Crystal Structure of Porcine Reproductive and Respiratory Syndrome Virus Leader Protease Nsp1α

Yuna Sun,1† Fei Xue,2† Yu Guo,3 Ming Ma,1 Ning Hao,1 Xuejun C. Zhang,1 Zhiyong Lou,4* Xuemei Li,1,* and Zihe Rao1,2,3

National Laboratory of Macromolecules, Institute of Biophysics, Chinese Academy of Science, Beijing 100101, China; Structural Biology Laboratory, Tsinghua University, Beijing 100084, China; and College of Life Sciences and Tianjin State Laboratory of Protein Science, Nankai University, Tianjin 300071, China

Received 15 December 2008/Accepted 4 August 2009

Porcine reproductive and respiratory syndrome (PRRS) virus (PRRSV), a positive-strand RNA virus that belongs to the Arteriviridae family of Nidovirales, has been identified as the causative agent of PRRS. Nsp1α is the amino (N)-terminal protein in a polyprotein encoded by the PRRSV genome and is reported to be crucial for subgenomic mRNA synthesis, presumably by serving as a transcription factor. Before functioning in transcription, nsp1α proteolytically releases itself from nsp1β. However, the structural basis for the self-releasing and biological functions of nsp1α remains elusive. Here we report the crystal structure of nsp1α of PRRSV (strain XII-GD) in its naturally self-processed form. Nsp1α contains a ZF domain (which may be required for its biological function), a papain-like cysteine protease (PCP) domain with a zinc ion unexpectedly bound at the active site (which is essential for proteolytic self-release of nsp1α), and a carboxyl-terminal extension (which occupies the substrate binding site of the PCP domain). Furthermore, we determined the exact location of the nsp1α self-processing site at Cys-Ala-Met180 Ala-Asp-Val by use of crystallographic data and N-terminal amino acid sequencing. The crystal structure also suggested an in cis self-processing mechanism for nsp1α. Furthermore, nsp1α appears to have a dimeric architecture both in solution and as a crystal, with a hydrophilic groove on the molecular surface that may be related to nsp1α’s biological function. Compared with existing structure and function data, our results suggest that PRRSV nsp1α functions differently from other reported viral leader proteases, such as that of foot-and-mouth disease.

The PRRS disease similarly poses a real threat to the worldwide swine industry.

PRRSV, a small enveloped RNA virus, is a positive-strand RNA virus and belongs to the genus Arterivirus, family Arteriviridae, and order Nidovirales. Three other members of the Arterivirus genus have been reported, namely Equine arteritis virus (EAV), Simian hemorrhagic fever virus (SHFV), and Lactate dehydrogenase-elevating virus (LDV). The single-stranded, positive-sense RNA genome of PRRSV is about 15 kb in size and encodes nine open reading frames (ORFs), including two large ones, ORF1a and ORF1b. The smaller ORFs encode structural proteins (nsps) (Fig. 1A). In particular, nsp1α, nsp1β, and nsp2 release themselves from pp1a first, and then nsp4 (a 3C-like protease) cleaves each of the remaining nsp’s from the polyproteins, including polymerase (RdRp/nsp9), helicase (Hel/nsp10), and Xenopus laevis homolog poly(U)-specific endoribonuclease (N/nsp11) (5, 10, 11, 37). These resulting mature nsps form a replication/transcription complex essential for viral RNA synthesis (6, 27).

Located at the N terminus of pp1a, PRRSV nsp1 is processed into two multifunctional proteins, nsp1α and nsp1β, each of which contains a papain-like cysteine protease (PCP) domain essential for self-release from the polyprotein. In particular, nsp1α contains an amino-terminal zinc finger (ZF) and the PCPs protease domain, while nsp1β contains PCPβ (3, 25, 32, 33). Nsp1β is responsible for the cleavage between nsp1β
and the downstream protein nsp2 at the cleavage site Trp-Tyr-Gly203 | Ala-Gly-Lys (numbered according to the nsp1β amino acid sequence), while the exact cleavage site of nsp1α has not yet been identified (16). In principle, such a self-processing event may occur either intra- or intermolecularly. A previous study of the crystal structure of the foot-and-mouth disease virus (FMDV) leader protease (LPP), which is a PCC, suggests that in cis self-processing may be favored (8, 9). However, such a process remains speculative for PRRSV nsp1. Although nsp1α of PRRSV shares low amino acid sequence identity with papain (≤15%), the characteristic residues of papain-like proteases are generally conserved in the primary sequences of nsp1α proteins of all reported arteriviruses, except EAV. Moreover, it has been shown that loss of PCPα activity totally abolishes the synthesis of subgenomic (sg) mRNA in PRSSV but has no effect on its genome replication, suggesting a role for nsp1α in transcription (31, 33). Furthermore, the homologue of PRRSV nsp1 in EAV was found to bind with p100, a transcription coactivator and thus affect sg mRNA synthesis (32).

To understand the PRRSV LPP processing mechanisms and to identify potential drug targets against the PRRS disease, we carried out structure-function studies of the nsp1α protein. We report here the three-dimensional structure of nsp1α of PRRSV (strain XH-GD; GenBank accession no. EU624117) and discuss its relationship with the papain superfamily. Moreover, we experimentally determined the exact self-processing site between nsp1α and -β. The crystallographic and biological data indicated a dimeric architecture for nsp1α, with a hydrophilic groove on the molecular surface that could be related to the biological function of nsp1α. In light of these available data, we also propose a probable mechanism of self-processing and substrate recognition for nsp1α.

MATERIALS AND METHODS

Protein expression, purification, and characterization. The primers 5’-CCG GAA TTC ATG TCT GGG ATA CTT GAT CG-3’ and 5’-CCG CTC GAG TTA ACC GTA CCA CTG ATG ACT GC-3’ were used to amplify the full-length nsp1 gene from the cDNA of the PRRSV XH-GD strain (EU624117). The primers included EcoRI and XhoI restriction sites (shown in bold). The nsp1 gene was amplified using PCR and then cloned into the pET-28a vector (Invitrogen), and the sequence of the insert was verified by DNA sequencing. The gene was amplified using PCR and then cloned into the pET-28a vector (Invisible H11032).

Transformed cells were then cultured at 37°C in LB medium containing 50 μg/ml kanamycin. When the culture density reached an A600 of 0.6, induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was performed, and cell growth continued for an additional 12 to 14 h at 16°C. Cells were harvested by centrifugation, then resuspended in lysis buffer (20 mM 2-(N-morpholino)ethanesulfonic acid [MES; pH 6.5], 300 mM NaCl, 10% [vol/vol] glycerol, and 50 mM imidazole), and the cell lysing was facilitated by sonication. The lysate was centrifuged at 20,000 × g for 30 min to remove cell debris. The supernatant was applied to a Ni2+ chelating column (1 mL Ni2+-NTA agarose) and nonspecific bound proteins were washed off with the lysis buffer. Presumably, self-processing occurred before the purification. The N-terminal His-tagged self-processed nsp1 protein was eluted with an eluting buffer (20 mM MES [pH 6.5], 500 mM NaCl, 10% [vol/vol] glycerol, and 200 mM imidazole). The purity of the eluted protein was estimated by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis to be greater than 95% and was transferred into a storage buffer of 50 mM MES (pH 6.0), 1 M NaCl and 50 mM dithiothreitol (DTT).

To determine the native cleavage site between nsp1α and nsp1β, we designed a construct of full-length nsp1 with both N- and C-terminal His tags in a pET-28a vector. The new construct contained a thrombin cleavage site between the N-terminal His tag and nsp1. After loading the samples to a Ni2+ chelating column (1 mL Ni2+-NTA agarose) and washing them with 20 ml MCAC buffer (20 mM MES, pH 6.5; 500 mM NaCl; and 10% [vol/vol] glycerol) containing 50 mM imidazole, thrombin was added to the column and incubated 12 h at 4°C to remove the autolice nsp1α while holding the nsp1β protein on the column through the C-terminal His tag. Both nsp1α and nsp1β were collected for analysis.

Crystalization. Crystalization of nsp1α was performed at 18°C using the hanging-drop vapor-diffusion method. A series of crystalization grids was prepared by mixing ~15 mg/ml nsp1α in the storage buffer with an equal volume of the reservoir solution containing 100 mM Tris-HCl (pH 8.5) and 25% (vol/vol) tert-butanol. Single crystals with a size of 50 by 50 by 100 μm were grown in 15 days and were soaked in a cryo-protectant solution consisting of the reservoir solution and 10% (vol/vol) glycerol. Crystals were flash-frozen in liquid nitrogen and then transferred into a dry nitrogen stream at 100 K for X-ray diffraction data collection.

Data collection, processing, and structure determination. Multiwavelength anomalous dispersion (MAD) data sets using the bound zinc were collected for native nsp1α at 100 K using a Mar555 flat panel detector on the 1W2 beamline at Beijing Synchrotron Radiation Facility (BSRF). Data were processed and scaled using the program package MOSFLM (19). Two zinc ions per asymmetric unit were identified using the MAD data sets with the program SHELXD (28), with a partial occupancy over 0.9 and a Cα,RMSD of 58.55/41.42. Initial phases were obtained using the computer program SOLVE (29) and were subsequently improved by using RESOLVE (29) to a mean overall figure of merit of 0.48. The electron density map at this stage was traceable with RESOLVE (29), and the automated model-building resulted in multiple peptide fragments. Manual model-building with the program O (12) and initial refinement with CNS software (2) generated a starting model for further refinement. The final model building and refinement were performed with the programs COOT (4) and Refmac (23). Solvent molecules were located from stereochemically reasonable peaks in the σA-weighted Fo-Fc difference electron density map. The quality of the final refined model was verified using the program PROCHECK (18). Final refinement statistics are summarized in Table 1. Struc-
The nsp1 monomer possesses a compact, elliptical structure with dimensions of 50 by 35 by 30 Å, consisting of five β strands and five α-helices. It can be divided into three parts (Fig. 1B): the N-terminal ZF domain, Pro66 to Gln166, and the C-terminal extension (CTE; Arg167 to Met180). The overall refined structure of the PRRSV nsp1 monomer is generally similar to that of the EAV nsp1 (Fig. 1C). A previous bioinformatics study and computer modeling predicted that the N-terminal ZF domain of nsp1 monomer possesses a compact, elliptical structure with dimensions of 50 by 35 by 30 Å, consisting of five β strands and five α-helices. It can be divided into three parts (Fig. 1B): the N-terminal ZF domain, Pro66 to Gln166, and the C-terminal extension (CTE; Arg167 to Met180). The overall refined structure was of good quality (Table 1).

The N-terminal ZF domain consisted mainly of “random” coils plus two short antiparallel β strands (β1 and β2) and a short α-helix (α1) (Fig. 1C). A previous bioinformatics study and computer modeling predicted that the N-terminal ZF domain of nsp1α belongs to the 4-Cys (C4) ZF superfamily (20, 33). Our crystal structure of nsp1α confirmed this prediction and further revealed that the topology of the ZF domain is generally similar to that of the ββα ZF family (Fig. 1C), which includes over 1,000 known transcription factors (20). As a protease domain is not common in transcription factors, the ZF domain of nsp1α could explain the previously reported (16) effect of nsp1α on sg mRNA synthesis in PRRSV. Moreover, the homologue of nsp1α in EAV was demonstrated to be unnecessary for replication, but crucial for transcription, and several mutations in the putative N-terminal ZF domain of EAV nsp1 selectively abolished transcription of certain mRNAs (31). The conservation of the ZF domain among PRRSV, LDV, SHFV, and EAV suggests that their N-terminal ZF domains play similar roles in Arterivirus (see Fig. 4B) (31). The PCPα domain of PRRSV nsp1α has a typical papain

### RESULTS AND DISCUSSION

#### 3D structure of the PRRSV nsp1α monomer.

In order to obtain naturally self-processed nsp1α, we expressed the full-length nsp1α of PRRSV (i.e., residues Met1 to Gly383 of the p1a polyprotein) in *Escherichia coli*, and the recombinant nsp1α protein that was self released from downstream nsp1β was purified. The crystal structure of the 19-kDa nsp1α (i.e., residues Met1 to Met180) was determined using the MAD method and refined at a 2.4-Å resolution, resulting in a final Rwork value of 22.5% (Rfree = 27.1%). The crystal form belongs to the space group P42121, and there is one nsp1α monomer per asymmetric unit, with a Matthews coefficient of 2.6 Å³/Da, corresponding to a 51% solvent content (21). The nsp1α monomer possesses a compact, elliptical structure with dimensions of 50 by 35 by 30 Å, consisting of five β strands and five α-helices. It can be divided into three parts (Fig. 1B): the N-terminal ZF domain, Pro66 to Gln166, and the C-terminal extension (CTE; Arg167 to Met180). The overall refined structure was of good quality (Table 1).

- **Table 1. Statistics of data collection and refinement**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Zn peak value(s)</th>
<th>Zn edge value(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength used (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of all reflections</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completeness of data (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average I/σ(I)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSD bond distance (Å)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSD bond angle (°)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of nonsolvent atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of solvent atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg overall B value (Å²)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Refinement statistics:

- No. of reflections used [σ(I) > 0]
- Rwork (%)                  |
- Rfree (%)                  |
- RMSD bond distance (Å)     |
- RMSD bond angle (°)        |
- No. of nonsolvent atoms    |
- No. of solvent atoms       |
- Avg overall B value (Å²)   |

Ramachandran plot (excluding Pro and Gly):

- No. of residues in most favored regions (%) 115 (81.0)
- No. of residues in additionally allowed regions (%) 22 (15.5)
- No. of residues in generously allowed regions (%) 5 (3.5)

Cross-linking gel assay.

Recombinant nsp1α was diluted to 5 mg/ml in phosphate buffered saline (pH 7.0), 1 M NaCl, and 1 mM DTT. Ethylene glycol disuccinate di(ethyl) ester (EGS) was dissolved in dimethyl sulfoxide to give a final EGS concentrations of 0, 0.315, 0.625, 1.25, 2.5, 5.0, and 10.0 mM individually. After the mixture was incubated at room temperature for 15 min, the reaction was quenched for 10 min by adding 1 M Tris-HCl (pH 7.5) to a final concentration of 50 mM. An equal volume of 2× SDS-polyacrylamide gel electrophoresis sample buffer was added to the reaction mixture, and a small amount was analyzed on a 15% SDS polyacrylamide gel.

**Protein structure accession number.** Coordinates and structure factors for the crystal structure of PRRSV nsp1α were deposited with the Protein Data Bank (PDB accession no. 3IFU).
fold (13), which is a compact global structure consisting of sequentially connected left (L) and right (R) parts in a so-called standard orientation (9) (Fig. 1B). The L subdomain of PCPa consists of four α-helices (α2 to α5), while the R subdomain is formed by three antiparallel β strands (β3 to β5). Residue Cys76, which is located at the N-terminal end of the longest helix, α2, in the L subdomain, and His146, which is located in the turn connecting the two longest β strands, β4 and β5, in the R subdomain, face each other at the L-R interface and form the catalytic center of the PCPa domain.

The structure of the 14-residue CTE region of PRRSV nsp1α corresponds to unambiguous electron densities in experimentally MAD-phased 2Fo-Fc maps. The average B factor of this domain was 62 Å², comparable with the overall average B factor of 58 Å². This CTE peptide inserts into a putative substrate binding site of the PCPa domain from the same protein subunit. This observation suggests that CTE likely serves as a natural substrate for PCPa through intramolecular interactions, implying an in cis self-processing mode for nsp1α.

At the amino acid sequence level, CTE is a characteristic feature of all nsp1α proteins of the Arteriviridae family, yet it has not been shown in a 3D structure before for any of the family members.

In the crystal structure of PRRSV nsp1α, we identified two zinc ions associated with each nsp1α subunit, one for the ZF domain and the other for the PCPa domain. In fact, the MAD signals from the zinc ions were successfully used to determine the initial phases of the crystal structure (Table 1). The zinc binding to nsp1α was further confirmed by atomic absorption spectrum and extended X-ray absorption fine structure spectra. In the 3D structure, both the N- and C-terminal zinc ions are tetrahedrally coordinated with Cys8, Cys10, Cys25, and Cys28 and with Cys70, Cys76, His146, and Met180, respectively. The N-terminal zinc coordination is consistent with the bioinformatics prediction of the C4 type of ZFs and was previously found to be related to sg mRNA synthesis (16). In contrast, the C-terminal-bound zinc ion is the first one among the reported 3D structures of papain-like proteases, and its precise biological role remains unclear (Fig. 2A and 2A'). Unlike the conserved ZF domain among the other arteriviruses genus members, key residues observed forming the zinc coordination in the PCP domain of EAV vary from those in SHFV, LDV, and PRRSV, consistent with an earlier observation that the PCP-like domain of EAV nsp1 lacks PCP activity due to the absence of the catalytic cysteine. This further suggests that SHFV and LDV share similar zinc binding modes and PCP activities (Fig. 4C) (33).

**PCPα active site bound with a zinc ion.** As a conserved functional feature of the papain superfamily, the active site of PRRSV nsp1α contains the catalytic residues Cys76 and His146 facing each other at the L-R interface. His146 of the His-Cys catalytic dyad is stabilized by an elaborate hydrogen bond network that includes Glu69 and Asn143, similar to two asparagines in other reported PCPs (Fig. 3D). The Oε1 and Nα2 atoms of Asn143 contacted the backbone amide group of His146 and the Oε2 atom of Glu69 at 2.8 Å and 3.0 Å, respectively. The Oε2 atom of Glu69 further stabilized the imidazole side chain of His146 via a 3.1-Å hydrogen bond with the Nα1 atom. Through this hydrogen bond network, the imidazole ring face and form the catalytic center of the PCPa domain.

The structure of this domain was 62 Å², comparable with the overall average electron density. One of its carboxylate oxygen atoms (with a B factor of 51.4 Å²) is located in the putative oxygen hole (formed by residues Ser71 to Ala75) and forms hydrogen bonds with both the main-chain carboxyl oxygen of Gly74 (2.8 Å) and the amide nitrogen of Cys76 (2.8 Å), whereas the second carboxylate oxygen (with a B factor of 51.6 Å²) coordinates the metal ion of the catalytic site (2.1 Å).

As mentioned above, the zinc ion bound in the PCP domain active site is a novel feature of PRRSV nsp1α that has not previously been reported in any other crystal structure of the papain superfamily. This metal ion was first identified as a Zn²⁺ ion during X-ray diffraction data collection and was later confirmed by ICP-MS performed with an ICP-MS XII instrument (Thermo). In this experiment, 95 nmol zinc ion was found in a 1-mg (~50-nmol) protein sample of PRRSV nsp1α. This suggests that there is one more Zn²⁺ ion in PRRSV nsp1α in addition to the well-known one associated with the ZF domain. Moreover, in an anomalous difference Fourier map that was calculated using the diffraction data collected at the zinc absorption edge (λ = 1.2821 Å), two distinct peaks were observed in each nsp1α monomer at a contour level as high as 6.5 standard deviations and thus were assigned as Zn²⁺ ions in the subsequent structure determination. To verify the importance of this Zn²⁺ ion to proteolytic activity, we synthesized a fluorescent substrate MCA-ECADVY-Lys(Dnp)-Lys-NH2 and measured the proteolytic activity of nsp1α in both the absence and presence of EDTA. However, the native full-length PRRSV nsp1α was totally inactive, which is not surprising, considering the above-mentioned CTE binding mode (also see below). In addition, all our attempts to construct nsp1α variants of varied C-terminal truncations to restore the activity of PCP failed, as none of them could be expressed in a soluble form. These negative results unfortunately prevented us from directly examining the proteolytic activity of PRRSV nsp1α and further investigating the effect of zinc ion on PCPa activity. Nevertheless, considering that there was no Zn²⁺ ion artificially added during the purification and crystallization of nsp1α, our crystal structure and ICP-MS data unambiguously suggest a native zinc binding site at the PCPα active site, which is quite novel among the reported crystal structures of papain-like protease. Its precise biological role needs to be further investigated.

**Homodimer assembly of PRRSV nsp1α.** The equilibrium between the monomer and dimer of PRRSV nsp1α was first observed in gel filtration chromatography during purification of nsp1α, even in the presence of 1 M NaCl (Fig. 1F). Subsequently, the results of EGS cross-linking and analytical ultracentrifugation also showed that nsp1α exists as a homodimer in solution and indicated that such a dimer remains associated even at high salt concentrations (Fig. 1F).

Consistently, two monomers formed a homodimer via the only available crystallographic twofold axis in the crystal form (Fig. 1D and E). This dimer possesses a hydrophilic, half-
cylinder-shaped, symmetrical, open channel with a curvature diameter of ~26 Å and length of ~50 Å on its molecular surface (Fig. 2B and C). Major contacts between the two subunits were provided by several hydrophobic residues of the PCPα domain, especially long loops connecting α1 and α2, β3 and β4, and β5 and CTE, while minor contributions came from residues of the ZF domain, including the region connecting β1 and β2 and some coil regions near the peptide N terminus. The two putative protease active sites of each homodimer are on opposite sides and exposed to solvent, permitting independent function if they are not blocked by CTE (also see below). The relatively hydrophobic contact surface between the two monomers was 2,400 Å² (25% of total) from each monomer, consistent with the strong dimerization in solution (Fig. 1G). In addition, we constructed several dimer-breaking point mutations but failed to express them as soluble proteins in E. coli, presumably because of unfavorable solvent-exposed hydrophobic surfaces. Together, both the structural and solution data suggest that the nsp1α homodimer is likely a biologically functional unit. In sharp contrast, a similar investigation of FMDV reveals that its corresponding protease L[^pro] functions as a monomer when it proteolytically cleaves the host cell protein eukaryotic initiation factor eIF4G (8, 9). Although the PCPα domain of PRRSV nsp1α shares high structural similarity with FMDV L[^pro], the N-terminal ZF domain is absent from the latter. Therefore, the additional structural features, including

**FIG. 2.** Functional implications of the N-terminal ZF domain in nsp1α. (A) Tetrahedral coordination of the zinc ion. The zinc ion and four coordinated cysteine residues are covered with a 2Fo-Fc electron density map phased with a MAD data set and contoured at 1.0 σ (cyan). (B and C) Top and side views of the hydrophilic groove on the molecular surface of the nsp1α homodimer.

**FIG. 3.** Active site of the C-terminal PCPα domain and its interaction with CTE. (A) Bound zinc ion and its coordination residues. The structural model is overlapped with a 2Fo-Fc electron density map phased with a MAD data set and contoured at 1.0 σ. (B) Uncleaved Gln166-Arg167 peptide bond. The 2Fc-Fc electron density map, contoured at 1.0 σ, shows inconsistency with a previously predicted cleavage site. (C) CTE binding to PCPα. PCPα is shown as a molecular surface model colored according to electrostatic potential (red for negatively charged regions and blue for positively charged regions). The CTE is shown as stick models, green for carbon, red for oxygen, and blue for nitrogen atoms. The zinc ion, key residues, and major binding pockets are labeled. (D) Hydrogen bond network around His146. Residues are colored according to their domain association (blue for ZF domain, purple for PCPα, and green for CTE). Hydrogen bonds are shown as red dashed lines and labeled with the distance.
FIG. 4. Structural and sequence comparisons of nsp1α homolog proteins. (A) Comparison of surface electrostatic potential of the PCP catalytic centers of papain (PDB accession no. 9PAP), FMDV Lpro (PDB accession no. 1QOL), and PRRSV PCPα (residues Pro66 to Gln166). Yellow arrows indicate the substrate binding grooves. (B) Sequence alignment of the N-terminal domains from PRRSV, EAV, LDV, and SHFV. Key residues of the C4 motif are marked with red arrows. Residue numbers at the top refer to the PRRSV sequence. (C) Sequence alignment of the C-terminal domains from PRRSV, EAV, LDV, SHFV, and FMDV. Key residues for the proteolytic activity of PCPs are marked with blue triangles at the bottom. Secondary-structure elements of PRRSV nsp1α are marked on top of the alignment; α-helices and β strands are presented as curves and arrows, respectively. Four residues upstream and downstream of the cleavage site of PRRSV nsp1α are highlighted with blue and red frames, and the cleavage site is marked with a red arrow on the top. The residue numbers on the top refer to the PRRSV sequence. The sequences were aligned with CLUSTALW (17), and the alignment was drawn with ESPript (7).
the ZF domain and dimeric architecture of PRRSV nsp1α, may bear important functions beyond the protease activity.

Identifying the autocleavage site. The crystal structure of PRRSV nsp1α showed the autocleavage site during nsp1 processing. PRRSV nsp1α releases itself from the parental polypeptide chain by a cleavage between its own C terminus and the N terminus of nsp1β, the downstream protein in the polyprotein peptide. During the early stages of model building, we built an initial model of nsp1α from Met1 to Gln166 into the experimental electron density map. However, unambiguous and consecutive electron density was found for an additional peptide C-terminal to Gln166. By carefully analyzing the primary sequence, we were able to fit 14 more residues (Arg167 to Met180) into this extra piece of electron density; the final model contained residues Met1 to Met180. Moreover, in an unbiased Fo-Fc difference Fourier map, the C-terminal residue Met180 had excellent electron density in the vicinity of the PCP catalytic center. Meanwhile, the peptide between Gln166 and Arg167 was covered with clear continuous electron density (Fig. 3B). These results suggest that the nsp1α peptide ends exactly at Met180 and that the CTE region is an intact part of nsp1α instead of an isolated ligand or substrate.

To further verify the biological relevance of the above structural observation, we constructed a full-length nsp1 fused with both an N- and C-terminal His tag and a thrombin cleavage site between the N-terminal His tag and nsp1. After thrombin cleavage and self-processing, nsp1α and nsp1β were separated, and we were able to purify the nsp1β protein using affinity chromatography and to analyze its N-terminal amino acid sequence. The result showed that the N-terminal sequence of nsp1β starts with Ala181-Val-Asp-Ala-Tyr. In addition, the purified nsp1α was analyzed by mass spectroscopy methods, and the result showed a molecular weight of 19,900 for the nsp1α peptide, consistent with cleavage after Met180. Moreover, residues in the downstream region of this self-cleavage site in PRRSV, LDV, and SFHV are relatively conserved, and they correspond to the N terminus of nsp1β after cleavage. Notably, residues in EAV corresponding to this conserved nsp1α cleavage site also show some recognizable similarity with other family members, although EAV nsp1α lacks proteolytic activity because of an inactive PCP activity center (33) (Fig. 4C). Thus, it is likely that most Arterivirus members share a self-processing site similar to that of nsp1α. Taking all these data together, we conclude that the exact native nsp1α autocleavage site is after Met180 in the sequence of Cys-Ala-Met180| Ala-Asp-Val (Fig. 3A).

Interaction between PCPα and CTE. The presence of CTE residues inside the substrate-binding pockets illustrates the natural form of substrate binding during self-processing. The last six residues of the CTE were in an extended conformation (Fig. 3C) similar to that observed in complexes of enzymes of the papain superfamily with their peptide-like inhibitors (35, 36). Three major sequence regions in PCPα contributed to the CTE binding. In Fig. 3D, the loop between helices α3 and α4 (Gly109 to Pro113) formed the left wall of the binding pocket, while the loop between strands β4 and β5 formed the right wall. The central helix, α2, plus some side chains from the two loops mentioned above formed the bottom of the binding pocket. The two walls sandwiched CTE tightly to stabilize its extended conformation, and the catalytic center was located at the top of the substrate binding pocket.

More details on the CTE side of the major interactions between CTE and PCPα were provided by residues Phe173 to Met180. Whereas Met180 occupied the S1 position as mentioned above, Ala179 occupied the S2 subsite. The PCPα loop regions that connect helices α3 and α4 (i.e., residues Gln110 to Thr112) and strands β4 and β5 (i.e., residues Leu145 to Val147) formed a narrow cleft shaping the S2 subsite, as in many other papain-like proteases (14). The putative P1 residue of CTE was surrounded by the hydrophobic side chains of Leu111, Val116, Val142, Val147, and Trp128, and therefore, the corresponding S3 site of the PCPα domain was a mostly hydrophobic subsite with a vacuole volume of ~ 350 Å³. The hydrophobic side chain of Phe176 from CTE completely buried in this subsite. This hydrophobic side is conserved to some extent in many papain-like proteases (Fig. 4B). The S3 subsite was at the end of the binding pocket and was occupied by Pro175. Phe173 and Pro175 were additionally stabilized by Trp128 and His157 through π-electron interaction. In addition, residues of CTE also established some main-chain hydrogen bonds with PCPα to stabilize its conformation (Fig. 3C and Table 2).

Nsp1α proteins of PRRSV, LDV, SFHV, and FMDV Lpro show recognizable sequence homology, including the residues involved in PCP activity (Fig. 4C). For example, besides the Cys76 and His146 catalytic dyad (PRRSV numbers) (14), Trp77, which shields the hydrogen bond between the dyad from bulk solvent, is also among the most conserved residues, indicating similar cleavage reaction mechanisms for these Lpros of the Arteriviridae and Picornaviridae viruses. Nevertheless, there are distinct differences between the crystal structures of PRRSV nsp1α and FMDV Lpro. In the crystal structure of FMDV Lpro, one of two CTEs in each crystallographic asymmetric unit is in an unbound and flexible state, while the other one is well stabilized by the PCP domain of an adjacent Lpro molecule. This observation illustrates that the contact between FMDV Lpro CTE and the PCP catalytic domain is not very stable in solution, which would be consistent with its alternative proteolytic activity toward the protein substrate eIF4G. In sharp contrast, the CTE peptide of PRRSV nsp1α is well stabilized by the PCP domain and does not appear to allow subsequent substrate replacement for any further protease activity after self-cleavage. This product inhibitor phenomenon suggests that the PCP domain of the Lpro from the Arteriviridae

### Table 2. Hydrogen bonds between CTE and PCPα residues

<table>
<thead>
<tr>
<th>CTE Residue</th>
<th>PCPα Residue</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met180 N</td>
<td>Leu145 O</td>
<td>3.3</td>
</tr>
<tr>
<td>Ala179 O</td>
<td>Leu111 N</td>
<td>3.1</td>
</tr>
<tr>
<td>C</td>
<td>Leu145 O</td>
<td>3.0</td>
</tr>
<tr>
<td>Cys178 O</td>
<td>Ser144 O</td>
<td>3.1 (2.6)</td>
</tr>
<tr>
<td>Gly59 N</td>
<td>Trp128 O</td>
<td>3.2</td>
</tr>
</tbody>
</table>

**Note:** These values are derived from the crystal structures and indicate the hydrogen bond distances between the CTE and PCPα residues.
family is unlikely to have specific proteolytic activity toward a host cell protein; instead, the self-released nsp1α is likely to be involved in nonproteolytic activities, such as acting as a platform for the N-terminal ZF domain to play its role in sg mRNA synthesis (16).

A putative self-processing mechanism. Although there is not yet sufficient biochemical data to firmly establish a model regarding the self-processing mechanism of nsp1α-like Lpro, several structural features of PRRSV nsp1α suggest that an in cis (intramolecular) processing mode may be favored in PRRSV nsp1α self-processing. First of all, in our crystal structure, the residues N-terminal to the cleavage site were well defined in the electron density map and were strongly stabilized by an intramolecular hydrogen bond network (Table 2). Notably, as mentioned above, one carboxylate oxygen atom of C-terminal Met180 pointed toward the “oxyanion hole” and was stabilized by the active site of this oxygen atom likely represents the actual product after peptide hydrolysis. In addition, as we mentioned above, the cleavage sites of the two subunits in the homodimer were on opposite sides from each other, with a linear distance of 45 Å. Considering the limited peptide length of the CTE region, it is unlikely that the two subunits perform self-processing by swapping their CTE peptides. Therefore, intramolecular self-processing seems the most probable mechanism for PRRSV nsp1α based on our current model. A similar conclusion was suggested by the crystallographic investigation of FMDV Lpro. Although an in trans binding of the CTE peptide was observed with the FMDV Lpro crystal structure, it was attributed to crystal packing, and an in cis cleavage mechanism is considered the most probable scenario in solution (9).

In order to explore the possibility of studying the quaternary structure of a full-length nsp1α (i.e., nsp1α-nsp1β) polyprotein, we constructed an active site point mutation, C76S, to abolish the PCPα protease activity. Although we attempted to express it in dozens of combinations of expression vectors and culture conditions, this mutation was always present only in inclusion bodies, which prompted us to speculate that accurate self-processing might be essential for the structural stability of nsp1α and nsp1β as well as their functions.

Conclusion. PRRSV nsp1α is the Lpro of the PRRSV pp1a polyprotein. Our crystal structure of nsp1α shows that it consists of an N-terminal ZF domain, which could be related to its biological role as a transcription factor; a PCP domain, which can self-release nsp1α from the polyprotein chain; and a CTE, which is the result of proteolytic self-processing and illustrates part of the substrate binding mode. Moreover, the crystal structure of nsp1α shows a dimeric architecture, which provides a putative hydrophilic surface channel and may be related to some important biological functions of nsp1α. The crystallographic and ICP-MS results also show a novel zirconium binding site at the active site of PCPα, which has not been reported before for any crystal structure of PCP. Furthermore, our crystal structure suggests that an in cis mode is the most likely mechanism of PCPα self-processing.

ACKNOWLEDGMENTS

We thank Liao M. of South China Agricultural University for kindly providing the PRRSV XH-GD genome. This work was supported by the National Natural Science Foundation of China (grant no. 30870486), 863 Project (China) (grant no. 2006AA02A322), the National Major Projects (grant no. 2009ZX09311-001), the Protein Studies Project (China) (grant no. 2006CB109001), the Ministry of Science and Technology (MOST) 973 Project (grant no. 2006CB806503 and 2007CB914304), and the PSA II Project from MOST and KNAW (grant no. 2008AA000238).

REFERENCES


