Brief Communication

Autocatalytic activity and substrate specificity of the pestivirus N-terminal protease N^{pro}

Keerthi Gottipati [a], Sudheer Acholi [a], Nicolas Ruggli [b], Kyung H. Choi [a,*]

[a] Department of Biochemistry & Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, The University of Texas Medical Branch, Galveston, TX 77555-0647, USA
[b] Institute of Virology and Immunology, CH-3147 Mittelhäusern, Switzerland

ABSTRACT

Pestivirus N^{pro} is the first protein translated in the viral polypeptide, and cleaves itself off co-translationally generating the N-terminus of the core protein. Once released, N^{pro} blocks the host's interferon response by inducing degradation of interferon regulatory factor-3. N^{pro}'s intracellular autocatalytic activity and lack of trans-activity have hampered in vitro cleavage studies to establish its substrate specificity and the roles of individual residues. We constructed N^{pro}-GFP fusion proteins that carry the authentic cleavage site and determined the autoproteolytic activities of N^{pro} proteins containing substitutions at the predicted catalytic sites Glu22 and Cys69, at Arg100 that forms a salt bridge with Glu22, and at the cleavage site Cys168. Contrary to previous reports, we show that N^{pro}'s catalytic activity does not involve Glu22, which may instead be involved in protein stability. Furthermore, N^{pro} does not have specificity for Cys168 at the cleavage site even though this residue is conserved throughout the pestivirus genus.

© 2014 Elsevier Inc. All rights reserved.

Introduction

Pestiviruses belong to the family *Flaviviridae*. The *Flaviviridae* also encompass the genus hepacivirus, which includes hepatitis C virus, and the genus flavivirus, which includes dengue, West Nile, and yellow fever viruses (Lindenbach et al., 2007). Pestiviruses are highly contagious animal pathogens and include bovine viral diarrhea virus (BVDV), classical swine fever virus (CSFV), and the border disease virus (BDV) of sheep. The pestiviruses are enveloped viruses whose genomes are comprised of positive-sense single strand RNA (ssRNA) of approximately 12.5 Kb (Meyers and Thiel, 1996). The RNA genome consists of a 5'-untranslated region (5'-UTR), a single open reading frame (ORF), and a 3'-untranslated region (3'-UTR). The 5'-UTR acts as an internal ribosomal entry site for initiation of cap-independent translation. The ORF encodes a large polyprotein of approximately 3900 amino acids that is processed by viral and cellular proteases into 12 mature proteins whose order is N^{pro}–C–E^{rns}–E1–E2–p7–NS2–NS3–NS4A–NS4B–NS5A–NS5B. The structural proteins, the core protein C and the envelope proteins E^{rns}, E1 and E2 form the virion along with the RNA genome. The rest of the proteins are non-structural (NS) proteins that are involved in polyprotein processing, viral replication, and evasion of the cellular antiviral defense.

The first protein encoded by pestiviruses is the N-terminal protease, N^{pro}. This protein has two known functions. First, it is a cysteine autoprotease that cleaves the bond between Cys168 and Ser169 of the viral polyprotein via *cis*-cleavage, generating the N-terminus of the core protein (Stark et al., 1993). After the initial cleavage, N^{pro} is no longer proteolytically active, and *trans*-activity has not been observed (Wiskerchen et al., 1991). Sequence alignment with other proteases shows no significant homology, so pestivirus N^{pro} constitutes its own peptidase subfamily, C53, within the cysteine protease family (Rawlings et al., 2012). The putative catalytic residues were originally identified as Glu22–His49–Cys69, as mutations of any of these residues blocks cleavage between N^{pro} and the core protein (Rumenapf et al., 1998). Additionally, a deletion of the N-terminal 22 residues, including the predicted Glu22 triad residue, abrogates N^{pro}'s ability to cleave itself off from the viral polyprotein (Ruggli et al., 2009; Rumenapf et al., 1998). Mutations in the C-terminus, i.e., the substrate of the protease, also affect cleavage. Substitutions of Trp164 to Ala or Cys168 to Glu abolish the self-cleavage reaction, indicating that the substrate binding site in N^{pro} spans at least 5 residues from the C-terminus (Wiskerchen et al., 1991). Recently, crystal structures of N^{pro} from CSFV and BVDV have been determined (Gottipati et al., 2013; Zogg et al., 2013). N^{pro} has a 'clam-shell'-like fold consisting

of a protease domain containing a Cys69-His49 catalytic diad and a zinc-binding domain. The C-terminus of the protein is bound in the protease domain, suggesting a mechanism for autoproteolysis and subsequent auto-inhibition. Particularly, the C-terminal β-strand (one half of the product peptide) remains buried in the active site pocket indicating that once cleaved, the C-terminus of N^{pro} acts as an intramolecular inhibitor and thus prevents *trans*-activity, i.e., the enzyme is inactive towards additional substrates. The structures also show that Glu22 is not in the active site, despite its importance in N^{pro}'s catalytic activity. A second function of N^{pro} is to prevent interferon-α/β induction in infected cells (Gil et al., 2006; Ruggli et al., 2003). N^{pro} induces proteasomal degradation of interferon regulatory factor 3 (IRF3), a transcription factor for the interferon-β gene. Thus the presence of N^{pro} significantly decreases the amount of available IRF3, disrupting the type I interferon induction (Bauhofer et al., 2007; Chen et al., 2007). Many residues in both the N- and C-terminal regions of N^{pro} are thought to be involved in the IRF3 antagonist function. In particular, the C-terminal zinc-binding site residues Cys112, Cys134, Asp136, and Cys138 are essential for the IRF3 interaction (Szymanski et al., 2009). N^{pro} has no counterpart in either hepatitis C virus or flaviviruses, and is not required for viral replication (Tratschin et al., 1998).

Although N^{pro} provides a unique and early target in the development of therapeutics against pestiviral infections, *in vitro* studies of its autoproteolytic activity and substrate specificity were difficult to perform due to N^{pro}'s intracellular autocatalytic activity and lack of *trans*-activity. We have thus constructed N^{pro}-green fluorescent protein (GFP) fusion proteins that contain an authentic *cis*-cleavage site between N^{pro} and GFP. To establish the roles of N^{pro} residues in catalysis, we separately introduced a 17-residue N-terminal deletion, and substitutions at the predicted active site Glu22 and Cys69 and cleavage site Cys168 within the N^{pro}-GFP protein. Combined with N-terminal sequencing and mass spectrometry, we show that the first 17 residues from the N-terminus are dispensable for the protease activity, and Cys69 is required for protease activity, consistent with previous cell-based assays. However, N^{pro}'s catalytic activity does not require Glu22. In addition, N^{pro} does not have specificity for Cys168 at the

cleavage site, and this site can accommodate small to medium sized amino acids that have non-branched beta carbon. Our results suggest new roles for Glu22 and Cys168 in stabilization of the protein fold and in the interferon subversion function.

Results

Construction of N^{pro}-GFP fusion proteins

N^{pro} only has a *cis*-cleavage activity at its C-terminus, and does not cleave peptide substrates *in trans*; thus synthetic peptide substrates cannot be used for activity assays. In order to establish an *in vitro* autoprotease assay, we have constructed an N^{pro} fusion protein with GFP at its C-terminus (Fig. 1A). The fusion protein consists of an N-terminal His$_6$-tag, N^{pro}, the N-terminal four residues of the core protein (Ser–Asp–Asp–Gly, the authentic N^{pro} cleavage site), GFP, and a C-terminal His$_6$-tag. Because the autoproteolytic process occurs during protein translation, the cleavage product must be stable until it is detected in our assay. We thus chose to add a protein rather than a short peptide at the C-terminus of N^{pro}. His$_6$-tags were incorporated at both the N- and C-termini of the fusion protein so that the intact precursor protein as well as the cleaved N^{pro} and GFP products could be captured using single-step immobilized metal affinity chromatography (IMAC). The molecular weights for intact N^{pro}-GFP fusion protein, N^{pro}, and GFP are 48, 19, and 29 kDa, respectively, so they could easily be separated and identified by SDS-PAGE.

We first introduced an N-terminal deletion of 17 amino acids (Δ17) in N^{pro} as a positive control for functional N^{pro}, relating to the N-terminally-truncated N^{pro} for which the crystal structure was determined (Gottipati et al., 2013). It has been reported that up to 21 amino acids from the N-terminus of N^{pro} can be deleted without adverse effect on proteolytic activity in a cell-based assay (Rumenapf et al., 1998). The Δ17 N^{pro} was indeed proteolytically active, and two cleaved protein bands, 19 kDa N^{pro} and 29 kDa GFP, were clearly visible on SDS-PAGE after single-step IMAC purification (Fig. 1B). No uncleaved N^{pro}-GFP fusion protein was observed in the soluble fraction. In order to identify the cleavage site in the
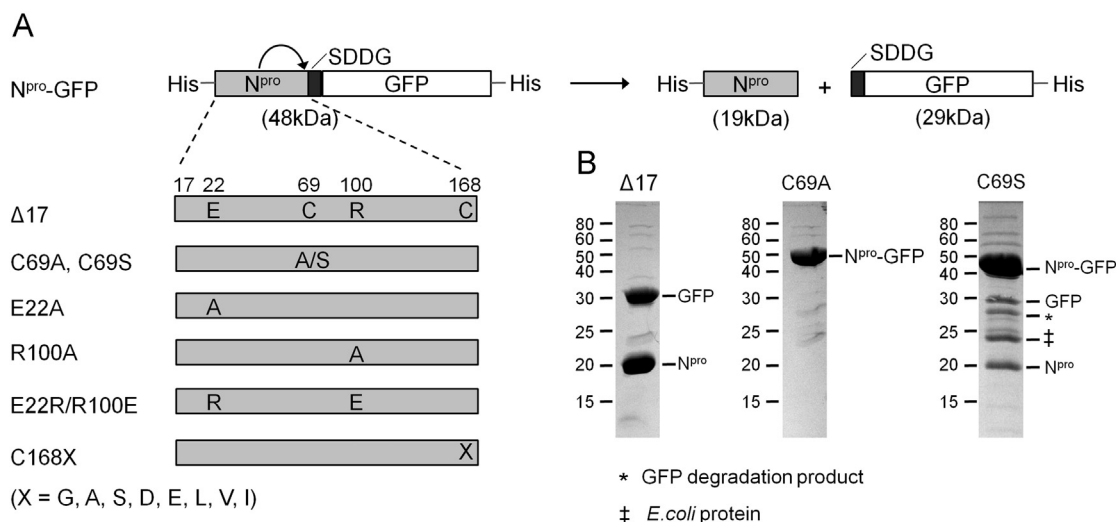

Fig. 1. N^{pro}-GFP fusion protein constructs. (A) Schematic of the N^{pro}-GFP fusion protein. The fusion protein consists of an N-terminal His$_6$-tag, N^{pro}, SDDG, GFP and a C-terminal His$_6$-tag. SDDG represents the first four residues (Ser–Asp–Asp–Gly) of the core protein, the authentic N-terminus of the natural cleavage product of N^{pro}. Upon autocleavage of N^{pro}, the N^{pro}-GFP fusion construct (48 kDa) will release full-length N^{pro} (19 kDa) and GFP (29 kDa). The autocleavage site is indicated with an arrow, and expected molecular weights of substrate and cleaved products are shown below the schematic. The various N^{pro} mutants used in this study are shown below the full-length N^{pro}. (B) Autoprotease activities of the Δ17, C69A, and C69S N^{pro}-GFP fusion proteins. The His$_6$-tagged proteins were purified by IMAC, eluted with 200 mM imidazole, and analyzed using SDS-PAGE. The asterisk (*) represents GFP degradation product that was likely generated by *E. coli* proteases. Bacterial protein identified by mass spectrometry is indicated with double dagger (‡).

N^{pro}-GFP fusion protein, the cleaved ~30 kDa GFP protein band was subjected to N-terminal peptide sequencing. The protein had a Ser–Asp–Asp–Gly sequence as its first four residues, confirming that N^{pro} cleaved between Cys168 and Ser169, as predicted. Thus the first 17 amino acids of N^{pro} are dispensable for proteolytic activity, consistent with the previous results. This N^{pro}-GFP fusion protein was used to introduce further substitutions in N^{pro}.

Cys69 is the catalytic nucleophile in N^{pro}

N^{pro} was predicted to harbor a cysteine protease triad composed of Glu22–His49–Cys69 using site-directed mutagenesis combined with translation in cell-free lysates (Rumenapf et al., 1998) and in cell-based assays (Ruggli et al., 2009). A Cys69 to Ala substitution abolished N^{pro}'s autoproteolytic activity, indicating that Cys69 is the catalytic nucleophile. However, a Cys69 to Ser substitution was reported to be either inactive in cell-free lysates (Rumenapf et al., 1998) or partially active in cell-based assays (Ruggli et al., 2009). We introduced Cys69 to Ala (C69A) or Ser (C69S) substitution in the N^{pro}-GFP fusion protein, and measured the *cis*-activities of the resulting proteins. The C69A N^{pro} was inactive; only the uncleaved ~50 kDa fusion protein was observed (Fig. 1B), consistent with the role of Cys69 as a catalytic nucleophile. In contrast, the C69S N^{pro}-GFP fusion protein produced both the uncleaved fusion protein (major product), and cleaved 19 kDa N^{pro} and 29 kDa GFP (Fig. 1B). Thus C69S N^{pro} was partially active, and Ser could replace Cys69 as the nucleophile to some extent. Additionally, two other protein bands at 27 and 23 kDa were apparent. Therefore the identities of the four protein bands at 19, 23, 27, and 29 kDa were determined by in-gel tryptic digestion followed by mass spectrometry. The 29 and 27 kDa protein bands were derived from GFP, and the 19 kDa protein band was N^{pro}. The 23 kDa protein was an *E. coli* bacterial protein. In order to determine whether the two GFP bands were produced by N^{pro} cleavage, the N-termini of both the 29 and 27 kDa proteins were sequenced. The 29 kDa protein had Ser–Asp–Asp–Gly as its first four residues, confirming that the C69S N^{pro} protein is partially active and able to cleave the peptide bond between C168 and S169 of the fusion protein. The 27 kDa protein band contained a mixed species whose N-terminus coincided with an internal GFP sequence, suggesting that the N^{pro}-GFP fusion protein containing C69S N^{pro} is unstable and is degraded by cellular proteases.

Glu22 is not essential for catalytic activity, but may be required for protein stability

Substitutions of Glu22 to either Val or Leu have been shown either to lose their proteolytic activity (Chen et al., 2007; Gil et al., 2006; Rumenapf et al., 1998) or to be partially active (Ruggli et al., 2009). In addition, deletion of the N-terminal 22 residues including Glu22 resulted in an inactive protein, while deletion of the N-terminal 21 residues resulted in an active protease (Rumenapf et al., 1998), suggesting that Glu22 may be essential for the catalytic mechanism. However, recent crystal structures of N^{pro} show that Glu22 is 25 Å away from Cys69, and is therefore unlikely to be involved in the catalysis (Fig. 2A). Instead, Glu22 forms a salt bridge with Arg100, suggesting that this interaction may be important for the stability of the protein. To determine whether Glu22 is essential for proteolytic activity, we introduced a Glu22 to Ala substitution in the N^{pro}-GFP fusion protein. The E22A fusion protein had the uncleaved N^{pro}-GFP protein as a major product, but many smaller molecular weight proteins between 20 and 30 kDa were also observed by SDS-PAGE (Fig. 2B). These proteins were identified by in-gel staining with a fluorescent solution (InVision™ His-tag In-gel stain) that detects His_6-tagged proteins. Most of the protein bands were stained, indicating that they are degradation products of the N^{pro}-GFP fusion protein containing a His_6-tag, which is found at both the N- and C-termini of the fusion protein (not shown). In-gel trypsin digestion followed by mass spectrometry showed that both the 29 and 27 kDa protein bands were derived from GFP, similar to the products of the C69S N^{pro}-GFP fusion protein. The N-terminal sequencing identified Ser–Asp–Asp–Gly as the first four residues of the 29 kDa protein, indicating that a small amount of the fusion protein containing Glu22 to Ala substitution was cleaved at the C-terminus of N^{pro}. However, the 19 kDa N^{pro} protein, the other product of the cleavage reaction, was not observed on the SDS gel, suggesting that N^{pro} containing the E22A substitution is partially active but unstable, likely to be further degraded by *E. coli* proteases. N-terminal sequencing of the 27 kDa protein indicated the internal cleavage site of GFP, again suggesting that the E22A fusion protein was degraded by cellular proteases. Because the E22A N^{pro} has a residual proteolytic activity, it seems that the reported loss of activity in Glu22 to Val- or Leu-substituted N^{pro} is not caused by the direct involvement of Glu22 in catalysis. Rather, the Glu22 mutations likely caused an unstable protein structure that led to degradation of the N^{pro} protein itself.

Because Glu22 forms a salt bridge with Arg100, we also introduced Arg100 to Ala mutation and tested its proteolytic activity (Fig. 2B). If the salt bridge between Glu22 and Arg100 is important for protein stability, the R100A N^{pro} would also exhibit low protease activity. As expected, the N^{pro}-GFP fusion protein containing the Arg100 to Ala substitution showed multiple protein bands, similar to those from E22A fusion protein. Thus, it seems likely that the salt bridge between Glu22 and Arg100 is important for protein stability, and not for catalytic activity. To test whether reversal of the positions of the two residues in the salt bridge would
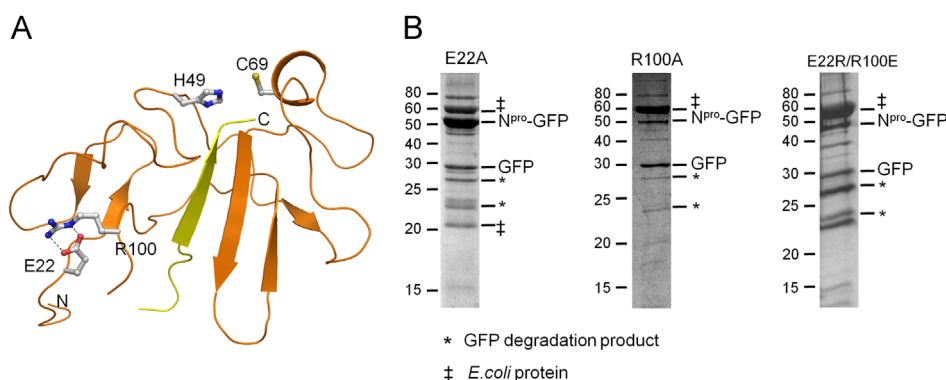

Fig. 2. Catalytic activities of N^{pro}-GFP fusion proteins containing substitutions at Glu22, Arg100, or both. (A) Structure of the N^{pro} protease domain. N^{pro} consists of a protease and a zinc-binding domain. Only the protease domain is shown for clarity (PDB accession code 4H9K). The locations of Glu22 and Arg100 in relation to the catalytic diad (Cys69 and His49) are indicated. Dashed lines indicate a salt bridge between Glu22 and Arg100. (B) Autoproteolytic activity of the E22A-, R100A-, and E22R/R100E-mutant N^{pro}-GFP fusion proteins. Asterisks (*) represent GFP degradation products and double daggers (‡) indicate bacterial proteins as identified by mass spectrometry.

maintain catalytic activity, we also introduced double mutations, Glu22 to Arg and Arg100 to Glu, and tested autoproteolytic activity of the resulting protein. The double mutant is also partially active, and the major product was the uncleaved full-length fusion protein with small amount of the cleaved GFP (Fig. 2B). As in the case of R100A, the cleaved N^{pro} was not observed. Thus, both Glu22 and Arg100 residues, and their relative geometry in the structure of N^{pro} seem to be important for stability of the protein.

N^{pro} is not specific for the Cys168 cleavage site

The cleavage site for a protease substrate is defined as …P3–P2–P1–P1'–P2'–P3'…, where cleavage occurs between the P1 and P1' residues (Schechter and Berger, 1967). The cleavage site of N^{pro} between Cys168 and Ser169 in the viral polyprotein is absolutely conserved in pestiviruses (Fig. 3A), suggesting that N^{pro} may be specific for the C-terminal Cys168 at the P1 site. Consistent with this notion, a Cys168 to Glu substitution abolishes autoproteolytic activity of N^{pro} (Achmuller et al., 2007). In order to determine whether N^{pro} requires Cys168 at the P1 site for cleavage, we constructed fusion proteins wherein Cys168 was mutated to Gly, Ala, Ser, Asp, Glu, Leu, Val or Ile and determined the autoproteolytic activities of the different mutants. Unexpectedly, mutation of the terminal cysteine residue did not always result in a complete loss of activity of the N^{pro} protease. The C168A and C168S fusion proteins were proteolytically active and produced only the cleaved N^{pro} and GFP proteins. Uncleaved N^{pro}-GFP fusion protein was not observed (Fig. 3B). To ensure that N^{pro} was cleaved at its C-terminus after residue 168, the 29 kDa GFP band was N-terminally sequenced. The first four residues of the cleaved protein were Ser–Asp–Asp–Gly, confirming the correct N^{pro} cleavage at its C-terminus. Therefore, N^{pro} is not specific for Cys168 at the cleavage site and can accommodate Ala or Ser at the P1 position. We also observed by mass spectrometry that the N^{pro} protein containing a substitution of Cys168 to Ala and insertion of ^{169}Ser–Asp–Asp–Gly^{172} at the C-terminus was cleaved after residue 168 (unpublished data), consistent with our observation that N^{pro} does not have a cleavage specificity for Cys168 at the P1 position.

The C168L, C168D and C168G fusion proteins were partially active and analysis of cleavage products indicated that cleaved N^{pro} and GFP proteins were present along with the uncleaved full-length N^{pro}-GFP fusion protein. On the other hand, analysis of the C168V, C168I, and C168E fusion proteins showed that the uncleaved protein was the major product, indicating that these proteins are largely inactive (Fig. 3B). In the case of C168I and C168E, small amounts of 27 kDa GFP (internal cleavage product of GFP) were also observed, along with the full-length 29 kDa GFP, suggesting that the Cys168 to Glu or Ile substitutions render the fusion protein unstable and subject to degradation by a cellular protease, similar to the C69S fusion protein.

Discussion

N^{pro} catalyzes self-cleavage during translation, and thus measuring its cleavage activity and identifying the cleavage products in the cellular extract were difficult. Thus the N^{pro} protease activity has been studied in a cell-free in vitro translation system, together with metabolic [^{35}S]-methionine labeling (Rumenapf et al., 1998) or using a denatured mutant N^{pro} that needs to be refolded to have protease activity (Achmuller et al., 2007). In order to readily address the protease activity without the use of radio-labeling or Western blot, we constructed an N^{pro}-GFP fusion protein in the E. coli expression system. Because N^{pro} processes self-cleavage during translation, we added His$_6$-tags at both the N- and C-termini so that the resulting proteins, whether cleaved or not, can be captured by a one-step purification using IMAC. This analysis enables us to identify both cleaved products, rather than
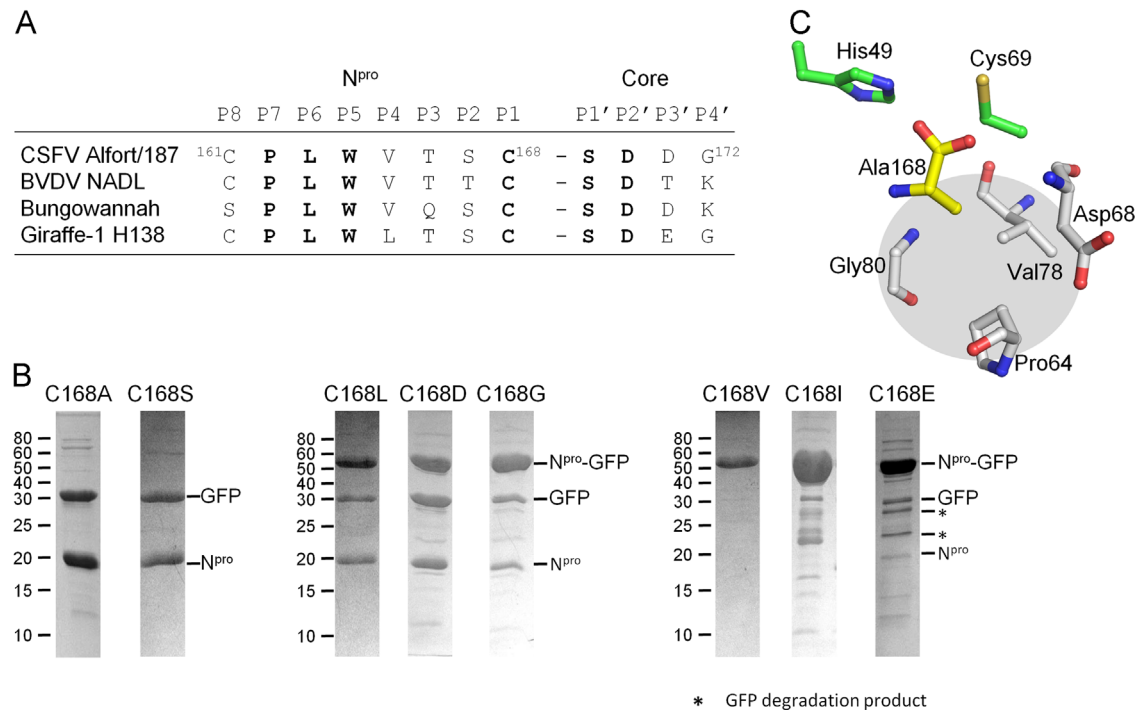

Fig. 3. N^{pro}'s autoproteolytic specificity for Cys168. (A) Alignment of the pestivirus N^{pro}-Core cleavage site sequences. The sequences of the CSFV strain Alfort/187 (GenBank accession number X87939), BVDV strain NADL (GenBank accession number M31182), Bungowannah virus (GenBank accession number DQ901403), and Giraffe-1 strain H138 (GenBank accession number NC_003678) are included in the alignment. (B) Autoproteolytic activities of N^{pro}-GFP fusion proteins containing Cys168 to Ala, Ser, Leu, Asp, Gly, Val, Ile, or Glu substitutions. Asterisks (*) represent GFP degradation products as identified by mass spectrometry. (C) Structure of the substrate binding pocket in C168A N^{pro}. Ala168 is bound in the shallow hydrophobic pocket (shaded in gray) consisting of residues Pro64, Asp68, Val78, and Gly80 (PDB accession code 4H9K). The residues Cys69 and His49 of the catalytic diad are shown in green, and the Ala168 is shown in yellow.

identifying a single product using an antibody against either N^{pro} or the attached protein. Our proteolysis results with the Δ17 N^{pro}-GFP fusion protein (positive control, functional N^{pro} protease) and the C69A N^{pro}-GFP fusion protein (negative control, non-functional N^{pro} protease) were consistent with the previous reports using the *in vitro* translation system (see below). Thus the N^{pro}-GFP fusion system provides a useful tool to study the autoproteolytic activity of N^{pro}.

So far, five residues in N^{pro} have been shown to be required for its protease activity. Substitutions of Glu22 (to Val or Leu), His49 (to Ala or Leu), Cys69 (to Ala or Ser), Trp164 (to Ala), and Cys168 (to Glu) resulted in a loss of proteolytic activity (Hilton et al., 2006; Rumenapf et al., 1998; Wiskerchen et al., 1991). In this study, we explored the roles of the predicted catalytic sites Glu22 and Cys69 (E22A, C69A, and C69S), Arg100 that forms a salt bridge with Glu22 (R100A, E22R/R100E) and cleavage site Cys168 (C168G, C168A, C168S, C168D, C168E, C168L, C168V and C168I) by introducing site-specific mutations into the N^{pro}-GFP fusion protein. The Δ17 N^{pro}-GFP fusion protein and the fusion proteins carrying the individual substitutions show different degrees of cleavage efficiencies. The individual mutations affect cleavage efficiencies in the following order: Δ17, C168A/S (complete cleavage)≈C168L/D/G > R100A, E22R/R100E ≈ E22A, C69S, C168E/I/V > C69A (no cleavage). Because the N^{pro}-GFP fusion protein assay does not measure cleavage rates, lethal mutants and significantly defective mutants would have similar cleavage results. The Δ17 N^{pro}-GFP fusion protein was fully active, and thus the first 17 residues are dispensable for proteolytic activity. Since the N-terminal deletion mutant that lacks the first 21 amino acids is active as a protease (Rumenapf et al., 1998), our result is in agreement with this report. The C69A-containing fusion protein was inactive, and no auto-cleavage was observed. This result is consistent with the N^{pro} crystal structure that established Cys69 as the catalytic nucleophile. Contradictory results have been reported for Cys69 to Ser substitution in that C69S N^{pro} led either to complete loss of the proteolytic activity using an *in vitro*-translated protein or partial activity in HEK 293T cells (Ruggli et al., 2009; Rumenapf et al., 1998). We found that the major product of the C69S-containing N^{pro}-GFP fusion protein is the uncleaved protein, indicating that auto-catalysis is significantly impaired by the Ser substitution (Fig. 1B). Nevertheless, small amounts of cleaved products were also observed, indicating that the C69S N^{pro} has partial activity. Thus, Ser can fill the role of the Cys69 nucleophile, as seen in some cysteine proteases (Lawson and Semler, 1991).

There are also contradictory reports on the role of Glu22, since substitution of Glu22 to Val or Leu led to either complete or partial inactivation of the protein (Chen et al., 2007; Gil et al., 2006; Ruggli et al., 2009; Rumenapf et al., 1998). As hydrophobic residues such as Val or Leu may have a negative influence on the Glu22 site, we replaced Glu22 with Ala in the N^{pro}-GFP fusion protein. The major product of the E22A-containing N^{pro}-GFP fusion protein is the uncleaved protein, although a small amount of correctly cleaved GFP was also produced (Fig. 2B). However, the cleaved N^{pro} was not observed in the gel, suggesting that degradation of E22A N^{pro} occurred during protein expression. This observation explains both reports on the activity of Glu22 mutants. Depending on the detection method, whether the cleavage was tested by the presence of N^{pro} or the linked protein, the activity of the Glu22 mutant would be inactive or partially active, respectively. As E22A N^{pro} has some marginal protease activity, Glu22 is unlikely to be involved in the catalysis. Rather, the substitution of Ala for Glu at residue 22 may destabilize the N^{pro}-GFP fusion protein, leading to non-specific degradation of GFP and N^{pro} during protein expression. This is also consistent with the cleavage result from Arg100 to Ala substitution. Glu22 and Arg100 form a salt bridge in the crystal structures (Gottipati et al., 2013; Zogg et al., 2013), and disruption of either residue led to the loss of

protease activity due to instability of the resulting proteins. It should be noted that when N^{pro}-GFP fusion proteins have low cleavage activity as in the case of E22A, C69S, R100A, E22R/R100A, C168I and C168E proteins, we often observe additional non-specifically cleaved protein products. Thus the stability of the N^{pro} protein is closely related to its catalytic activity.

The sequences of pestivirus N^{pro} from the P7 to P1 positions at its C-terminus are highly conserved, and thus N^{pro} was thought to have high cleavage specificity (Fig. 3A). Pro162, Leu163, Trp164, and Cys168 are absolutely conserved in the 36 sequences we examined, and the other three positions—Val165, Thr166, and Ser167—only allow similar substitutions. In addition, Trp164 to Ala and Cys168 to Glu substitutions abolished enzymatic activity, emphasizing the importance of the C-terminal residues in proteolysis (Achmuller et al., 2007; Wiskerchen et al., 1991). Surprisingly, substitutions of Cys168 to Ala or Ser in the N^{pro}-GFP fusion protein did not alter catalytic activity of N^{pro}. C168A and C168S N^{pro} are fully active, and the C168A- or C168S-containing N^{pro}-GFP fusion protein was completely cleaved into the individual N^{pro} and GFP proteins. No uncleaved fusion protein was observed after auto-catalysis. Therefore, N^{pro} does not recognize Cys168 for cleavage, and Cys168 is not required for the catalytic activity or substrate specificity. This contrasts with the recent result that N^{pro} has a preference for Cys over Ala at the P1 position (Zogg et al., 2013). When N^{pro} fusion protein was refolded from the inclusion bodies, approximately 40% of the C168A protein showed cleaved product, while the wild-type protein showed ~70% cleaved product (Zogg et al., 2013). Since the denatured protein (in 8 M guanidinium HCl) is incubated in a native buffer for 15 h at room temperature, C168A protein could potentially misfold and lose the activity during the refolding process. When the autoprotease activity was measured from the native protein, we do not see any difference between the wild-type (Δ17) and C168A N^{pro} (Figs. 1 and 3B).

We have determined the crystal structures of wild-type and C168A N^{pro}, providing structural information consistent with our biochemical results described here (Gottipati et al., 2013). The C-terminus of N^{pro} from the P7 to P1 sites forms a β-strand that is bound in the active site. Thus the C-terminal residues have to maintain main chain backbone interactions as well as the specific side-chain interactions, restricting the residues that can substitute for the C-terminal residues. In particular, the P1 site Ala168 in C168A N^{pro} is in a shallow hydrophobic binding pocket consisting of Pro64, Asp68, Val78, and Gly80, which likely allows only small hydrophobic residues at the P1 site (Fig. 3C). Cys or near-isosteric Ser residues would also fit into the P1 binding site without steric clashes.

We then tested whether N^{pro} can tolerate other substitutions at the P1 position. The substitution of Cys168 to Gly in the N^{pro}-GFP fusion protein resulted in partial loss of protease activity of N^{pro}. A Gly residue would create an empty P1 binding pocket along with increased flexibility of the substrate peptide, which may contribute to inefficient catalysis of the fusion protein cleavage. The substitution of Cys168 to Leu also resulted in partial activity, indicating that N^{pro} can accommodate moderately sized hydrophobic residues. However, mutation of Cys168 to Val or Ile resulted in a greater loss of activity than the Leu substitution, and the major product of C168V- or C168I-containing N^{pro}-GFP fusion protein was the uncleaved fusion protein (Fig. 3B). Compared to Leu, both Val and Ile are branched at the beta carbon (Cβ), and the bulky, Cβ-branched side chains likely restrict their access to the P1 binding pocket. Consistent with the previous result, the C168E-containing N^{pro}-GFP fusion protein has negligible protease activity. The uncleaved fusion protein was the major product found after autocleavage, with a small amount of cleaved GFP and N^{pro} proteins. However, the C168D substituted fusion protein is partially active and both cleaved products, GFP and N^{pro},

were observed (Fig. 3B). Since both Asp and Glu residues have negatively charged side chains and unbranched Cβ, the loss of enzymatic activity in C168E N^{pro} is likely because the long side chain of Glu168 (and not its negative charge) prevents its binding in the Cys168 binding pocket. Taken together, these results indicate that N^{pro} can accommodate small to medium size amino acids with non-branched Cβ at the P1 position. Although N^{pro} can accommodate other residues at position 168 without affecting catalytic activity, such as Ala and Ser, pestiviruses have conserved Cys168 at the P1 position throughout evolution. It is thus possible that the Cys residue may have a role in N^{pro}'s function as an IRF3 antagonist, or in viral pathogenesis.

Materials and methods

Construction of the N^{pro}-GFP fusion protein and site-directed mutagenesis

The N-terminal deletion mutant of N^{pro} lacking the first 17 amino acids (Δ17) was amplified from the full-length construct using the oligonucleotide primers 5' ggcagccatatgggagtggaggaaccggtatac 3' (forward) and 5' cgaattcggatccgccatcatcagagcaactggtaac 3' (reverse). The DNA fragment containing the N^{pro} gene was sub-cloned into the NdeI and BamHI restriction sites of the vector pET-28b to create pET28b+N^{pro}. A plasmid containing GFPuv, the "cycle 3" variant of green fluorescent protein from Aequorea victoria (jellyfish) described by Crameri et al. (1996), was obtained from Dr. Jose Barral at UTMB (Crameri et al., 1996; Fukuda et al., 2000). GFPuv was amplified using the oligonucleotide primers: 5' ggatccgaattcatggctagcaaaggagaagaacttttcac 3' (forward) and 5' cgagtgcggccgctttgtagagctcatccatgccatg 3' (reverse) that contain EcoRI and NotI restriction sites, respectively. GFPuv was inserted into the fusion vector pET28b+N^{pro} and digested with EcoRI and NotI to generate the final N^{pro}-GFP fusion construct: NdeI–N^{pro}–BamH1–EcoRI–GFPuv–NotI. The resulting N^{pro}-GFP fusion protein consists of an N-terminal cleavable hexahistidine (His$_6$) tag, N^{pro} residues from 18 to 168, a Ser–Asp–Asp–Gly sequence, GFPuv, and a C-terminal His$_6$-tag (Fig. 1). All mutant proteins were created using the N^{pro}-GFP fusion protein. The catalytic site C69A N^{pro}-GFP fusion protein was amplified from the full-length construct using the same oligonucleotide primers as for Δ17N^{pro}. The single mutants C69S, E22A, R100A, C168X (X=G, A, S, D, E, L, V or I) and the double mutants E22R/R100E of the N^{pro}-GFP fusion proteins were generated using QuikChange^{TM} site-directed mutagenesis kit (Stratagene) per the manufacturer's protocol. All plasmid constructs were verified by DNA sequencing in UTMB's molecular genomics core.

Measurements of autoprotease activity of N^{pro}-GFP fusion proteins

The N^{pro}-GFP plasmids containing the various mutations in N^{pro} were transfected into Escherichia coli Rosetta^{TM} cells (Novagen). For activity assays, 500 ml E. coli cultures were grown to an A_{600} of 0.6–0.8 in Luria-Bertani (LB) medium supplemented with 50 μg/ml ampicillin and 33 μg/ml chloramphenicol at 37 °C, at which point they were induced with 0.5 mM IPTG and growth was continued at 18 °C overnight. The cells were pelleted and stored at −20 °C until use.

The intracellular autocatalytic release of N^{pro} from the N^{pro}-GFP fusion protein precursor was monitored by immobilized metal affinity chromatography (IMAC) using His$_6$-tags. The harvested cells were suspended in lysis buffer [50 mM sodium phosphate (pH 7.0), 300 mM NaCl, and 5 mM β-mercaptoethanol] mixed with an EDTA-free protease inhibitor cocktail tablet (Roche). Cells were lysed by sonication with a Misonix XL-2000 sonicator, followed by

centrifugation to remove cellular debris. The soluble fraction was applied to 1 ml of pre-equilibrated Talon^{TM} (Clontech) metal affinity chromatography resin. After a 1 h incubation at 4 °C, the resin was washed with Talon-wash buffer (50 mM sodium phosphate buffer, pH 8.0 containing 0.5 M NaCl), and then with 5 mM imidazole in Talon-wash buffer to remove non-specifically bound proteins. His$_6$-tagged proteins were then eluted with a stepwise gradient of 10, 20, and 150–200 mM imidazole in Talon wash buffer, and analyzed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The intact N^{pro}-GFP fusion protein, as well as the cleaved protein products, was identified by their molecular sizes. In some cases the gel was also stained with InVision^{TM} His-tag In-gel stain (Invitrogen) according to the manufacturer's instructions, to identify proteins containing the His$_6$-tag. This stain contains a fluorescent dye attached to a Ni^{2+} complex, and binds to the His$_6$-tagged proteins. The stained gel was visualized on ImageQuant (GE healthcare) imaging system.

Identification of proteins by mass spectrometry and N-terminal peptide sequencing

The identities of protein bands on SDS gels were determined by MALDI-TOF at the UTMB mass spectrometry core facility. Briefly, protein bands were excised from the SDS-PAGE gel and incubated with a trypsin solution for 6 h at 37 °C. The resulting solution was loaded onto a mass spectrometry plate for MALDI-TOF analysis. The cleavage site in the N^{pro}-GFP fusion protein was determined by N-terminal peptide sequencing in the protein chemistry laboratory at UTMB. Following electrophoresis, the SDS gel was blotted onto PVDF membrane using Mini Trans-Blot cell (Bio-Rad). After Coomassie staining of the membrane, the protein bands were excised and used for N-terminal peptide sequencing.

References

Achmuller, C., Kaar, W., Ahrer, K., Wechner, P., Hahn, R., Werther, F., Schmidinger, H., Cserjan-Puschmann, M., Clementschitsch, F., Striedner, G., Bayer, K., Jungbauer, A., Auer, B., 2007. N(pro) fusion technology to produce proteins with authentic N termini in E. coli. Nat. Methods 4, 1037–1043.

Bauhofer, O., Summerfield, A., Sakoda, Y., Tratschin, J.D., Hofmann, M.A., Ruggli, N., 2007. Classical swine fever virus Npro interacts with interferon regulatory factor 3 and induces its proteasomal degradation. J. Virol. 81, 3087–3096.

Chen, Z., Rijnbrand, R., Jangra, R.K., Devaraj, S.G., Qu, L., Ma, Y., Lemon, S.M., Li, K., 2007. Ubiquitination and proteasomal degradation of interferon regulatory factor-3 induced by Npro from a cytopathic bovine viral diarrhea virus. Virology 366, 277–292.

Crameri, A., Whitehorn, E.A., Tate, E., Stemmer, W.P., 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat. Biotechnol. 14, 315–319.

Fukuda, H., Arai, M., Kuwajima, K., 2000. Folding of green fluorescent protein and the cycle3 mutant. Biochemistry 39, 12025–12032.

Gil, L.H., Ansari, I.H., Vassilev, V., Liang, D., Lai, V.C., Zhong, W., Hong, Z., Dubovi, E.J., Donis, R.O., 2006. The amino-terminal domain of bovine viral diarrhea virus Npro protein is necessary for alpha/beta interferon antagonism. J. Virol. 80, 900–911.

Gottipati, K., Ruggli, N., Gerber, M., Tratschin, J.D., Benning, M., Bellamy, H., Choi, K.H., 2013. The structure of classical swine fever virus N(pro): A novel cysteine autoprotease and zinc-binding protein involved in subversion of type I interferon induction. PLoS Pathog. 9, e1003704.

Hilton, L., Moganeradj, K., Zhang, G., Chen, Y.H., Randall, R.E., McCauley, J.W., Goodbourn, S., 2006. The NPro product of bovine viral diarrhea virus inhibits

DNA binding by interferon regulatory factor 3 and targets it for proteasomal degradation. J. Virol. 80, 11723–11732.

Lawson, M.A., Semler, B.L., 1991. Poliovirus thiol proteinase 3C can utilize a serine nucleophile within the putative catalytic triad. Proc. Natl. Acad. Sci. USA 88, 9919–9923.

Lindenbach, B.D., Thiel, H.J., Rice, C.M., 5th ed. 2007. Flaviviridae: the viruses and their replication. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology. Lippincott-Raven Publishers, Philadelphia, PA, pp. 1101–1152

Meyers, G., Thiel, H.J., 1996. Molecular characterization of pestiviruses. Adv. Virus Res. 47, 53–118.

Rawlings, N.D., Barrett, A.J., Bateman, A., 2012. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucl. Acids Res. 40, D343–350.

Ruggli, N., Summerfield, A., Fiebach, A.R., Guzylack-Piriou, L., Bauhofer, O., Lamm, C.G., Waltersperger, S., Matsuno, K., Liu, L., Gerber, M., Choi, K.H., Hofmann, M.A., Sakoda, Y., Tratschin, J.D., 2009. Classical swine fever virus can remain virulent after specific elimination of the interferon regulatory factor 3-degrading function of Npro. J. Virol. 83, 817–829.

Ruggli, N., Tratschin, J.D., Schweizer, M., McCullough, K.C., Hofmann, M.A., Summerfield, A., 2003. Classical swine fever virus interferes with cellular antiviral defense: evidence for a novel function of N(pro). J. Virol. 77, 7645–7654.

Rumenapf, T., Stark, R., Heimann, M., Thiel, H.J., 1998. N-terminal protease of pestiviruses: identification of putative catalytic residues by site-directed mutagenesis. J. Virol. 72, 2544–2547.

Schechter, I., Berger, A., 1967. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. 27, 157–162.

Stark, R., Meyers, G., Rumenapf, T., Thiel, H.J., 1993. Processing of pestivirus polyprotein: cleavage site between autoprotease and nucleocapsid protein of classical swine fever virus. J. Virol. 67, 7088–7095.

Szymanski, M.R., Fiebach, A.R., Tratschin, J.D., Gut, M., Ramanujam, V.M., Gottipati, K., Patel, P., Ye, M., Ruggli, N., Choi, K.H., 2009. Zinc binding in pestivirus N(pro) is required for interferon regulatory factor 3 interaction and degradation. J. Mol. Biol. 391, 438–449.

Tratschin, J.D., Moser, C., Ruggli, N., Hofmann, M.A., 1998. Classical swine fever virus leader proteinase Npro is not required for viral replication in cell culture. J, Virol, 72, 7681–7684.

Wiskerchen, M., Belzer, S.K., Collett, M.S., 1991. Pestivirus gene expression: the first protein product of the bovine viral diarrhea virus large open reading frame, p20, possesses proteolytic activity. J, Virol, 65, 4508–4514.

Zogg, T., Sponring, M., Schindler, S., Koll, M., Schneider, R., Brandstetter, H., Auer, B., 2013. Crystal structures of the viral protease npro imply distinct roles for the catalytic water in catalysis. Structure 21, 929–938.