Characterization of Self-processing Activities and Substrate Specificities of Porcine Torovirus 3C-like Protease

Shangen Xu[a,b], Junwei Zhou[a,b], Yingjin Chen[a,b], Xue Tong[a,b], Zixin Wang[a,b], Jiahui Guo[a,b], Jiyao Chen[a,b], Liurong Fang[a,b], Dang Wang[a,b#], Shaobo Xiao[a,b#]

[a]State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, China

[b]The Key Laboratory of Preventive Veterinary Medicine in Hubei Province, Cooperative Innovation Center for Sustainable Pig Production, Wuhan 430070, China

[#]Corresponding authors. Laboratory of Animal Virology, College of Veterinary Medicine, Huazhong Agricultural University, 1 Shi-zi-shan Street, Wuhan 430070, Hubei, PR China. E-mails: wangdang@mail.hzau.edu.cn (Dang Wang); vet@mail.hzau.edu.cn (Shaobo Xiao)

Running title: ***Characterization of PToV 3CL^{pro}***

Abstract word count: 225

Main text word count: 7053

ABSTRACT

The 3C-like protease ($3CL^{pro}$) of Nidovirus plays an important role in viral replication and manipulation of host antiviral innate immunity, which makes it an ideal antiviral target. Here, we characterized that porcine torovirus (PToV, family Tobaniviridae, order Nidovirales) $3CL^{pro}$ autocatalytically release itself from the viral precursor protein by self-cleavage. Site-directed mutagenesis suggested that PToV $3CL^{pro}$, as a serine protease, employed His53 and Ser160 as the active-site residues. Interestingly, unlike most nidovirus $3CL^{pro}$, the P1 residue plays a less essential role in N-terminal self-cleavage of PToV $3CL^{pro}$. Substituting either P1 or P4 residue of substrate alone has little discernible effect on N-terminal cleavage. Notably, replacement of the two residues together completely blocks N-terminal cleavage, suggesting that N-terminal self-cleavage of PToV $3CL^{pro}$ is synergistically affected by both P1 and P4 residues. Using a cyclized luciferase-based biosensor, we systematically scanned the polyproteins for cleavage sites, and identified (FXXQ↓A/S) as the main consensus sequences. Subsequent homology modeling and biochemical experiments suggested that the protease formed putative pockets S1 and S4 between the substrate. Indeed, mutants of both predicted S1 (D159A, H174A) and S4 (P62G/L185G) pockets completely lost the ability of cleavage activity of PToV $3CL^{pro}$. In conclusion, the characterization of self-processing activities and substrate specificities of PToV $3CL^{pro}$ will offer helpful information for the mechanism of nidovirus 3C-like proteinase's substrate specificities and the rational development of the anti-nidovirus drugs.

KEYWORDS: Porcine torovirus, 3C-like protease ($3CL^{pro}$), self-processing activity, substrate specificity, pocket

IMPORTANCE

Currently, the active-site residues and substrate specificities of 3C-like protease (3CL^{pro}) differ among nidoviruses, and the detailed catalytic mechanism remains largely unknown. Here, porcine torovirus (PToV) 3CL^{pro} cleaves 12 sites in the polyproteins, including its N- and C-terminal self-processing sites. Unlike coronaviruses and arteriviruses, PToV 3CL^{pro} employed His53 and Ser160 as the active-site residues that recognize a Glutamine (Gln) at the P1 position. Surprisingly, mutations of P1-Gln impaired the C-terminal self-processing but did not affect N-terminal self-processing. The "noncanonical" substrate specificity for its N-terminal self-processing was attributed to the Phenylalanine (Phe) residue at the P4 position in the N-terminal site. Furthermore, a double glycine (neutral) substitution at the putative P4-Phe-binding residues (P62G/L185G) abolished the cleavage activity of PToV 3CL^{pro} suggested the potential hydrophobic force between the PToV 3CL^{pro} and P4-Phe side chains.

Introduction

Toroviruses (order Nidovirales; suborder Tornidovirineae; family Tobaniviridae; subfamily Torovirinae; genus Torovirus) are enveloped positive-stranded RNA viruses that cause a variety of diseases in animals and humans, with genomes ranging from 23 to 33 kb (1-4). Toroviruses (ToVs) are potentially zoonotic, although the precise mechanisms of transmission are still unknown. The existence of human ToV or ToV-like particles and the emergence of coronavirus disease 2019 (COVID-19) in humans, another member of the Coronaviridae family, highlights the need to further understand the biology of ToVs (3, 5-7).

Porcine torovirus was first detected and characterised in faeces samples from piglets in the Netherlands by RT-PCR and immunoelectron microscopy in 1998 (8). Limited epidemiological studies suggest that PToV appears to be widespread in the global pig population and swine herds have a high seroprevalence although its pathogenicity remains unclear (9-15). Although PToV is clinically believed to cause some subclinical symptoms and has not yet caused significant economic damage, recent studies have shown that the PToV genome is highly variable and susceptible to recombination with viruses of other genera, poses a risk of cross-species transmission, making the virus of increasing concern (16-23). The inability to isolate PToV in cell culture severely limits our current understanding of the biological properties and pathogenicity of the disease. Before the successful isolation of the virus achieved, the pathogenicity of PToV could be better understood with detailed research on the function of the PToV-encoded proteins.

As with other nidoviruses, two-thirds of the PToV genome contains two long open reading frames (ORFs), ORF1a and ORF1ab, which encode the replicative polyproteins (pp1a and pp1ab) (24). The polyproteins pp1a and pp1ab are cleaved post-translationally by virus-encoded 3C-like protease ($3CL^{pro}$, also known as main protease) to produce most of the functional nonstructural proteins that play important roles in regulating viral replication and pathogenesis (25, 26). Given its crucial role in viral replication, $3CL^{pro}$ is an important target for the design and development of potent antiviral agents (27). However, the $3CL^{pro}$ sequences of nidoviruses vary widely among themselves and are only reasonably conserved in the region of the active site (28-31). For example, arterivirus $3CL^{pro}$s employ a canonical Ser-His-Asp triad, which is responsible for the recognition of cleavage sites with a glutamic acid (Glu, E) at the P1 position (25, 32-34). However, coronavirus $3CL^{pro}$s use a Cys-His catalytic dyad that recognizes a glutamine (Gln, Q) at the P1 position (35-37). In this study, we characterized the self-processing activities and substrate specificities of the PToV $3CL^{pro}$. In contrast to coronaviruses and arterivirus, PToV $3CL^{pro}$ employed His53 and Ser160 as the active-site residues that recognizes a glutamine (Gln, Q) at the P1 position. Interestingly, the phenylalanine (Phe, F) at P4 position and the Gln at P1 position had important synergistic effects on the N-terminal self-cleavage of PToV $3CL^{pro}$, as mutation of either residue (Q-P1-A or F-P4-A) alone affecting only partial cleavage activity. However, replacement of the two residues (Q-P1-A/F-P4-A) together completely block N-terminal cleavage. Subsequently, homology modeling and biochemical experiments showed that the potential interactions between the

protease's S1/S4 pockets and P1/P4 positions of the substrate are related to the loss of the N-terminal self-cleavage of PToV 3CL^{pro}.

RESULTS

Complete genome sequencing and genetic analysis of porcine torovirus

To characterize the PToV 3CL^{pro} and the proteolytic processing of PToV polyproteins, we firstly collected PToV RT-PCR-positive fecal samples from pigs. Briefly, total RNA was extracted from pig feces and the complete PToV genome was determined from a collection of cDNA clones derived from reverse-transcribed RNA. The whole genome sequence comprises 28, 276 nucleotides excluding the 3'-terminal poly(A) tail and has been deposited in the GenBank database under accession number MH603532 (named HB-1, Fig. 1A). Sequence similarity analysis showed that the complete genome of PToV HB-1 strain shared an 88.24%-92.07% nucleotide similarity with those of previously sequenced PToVs (Genebank Accessions: JQ860350, KM403390, LC483442 and LT900503). Furthermore, we evaluated the phylogenetic position of PToV in relation to other nidoviruses. Comparative sequence and phylogenetic analyses revealed PToV and other toroviruses formed a separate lineage (suborder Tornidovirineae) that separating viruses of the suborders Coronavirieae and Arterivirieae (Fig. 1B), which led us to believe that toroviruses might represent a separate cluster of nidoviruses that linked arteriviruses and coronaviruses.

Identification of a PToV ORF1a-encoded 3C-like serine protease

Previous studies have shown that the central and C-terminal portions of nidovirus replicase polyproteins are processed extensively by the viral 3CL^{pro} (32, 38, 39). Within the replicase polyprotein, nidovirus 3C-like proteases (also known as main proteases, M^{pro}) are generally flanked by two transmembrane (TM)-spanning domains (TM-3CL^{pro}-TM) (40, 41). When transmembrane prediction was performed using TMHMM on the pp1a region (encoded by ORF1a) based on the multiple alignment of toroviruses sequences, some conserved transmembrane regions were observed (data no shown). For putative PToV 3CL^{pro}, two transmembrane domains (show as red) were predicted at its N- and C-terminal borders (Fig. 2A), which is consistent with previous studies (40). Firstly, we employed a ‘self-cleavage' approach to identified the PToV 3CL^{pro} domain. To avoid the possible cytotoxicity caused by the expression of hydrophobic sequences in E. coli, we expressed a fusion protein comprising PToV pp1a residues 3012 to 3392 (pp1a-3012-3392, which represent the putative 3CL^{pro} domain together with the N- and C-terminal flanking sequences except for the predicted transmembrane domains). Considering the expected small size of the potential C-terminal autocleavage product probably prevented its detection, a C-terminal glutathione S-transferase (GST) was fused to pp1a-3012-3392, resulting in the expression construct pp1a-3012-3392-GST (named WT-GST, Fig. 2B). The predicted molecular mass of this fusion protein is 69 kDa, of which 26 kDa is contributed by the GST protein. Based on the conservation of putative active-site residues in many viral 3C and 3C-like proteases (including GX(C/S)G motif), the

146 active-site nucleophile Ser-3325 was predicted in PToV 3CL^{pro} domain. To confirm
147 that any detected cleavage products were due to the proteolytic activity of the 3CL^{pro},
148 we also generated a mutant derivative of the pp1a-3012-3392-GST construct in which
149 the conserved serine of the active-site nucleophile (Ser-3325) was mutated to alanine
150 (named S3225A-GST, and served as a control). As a second control, the GST fusion
151 protein encoded by the pET-28a plasmid (empty vector) was expressed. When the
152 expression of the 3CL^{pro} WT-GST construct was induced by isopropyl
153 β-d-1-thiogalactopyranoside (IPTG), we observed three proteins of about 10 kDa
154 (predicted N-terminal cleavage product, N-term CP), 27 kDa (predicted C-terminal
155 cleavage product, C-term CP), and 32 kDa (predicted mature 3CL^{pro}) in the
156 Coomassie-stained SDS-polyacrylamide gel, respectively (Fig. 2C, Lane 6). These
157 cleavage products were not detected in pET-28a-GST, 3CL^{pro} mut-S3225A-GST, and
158 noninduced cells (the controls). In contrast, the induction of S3225A-GST protein
159 expression led to one additional overexpressed protein with the expected size of 69
160 kDa for the unprocessed pp1a-3012-3392-GST fusion protein (Fig. 2C, lane 7).
161 Collectively, these results indicate that the PToV-3CL^{pro} wild-type WT-GST construct
162 had the predicted proteolytic activity, while this proteolytic activity was abolished
163 after the assumed active-site nucleophile, Ser3225, was replaced with alanine (Ala).

164 To further determine the N- and C-terminal PToV 3CL^{pro} self-processing sites,
165 the predicted mature 3CL^{pro} (32 kDa) and C-terminal cleavage product (27 kDa) were
166 cut from the SDS-PAGE gel and N-terminal sequencing was performed by Edman
167 degradation. The mature 3CL^{pro} protein contained the sequence

168 Ser3096-Val-Phe-Ser-Lys at its N termini (Fig. 3A), indicating that PToV 3CL^{pro}
169 N-terminal cleavage occurred between Gln3095 and Ser3096 in the sequence context
170 SNFSFQ^{3095}↓S^{3096}VFSKV. In addition, the sequence Ser3383-Val-Ser-Thr-Asn at its
171 N termini of PToV 3CL^{pro} C-terminal cleavage product indicated that the 3CL^{pro}
172 C-terminal cleavage in the PToV pp1a occurred at QPFKKQ^{3382}↓S^{3383}VSTNV (Fig.
173 3B). Taken together, these data suggest that PToV pp1a/pp1ab residues 3096 to 3382
174 region contain an active serine protease that undergoes self-processing at both N- and
175 C-terminal cleavage sites (SNFSFQ^{3095}↓S^{3096}VFSKV and QPFKKQ^{3382}↓S^{3383}VSTNV,
176 respectively) with a conserved P1-Gln residue.
177
178 **Identification of Catalytic residues in the PToV 3CL^{pro} active site**
179 Previous studies suggested that the nidoviruses 3CL^{pro} are serine/cysteine proteases
180 that mostly utilize a Ser-His-Glu/Asp catalytic triad or a Cys-His catalytic dyad (30,
181 33, 42-45). For the PToV 3CL^{pro}, a comparative sequence analysis of nidoviruses
182 3CL^{pro} together with the identified PToV 3CL^{pro} was performed to further identify the
183 active site residues of PToV 3CL^{pro}. As shown in Figure 4A, residues H53, S160 and
184 E91 (the numbers indicated here represent their positions in mature PToV 3CL^{pro})
185 were conserved with the identify active-site Ser, His, and Asp/Glu of other nidovirus
186 3C-like proteases, suggesting that these three residues might be catalytic residues in
187 the PToV 3CL^{pro} active site. To verify these predictions, the PToV 3CL^{pro} domain
188 together with flanking pp1a sequence (pp1a-3012-3392) was fused with both N- and
189 C terminal GST tag using a mammalian expression vector pCAGGS as a backbone

(Fig. 4B). Then, the 3CL^{pro} domain was expressed in HEK-293T cells, and proteolytic processing was analyzed by Western blot analysis of the total lysates using GST antibody. As expected, two faster-migrating protein bands were detected in wild-type 3CL^{pro} (GST-WT-GST)-transfected cells at 34 kDa (predicted N-terminal self-cleavage product) and 27 kDa (predicted C-terminal self-cleavage product) (Fig. 4C, lane 2). However, the substitution of Ser160 and His53 with Ala resulted in an unprocessed protein (corresponding with the molecular mass of 92 kDa) that could be detected by Western blotting using a GST-specific antibody (Fig. 4C, lane 3 and 5), which is consistent with the results obtained from another fusion protein construct expressed in E.coli expression system (Fig. 2C). In contrast, the mutation of Glu91 (E91A) did not seem to influence the auto-processing release of the 3CL^{pro} domain from the GST-E91A-GST fusion protein (Fig. 4C, lane 4), which contradicted its assumed role as a third catalytic residue. Interestingly, some additional smaller processing products of less than 92 kDa (unprocessed protein) were also observed in GST-H53A-GST and GST-S160A-GST-transfected cells using a GST-specific antibody by Western blot, as previously shown for Alphamesonivirus 3CL^{pro} (28). These results may due to non-specific proteolytic cleavage by host cellular proteases and deserve further investigation.

We further performed a *trans*-cleavage assay using a GST-S160A-GST as a substrate to verify catalytic residues (H53 and S160) in the active site of PToV 3CL^{pro}. To this end, the wild type and mutants of putative catalytic residues in mature 3CL^{pro} (pp1a-3096-3382) was fused with pCAGGS-HA (N-terminal HA tag) and transfected

212 together with the GST-S160A-GST construct in HEK-293T cells. In contrast to 3CL^{pro}
213 mutants (H53A and S160A), the overexpression of WT and E91A mutant of 3CL^{pro}
214 significantly induced GST-S160A-GST cleavage (Fig. 4D). Together, the data showed
215 that the PToV 3CL^{pro} uses Ser160 and His53 as active-site residues in both *cis*- and
216 *trans*-cleavage.
217
218 **Characterization of self-cleavage substrate specificities of PToV 3CL^{pro}**
219 To better understand the specific substrate properties for efficient cleavage by the
220 PToV 3CL^{pro}, mutagenesis of the N- and C-terminal self-processing sites were
221 performed. Since most nidovirus 3CL^{pro}s have a marked preference for Gln/Glu at the
222 P1 position of the substrate (24, 25, 29, 47-49), we sought to abolish PToV 3CL^{pro}
223 self-cleavage by replacement of the P1-Gln residue by Ala (Q3095A and Q3382A).
224 As a control, the two other glutamines around the N- and C-terminal PToV 3CLpro
225 self-processing sites were also replaced with alanine (Q3088A and Q3377A) (Fig. 5A).
226 As expected, the expression of the fusion proteins GST-WT-GST, GST-Q3088A-GST
227 and GST-Q3377A-GST led to two faster migrating protein bands that migrated with
228 the marker proteins GST-N and C-GST (Fig. 5B and 5C). Moreover, the fusion
229 proteins GST-Q3382A-GST resulted in a putative C-terminal unprocessed protein
230 (Fig. 5B, lane 6), suggesting that Gln3382 at the C-terminal P1 position caused the
231 loss of its C-terminal self-processing site. Surprisingly, although a putative N-terminal
232 unprocessed protein (corresponding to the molecular mass of 66 kDa) was detected
233 from the GST-Q3095A-GST protein, N-terminal cleavage also occurred with nearly

the same efficiency as the wild-type construct (Fig. 5B, lane 4), suggesting that the N-terminal P1-Gln residue might be not a sufficient determinant for N-terminal cleavage by PToV $3CL^{pro}$.

To further investigate the possible influences of single or multiple residues at the cleavage sites, we expressed a series of mutant derivatives at the N-terminal self-cleavage of PToV $3CL^{pro}$ containing appropriate amino acid substitutions (Fig. 6A, 6B). As shown in Fig. 6C, the N-terminal self-cleavage products was no longer detectable after deletion of P6-P6' or P6-P1 at the N-terminal self-cleavage region (Fig. 6C, lanes 3 and 4). In contrast, the deletion of P1'-P6' did not prevent the release of N-terminal self-cleavage product (Fig. 6C, lane5), speculating that one or more residues at P6-P1 positions might be sufficient determinants for the N-terminal self-cleavage of PToV $3CL^{pro}$. To test this hypothesis, consecutive Ala substitutions were introduced at the P6-P1 positions, named (P2-P1)-A, (P3-P1)-A, (P4-P1)-A, (P5-P1)-A, and (P6-P1)-A (Fig. 6D). In contrast with consecutive Ala substitutions at the P2-P1 and P3-P1 positions, the (P4-P1)-A mutant prevented the release of N-terminal cleavage products (Fig. 6D, lane 6), suggesting the Phe at P4 position might be important for the effective N-terminal cleavage. Similar results were obtained from constructs containing additional substitutions at the P5 and P6 positions (Fig. 6D, lanes 7-8). To further test whether Phe at P4 position is a sufficient determinant for N-terminal self-cleavage, a single Ala substitution at P4 position (F-P4-A) and a double Ala substitution at P1 and P4 position (F-P4-A/Q-P1-A) were constructed. Single point mutations at P4 or P1 position only partially affect, however,

256 double point mutations at P4 and P1 position abolish this ability of N-terminal
257 self-cleavage (Fig. 6E). Similar results were obtained in the *trans*-cleavage assay,
258 which suggested that Phe at P4 and Gln at P1 position synergistically control cleavage
259 at the N-terminal 3CL^{pro} self-processing site (Fig. 6F).
260
261 **Identification of cleavage sites recognized by PToV 3CL^{pro} within polyproteins**
262 To further investigate whether proteolytic processing can occur at other PToV
263 polyprotein cleavage sites, 12 putative cleavage sites of PToV 3CL^{pro} were predicted
264 based on alignments of pp1ab polyproteins amongst PToV isolates and other related
265 viruses in the Torovirus genus. Subsequently, we performed *trans*-cleavage assay
266 using the cyclized luciferase reporter gene vector, which has been used for
267 identification of cleavage sites recognized by viral protease in previous studies (46).
268 To this end, the P6-P6' positions of PToV 3CL^{pro} potential recognition regions were
269 introduced into the cyclized luciferase reporter. The identified PToV 3CL^{pro}
270 N-terminal auto-cleavage sequence (named CP4/nsp2) and C-terminal auto-cleavage
271 sequence (named nsp2/nsp3) were served as positive controls, while the Tobacco Etch
272 Virus (TEV) protease recognition sequence ENLYFQYS (named 233D), and the
273 putative PToV 3CL^{pro} internal recognition sequence (named nsp2 internal, the same
274 location which can be cleaved by EToV 3CL^{pro}) were used as negative controls (30).
275 As shown in Figure 7A, the luciferase activity of positive controls (CP4/nsp2 and
276 nsp2/nsp3) were markedly induced in PToV 3CL^{pro}-cotransfected cells, whereas no
277 activity of negative controls was detected. Also, ten additional putative cleavage sites

were identified to be cleaved by PToV 3CL^{pro} (Fig. 7B), as determined by the cyclized luciferase reporters. Consistent with previous study, the 12 cleavage sites of PToV 3CL^{pro} shared a high similarity with putative cleavage sites in pp1a/pp1ab for the EToV and BRV 3CL^{pro} (Fig. 7B). Interestingly, a predicted cleavage site (CP3/CP4, Fig. 7A and 7B) is cleaved by PToV 3CL^{pro}, which is typically cleaved by the papain-like protease. This cleavage site might be a conserved cleavage site for torovirus 3CL^{pro}s, as it was also predicted to be cleaved by EToV 3CL^{pro} (30, 47).

Comparative sequence analysis of the 12 PToV 3CL^{pro} cleavage sites in pp1ab revealed that Gln is strictly conserved at the P1 position. The logo representations revealed Phe at the P4 position was also highly conserved (Fig. 7C) but replaced with Tyr, Leu, Ile or Met in some of the PToV 3CL^{pro} cleavage sites. To investigate the possible effects of the Phe-to-Tyr/Leu/Ile/Met substitution at P4, different single amino acid substitutions at P4 position were constructed using the CP4/nsp2 reporter as a backbone (F-P4-Y, F-P4-L, F-P4-I, F-P4-M, and F-P4-A). As shown in Figure 7D, this pronounced preference for Phe at P4 position seems to in favor of efficient cleavage, since substitution of P4-Phe by Ala yielded a 10-fold reduction in cleavage rate, whereas introduction of natural amino acids Tyr, Leu, Ile or Met at P4 reduced the rate of cleavage by 2-fold. Taken together, the data suggested that macrobenzene side chains of hydrophobic residues at the P4 position might be preferred in substrates of PToV 3C^{pro}.

Investigation of the putative S1 and S4 pockets within PToV 3CL^{pro}

To provide a structural rationale for PToV 3CL^{pro}'s preference in P4 and P1 positions, we attempted to build a structural model of PToV 3CL^{pro}. As PToV 3CL^{pro} and EAV 3CL^{pro} share the highest amino acid sequence similarities (30.1%), we generated a homology model of PToV 3CL^{pro} in Modeller 9 using the X-ray crystal structure of EAV 3CL^{pro} (PDB accession number 1MBM) as a template. The model showed that PToV 3CL^{pro} adopts a fold similar to that of EAV 3CL^{pro}, and is predicted to has a three-domain structure, with a two-β-barrel structure with His53 and Ser160 as the active-site residues (Fig. 8A). To study the molecular mechanism of PToV 3CL^{pro}'s preference for P4 and P1 positions, the structure of Streptomyces griseus proteinase E (SGPE) in complex with a tetrapeptide (PDB code: 1HPG) was used as the reference structure for PToV 3CL^{pro} complex construction, as among all known crystal complex structure, EAV 3CL^{pro} shared a high similarity with SGPE structure, which is reported in previous study (32, 48). Then we established a protease-substrate complex system consisting of PToV 3CL^{pro} and N-terminal self-cleavage sequence SNFSFQ↓SVFSKV, and performed molecular dynamics simulations to improve structure accuracy and study the binding model between PToV 3CL^{pro} and substrate. As shown in Figure 8B, two putative pockets S1 (Ser-153, Asp-159 and His-174) and S4 (Lys-61, Pro-62, Leu-59, Leu-185, Gly-176) are formed between PToV 3CL^{pro} and the substrate, where the S4 pocket is composed mainly of several hydrophobic amino acids (Fig. 8C and 8D). To test the coupling of atomic fluctuations in the S4 and S1 pockets, we calculated the correlation coefficient between atomic fluctuations in the main chain of PToV 3CL^{pro} and the P4 and P1 positions of substrate. As shown in Figure 8E, the

dynamics of the P4 and P1 positions are highly correlated with the movement of amino acids in the S4 and S1 pockets, which proves the stability of the pockets.

To further verify the results of the molecular dynamics simulations, several amino acids in the S1 and S4 pockets of PToV 3CL^{pro} were selected and mutated to alanine. Using pCAGGS-PToV 3CL^{pro} as a backbone, the putative P1-Gln-binding residues in S1 pocket, Glu159 and His174, were mutated to Ala and studied in a peptide-based cyclized luciferase reporter system (contain N-terminal self-cleavage sequences SNFSFQ↓SVFSKV). When analyzed by Dual luciferase assay and Western blot, in contrast to the wild PToV 3CL^{pro} (named WT), the D159A and H174A mutants were virtually inactive to induce the activity of luciferase and cleave the substrate, which confirmed that D159 and H174 in S1 pocket affect the effectiveness of protease cleavage (Fig. 9A). To further verify the role of hydrophobic forces in the S4 pocket, the effect of substitutions of the putative P4-Phe-binding residues, Pro62 and Leu185, with Ala (hydrophobic) or Gly (neutral) were also analyzed in this experiment. As shown in Figure 9B and 9C, the mutants (P62A, L185A, P62G, L185G, and P62A/L185A) exhibited lower cleavage activities compared with the wild 3CL^{pro}. Nevertheless, the double mutant (P62G/L185G) completely abolished the ability of cleavage activity of PToV 3CL^{pro} (Fig. 9C). This means that the potential hydrophobic force between the PToV 3CL^{pro} and P4-Phe side chains is critical for substrate binding, for a double mutant of P62 and L185 to Gly will destroy the substrate binding between PToV 3CL^{pro} and P4-Phe side chains (Fig. 9C). To further investigate the role of potential interaction in the S4 pocket or S1 pocket in binding

different substrate peptide, nsp4/5 (PTIMWQ↓SDDVAD) and nsp7/9 (PIYQPQ↓SVRYAD) of PToV 3CL^{pro} was used for further assay. As shown in Figure 10 and 11, the protease loses the ability to cleave the substrate nsp4/5 and nsp7/9 after mutating the S1 and S4 pocket key amino acids.

DISCUSSION

In nidoviruses, 3CL^{pro} cleaves the viral polyproteins at several conserved sites and is involved in translational processing of the viral non-structural proteins (25, 49). It has therefore also been referred to as a potential target for the development of antiviral agents. In this study, we sequenced the complete genome of PToV strain (HB-1), and first sought to refine the putative PToV 3CL^{pro}. We confirmed the *cis*- and *trans*-activity of PToV 3CL^{pro} and identified 12 cleavage sites on the polyproteins of PToV. The results identify a consensus cleavage sequence for PToV 3CL^{pro} and help clarify the genomic organization of PToV.

Using a bacterial expression system, we confirmed the putative self-processing activity of the PToV 3CL^{pro} domain and identified N- and C-terminal 3CL^{pro} self-processing sites by N-terminal sequencing of candidate processing products. Sequence comparisons suggested that the PToV 3CL^{pro} active site may employs a catalytic triad comprising the putative Ser-His-Asp residues. However, our results suggested that 3CL^{pro} employs Ser-160 and His-53 as the active-site residues. The mutagenesis data showed that the replacement of the presumed third catalytic residue, Asp-91, had no detectable effect on PToV 3CL^{pro} activity. It should also be noted that

366 the third catalytic residue of the chymotrypsin catalytic triad is absent in
367 coronaviruses 3CL^{pro} (45, 50). In fact, catalytic residues in different nidoviruses
368 3CL^{pro} active site differ greatly between them. For example, coronaviruses 3CL^{pro} is a
369 catalytic dyad (Cys-His) with a conserved P1-Gln residue (35), while arteriviruses
370 3CL^{pro} is a catalytic trimer (Ser-His-Asp) with a conserved P1-Glu residue (34). In
371 this study, we found that PToV 3CL^{pro} is a serine protease (as is the arteriviruses
372 3CL^{pro}) that employs Ser-160 and His-53 as the active-site residues with a conserved
373 P1-Gln residue (similar to the coronaviruses 3CL^{pro}). Interestingly,
374 phylogenetic analysis of polyproteins among nidoviruses revealed that the
375 coronaviruses represent the most closely related family to PToV (41, 51, 52). In the
376 case of the 3CL^{pro}, however, the most significant matches were found in homologs
377 from arteriviruses. Taken together, our results provide an evolutionarily link between
378 the 3CL^{pro} of arteriviruses, toroviruses and coronaviruses.

379 Similar to some other nidovirus 3C-like proteases, PToV 3CL^{pro} have a
380 pronounced substrates preference for Gln at the P1 position. Surprisingly, Ala
381 substitutions at P1-Gln impaired the C-terminal self-processing, but only slightly
382 affect N-terminal self-cleavage. The subsequent mutagenesis of substrate revealed
383 that this “noncanonical” substrate specificity for its N-terminal self-processing
384 required both P1-Gln and P4-Phe. Interestingly, the Phe mediated-noncanonical
385 substrate specificity was not unique to torovirus. Previous studies have shown that
386 Phe is required at both the P2 position and the P3′ position for the C-terminal
387 self-cleavage of SARS-CoV 3CL^{pro}, which is also conserved in SARS-CoV-2 3CL^{pro}

(53, 54). Unlike SARS-CoV or SARS-CoV-2, based on the 12 cleavage sites of PToV 3CL^{pro} determined in this study, Phe was also found to be highly conserved at P4 position recognized by PToV 3CL^{pro} within polyproteins. Moreover, it has been shown that EToV 3CL^{pro} cleaves its own internal self-sequence SEFATQ↓AWQTVN, and this cleavage apparently down-regulates protease activity, but this internal self-cleavage is not present in PToV 3CL^{pro} (Fig. 2C and 7A) (55). Sequence comparison showed a high degree of protease similarity within torovirus, however in the same region the PToV 3CL^{pro} base sequence was SELATQ↓AWQTVN, where the Phe at P4 position of EToV 3CL^{pro} became the Leu in PToV 3CL^{pro} (data not shown). This mutation might be responsible for the disappearance of internal cleavage of PToV 3CL^{pro}, which may be beneficial for the proliferation of PToV. Consistent with our results, in the case of coronavirus, mesoniviruses and ronivirus 3CL^{pro} (25, 36, 56-58), residues at the P4 position is also thought to be additional specificity determinants.

The general mechanism of catalysis of 3CL^{pro} relies on the nucleophilic attack of the peptide bond of the substrate by a catalytic Cys/Ser residue in the active site (37, 59). The substrate specificity of the 3CL^{pro} on the other hand is primarily determined by the interactions between the side chains of the P1 residue of the substrate that is accommodated in the S1 pocket of the active site (26, 42). With respect to conserved residues potentially involved in forming important specificity pocket, mutagenesis and molecular dynamics investigation in this study suggested that the Asp-159 and His174 residues in PToV 3CL^{pro} might play an important in binding potential P1-Gln

410 residues. In addition to the conserved residues in putative S1 pocket, many viral
411 3CL^{pro} possess other conserved residues that assists in the binding of other residue in
412 substrate (25, 26, 28, 60-62). In case of PToV 3CL^{pro}, the conserved residues in
413 putative S4 pocket (Pro-62, Leu-185) are also considered an important determinant
414 for the PToV 3CL^{pro}-mediated cleavage. All of these residues in the putative S1 and
415 S4 pockets are highly conserved in 3CL^{pro} of toroviruses, but not coronavirues and
416 arteriviruses (data not shown). The composition of different residues in the S1 and S4
417 pockets may be a key factor in the substrate specificities of different residues at P1
418 and P4 position by different nidoviruses 3CL^{pro}s. It should be noted that although
419 alanine scanning has been widely used to identify side chains that play important roles
420 in protein–peptide interactions, alanine substitutions at key residues within the
421 protease has the potential to affect its structural integrity. Thus, relative to alanine
422 scanning, other amino acid residues scanning (e.g., hydrophile scanning) may provide
423 new insight regarding the interactions experienced by specific side chains upon
424 protease-substrate complex. Meanwhile, the determination of crystal structures of
425 PToV 3CL^{pro}-substrate will provide more details on the interactions of PToV
426 3CL^{pro}-substrate and the molecular determinants of substrate specificities of PToV
427 3CL^{pro}, which deserve further investigation.

428 In summary, this first characterization of the PToV 3CL^{pro} supported and
429 developed previous findings for other nidoviruses. It appears that torovirus 3CL^{pro}s
430 have evolved properties that distinguish them from other nidovirus proteases. In
431 particular, this applies to the unique substrate specificity identified for the PToV

3CL^{pro} in this study. The use of Ser-160 and His-53 as the active-site residues, and a P4 residue with Phe as an additional specificity determinant of PToV 3CL^{pro} provides additional insights for understanding the diversity of nidoviruses 3CL^{pro}.

MATERIALS AND METHODS

Sequence analysis of the PToV genome. The positive PToV faeces samples from pigs with severe diarrhoea were diluted in phosphate-buffered saline solution (PBS, pH 7.4) and finally clarified by centrifugation at low speed (4, 000 ×g) for 10 min. The supernatants were then collected and subjected to RNA extraction. The total RNA was extracted with the RNA Solv Reagent (OMEGA BIO-TEK) and dissolved in DEPC-treated water. The cDNA was synthesised by reverse transcriptase (RT) using the Roche reverse transcription system. For genome sequence analysis, oligo-primer pairs targeting the genome of PToV were used, and the resulting amplicons were cloned using pGEM®-T Vector according to the manufacturer's protocols (Promega).

Inferring virus evolutionary history. The virus 3CL^{pro} were aligned by using the E-INS-i algorithm implemented in the program MAFFT (63). The evolutionary history was then inferred by using the Maximum Likelihood method and Le_Gascuel_2008 model (64). The tree with the highest log likelihood (-36582.01) is shown. Initial tree (s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among

454 sites (5 categories (+G, parameter = 1.9218)). The rate variation model allowed for
455 some sites to be evolutionarily invariable ([+I], 0.62% sites). The tree is drawn to
456 scale, with branch lengths measured in the number of substitutions per site. This
457 analysis involved 90 amino acid sequences. There was a total of 633 positions in the
458 final dataset. Evolutionary analyses were conducted in MEGA X (65).

459 **Plasmids.** The complete genomic sequence of porcine torovirus isolate HB-1
460 (accession numbers, MH_603532) was used for the source for all constructs described
461 below. Firstly, to construct pET28a-GST, the GST coding sequence GST tag was
462 amplified from pET-42b and cloned into the C-terminal of the pET28a polyclonal site.
463 Then, to generate pET28a-WT-GST and pET28a-S3225A-GST, the predicted coding
464 sequence of PToV $3CL^{pro}$, together with short flanking sequences (pp1a/pp1ab
465 3012-3392), were amplified by PCR and subclone into pET28a-GST.

466 To construct p-GST-GST, the GST tag was amplified from pET-42b and then
467 cloned into pCAGGS vector. Subsequently, the pp1a/pp1ab 3012-3392 was amplified
468 from HB-1 and the resulting PCR products were ligated with p-GST-GST by using T4
469 DNA ligase to generate p-GST-WT-GST. To construct pCAGGS-PToV-$3CL^{pro}$, the
470 coding region for pp1a/pp1ab amino acids Ser-3096 to Gln-3382 was amplified from
471 p-GST-GST-WT and subcloned into pCAGGS vector.

472 To construct a total of 13 reporter plasmids, the predicted peptides were
473 amplified from pp1a/pp1ab of PToV HB-1 and subcloned into the 233D reporter
474 vector, including CP3/CP4, CP4/nsp2, nsp2/nsp3, nsp3/nsp4, nsp4/nsp5, nsp5/nsp6,
475 nsp6/nsp7, nsp7/nsp9, nsp9/nsp10, nsp10/nsp11, nsp11/nsp12, nsp12/nsp13 and nsp2

476 internal.

477 **Bacterial expression and N-terminal sequencing.** To express recombinant proteins
478 encoded by pET28a-GST, pET28a-WT-GST and pET28a-S160A-GST, the E. coli
479 BL21 (DE3) strain was used. Freshly transformed cells were cultured in LB medium
480 containing 50 mg/l kanamycin until an optical density at 600 nm (OD600) reached 0.6
481 to 0.9. The cultures were then separated, and protein expression was induced with 0.8
482 mM isopropyl-D-thiogalactopyranoside (IPTG) in one of the cultures. The induced
483 and uninduced cultures were incubated at 37°C for an additional 5 h under severe
484 shaking (225 rpm). After 5 h incubation cells were collected in PBS and lysed by
485 sonication. The total cell lysates were obtained by adding Laemmli sample buffer and
486 incubated at 95°C for 5 min. Protein expression was analyzed by SDS-PAGE and
487 Coomassie staining.

488 An appropriate amount of pET-WT-28a protein sample was separated onto a 12%
489 SDS-polyacrylamide gel and transferred onto polyvinylidene fluoride membranes
490 (Millipore, Burlington, Massachusetts, USA). The membranes were then stained with
491 0.025% Komas Blue R-250 containing 40% methanol and subsequently decolorized
492 with 50% methanol. The area containing the desired protein was excised, and the
493 protein was subjected to N-terminal sequencing by Edman degradation (Biotech
494 Bioengineering (Shanghai) Co).

495 **Dual luciferase assay.** HEK-293T cells were obtained from the China Typical
496 Culture Collection Center (Wuhan, China) and grown at 37°C, 5% CO_2 in Dulbecco's
497 modified Eagle medium (Invitrogen, Madison, WI, USA), to which 10% fetal bovine

498 serum was added. The luciferase reporter constructs and negative controls (233D and
499 nsp2 internal, respectively) were used to identified the putative recongination sites of
500 PToV 3CL^{pro}. Briefly, HEK-293T cells in 24-well plates were transfected with various
501 luciferase reporter plasmids and pRL-TK (Promega), together with 3CL^{pro} expression
502 plasmids or the empty control plasmid, At 28 h post-transfection, the cells were lysed,
503 and a luciferase reporter assay system (Promega) was used to test the luciferase
504 activities in the lysed cells. The activities were normalized to the corresponding
505 Renilla luciferase activities.

506 **Western blot.** HEK-293T cells cultured in 6-well plates were transfected with various
507 plasmids. After 28 h transfection. The cells were collected by lysis buffer (Beyotime)
508 and subsequently added to sample upload buffer (Beyotime) and boiled for 10 min.
509 Subsequently, samples were gummed with SDS-PAGE and transferred onto
510 polyvinylidene fluoride membranes (Millipore, Burlington, Massachusetts, USA).
511 Anti-GST antibody (MBL, Nagoya, Japan), anti-β-actin antibody (Antgene, Wuhan,
512 China) and anti-HA antibody (MBL, Nagoya, Japan) were used to detect the
513 respective proteins.

514 **Homology modelling.** Homology model of PToV 3CL^{pro} was generated with
515 Modeller 9 using EAV 3CL^{pro} (PDB code: 1MBM) as the principal template (34, 66).
516 The PToV 3CL^{pro} in complex with substrate peptides was generated using the
517 structure of SGPE in complex with a tetrapeptide (PDB code: 1HPG) as the reference
518 structure, because EAV 3CL^{pro} shared a high similarity with SGPE structure among all
519 known crystal complex structure (32, 48). The substrate in the complex was replaced

with (CP4/nsp2 auto-cleavage sequence of PToV 3CL^{pro} SNFSFQ↓SVFSKV) using SYBYL-X (v.2.0; https://omictools.com/sybyl-x-tool). In order to improve structure accuracy, MD simulations were performed using the Amber ff14SB force field implemented in the GROMACS 2018 software package as previously described (67, 68).

Data availability. The complete genomic nucleotide sequence of porcine torovirus isolate HB1 has been deposited in GenBank with accession number MH603532.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Plan of China (2017YFD0501101), the National Natural Science Foundation of China (31872485, 31941005 and 31672566), and the Major S&T Project of Hubei Province (2017ABA138).

REFERENCES

1. **Weiss M, Steck F, Horzinek MC.** 1983. Purification and Partial Characterization of a New Enveloped RNA Virus (Berne Virus). Journal of General Virology **64:**1849-1858.
2. **Woode GN, Reed DE, Runnels PL, Herrig MA, Hill HT.** 1982. Studies with an unclassified virus isolated from diarrheic calves. Veterinary Microbiology **7:**221-240.
3. **Beards GM, Green J, Hall C, Flewett TH, Lamouliatte F, Dupasquier P.** 1984. An Enveloped Virus in Stools of Children and Adults With Gastroenteritis That Resembles the Breda Virus of Calves. Lancet **1:**1050-1052.
4. **Stenglein MD, Jacobson ER, Wozniak EJ, Wellehan JF, Kincaid A, Gordon M, Porter BF, Baumgartner W, Stahl S, Kelley K, Towner JS, DeRisi JL.** 2014. Ball python nidovirus: a candidate etiologic agent for severe respiratory disease in Python regius. MBio **5:**e01484-01414.
5. **Weiss M, Steck F, Kaderli R, Horzinek MC.** 1984. Antibodies to Berne Virus in Horses and Other Animals. Veterinary Microbiology **9:**523-531.
6. **Koopmans M, Petric M, Glass RI, Monroe SS.** 1993. Enzyme-linked immunosorbent assay reactivity of torovirus-like particles in fecal specimens from humans with diarrhea. J Clin Microbiol **31:**2738-2744.
7. **Brown DWG, Beards GM, Flewett TH.** 1987. Detection of Breda Virus Antigen and Antibody in Humans and Animals by Enzyme Immunoassay. Journal of Clinical Microbiology **25:**637-640.
8. **Kroneman A, Cornelissen L, Horzinek MC, de Groot RJ, Egberink HF.** 1998. Identification and characterization of a porcine Torovirus. Journal of Virology **72:**3507-3511.
9. **Durham PJK, Hassard LE, Norman GR, Yemen RL.** 1989. Viruses and Virus-Like Particles Detected During Examination of Feces From Calves and Piglets With Diarrhea. Canadian Veterinary Journal-Revue Veterinaire Canadienne **30:**876-881.
10. **Penrith ML, Gerdes GH.** 1992. Breda Virus-Like Particles in Pigs in South Africa. Journal of the South African Veterinary Association-Tydskrif Van Die Suid-Afrikaanse Veterinere Vereniging **63:**102-102.
11. **Pignatelli J, Jimenez M, Luque J, Rejas MT, Lavazza A, Rodriguez D.** 2009. Molecular characterization of a new PToV strain. Evolutionary implications. Virus Res **143:**33-43.
12. **Pignatelli J, Jimenez M, Grau-Roma L, Rodriguez D.** 2010. Detection of porcine torovirus by real time RT-PCR in piglets from a Spanish farm. Journal of Virological Methods **163:**398-404.
13. **Shin D-J, Park S-I, Jeong Y-J, Hosmillo M, Kim H-H, Kim H-J, Kwon H-J, Kang M-I, Park S-J, Cho K-O.** 2010. Detection and molecular characterization of porcine toroviruses in Korea. Archives of Virology **155:**417-422.
14. **Shin DJ, Park SI, Jeong YJ, Hosmillo M, Kim HH, Kim HJ, Kwon HJ, Kang MI, Park SJ, Cho KO.** 2010. Detection and molecular characterization of porcine toroviruses in Korea. Arch Virol **155:**417-422.
15. **Fujii Y, Kashima Y, Sunaga F, Aoki H, Imai R, Sano K, Katayamas Y, Omatsu T, Oba M, Furuya T, Tsuzuku S, Ouchi Y, Shirai J, Mizutani T, Oka T, Nagai M.** 2020. Complete genome sequencing and genetic analysis of a Japanese porcine torovirus strain detected in swine feces. Archives of Virology **165:**471-477.

16. **Ito M, Tsuchiaka S, Naoi Y, Otomaru K, Sato M, Masuda T, Haga K, Oka T, Yamasato H, Omatsu T, Sugimura S, Aoki H, Furuya T, Katayama Y, Oba M, Shirai J, Katayama K, Mizutani T, Nagai M.** 2016. Whole genome analysis of Japanese bovine toroviruses reveals natural recombination between porcine and bovine toroviruses. Infect Genet Evol **38:**90-95.

17. **Smits SL, Lavazza A, Matiz K, Horzinek MC, Koopmans MP, de Groot RJ.** 2003. Phylogenetic and evolutionary relationship among torovirus field variants: Evidence for multiple intertypic recombination events. Journal of Virology **77:**9567-9577.

18. **Lee S, Lee C.** 2019. First detection of novel enterovirus G recombining a torovirus papain-like protease gene associated with diarrhoea in swine in South Korea. Transboundary and Emerging Diseases **66:**1023-1028.

19. **Tsuchiaka S, Naoi Y, Imai R, Masuda T, Ito M, Akagami M, Ouchi Y, Ishii K, Sakaguchi S, Omatsu T, Katayama Y, Oba M, Shirai J, Satani Y, Takashima Y, Taniguchi Y, Takasu M, Madarame H, Sunaga F, Aoki H, Makino S, Mizutani T, Nagai M.** 2018. Genetic diversity and recombination of enterovirus G strains in Japanese pigs: High prevalence of strains carrying a papain-like cysteine protease sequence in the enterovirus G population. Plos One **13**.

20. **Knutson TP, Velayudhan BT, Marthaler DG.** 2017. A porcine enterovirus G associated with enteric disease contains a novel papain-like cysteine protease. Journal of General Virology **98:**1305-1310.

21. **Ito M, Tsuchiaka S, Naoi Y, Otomaru K, Sato M, Masuda T, Haga K, Oka T, Yamasato H, Omatsu T, Sugimura S, Aoki H, Furuya T, Katayama Y, Oba M, Shirai J, Katayama K, Mizutani T, Nagai M.** 2016. Whole genome analysis of Japanese bovine toroviruses reveals natural recombination between porcine and bovine toroviruses. Infection Genetics and Evolution **38:**90-95.

22. **Conceição-Neto N, Theuns S, Cui T, Zeller M, Yinda CK, Christiaens I, Heylen E, Van Ranst M, Carpentier S, Nauwynck HJ, Matthijnssens J.** 2017. Identification of an enterovirus recombinant with a torovirus-like gene insertion during a diarrhea outbreak in fattening pigs. Virus Evol **3:**vex024.

23. **Tsuchiaka S, Naoi Y, Imai R, Masuda T, Ito M, Akagami M, Ouchi Y, Ishii K, Sakaguchi S, Omatsu T, Katayama Y, Oba M, Shirai J, Satani Y, Takashima Y, Taniguchi Y, Takasu M, Madarame H, Sunaga F, Aoki H, Makino S, Mizutani T, Nagai M.** 2018. Genetic diversity and recombination of enterovirus G strains in Japanese pigs: High prevalence of strains carrying a papain-like cysteine protease sequence in the enterovirus G population. PLoS One **13:**e0190819.

24. **Sun H, Lan D, Lu L, Chen M, Wang C, Hua X.** 2014. Molecular characterization and phylogenetic analysis of the genome of porcine torovirus. Archives of Virology **159:**773-778.

25. **Ziebuhr J, Snijder EJ, Gorbalenya AE.** 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. Journal of General Virology **81:**853-879.

26. **Gorbalenya AE, Donchenko AP, Blinov VM, Koonin EV.** 1989. Cysteine Proteases of Positive Strand RNA Viruses and Chymotrypsin-Like Serine Proteases. A Distinct Protein Superfamily With a Common Structural Fold. Febs Letters **243:**103-114.

27. **Pillaiyar T, Manickam M, Namasivayam V, Hayashi Y, Jung SH.** 2016. An Overview of Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. J Med Chem **59:**6595-6628.

28. **Blanck S, Stinn A, Tsiklauri L, Zirkel F, Junglen S, Ziebuhr J.** 2014. Characterization of an alphamesonivirus 3C-like protease defines a special group of nidovirus main proteases. J Virol **88:**13747-13758.

29. **Schutze H, Ulferts R, Schelle B, Bayer S, Granzow H, Hoffmann B, Mettenleiter TC, Ziebuhr J.** 2006. Characterization of White bream virus reveals a novel genetic cluster of nidoviruses. J Virol **80:**11598-11609.

30. **Smits SL, Snijder EJ, de Groot RJ.** 2006. Characterization of a torovirus main proteinase. Journal of Virology **80:**4157-4167.

31. **Ulferts R, Mettenleiter TC, Ziebuhr J.** 2011. Characterization of Bafinivirus main protease autoprocessing activities. J Virol **85:**1348-1359.

32. **Tian X, Lu G, Gao F, Peng H, Feng Y, Ma G, Bartlam M, Tian K, Yan J, Hilgenfeld R, Gao GF.** 2009. Structure and cleavage specificity of the chymotrypsin-like serine protease (3CLSP/nsp4) of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV). J Mol Biol **392:**977-993.

33. **Snijder EJ, Wassenaar AL, van Dinten LC, Spaan WJ, Gorbalenya AE.** 1996. The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. J Biol Chem **271:**4864-4871.

34. **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.

35. **Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R.** 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. Embo j **21:**3213-3224.

36. **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.

37. **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.

38. **Ziebuhr J, Siddell SG.** 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.

39. **van Aken D, Zevenhoven-Dobbe J, Gorbalenya AE, Snijder EJ.** 2006. Proteolytic maturation of replicase polyprotein pp1a by the nsp4 main proteinase is essential for equine arteritis virus replication and includes internal cleavage of nsp7. J Gen Virol **87:**3473-3482.

40. **Pinon JD, Mayreddy RR, Turner JD, Khan FS, Bonilla PJ, Weiss SR.** 1997. Efficient autoproteolytic processing of the MHV-A59 3C-like proteinase from the flanking hydrophobic domains requires membranes. Virology **230:**309-322.

41. **Nga PT, Parquet Mdel C, Lauber C, Parida M, Nabeshima T, Yu F, Thuy NT, Inoue S, Ito T, Okamoto K, Ichinose A, Snijder EJ, Morita K, Gorbalenya AE.** 2011. Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. PLoS Pathog **7:**e1002215.

42. **Bazan JF, Fletterick RJ.** 1988. Viral Cysteine Proteases Are Homologous to the Trypsin-Like Family of Serine Proteases: Structural and Functional Implications. Proceedings of the

National Academy of Sciences of the United States of America **85:**7872-7876.

43. **Gorbalenya AE, Blinov VM, Donchenko AP.** 1986. Poliovirus-encoded Proteinase 3C: A Possible Evolutionary Link Between Cellular Serine and Cysteine Proteinase Families. Febs Letters **194:**253-257.

44. **Malcolm BA.** 1995. The Picornaviral 3C Proteinases: Cysteine Nucleophiles in Serine Proteinase Folds. Protein Science **4:**1439-1445.

45. **Ziebuhr J, Heusipp G, Siddell SG.** 1997. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. J Virol **71:**3992-3997.

46. **Zhou J, Wang D, Xi Y, Zhu X, Yang Y, Lv M, Luo C, Chen J, Ye X, Fang L, Xiao S.** 2017. Assessing activity of Hepatitis A virus 3C protease using a cyclized luciferase-based biosensor. Biochem Biophys Res Commun **488:**621-627.

47. **Draker R, Roper RL, Petric M, Tellier R.** 2006. The complete sequence of the bovine torovirus genome. Virus Res **115:**56-68.

48. **Shi Y, Lei Y, Ye G, Sun L, Fang L, Xiao S, Fu ZF, Yin P, Song Y, Peng G.** 2018. Identification of two antiviral inhibitors targeting 3C-like serine/3C-like protease of porcine reproductive and respiratory syndrome virus and porcine epidemic diarrhea virus. Vet Microbiol **213:**114-122.

49. **Li Y, Tas A, Sun Z, Snijder EJ, Fang Y.** 2015. Proteolytic processing of the porcine reproductive and respiratory syndrome virus replicase. Virus Res **202:**48-59.

50. **Hegyi A, Friebe A, Gorbalenya AE, Ziebuhr J.** 2002. Mutational analysis of the active centre of coronavirus 3C-like proteases. Journal of General Virology **83:**581-593.

51. **Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ.** 2006. Nidovirales: Evolving the largest RNA virus genome. Virus Research **117:**17-37.

52. **Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ.** 2016. Redefining the invertebrate RNA virosphere. Nature **540:**539-543.

53. **Muramatsu T, Takemoto C, Kim YT, Wang H, Nishii W, Terada T, Shirouzu M, Yokoyama S.** 2016. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. **113:**12997-13002.

54. **Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Becker S, Rox K, Hilgenfeld R.** 2020. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science **368:**409-412.

55. **Smits SL, Snijder EJ, de Groot RJ.** 2006. Characterization of a torovirus main proteinase. J Virol **80:**4157-4167.

56. **Blanck S, Stinn A, Tsiklauri L, Zirkel F, Junglen S, Ziebuhr J.** 2014. Characterization of an Alphamesonivirus 3C-Like Protease Defines a Special Group of Nidovirus Main Proteases. Journal of Virology **88:**13747-13758.

57. **Ziebuhr J, Bayer S, Cowley JA, Gorbalenya AE.** 2003. The 3C-like proteinase of an invertebrate nidovirus links coronavirus and potyvirus homologs. Journal of Virology **77:**1415-1426.

58. **Hegyi A, Ziebuhr J.** 2002. Conservation of substrate specificities among coronavirus main proteases. Journal of General Virology **83:**595-599.

59. **Wang H, He S, Deng W, Zhang Y, Guobang L, Sun J, Zhao W, Guo Y, Yin Z, Li D, Shang L-Q.** 2020. Comprehensive Insights into the Catalytic Mechanism of Middle East Respiratory

Syndrome 3C-Like Protease and Severe Acute Respiratory Syndrome 3C-Like Protease. ACS Catalysis **XXXX**.

60. **Kanitz M, Blanck S, Heine A, Gulyaeva AA, Diederich WE.** 2019. Structural basis for catalysis and substrate specificity of a 3C-like cysteine protease from a mosquito mesonivirus. Virology **533**.

61. **Blanck S, Ziebuhr J.** 2016. Proteolytic processing of mesonivirus replicase polyproteins by the viral 3C-like protease. J Gen Virol **97:**1439-1445.

62. **Krausslich HG, Wimmer E.** 1988. Viral Proteinases. Annual Review of Biochemistry **57:**701-754.

63. **Katoh K, Rozewicki J, Yamada KD.** 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform **20:**1160-1166.

64. **Le SQ, Gascuel O.** 2008. An improved general amino acid replacement matrix. Mol Biol Evol **25:**1307-1320.

65. **Sudhir K, Glen S, Li M, Christina K, Koichiro T.** 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology & Evolution**:**6.

66. **Sali A, Blundell TL.** 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol **234:**779-815.

67. **Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E.** 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX **1-2:**19-25.

68. **Zhou J, Fang L, Yang Z, Xu S, Lv M, Sun Z, Chen J, Wang D, Gao J, Xiao S.** 2019. Identification of novel proteolytically inactive mutations in coronavirus 3C-like protease using a combined approach. Faseb j **33:**14575-14587.

69. **Fábio M, mi PY, Joon L, Nicola B, Tamer G, Nandana M, Prasad B, N TAR, C PS, D FR.** 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research**:**W1.

70. **Robert X, Gouet P.** 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research **42:**W320-W324.

71. **Krystkowiak I, Manguy J, Davey NE.** 2018. PSSMSearch: a server for modeling, visualization, proteome-wide discovery and annotation of protein motif specificity determinants. Nucleic Acids Res **46:**W235-w241.

FIGURE LEGENDS

Figure 1. Complete genome sequencing and genetic analysis of porcine torovirus. **(A)** Schematic diagram of the PToV (HB-1) genome structure, the genome contains six open reading frames (ORFs) flanked by a long 5′ untranslated region (UTR) (796-nt) and a short 3′ UTR (194-nt). The predicted PToV 3CL^{pro} specific region marked in dark, with all the predicted cleavage sites are listed above are shown in green triangle. **(B)** 3CL^{pro}-based phylogeny of representatives of established nidovirus species. The evolutionary history was inferred by using the Maximum Likelihood method and Le_Gascuel_2008 model. The amino acid sequences of nidovirus 3CL^{pro} were aligned using MAFFT. Different viruses are indicated at the terminal nodes, different colors represent different suborders: Ronidovirineae & Mesnidovirineae (orange), Cornidovirineae (blue), Tornidovirineae (red), Arnidovirineae (green), Monidovirineae (purple).

Figure 2. Identification of a PToV ORF1a-encoded 3C-like serine protease. (A) As predicted by TMHMM, the predicted 3CL^{pro} region was flanked by two transmembrane domains (red). **(B)** Schematic representation of PToV pp1a/1ab 3012-3392 and fusion protein constructs used in this experiment. The putative 3CL^{pro} domain is indicated in yellow, and flanking regions are shown in white ,the GST tag is shown in green. **(C)** Prokaryotic expression analysis of the GST (lanes 2 and 5), WT-GST (lanes 3 and 6), or S160A-GST (lanes 4 and 7) fusion proteins. The bacteria were mock induced (lanes 2, 3 and 4) or induced with 0.8 mM IPTG for 5 h at 37°C (lanes 5, 6 and 7). Total cell lysates were separated by sodium dodecyl

761 sulfate-polyacrylamide gel electrophoresis in a 12% polyacrylamide gel and stained
762 with Coomassie brilliant blue R-250 , then destaning with 50% methanol in a short
763 time (Fig. 2C upper) or long time (Fig. 2C lower). The positions of the fusion protein
764 and cleavage products are indicated by arrowheads.

765 **Figure 3. Characterization of the N-terminal and C-terminal PToV 3CL^{pro}**
766 **self-processing site by protein sequencing. (A-B)** Results of N-terminally sequenced.
767 The Fig. A and B indicate N-terminal and C-terminal self-processing site by protein
768 sequencing, respectively. Chromatograms of PTH amino acids from reaction cycles 1
769 to 5. Specific peaks are indicated by the single-letter code.

770 **Figure 4. Identification of catalytic residues in the PToV 3CL^{pro} active site. (A)**
771 Multiple sequence alignment of arterivirus and torovirus 3CL^{pro} domains. Clustal
772 Omega was used to multiple sequence alignment of torovirus and arterivirus 3CL^{pro}s
773 (69). Abbreviations of virus names and GenBank accession numbers for the sequences
774 are as follows: EAV, equine arteritis virus (NC_002532); LDV; lactate
775 dehydrogenase-elevating virus neurovirulent type C (NC_02534); and PRRSV,
776 porcine reproductive and respiratory syndrome virus strain Lelystad (AY588319);
777 EToV, equine torovirus strain Berne (X52374); PToV, porcine torovirus strain HB-1
778 (MH_603532) ; BRV, Breda virus (NC_007447). Secondary structure element of the
779 EAV 3CL^{pro} is shown above the sequence (PDB code: 1MBM). This was extracted
780 from the PDB file and added to the alignment using ESPript 3.0 (70). Residues shared
781 by porcine torovirus 3CL^{pro} and any of the other sequences are indicated by dark
782 arrowheads. **(B)**Schematic representation of the eukaryotic expression construct used

783 in this experiment. The putative 3CL^{pro} domain is indicated in yellow, and flanking
784 regions are shown in white. The two GST tag between PToV pp1a/1ab 3012-3392 are
785 shown in blue. 3CL^{pro} mut, putative 3CL^{pro} domain containing a Ser-160-to-Ala
786 change in the expressed sequence. **(C)** HEK-293T cells were transfected with
787 wild-type p-GST-WT-GST or p-GST-WT-GST mutants, including p-GST-H53A-GST,
788 p-GST-S160A-GST, and p-GST-E91A-GST. Cells were then lysed after 28 h
789 transfection and evaluated by western blotting. **(D)** HEK-293T cells were transfected
790 with p-GST-S160A-GST along with HA-tagged PToV 3CL^{pro} expression plasmid or
791 PToV 3CL^{pro} mutants, including pH53A, pS160A, and pE91A. Cells were then lysed
792 after 28 h transfection and evaluated by western blotting.

793 **Figure 5. Characterization of self-cleavage substrate specificities of PToV 3CL^{pro}.**
794 **(A)** Schematic representation of the eukaryotic expression constructs used in this
795 experiment. Positions of individual cleavage sites are indicated, and calculated
796 molecular masses of processing products are given. Cleavage products are represented
797 by symbols on the right of each lane as shown in the schematic representation. **(B)**
798 HEK-293T cells were transfected with wild-type p-GST-WT-GST or p-GST-WT-GST
799 mutants, including p-GST-Q3088A-GST, p-GST-Q3095A-GST, p-GST-Q3377A-GST
800 and p-GST-Q3382A-GST. Cells were then lysed after 28 h transfection and evaluated
801 by western blotting. **(C)** As shown in Figure A, two p-GST-WT-GST truncates were
802 constructed, including GST-N and C-GST, then the truncates were transfected with
803 HEK-293T cells, and lysed for Western blot after 28 h post-transfection.

804 **Figure 6. P4-F plays a key role in the N-terminal self-cleavage of PToV 3CL^{pro}.**

805 **(A)** Schematic representation of the eukaryotic expression constructs used in this
806 experiment. Positions and sequences of individual cleavage sites are indicated in left
807 and calculated molecular masses of processing products are given. Cleavage products
808 are represented by symbols on the right of each lane. **(B)** Schematic representation the
809 sequences of deletions or mutations in the N-terminal cleavage region of PToV 3CL^{pro}.
810 **(C-E)** The p-GST-WT-GST was selected as the template, and various p-GST-WT-GST
811 mutants were constructed as shown in Figure B**.** HEK-293T cells were transfected
812 with p-GST-WT-GST or various p-GST-WT-GST mutants. Cells were lysed after 28 h
813 transfection and then analyzed by western blotting. **(F)** The p-GST-S160A-GST was
814 selected as the template, and the P1 or P4 position was single-site mutated or two-site
815 mutated to A**.** HEK-293T cells were transfected with HA-tagged pCAGGS-PToV
816 3CL^{pro} along with p-GST-S160A-GST or p-GST-S160A-GST mutants, including
817 S160A+F-P4-A, S160A+Q-P1-A and S160A+(F-P4-A/Q-P1-A). Cells were then
818 lysed after 28 h transfection and evaluated by western blotting.

819 **Figure 7. Identification of cleavage sites recognized by PToV 3CL^{pro} within**
820 **polyproteins. (A)** HEK-293T cells cultured in 24-well plates were transfected with
821 various luciferase reporter plasmids and pRL-TK (Promega), together with 3CL^{pro}
822 expression plasmid pCAGGS-PToV 3CL^{pro} or the empty control plasmid. At 28 h
823 post-transfection, the cells were lysed for dual luciferase assay. **(B)** The results show
824 the identified sequence recognized by PToV 3CL^{pro}. Amino acids highlighted in red
825 represents highly conserved positions. **(C)** The identified sequences in Fig. B were
826 submitted to PSSMSearch for modeling and visualization of the determinants of

substrate specificities of the identified proteins (71). **(D)** HEK-293T cells in 24-well plates were transfected with CP4/nsp2 reporter or CP4/nsp2 mutants with single amino acid substitutions at P4 position and pRL-TK, together with 3CL^{pro} expression plasmid pCAGGS-PToV 3CL^{pro} or the empty control plasmid (named HA). At 28 h post-transfection, the cells were lysed for dual luciferase assay.

Figure 8. Investigation of the putative S1 and S4 pockets within PToV 3CL^{pro}. (A) Homology model of PToV 3CL^{pro} was generated in Modeller 9 using EAV 3CL^{pro} (PDB accession number 1MBM) as a template. **(B)** For ease of observation, the simulated substrate in the complex is shown in green. **(C)** Interaction between the P1 position and each amino acid in the S1 pocket. Light blue or green carbon atoms represent proteases and substrates, respectively. **(D)** Interactions between the P4 position and each amino acid in the S4 pocket. Red represents Van der Waals force and blue represents static electricity. **(E)** Correlation between atomic fluctuations of protease active site residues and the substrate. A warm color in the matrix indicates an increase in correlation. The color of the surface residue is used to indicate its position in the active site.

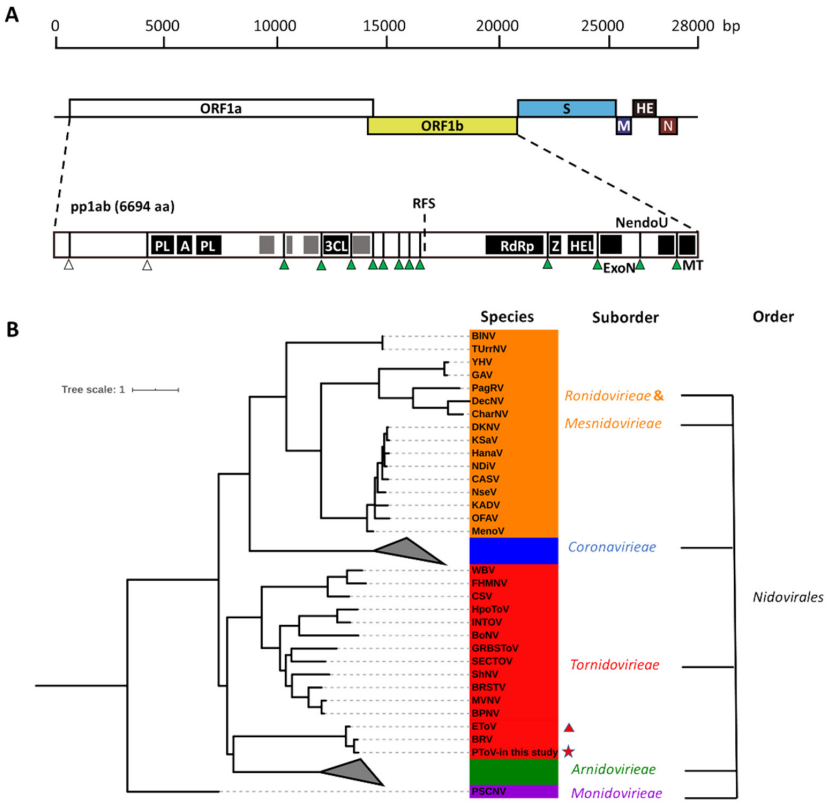
Figure 9. Identification of key amino acid within putative PToV 3CL^{pro} S1 and S4 pockets. (A) HEK293T cells were co-transfected with CP4/nsp2 and pRL-TK along with pCAGGS-PToV 3CL^{pro} (named WT) or S1 pocket key amino acids mutants D159A, H174A. At 28 h post-transfection, the cells were lysed for dual luciferase assay and Western Blot. **(B)** HEK293T cells were co-transfected with CP4/nsp2 and pRL-TK along with pCAGGS-PToV 3CL^{pro} (named WT) or S4 pocket key amino

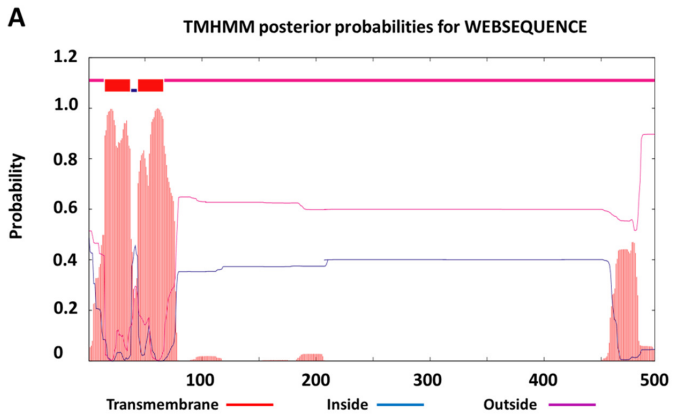
849 acids mutants P62A, L185A, P62G and L185G. At 28 h post-transfection, the cells
850 were lysed for dual luciferase assay and Western Blot. **(C)** HEK293T cells were
851 co-transfected with CP4/nsp2 and pRL-TK along with pCAGGS-PToV 3CL^{pro}
852 (named WT) or S4 pocket key amino acids mutants P65A/L185A and P65G/L185G.
853 At 28 h post-transfection, the cells were lysed for dual luciferase assay and Western
854 Blot.

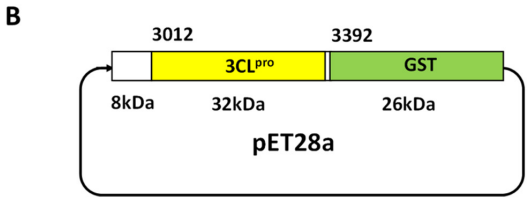
855 **Figure 10. Identification of key amino acid within PToV 3CL^{pro} S1 and S4**
856 **pockets in binding substrate peptide nsp4/5. (A)** The luciferase vector contains
857 substrate peptide nsp4/5 (PTIMWQ↓SDDVAD) was co-transfected with pRL-TK
858 along with pCAGGS-PToV 3CL^{pro} (named WT) or S1 pocket key amino acids
859 mutants D159A, H174A into HEK293T cells. At 28 h post-transfection, the cells were
860 lysed for dual luciferase assay and Western Blot. **(B-C)** The luciferase vector contains
861 substrate peptide nsp4/5 (PTIMWQ↓SDDVAD) was co-transfected with pRL-TK
862 along with pCAGGS-PToV 3CL^{pro} (named WT) or S4 pocket key amino acids
863 mutants P62A, L185A, P62G, L185G, P65A/L185A and P65G/L185G into HEK293T
864 cells. At 28 h post-transfection, the cells were lysed for dual luciferase assay and
865 Western Blot.

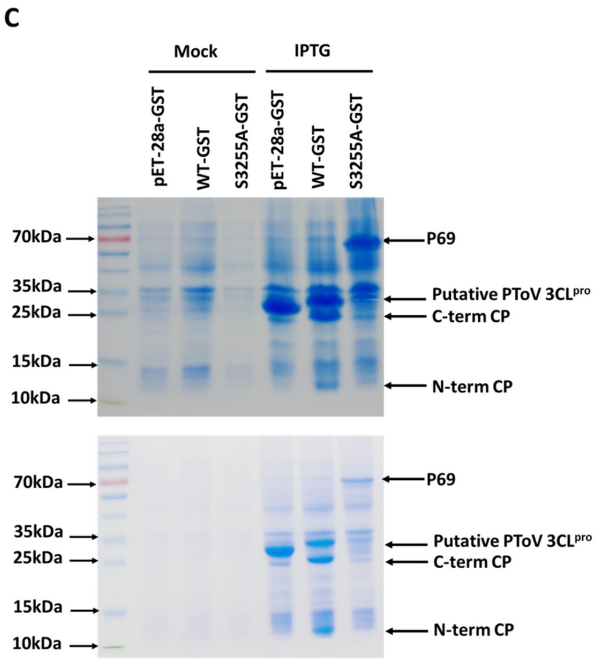
866 **Figure 11. Identification of key amino acid within PToV 3CL^{pro} S1 and S4**
867 **pockets in binding substrate peptide nsp7/9. (A)** The luciferase vector contains
868 substrate peptide nsp7/9 (PIYQPQ↓SVRYAD) was co-transfected with pRL-TK along
869 with pCAGGS-PToV 3CL^{pro} (named WT) or S1 pocket key amino acids mutants
870 D159A, H174A into HEK293T cells. At 28 h post-transfection, the cells were lysed

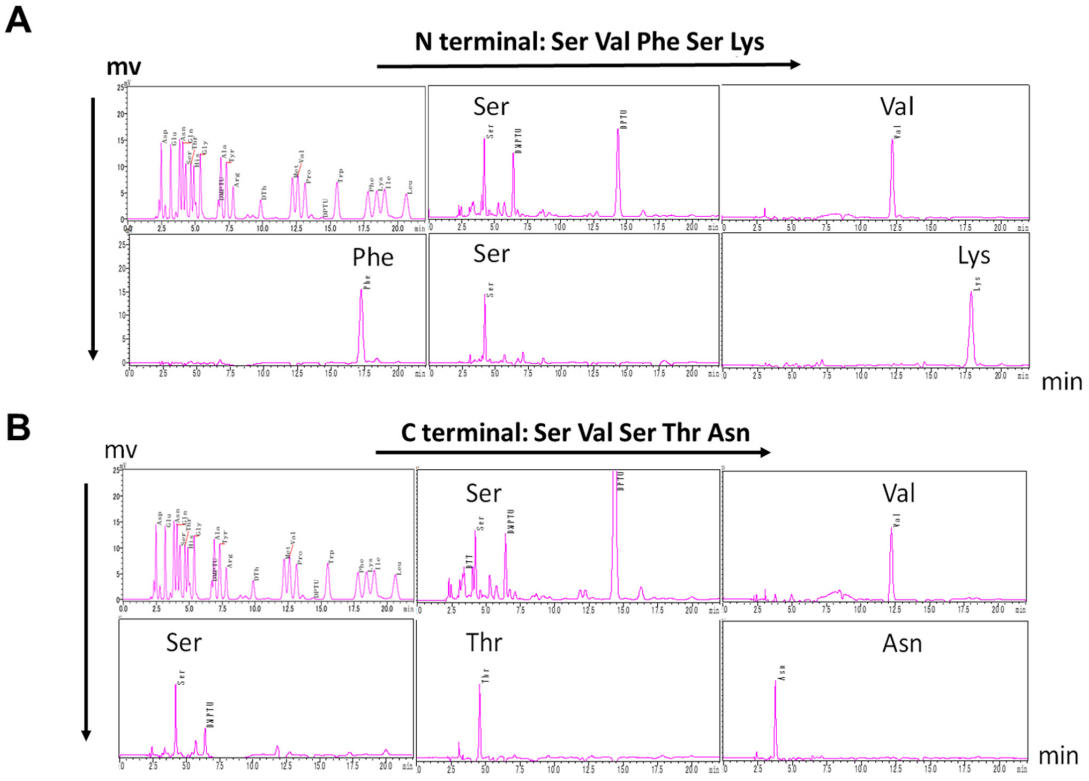
for dual luciferase assay and Western Blot. **(B-C)** The luciferase vector contains substrate peptide nsp7/9 (PIYQPQ↓SVRYAD) was co-transfected with pRL-TK along with pCAGGS-PToV $3CL^{pro}$ (named WT) or S4 pocket key amino acids mutants P62A, L185A, P62G, L185G, P65A/L185A and P65G/L185G into HEK293T cells. At 28 h post-transfection, the cells were lysed for dual luciferase assay and Western Blot.

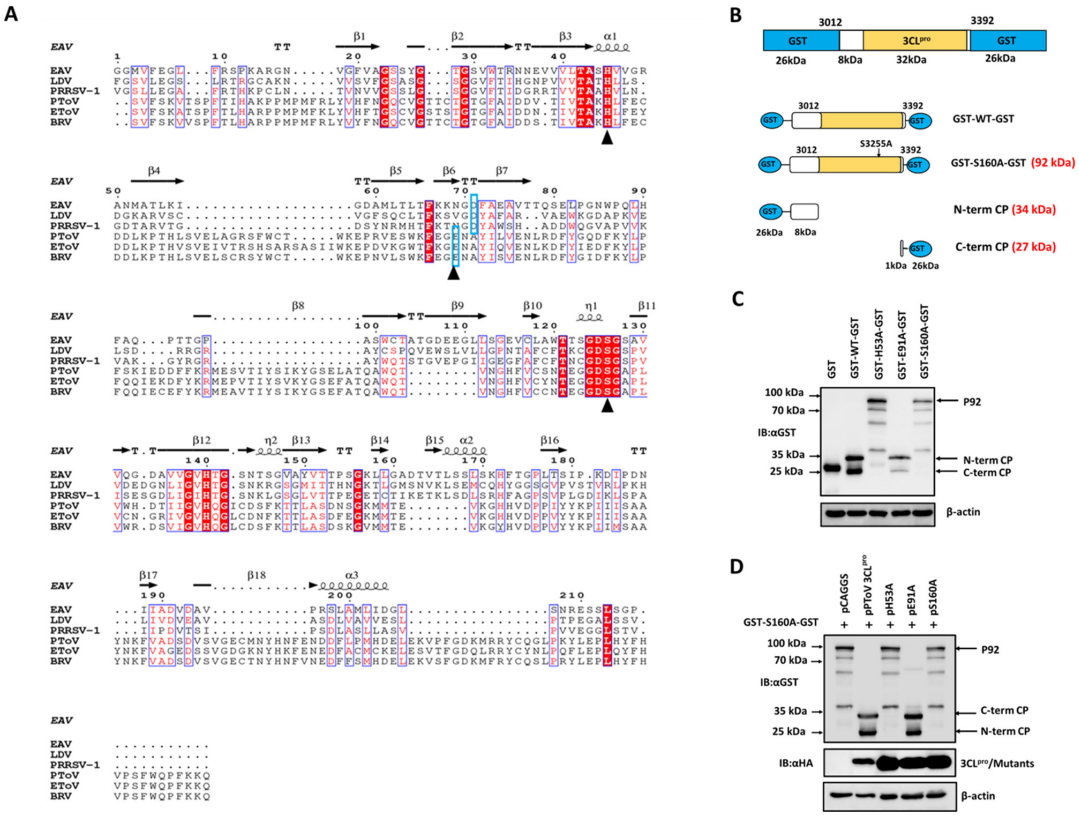


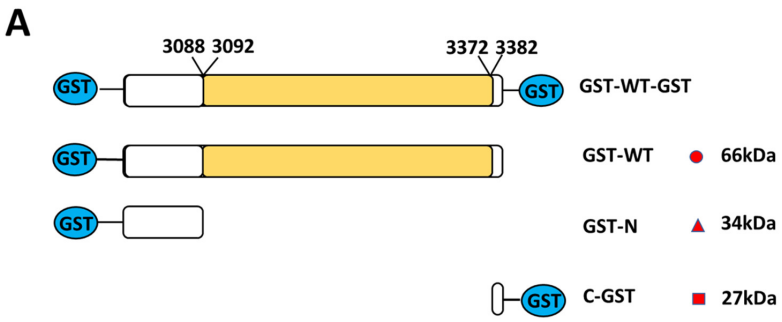


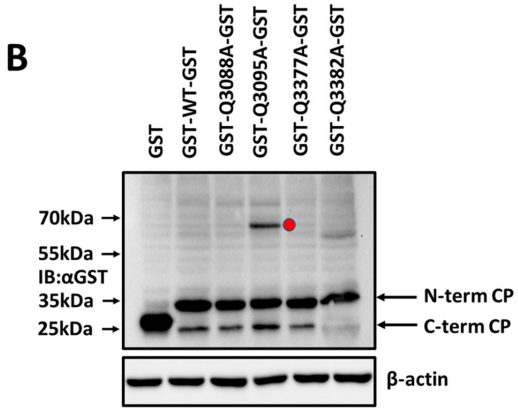


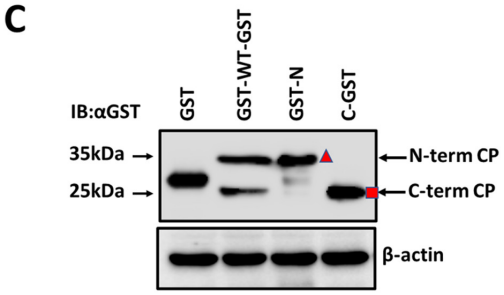


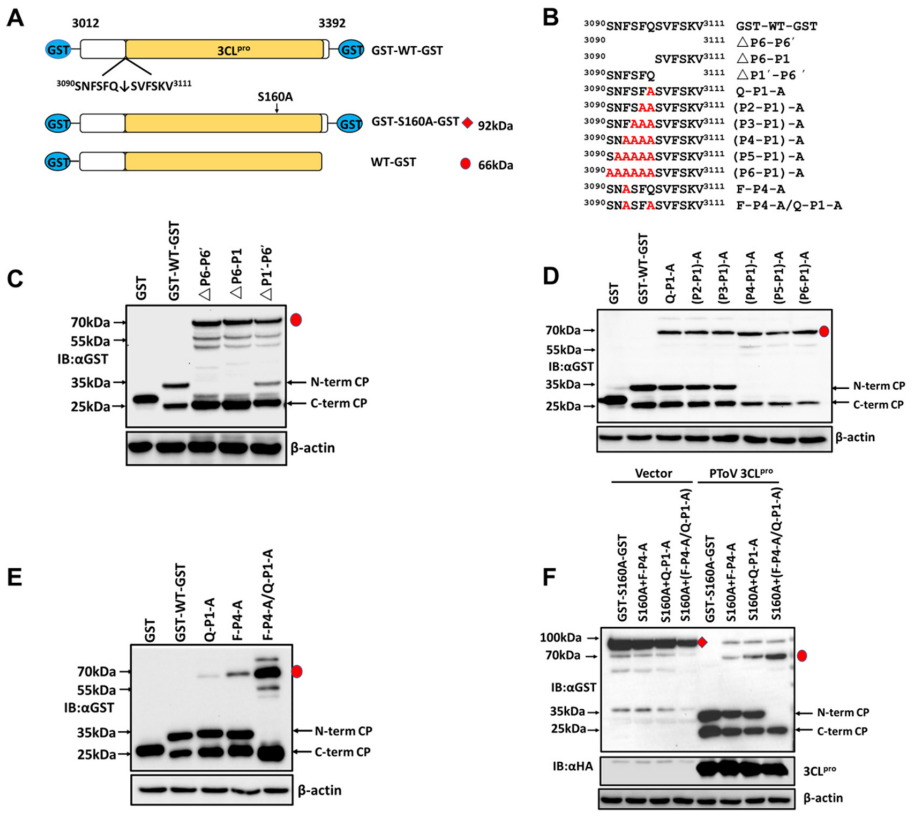


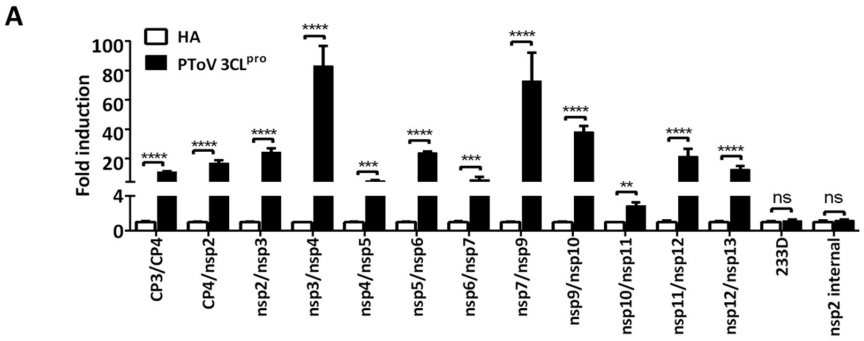












B

3CLpro Cleavage Site	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'	P6'
CP3/CP4	G	K	F	K	K	Q	S	I	A	Y	F	S
CP4/nsp2	S	N	F	S	F	Q	S	V	F	S	K	V
nsp2/nsp3	Q	P	F	K	K	Q	S	V	S	T	N	V
nsp3/nsp4	E	V	F	S	P	Q	A	G	T	Y	N	V
nsp4/nsp5	P	T	I	M	W	Q	S	D	D	V	A	D
nsp5/nsp6	G	V	F	K	Q	Q	A	N	I	G	Y	K
nsp6/nsp7	R	R	L	R	Y	Q	G	V	R	L	G	G
nsp7/nsp9	P	I	Y	Q	P	Q	S	V	R	I	A	D
nsp9/nsp10	C	E	F	E	Q	Q	S	S	L	D	K	V
nsp10/nsp11	F	G	M	E	K	Q	S	D	F	N	I	V
nsp11/nsp12	S	D	Y	V	E	Q	S	L	S	P	A	A
nsp12/nsp13	Q	T	F	Y	P	Q	K	D	F	I	R	S