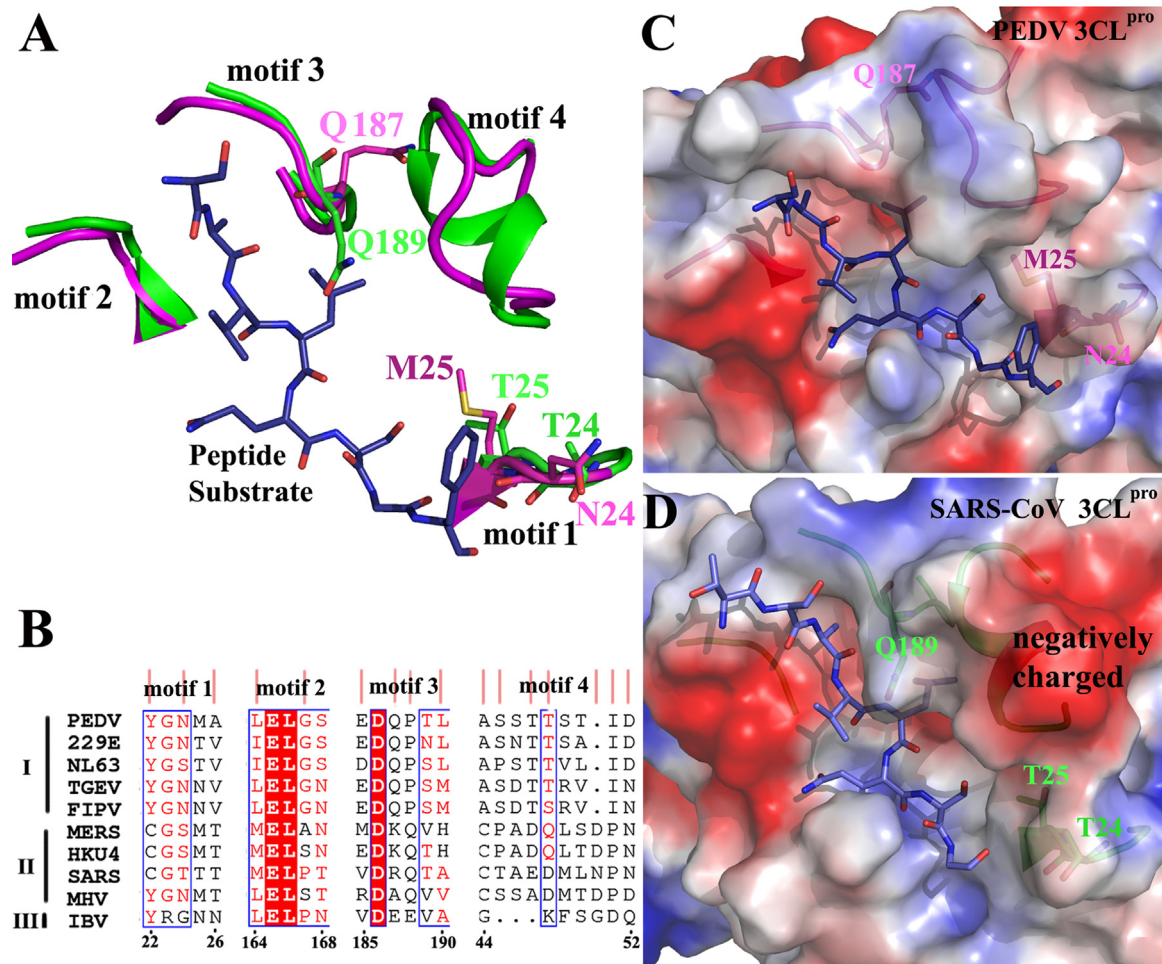


Fig. 8. Comparison of the potential substrate-binding motifs of the PEDV and SARS-CoV 3CL^{pro}s. **A.** A cartoon view of the four motifs around the substrate. The PEDV 3CL^{pro} complex structure (magenta) is superimposed with the complex structure of the SARS-CoV 3CL^{pro} variant (C144A) (PDB identifier 2Q6G) (green). The peptide substrate is colored blue. The four motifs and some residues shown as stick diagrams are labeled. **B.** Sequence alignment of the four motifs of 3CL^{pro}s of the three coronavirus groups. The residues that are relatively conserved in *Alphacoronavirus* 3CL^{pro} but differ from those of *Betacoronavirus* 3CL^{pro} are marked with pink vertical lines. **C.** Electrostatic surface potential of the PEDV 3CL^{pro} complex structure. The four motifs are shown as a magenta cartoon and the substrate is shown as blue sticks. Three residues shown as stick diagrams are labeled. **D.** Electrostatic surface potential of the SARS-CoV 3CL^{pro} complex structure. The four motifs are shown as a green cartoon and the substrate is shown as blue sticks. Three residues shown as stick diagrams are labeled, and the negatively charged area (red) of motif4 is labeled.

3CL^{pro} (“RVCYGNMA”19–26“QVTCGTTT”) and also performed the point mutation M25T. It was confirmed that both the variants performed obvious cleavage of the SARS-CoV 3CL^{pro}-derived substrate (data not shown); however, they failed to cleave the NEMO-derived substrate (Fig. 7(B)). It indicates that the motif (19–26), especially the M25 residue, is important for the cleavage of the NEMO-derived substrate by PEDV 3CL^{pro}. And PEDV 3CL^{pro} appears to bear residues with larger side chains at the S1' sub-site than SARS-CoV 3CL^{pro} does.

The conserved residues of the sub-sites result in relatively conserved substrate specificity of coronavirus 3CL^{pro}s. However, the non-conserved motifs of the pockets may also influence the recognition of the substrate. Except for the motif 1 (residues 22–26) mentioned above, three more motifs are found partly different between PEDV and SARS-CoV 3CL^{pro}s (PDB identifier 2Q6G) (Fig. 8). In motif 1, the two residues (N24 and M25) of PEDV 3CL^{pro} carry different side chains than those (T24 and T25) of SARS-CoV 3CL^{pro}, which result in differences in the pockets mouth (Fig. 8 (C) and (D)). Motif 2 of the two proteases consists of partly different residues. In motif 3, the aspartic acid side chains exhibit opposite orientations, which also result in differences in the pocket mouths. The residues of motif 4 obviously differ between the two proteases; they exhibit nearly no sequence homology (Fig. 8(B)), and the SARS-CoV 3CL^{pro} motif 4 contains two

negatively charged residues (E47 and D48) (Fig. 8(B) and (C)). We further analyzed the four motifs sequences between *Alphacoronavirus* and *Betacoronavirus* 3CL^{pro}s (Fig. 8(B)). The four motifs are relatively well conserved in *Alphacoronavirus* 3CL^{pro}s and differ from those of *Betacoronavirus*. These differences may also lead to the different sizes or depths of the pockets, thus causing the differences in recognition of the substrates.

3. Conclusion

In summary, we have determined the crystal structure of PEDV 3CL^{pro} alone and in complex with a peptide substrate that contains the N-terminal cleavage site. The strategy for mimicking the peptide substrate side chains to accommodate the corresponding sub-sites can be utilized to design inhibitors (Yang et al., 2005). Thus, our structures may provide the basis for the design of inhibitors against PEDV 3CL^{pro}. The biochemical and structural analyses revealed that PEDV 3CL^{pro} forms a tight dimer, and no variants caused a dimer-to-monomer switch except the N-finger deletion variant. The interaction networks involved in dimer formation of 3CL^{pro} are relatively well conserved among *Alphacoronavirus* but somewhat different from those of SARS-CoV 3CL^{pro} in *Betacoronavirus*. Finally, we find that non-conserved motifs in the pockets cause different cleavage of

substrate between the PEDV and SARS-CoV 3CL^{pro}s; we further analyzed the differences of the motifs between 3CL^{pro}s from *Alphacoronavirus* and *Betacoronavirus*.

4. Materials and methods

4.1. Cloning, protein expression and purification

The PEDV nsp5 (encoding 3CL^{pro}) gene was cloned from the viral cDNA of the FJZZ strain (GenBank: KC140102.1) via PCR. The forward and reverse primers contained NdeI and XhoI restriction sites, respectively. The PCR product contained a C-terminal His-tag for further protein purification. The PCR product was cloned into the pET42b expression vector. Multiple truncation recombinant plasmids and variants were also constructed. NcoI and XhoI restriction sites were used for the pET28a vector. BamHI and XhoI restriction sites were used for the pET28a-sumo vector. The recombinant WT- Δ 3aa-His (the last three amino acids, NLQ, were deleted, and a His-tag was added to the C-terminus) plasmid was used as a template for the following variants: M25T, H41A, H162A, C144A, C144S, D186A and the motif variant. The recombinant His-sumo-WT (a His-sumo fusion tag was added to the N-terminus) plasmid was used as a template for the following site mutations: R4A, S138A, R294A and Q295A. Then, overlap-extension PCR was performed (Ho et al., 1989). All of the recombinant expression plasmids were sequenced, and no unexpected mutations occurred.

For protein expression, the recombinant plasmids were transformed into *E. coli* BL21 (DE3) and then cultured at 37 °C in LB medium until the OD₆₀₀ reached 0.6–0.8. Then, 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) was added to induce protein expression. The cells were harvested after incubation at 27 °C for 7 h, resuspended in PBS and disrupted (ATS AH-1500). The supernatant was filtered and loaded onto a His Trap™ HP column (GE Healthcare), and the C-terminal His-tagged protein was finally eluted using a linear gradient between the binding buffer and elution buffer A (20 mM Tris, pH 7.4, 500 mM NaCl, and 500 mM imidazole), followed by another 30 ml of 100% elution buffer A. The target protein was further purified by a 120 ml Superdex 200 (GE Healthcare) column with elution buffer B (20 mM Tris, pH 7.4, 200 mM NaCl). The N-terminal His-sumo-tagged protease was digested with sumo protease to remove the tag and then loaded onto a 120 ml Superdex 200 column. The proteins were again eluted with elution buffer B and concentrated after SDS-PAGE analysis. All of the purification procedures were performed at 4 °C to avoid unexpected degradation.

4.2. Crystallization and structure determination

PEDV 3CL^{pro} was crystallized via the sitting drop vapor diffusion method at 20 °C. The best crystals were produced under the following conditions: 0.1 M NaCl, 0.2 M Na₂HPO₄, citric acid (pH 4.2) and 5% (w/v) PEG 3350. PEDV 3CL^{pro} variant (C144A) crystals were obtained under the same conditions. The 11-amino-acid peptide substrate (TSAVLQ↓SGFRK, that contains the N-terminal auto-cleavage site of SARS-CoV 3CL^{pro}) was dissolved at a 20 mM in elution buffer B as a stock solution. For crystal soaking, the solution was diluted to 10 mM and then added to the variant (C144A) crystallization drop at an equal volume. The crystals were soaked for 14 h before data collection. The single crystals were first washed with 5%, 10%, 15%, and 30% ethylene glycol (v/v) as a cryoprotectant and then flash-frozen in liquid nitrogen. All data collection was performed at beamline BL17U at the Shanghai Synchrotron Radiation Facility (SSRF) using a MAR 225 CCD detector (MAR Research). All of the obtained data sets were indexed, integrated and scaled using HKL-3000 (Otwinowski and Minor,

1997). The structure was solved by molecular replacement with PHASER (McCoy et al., 2007) using the structure of human coronavirus 229E 3CL^{pro} (PDB identifier 1P9S) as a starting model. The structure of the complex was determined via molecular replacement using the previously determined PEDV 3CL^{pro} structure as the search model. Manual model building was performed using Coot (Emsley and Cowtan, 2004), and the structure was refined with Phenix (Adams et al., 2002). Refinement statistics are shown in Table 1. All of the structural figures were drawn using PyMOL (Schrödinger, 2006).

4.3. FRET-based assays for enzymatic characteristics

Based on the N-terminal cleavage site of PEDV 3CL^{pro}, we designed the peptide substrate Dabcyl-YNSTLQ↓AGLRKM-E-Edans (Nanjing GenScript Company). The two fluorophores formed a quenching pair and exhibited fluorescence resonance energy transfer (FRET) within the peptide (Matayoshi et al., 1990). The increase in fluorescence upon cleavage of the fluorogenic peptide substrate was monitored every minute for 1 h at 485 nm, with excitation at 340 nm, using a fluorescence plate reader (Kuo et al., 2004). The WT- Δ 3aa-His protein was used in the assays. All reactions were performed in 20 mM HEPES, 50 mM NaCl, 0.4 mM EDTA, 30% glycerol and 4 mM DTT at pH 8.0 in a total volume of 100 μ l. The enzyme concentration used in the FRET assay was 200 nM, and the substrate concentration was 0–200 μ M. The initial rates were used to calculate the kinetic parameters by fitting with the Michaelis-Menten equation using GraphPad Prism5. For the variants, all of the assay conditions were identical, but the assays were performed at a fixed substrate concentration.

4.4. Size-exclusion chromatography (SEC) and analytical ultracentrifugation (AUC) analysis

The residue-residue interactions at the dimer interface of PEDV 3CL^{pro} are listed in Table 2. Four variants (R4A, S138A, R294A and Q295A) and an N-finger deletion variant (Δ 8aa) were selected for oligomeric state analysis. All of the mutated proteins were stored in elution buffer B at –80 °C. A Superdex 75 10/300 GL column (GE Healthcare) was used for the analysis of the WT and protein variants. The column was equilibrated with elution buffer B (20 mM Tris, pH 7.4, and 200 mM NaCl), and 250 μ g of protein was then loaded onto the column and eluted with elution buffer B. Equal volumes of size exclusion standards (Conalbumin, 75 kDa; carbonic anhydrase, 29 kDa; equine myoglobin, 17 kDa; ribonuclease A, 13.7 kDa; aprotinin, 6.5 kDa; vitamin B12, 1.35 kDa) (Bio-Rad and GE Healthcare) were used to calibrate the column. The 280 nm absorbance peaks were generated and overlaid using Bio-Rad NGC software.

To further confirm the oligomeric state and determine the effect of the variants on the monomer-dimer equilibrium of PEDV 3CL^{pro}, sedimentation velocity experiments were conducted at 18 °C on the XL-A instrument (Beckman) at 50,000 rpm. Proteins were prepared at the concentration of approximately 0.9 mg/ml in elution buffer B. The sedimentation boundary was monitored every 3 min for a total of 110 scans using the absorbance optics of 280 nm. The software Sedfit was used to fit the data to a distribution of Lamm equation solutions c(s) model (Schuck, 2000).

Accession numbers

Coordinate and structure factors have been submitted to the PDB (accession numbers 4XFQ and 4ZUH).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 31372440) and the Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (Program Nos. 2012RC008, 2013PY031 and 2662015JQ003). We thank Joe Pasquarella and Dr. Fang Li for discussion and the staff at the SSRF BL-17U beamline for assistance in data collection. Moreover, we also thank research associates at Center for Protein Research (CPR), Huazhong agricultural University, for technical support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.virol.2016.04.018>.

References

- Adams, M.J., Carstens, E.B., 2012. Ratification vote on taxonomic proposals to the international committee on taxonomy of viruses. *Arch. Virol.* 157, 1411–1422.
- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., Terwilliger, T.C., 2002. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1948–1954.
- Anand, K., Palm, G.J., Mesters, J.R., Siddell, S.G., 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.
- Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R., Hilgenfeld, R., 2003. Coronavirus main proteinase (3CL^{pro}) structure: basis for design of anti-SARS drugs. *Science* 300, 1763–1767.
- Chuck, C.P., Chong, L.T., Chen, C., Chow, H.F., Wan, D.C., Wong, K.B., 2010. Profiling of substrate specificity of SARS-CoV 3CL. *Plos One* 5, e13197.
- Chuck, C.P., Chow, H.F., Wan, D.C., Wong, K.B., 2011. Profiling of substrate specificities of 3C-like proteases from group 1, 2a, 2b, and 3 coronaviruses. *Plos One* 6, e27228.
- Collin, E.A., Anbalagan, S., Okda, F., Batman, R., Nelson, E., Hause, B.M., 2015. An inactivated vaccine made from a U.S. field isolate of porcine epidemic disease virus is immunogenic in pigs as demonstrated by a dose-titration. *BMC Vet. Res.* 11, 62.
- Chan, J.F.W., Lau, S.K.P., Woo, P.C.Y., 2013. The emerging novel Middle East respiratory syndrome coronavirus: the “knowns” and “unknowns”. *J. Formos. Med. Assoc.* 112, 372–381.
- Chen, S., Zhang, J., Hu, T., Chen, K., Jiang, H., Shen, X., 2008. Residues on the dimer interface of SARS coronavirus 3C-like protease: dimer stability characterization and enzyme catalytic activity analysis. *J. Biochem.* 143, 525–536.
- de Groot, R.J., Baker, S.C., Baric, R., Enjuanes, L., Gorbalenya, A.E., Holmes, K.V., Perlman, S., Poon, L., Rottier, P.J.M., Talbot, P.J., Woo, P.C.Y., Ziebuhr, J., 2012. Family Coronaviridae. In: King, A.M.Q., Lefkowitz, E., Adams, M.J., Carstens, E.B. (Eds.), Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, Oxford, pp. 806–828.
- Emmsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Garwes, D.J., 1988. Transmissible gastroenteritis. *Vet. Rec.* 122, 462–463.
- Hegyí, A., Friebe, A., Gorbalenya, A.E., Ziebuhr, J., 2002. Mutational analysis of the active centre of coronavirus 3C-like proteases. *J. Gen. Virol.* 83, 581–593.
- Hegyí, A., Ziebuhr, J., 2002. Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* 83, 595–599.
- Hilgenfeld, R., 2014. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J.* 281, 4085–4096.
- Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R., 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77, 51–59.
- Hsu, W.-C., Chang, H.-C., Chou, C.-Y., Tsai, P.-J., Lin, P.-I., Chang, G.-G., 2005. Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease. *J. Biol. Chem.* 280, 22741–22748.
- Kuo, C.-J., Chi, Y.-H., Hsu, J.T.-A., Liang, P.-H., 2004. Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. *Biochem. Biophys. Res. Commun.* 318, 862–867.
- Liu, D.X., Brown, T.D., 1995. Characterisation and mutational analysis of an ORF 1a-encoding proteinase domain responsible for proteolytic processing of the infectious bronchitis virus 1a/1b polyprotein. *Virology* 209, 420–427.
- Lu, Y., Denison, M.R., 1997. Determinants of mouse hepatitis virus 3C-like proteinase activity. *Virology* 230, 335–342.
- Li, C., Qi, Y., Teng, X., Yang, Z., Wei, P., Zhang, C., Tan, L., Zhou, L., Liu, Y., Lai, L., 2010. Maturation mechanism of severe acute respiratory syndrome (SARS) coronavirus 3C-like proteinase. *J. Biol. Chem.* 285, 28134–28140.
- Lu, Y., Lu, X., Denison, M.R., 1995. Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *J. Virol.* 69, 3554–3559.
- Marianayagam, N.J., Sunde, M., Matthews, J.M., 2004. The power of two: protein dimerization in biology. *Trends Biochem. Sci.* 29, 618–625.
- Matayoshi, E.D., Wang, G.T., Krafft, G.A., Erickson, J., 1990. Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer. *Science* 247, 954–958.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.
- Needle, D., Lountos, G.T., Waugh, D.S., 2015. Structures of the Middle East respiratory syndrome coronavirus 3C-like protease reveal insights into substrate specificity. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 71, 1102–1111.
- Oldham, J. Letter to the editor. *Pig Farming*, 1972, 10, 723.
- Otwinowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326.
- Pasick, J., Berhane, Y., Ojick, D., Maxie, G., Embury-Hyatt, C., Swekla, K., Handel, K., Fairles, J., Alexandersen, S., 2014. Investigation into the role of potentially contaminated feed as a source of the first-detected outbreaks of porcine epidemic diarrhoea in Canada. *Transbound. Emerg. Dis.* 61, 397–410.
- Pensaert, M.B., de Bock, P., 1978. A new coronavirus-like particle associated with diarrhoea in swine. *Arch. Virol.* 58, 243–247.
- Robert, X., Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324.
- Schechter, I., Berger, A., 1967. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27, 157–162.
- Song, D., Moon, H., Kang, B., 2015. Porcine epidemic diarrhoea: a review of current epidemiology and available vaccines. *Clin. Exp. Vaccin Res.* 4, 166–176.
- Shi, J., Sivaraman, J., 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.
- Song, D., Park, B., 2012. Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines. *Virus Genes* 44, 167–175.
- Schuck, P., 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* 78, 1606–1619.
- Schrödinger, L.L.C., 2006. The PyMOL Molecular Graphics System, Version 0.99r1.
- Tan, J.Z., George, S., Kusov, Y., Perbandt, M., Anemuller, S., Mesters, J.R., Norder, H., Coutard, B., Lacroix, C., Leyssen, P., Neyts, J., Hilgenfeld, R., 2013. 3C Protease of enterovirus 68: structure-based design of michael acceptor inhibitors and their broad-spectrum antiviral effects against picornaviruses. *J. Virol.* 87, 4339–4351.
- Thiel, V., Herold, J., Schelle, B., Siddell, S.G., 2001. Viral replicase gene products suffice for coronavirus discontinuous transcription. *J. Virol.* 75, 6676–6681.
- Thiel, V., Ivanov, K.A., Putics, A., Hertzog, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E.J., Rabenau, H., Doerr, H.W., Gorbalenya, A.E., Ziebuhr, J., 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* 84, 2305–2315.
- Tomar, S., Johnston, M.L., John, St.S.E., Osswald, H.L., Nyalapatla, P.R., Paul, L.N., Ghosh, A.K., Denison, M.R., Mesecar, A.D., 2015. Ligand-induced dimerization of MERS coronavirus nsp5 protease (3CL^{pro}): implications for nsp5 regulation and the development of antivirals. *J. Biol. Chem.*
- Vlasova, A.N., Marthaler, D., Wang, Q.H., Culhane, M.R., Rossow, K.D., Rovira, A., Collins, J., Saif, L.J., 2014. Distinct characteristics and complex evolution of PEDV strains, North America, May 2013–February 2014. *Emerg. Infect. Dis.* 20, 1620–1628.
- Wang, L.Y., Byrum, B., Zhang, Y., 2014b. New variant of porcine epidemic diarrhoea virus, United States, 2014. *Emerg. Infect. Dis.* 20, 917–919.
- Wang, D., Fang, L.R., Shi, Y.L., Zhang, H., Gao, L., Peng, G.Q., Chen, H.C., Li, K., Xiao, S. B., 2016. Porcine epidemic diarrhoea virus 3C-Like protease regulates its interferon antagonism by cleaving NEMO. *J. Virol.* 90, 2090–2101.
- Wei, P., Fan, K.Q., Chen, H., Ma, L., Huang, C.K., Tan, L., Xi, D., Li, C.M., Liu, Y., Cao, A.N., Lai, L.H., 2006. The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochem. Biophys. Res. Commun.* 339, 865–872.
- Weiss, S.R., Navas-Martin, S., 2005. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol. Mol. Biol. Rev.* 69, 635–+.
- Wood, E., 1977. An apparently new syndrome of porcine epidemic diarrhoea. *Vet. Rec.* 100, 243–244.
- Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., 2005. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *Plos Biol.* 3, e324.
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci.* 100, 13190–13195.
- Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E., 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* 81, 853–879.
- Ziebuhr, J., Siddell, S.G., 1999. Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab. *J. Virol.* 73, 177–185.
- Ziebuhr, J., 2005. The coronavirus replicase. *Curr. Top. Microbiol. Immunol.* 287, 57–94.