

# Structure of the foot-and-mouth disease virus leader protease: a papain-like fold adapted for self-processing and eIF4G recognition

Alba Guarné, José Tormo,  
Regina Kirchweger<sup>1,2</sup>, Doris Pfistermueller<sup>1</sup>,  
Ignasi Fita and Tim Skern<sup>1,3</sup>

Centre d'Investigació i Desenvolupament (CSIC), Jordi Girona Salgado 18–26, E-08034 Barcelona, Spain and <sup>1</sup>Institute of Biochemistry, Medical Faculty, University of Vienna, Dr Bohr-Gasse 9/3, A-1030 Vienna, Austria

<sup>2</sup>Present address: Department of Cell Biology and Genetics, Sloan Kettering Memorial Cancer Center, East 68th Street, New York, NY, USA

<sup>3</sup>Corresponding author  
e-mail: [timothy.skern@univie.ac.at](mailto:timothy.skern@univie.ac.at)

**The leader protease of foot-and-mouth disease virus, as well as cleaving itself from the nascent viral polyprotein, disables host cell protein synthesis by specific proteolysis of a cellular protein: the eukaryotic initiation factor 4G (eIF4G). The crystal structure of the leader protease presented here comprises a globular catalytic domain reminiscent of that of cysteine proteases of the papain superfamily, and a flexible C-terminal extension found intruding into the substrate-binding site of an adjacent molecule. Nevertheless, the relative disposition of this extension and the globular domain to each other supports intramolecular self-processing. The different sequences of the two substrates cleaved during viral replication, the viral polyprotein (at LysLeuLys↓GlyAlaGly) and eIF4G (at AsnLeuGly↓ArgThrThr), appear to be recognized by distinct features in a narrow, negatively charged groove traversing the active centre. The structure illustrates how the prototype papain fold has been adapted to the requirements of an RNA virus. Thus, the protein scaffold has been reduced to a minimum core domain, with the active site being modified to increase specificity. Furthermore, surface features have been developed which enable C-terminal self-processing from the viral polyprotein.**

**Keywords:** crystallography/cysteine proteases/picornavirus/RNA viruses/translational control

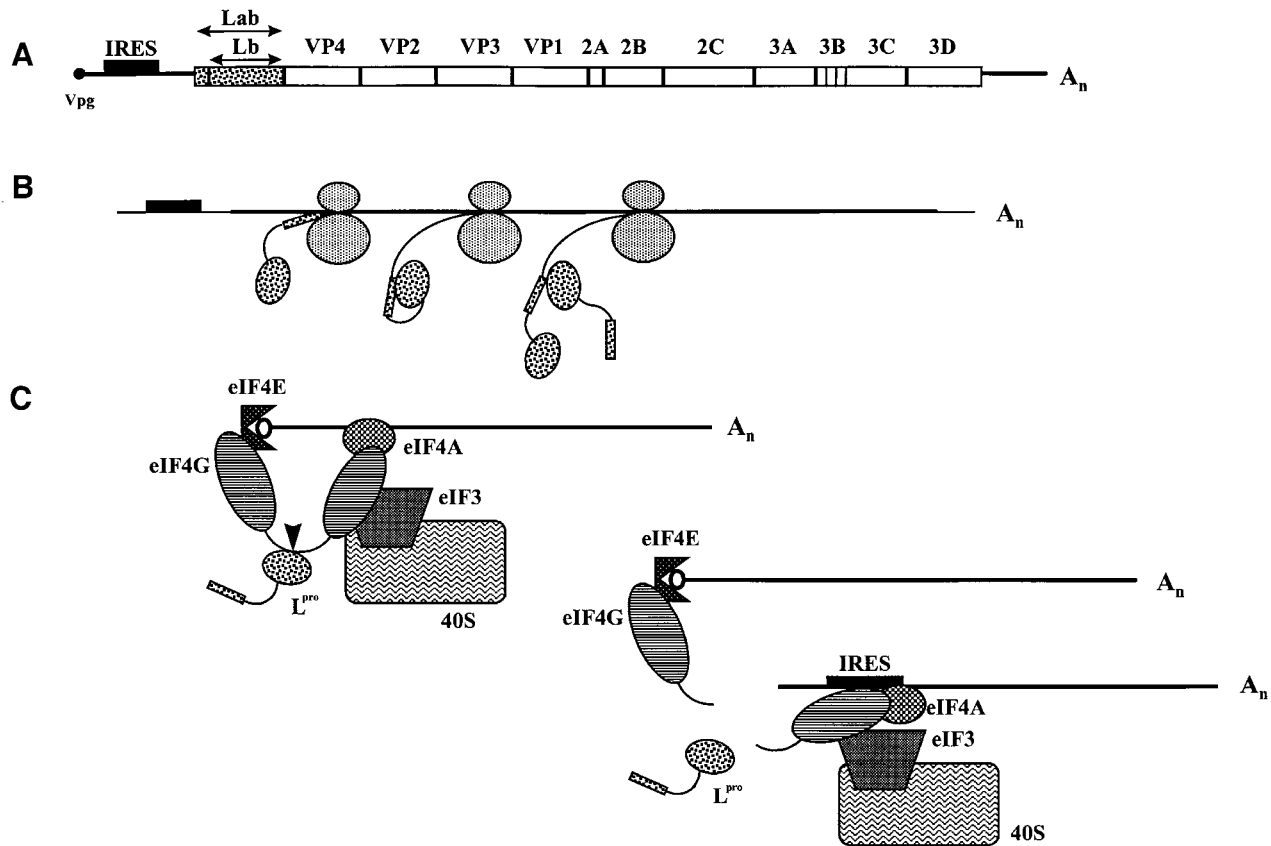
## Introduction

The structures of a number of viral proteases have been determined recently (Babé and Craik, 1997). Comparison of these viral structures with those of prototypes of the four classes of proteases reveals similarities, but also distinct differences in mechanism and structure. Thus, the picornaviral 3C cysteine protease and the NS3 serine protease of hepatitis C virus have a similar fold to the serine protease trypsin (Allaire *et al.*, 1994; Matthews *et al.*, 1994; Kim *et al.*, 1996; Love *et al.*, 1996). In contrast, the cytomegalovirus protease possesses a single

β-barrel structure, instead of the classic two β-barrel motif of trypsin, and a novel catalytic triad of His/His/Ser (Chen *et al.*, 1996; Tong *et al.*, 1996). Furthermore, the adenovirus protease, a viral cysteine protease, also has a novel fold, although it employs a catalytic triad Asp/His/Cys reminiscent of that found in the cysteine protease papain (Ding *et al.*, 1996). Papain-like cysteine proteases are characterized by a tryptophan residue following the nucleophilic cysteine, and two hydrophobic residues following a conserved histidine which increases the nucleophilic character of the catalytically active cysteine (Berti and Storer, 1995). Indeed, a number of viral proteolytic enzymes have been proposed to be papain-like cysteine proteases, based mostly on the presence of such conserved patterns (Gorbalenya *et al.*, 1991). However, the lack of direct structural information on these viral cysteine proteases has prevented the establishment of further relationships between papain and its putative viral relatives.

The leader protease (L<sup>pro</sup>) of foot-and-mouth disease virus (FMDV), an animal pathogen of global importance, is a proposed papain-like viral cysteine protease. Even though sequence identity between the FMDV L<sup>pro</sup> and papain is no higher than 15% (Gorbalenya *et al.*, 1991; Skern *et al.*, 1998), the characteristic residues of papain-like proteases are conserved in the primary sequences of the L<sup>pro</sup> from both FMDV and the related equine rhinoviruses (ERVs). L<sup>pro</sup> is the first protein encoded on the FMDV polyprotein (Figure 1A). Its sole role in viral maturation is to free itself from the polyprotein by cleavage between its own C-terminus and the N-terminus of VP4 (Figure 1B) at the sequence ArgLys LeuLys↓GlyAlaGlySer. Theoretically, this self-processing event can occur either intra- or intermolecularly (Figure 1B); expression of the L<sup>pro</sup> in various *in vivo* and *in vitro* systems has provided evidence for both types of reaction (Belsham *et al.*, 1990; Medina *et al.*, 1993; Piccione *et al.*, 1995b). Nevertheless, it has not yet been established which of the two mechanisms for self-processing is preferred *in vivo*. As initiation of protein synthesis on the FMDV genome occurs at one of two AUG codons lying 84 nucleotides apart, two species of L<sup>pro</sup> (known as Lab<sup>pro</sup> and Lb<sup>pro</sup>, depending on whether protein synthesis initiates at the first or second AUG codon, respectively) have been identified in infected cells and have been shown to possess the same enzymatic properties (Medina *et al.*, 1993). In addition to Lb<sup>pro</sup>, FMDV encodes two other proteolytic activities, the 2A peptide which has been proposed to cleave autocatalytically between its own C-terminus and the N-terminus of 2B, and the 3C protease (Ryan and Flint, 1997). This enzyme carries out all remaining cleavages, with the exception of that between VP4 and VP2 which occurs in the maturing capsid by an as yet unknown mechanism.

After the single self-processing event, L<sup>pro</sup> then plays a central role in FMDV replication by specifically cleaving



**Fig. 1.** Schematic drawing of the biological activities of  $L^{pro}$ . (A) The RNA genome of FMDV. The single open reading frame is shown as an open box (with the mature viral proteins indicated), non-coding regions by a line and the IRES with a closed box.  $L^{pro}$  forms are stippled. (B) Protein synthesis on FMDV mRNA;  $L^{pro}$  self-processing is indicated as an intra- or intermolecular event. (C) Role of eIF4 proteins in initiation of protein synthesis and cleavage of eIF4G by  $L^{pro}$ . eIF4 proteins involved in protein synthesis and the 40S ribosomal subunit are indicated. The  $m^7GDP$  5' cap structure of cellular mRNAs (open circle) and the  $L^{pro}$  cleavage site (arrow) are indicated. The eIF4G C-terminal domain still forms an initiation complex with IRES-containing mRNAs.

the host cell protein eukaryotic initiation factor eIF4G at the sequence AlaAsnLeuGly↓ArgThrThrLeu (Figure 1C; Kirchweger *et al.*, 1994). As a result, the domain of eIF4G which binds the cap-binding protein eIF4E is separated from the domain of eIF4G which binds eIF3, so that the infected cell is unable to recruit its own capped mRNA to the 40S ribosome. The viral mRNA is unaffected as it initiates protein synthesis internally via an IRES (Figure 1C). This shuts off cellular mRNA translation in favour of the viral counterpart.

We report here the three-dimensional structure of two variants of the FMDV  $L^{pro}$ , present their relationship to the papain superfamily and discuss mechanisms of self-processing and the basis of substrate recognition for specific cleavage on two different proteins.

## Results and discussion

### Structure determination

Wild-type  $L^{pro}$  crystals could not be obtained. However, substitution of the active site nucleophile Cys51 (Piccione *et al.*, 1995a; Roberts and Belsham, 1995; Ziegler *et al.*, 1995) with Ala allowed crystals to be grown as reported (Guarné *et al.*, 1996). These contained eight  $L^{pro}$  molecules in the asymmetric unit but were difficult both to reproduce and to manipulate. Thus, although a native diffraction data set from these crystals was obtained up

to 3.0 Å resolution, the corresponding phases could not be determined. We speculated that the properties of these crystals resulted from interactions between the C-terminus of one molecule and the active site of a neighbour, and therefore prepared a truncated form of the inactive  $L^{pro}$  lacking six C-terminal amino acids (termed  $sL^{pro}$ ). New crystals were obtained containing two  $sL^{pro}$  molecules in the asymmetric unit and the structure was solved by a combination of isomorphous replacement and density modification techniques. The  $L^{pro}$  Cys51Ala mutant was then determined by molecular replacement using the  $sL^{pro}$  structure as a search model (Table I; see Materials and methods).

### Description of the overall $L^{pro}$ structure

The  $L^{pro}$  structure (Figure 2A–C) presents a compact, globular region, ranging from Met29 to Tyr183 with an overall cubic shape of approximate edge dimensions of 30 Å, from which a flexible C-terminal extension (CTE) ranging from Asp184 to Lys201 extrudes. The globular region is divided into two subdomains, with the catalytically essential residues Cys51 (replaced by Ala in the  $L^{pro}$  structure) and His148 located at the interface. The first, N-terminal subdomain contains four  $\alpha$ -helices ( $\alpha$ 1,  $\alpha$ 2,  $\alpha$ 3 and  $\alpha$ 4) and two short antiparallel  $\beta$ -strands ( $\beta$ 1 and  $\beta$ 2) comprising only residues Glu30–Thr32 and Lys38–Thr40, respectively. The longest  $\alpha$ -helices  $\alpha$ 1 and

**Table I.** X-ray parameters and refinement statistics

	Lb <sup>pro</sup>	sLb <sup>pro</sup>		
		Native	HgCl <sub>2</sub>	K <sub>2</sub> PtCl <sub>4</sub>
Data collection #				
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C222 <sub>1</sub>	C222 <sub>1</sub>	C222 <sub>1</sub>
Resolution (Å)	25–3.0	25–3.0	25–2.7	25–2.8
No. of unique reflections	35 080	7339	9335	9083
Completeness (%)	92 (75)	84 (90)	82 (89)	94 (90)
$\langle I/\sigma I \rangle$	9.5 (3.0)	9.9 (4.4)	9.9 (7.4)	7.4 (4.1)
Multiplicity	7.7	6.0	8.3	9.6
$R_{\text{sym}}$ (%)	10.3 (26.3)	9.3 (19.9)	8.5 (17.2)	9.7 (19.1)
MIR phasing (25.0–3.0 Å) <sup>a</sup>				
$R_{\text{iso}}$ (%)			32.3 (38.4)	23.4 (30.3)
Sites			5	2
Phasing power			1.57	1.49
$R_{\text{Cullis}}$			0.68	0.79
Mean figure of merit			0.37	0.37
Model refinement (20.0–3.0 Å)				
$R_{\text{factor}}$ % ( $R_{\text{free}}$ , %)	25.8 (29.6)	23.5 (28.8)		
No. of reflections	34 977	7320		
Total no. of non-hydrogen atoms	11 190	2565		
No. of solvent molecules	3	20		
R.m.s. deviation in bonds (Å)	0.005	0.005		
R.m.s. deviation in angles (°)	1.2	1.3		
R.m.s. deviation in temperature (bonded atoms, Å <sup>2</sup> )	2.4	1.6		
Anisotropic correction				
( $a^*$ , $b^*$ , $c^*$ , Å <sup>2</sup> )	15.6; –16.9; 4.3	–1.3; 10.8; 9.2		
Average $B$ -factor (Å <sup>2</sup> )	38.8	25.6		

<sup>a</sup>Numbers in parentheses correspond to the highest resolution shell (3.05–3.00 Å).

$R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$ , where  $I$  = observed intensity and  $\langle I \rangle$  = averaged intensity obtained by multiple observations of symmetry related reflections.

$R_{\text{iso}} = \sum (|F_{\text{PH}}| - |F_{\text{P}}|) / \sum |F_{\text{P}}|$ , where  $|F_{\text{P}}|$  = protein structure factor amplitude and  $|F_{\text{PH}}|$  = heavy-atom derivative structure factor amplitude.

$R_{\text{Cullis}} = \sum |F_{\text{PH}} \pm F_{\text{P}}| - F_{\text{H(calc)}} / \sum |F_{\text{PH}} \pm F_{\text{P}}|$  for all centric reflections.

$\alpha 3$  (comprising residues Asn50–Glu64 and Leu78–Gly91, respectively) run perpendicular to each other, with the catalytic Cys51 being located towards the N-terminal side of helix  $\alpha 1$ . The shortest helix  $\alpha 2$  spans only six residues (Phe68–Ser73) and runs almost parallel to  $\alpha 3$ . The second Lb<sup>pro</sup> subdomain displays a fold belonging to the all  $\beta$ -family of proteins, as the only regular secondary structure elements are contained in a mixed  $\beta$ -sheet formed by one parallel ( $\beta 3$  with  $\beta 4$ ) and six antiparallel ( $\beta 4$ – $\beta 9$ )  $\beta$ -strands (Figure 2A–C). The essential His148 is located on the turn connecting the longest strands  $\beta 5$  and  $\beta 6$  (residues Phe137–Leu143 and Ala149–Thr155) which occupy a central position in the sheet.

The main difference between the structures of Lb<sup>pro</sup> and sLb<sup>pro</sup> is that, in the latter, the 12 residues remaining in the CTE are disordered in both molecules present in the asymmetric unit (from residue Asp184 onwards). The superimposition of the globular regions of Lb<sup>pro</sup> and sLb<sup>pro</sup> models gives an r.m.s. deviation of 0.4 Å (performed using the program SHP; Stuart *et al.*, 1979), which can be considered as an upper limit for the coordinate errors in the compact region of the two structures.

Several groups proposed a papain-like fold for the picornaviral L<sup>pro</sup> structure (Gorbalenya *et al.*, 1991; Piccione *et al.*, 1995a; Skern *et al.*, 1998). The papain-fold consists of left- (L) and right-hand (R) domains [standard papain-fold view (Kamphuis *et al.*, 1985)] which are structurally equivalent to the two subdomains described

for Lb<sup>pro</sup>. Superimposition of main-chain C $\alpha$  atoms from Lb<sup>pro</sup> with papain (Figure 2A and D) gives an averaged r.m.s. deviation of 1.3 Å for 76 equivalent residues. As the globular region of Lb<sup>pro</sup> represents the smallest polypeptide fragment with a papain-like topology, it lacks most of the decoration found in papain, including the prosegment-binding loop (PBL; Coulombe *et al.*, 1996; Figure 2D). Regions most conserved between the papain-like proteases and Lb<sup>pro</sup> structures are located around the active centre, especially secondary elements  $\alpha 1$  and  $\beta 5$ – $\beta 6$ , containing the catalytic cysteine and histidine residues (Figure 2B and C). Papain-like proteases have, however, no equivalent to the Lb<sup>pro</sup> CTE.

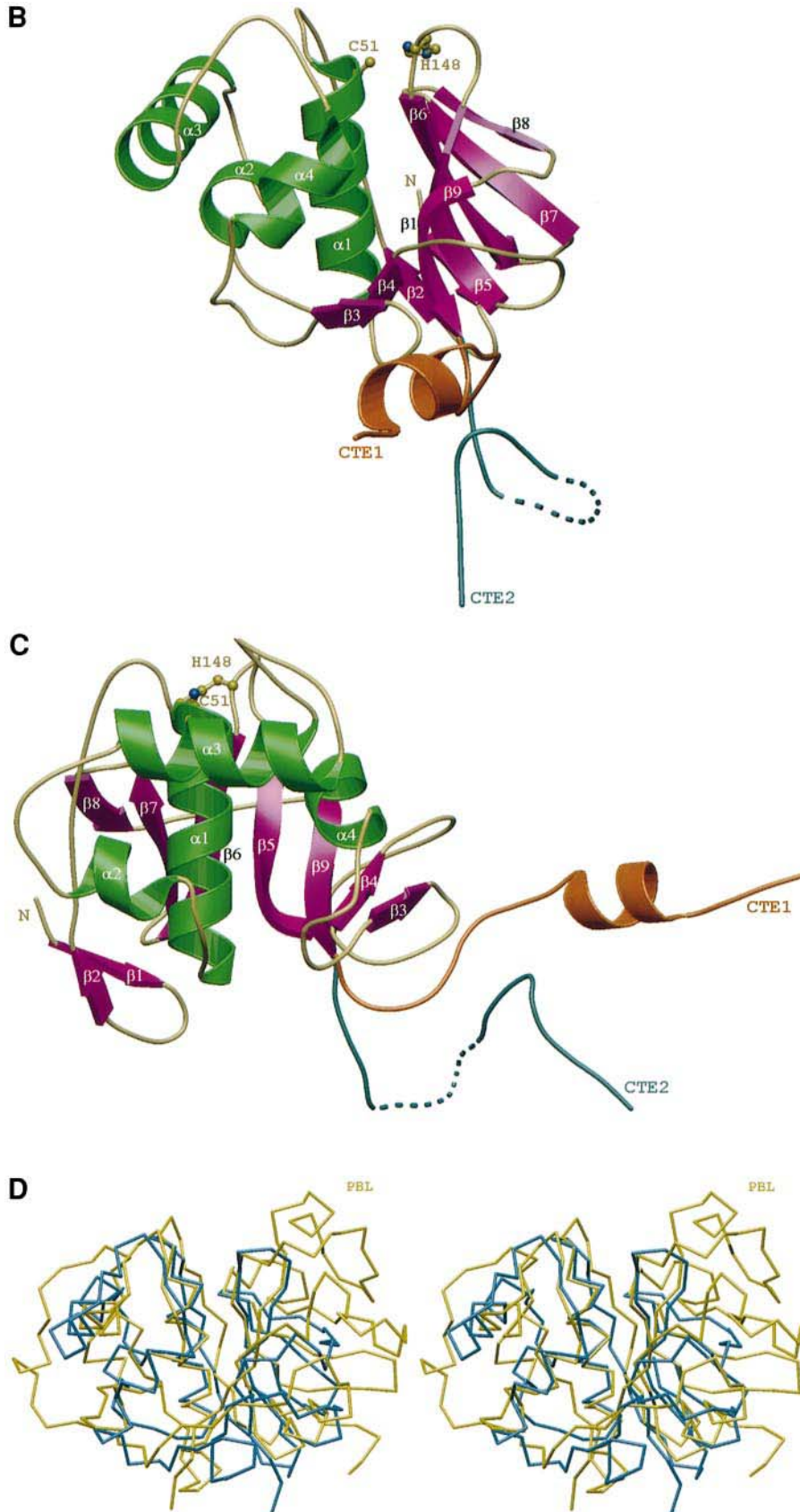
The extended conformations of the Lb<sup>pro</sup> CTEs are stabilized by a network of intermolecular interactions originating from the exchange of the CTEs between neighbouring molecules. In four out of the eight Lb<sup>pro</sup> molecules in the asymmetric unit, all residues in the CTE are visible (Figures 2C and 3) while, in the other four molecules, there is some disorder, and residues Glu186–Glu191 have not been traced (Figures 2C and 3). In the four molecules containing the disorder, the conformation of the visible residues is closely related to, though subtly different from that found in the other four (Figures 2B, C and 3). The 10 C-terminal amino acids, residues Trp192–Lys201, are nevertheless well defined in all eight crystallographically independent subunits and present identical interactions with the substrate-binding pocket of adjacent molecules in the crystal.



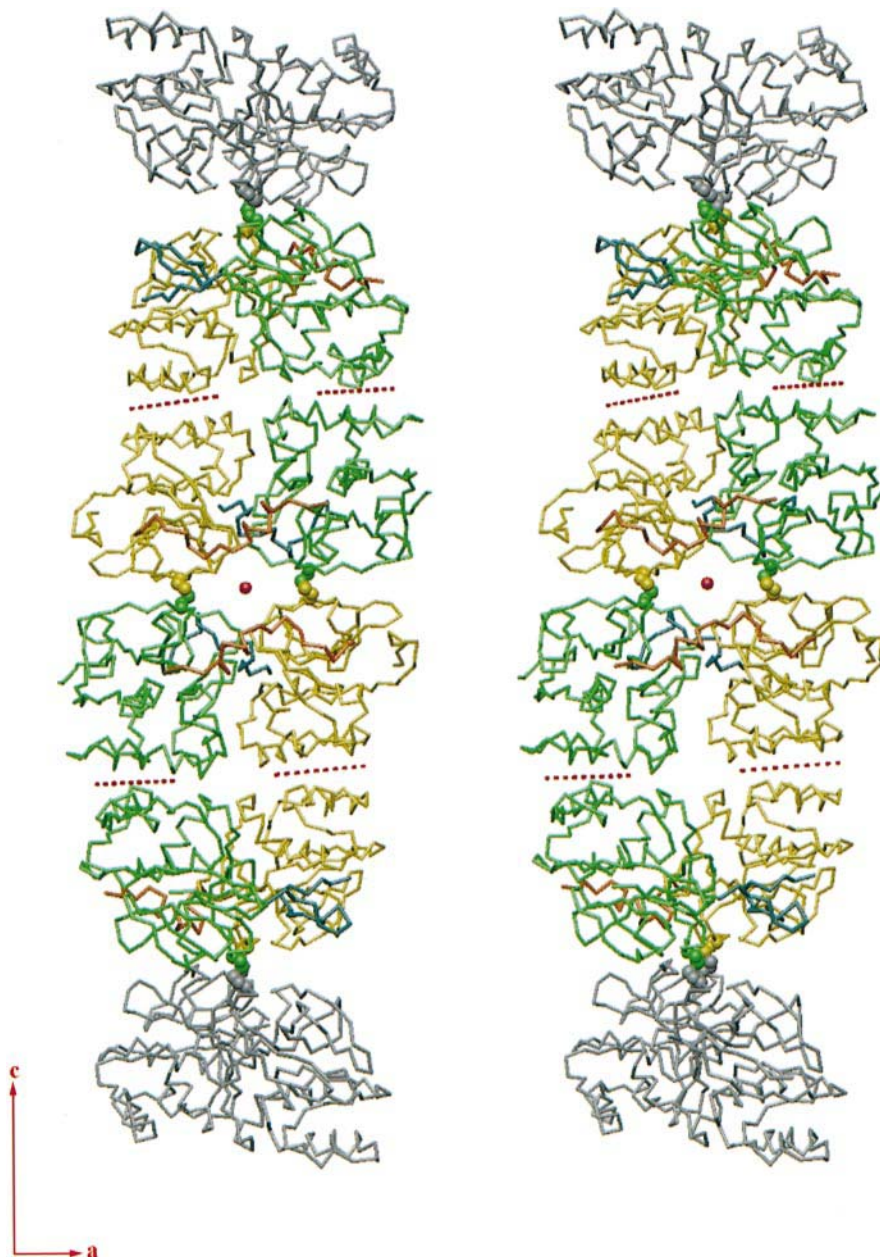
**The interaction between *L<sup>pro</sup>* and its C-terminus**

FMDV *L<sup>pro</sup>* frees itself from the growing polypeptide chain by specific cleavage at its own C-terminus (Figures 1A,B and 2A). Thus, the presence of CTE residues

inside the substrate-binding pockets of adjacent molecules illustrates substrate recognition during self-processing and represents, in fact, the P side of the substrate in the self-processing reaction. The peptide backbone of the final





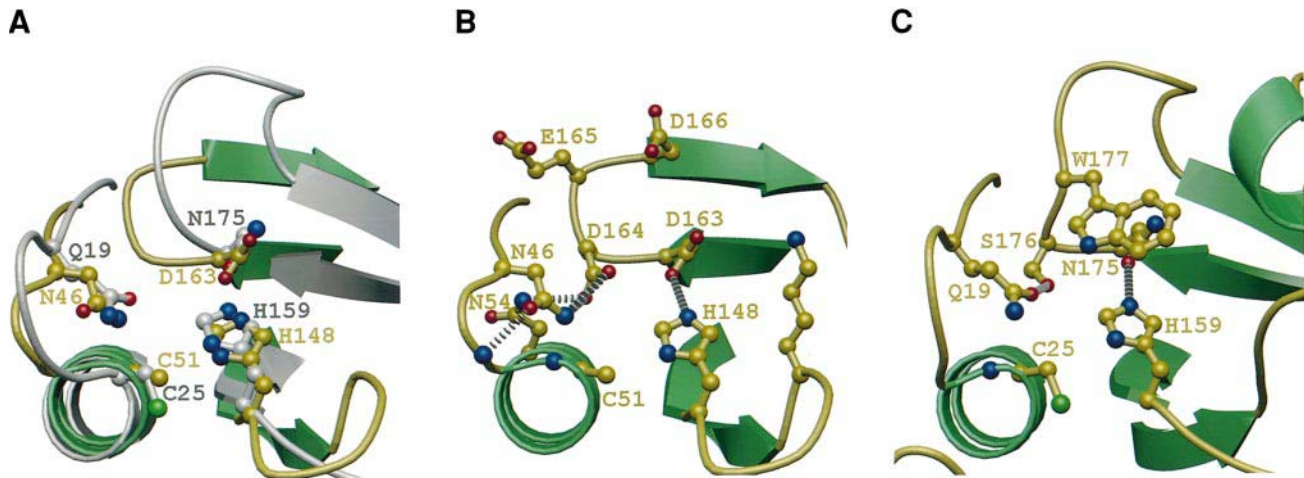


**Fig. 3.** Stereo view of the disposition of the Lb<sup>PRO</sup> molecules in the crystal. The eight Lb<sup>PRO</sup> molecules contained in the asymmetric unit are coloured. The orientation of the unit cell axis is also indicated. Four molecules, with the ordered CTEs, are shown in yellow, while the remaining four, with the disordered CTE, are shown in orange and blue, respectively. The corresponding CTEs are shown in orange and blue, respectively. Six intermolecular disulfide bridges, represented as thick balls, can also be seen, four of them with molecules of adjacent asymmetric units (in grey). The quasipolymeric character of the packing, with extensive non-covalent interactions alternating with disulfide bridges, is apparent. Two types of non-crystallographic symmetries are present. The first (dashed red lines) relates pairs of neighbouring molecules, whilst the second (red spot) relates four molecules of the asymmetric unit with the other four.

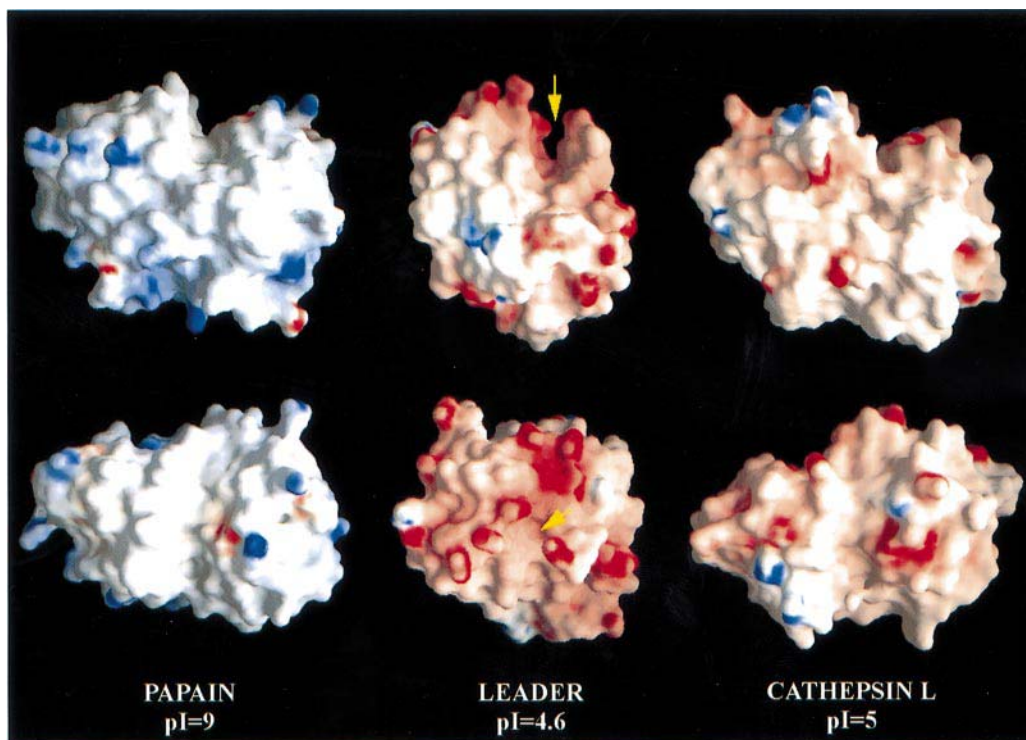
residues of the CTE is in an extended conformation (Figure 3) similar to that observed in complexes of enzymes of the papain superfamily with peptide-like inhibitors (Yamamoto *et al.*, 1991, 1992).

The main interactions between the CTE and the substrate-binding site (Figure 6) are provided by Lys201' and Leu200', with minor contributions from residues Lys199'–Val196' (residues in the CTE of a symmetry-related molecule are labelled with primed numbers). The final CTE residue, Lys201', is positioned close to the active site Cys51 (replaced by Ala in this structure) as if catalysis had been completed. One of its carboxylate oxygens,

located in the oxyanion hole, is hydrogen bonded to the side chain of Asn46 and the main-chain nitrogen of Cys51, whereas the second carboxylate oxygen accepts a hydrogen bond from the imidazole ring of the catalytic His148 (Figure 6), expected to be protonated as described for papain (Yamamoto *et al.*, 1991; Brocklehurst *et al.*, 1998). The CTE establishes additional interactions with the substrate-binding site through hydrogen bonds between main-chain atoms. Thus, the main-chain nitrogen of Lys201' forms a hydrogen bond with the main-chain carbonyl of Glu147; the main-chain oxygen and nitrogen atoms of Leu200' are hydrogen bonded to the main-chain



**Fig. 4.** Active site of Lb<sup>pro</sup>. (A) Arrangement of amino acid side chains around the active site (viewed down through the central helix  $\alpha 1$ ) after superimposition of papain (grey) and Lb<sup>pro</sup> (green) using the program SHP (Stuart *et al.*, 1979). Catalytic residues of Lb<sup>pro</sup> (Asn46, Cys51, His148 and Asp163) are in yellow, those of papain (Gln P-19, Cys P-25, His P-159 and Asn P-175) in grey. (B) Network of hydrogen bonds in the Lb<sup>pro</sup> active centre. The acidic cluster (Asp163–Asp166) in the S′-binding region is also shown. The orientation of the amide group of Asn46 is maintained by hydrogen bond interactions with the main-chain nitrogen of Asp49 and the side chains of Asn54 and Asp164. (C) Hydrogen bonds in the papain active centre. The fully conserved Trp P-177, in papain-like enzymes, covering the hydrogen bond between essential catalytic residues Asn P-175 and His P-159 is also shown.



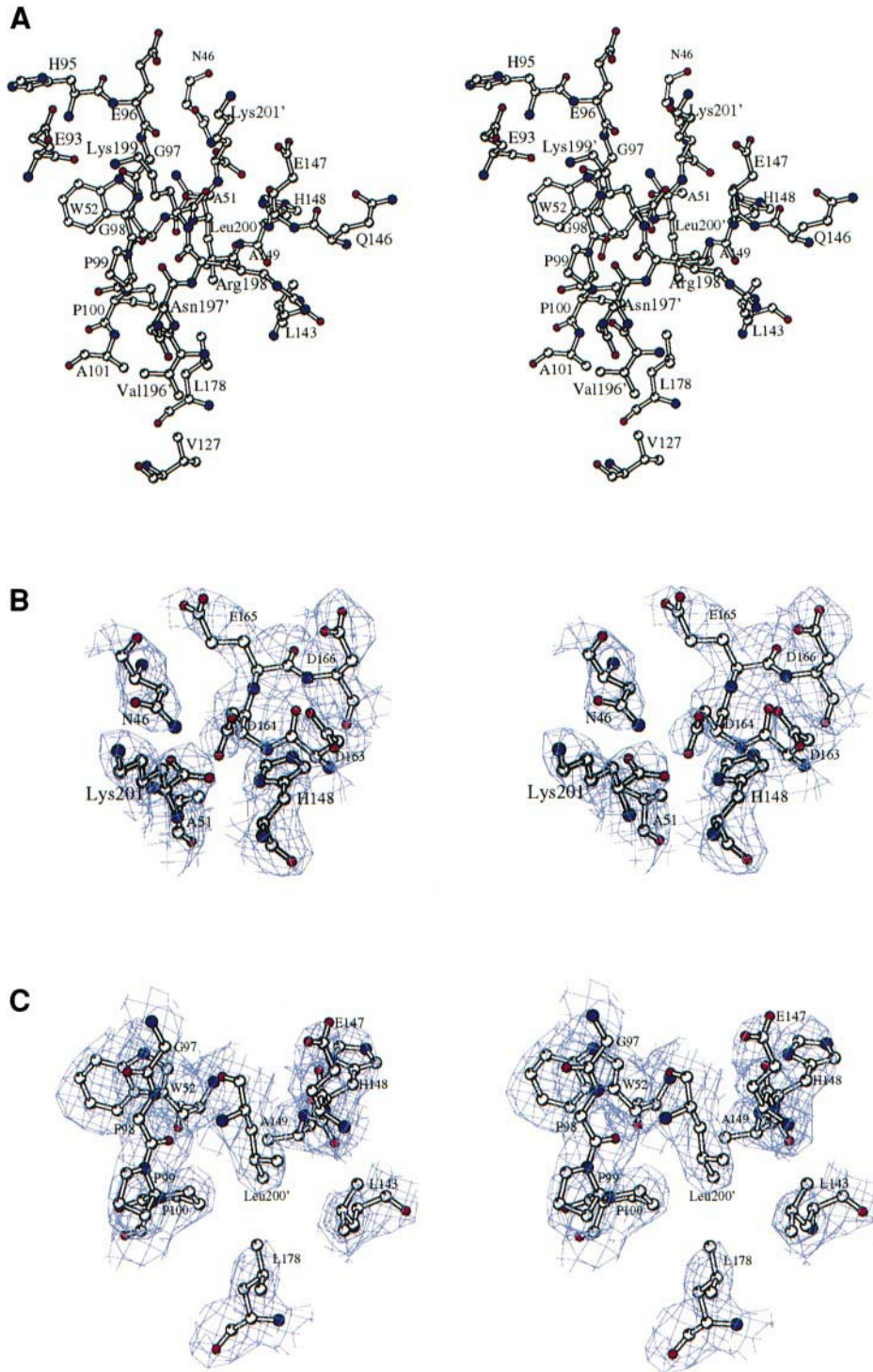
**Fig. 5.** Electrostatic potential surfaces of, left to right, papain (Kamphuis *et al.*, 1984; 1pap PDB code), Lb<sup>pro</sup> and cathepsin L (Coulombe *et al.*, 1996; 1aec PDB code). The yellow arrow indicates the Lb<sup>pro</sup> active centre. The standard papain view is used in the top panel; that in the lower panel, obtained by a simple 90° rotation of the upper one, is down the  $\alpha 1$  helix. The apparent differences in charge distribution reflect the indicated differences in isoelectric points, but also the different specificity of S and S′ subsites. [Figures 5 and 7 were generated by GRASP (Nicholls *et al.*, 1991)].

nitrogen and oxygen, respectively, of Gly98, building a short antiparallel  $\beta$ -sheet. These hydrogen bond interactions are a conserved feature in substrate binding by papain superfamily members.

The S<sub>1</sub> subsite in Lb<sup>pro</sup> is a narrow cleft bounded by the loop preceding the central helix  $\alpha 1$  on one side, the  $\beta$ -turn connecting strands  $\beta 5$  and  $\beta 6$  on the other side and

the active site at the bottom (Figures 2B and 6A,B). In the Lb<sup>pro</sup> structure, the aliphatic portion of the side chain of Lys201′, which occupies the S<sub>1</sub> subsite, is sandwiched between the main chain of residues His95–Glu96 and the side chain of Glu147, while its amino group establishes electrostatic interactions with the carboxylates of Glu96 and Glu147 (Figure 6A). The S<sub>1</sub> subsite in papain and





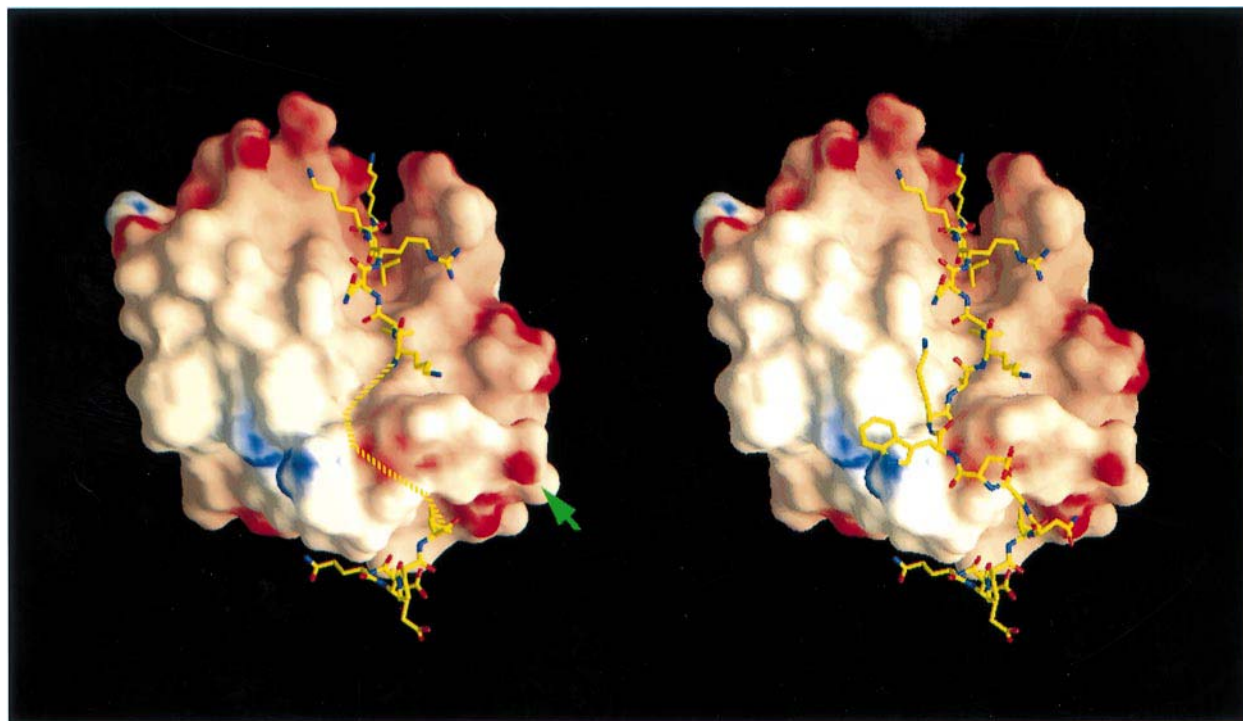
**Fig. 6.** Stereo view of the interactions between the CTE and the S site. (A) Residues involved in the formation of subsites  $S_1$ – $S_6$  are shown as balls and sticks, with nitrogen atoms coloured in blue and oxygen atoms in red. Residues of the CTE are also shown and labelled Val196'–Lys201'. (B and C) Residues involved in the formation of the  $S_1$  and  $S_2$  subsites, respectively, are shown as balls and sticks with their corresponding electron density in light blue. The  $P_1$  (Lys201') and  $P_2$  (Leu200') residues of the CTE are also shown.

other family members is a wide, unrestricted pocket which exerts relatively little influence on the substrate specificity (Figure 5). In Lb<sup>pro</sup>, which clearly prefers lysine at  $P_1$  in the self-processing reaction (Figure 2A), this subsite has become narrower and deeper due to a rearrangement of the loop connecting strands  $\beta_5$  and  $\beta_6$  on the R domain. Amino acid sequence alignments and modelling of the ERV1 L<sup>pro</sup> imply a correlation between the side chain at

$P_1$  and that of residue 147. Thus, FMDV enzymes, with lysine in  $P_1$ , have a negatively charged glutamate at position 147; in contrast, the corresponding residue in ERV1, with serine at  $P_1$ , is Gly149, shorter and not charged (Figure 2A).

The side chain of Leu200' ( $P_2$ ) is completely buried in a hydrophobic  $S_2$  pocket formed by Trp52, Gly97–Pro100, Leu143, Glu147–Ala149 and Leu178 (Figure 6A and C).





**Fig. 7.** Model for self-processing of the Lb<sup>P<sup>ro</sup></sup> *in cis*. The electrostatic potential surface of Lb<sup>P<sup>ro</sup></sup> with CTE residues Asp184–Asn189 and Lys195–Lys201 as sticks. Coordinates from Lys195 to Lys201 correspond to a neighbouring molecule in the crystal. However, the orientation and the proximity of this fragment to Asn189 points towards a simple, direct connection within the same molecule of Lb<sup>P<sup>ro</sup></sup>. This joining is indicated by dashes in the left panel and modelled by explicitly including residues Gly189–Ala194 in the right one. Polar amino acids would be exposed to the solvent, and only Trp192, pointing towards Trp105, appears to require minor rearrangements of the flexible and highly exposed Arg109. The green arrow indicates the position of the side chain of Cys133.

The architecture of the Lb<sup>P<sup>ro</sup></sup> S<sub>2</sub> subsite is very similar to that of other papain superfamily proteases; in fact, the residues defining the pocket in papain are identical to those in Lb<sup>P<sup>ro</sup></sup>, with the exception of Leu143 which is equivalent to Val P-133.

CTE residues Lys199' (P<sub>3</sub>) and Arg198' (P<sub>4</sub>) occupy loose pockets on opposite faces of the cleft, in the S<sub>3</sub> and S<sub>4</sub> subsites, respectively. The aliphatic portion of the side chain of Lys199' makes van der Waals contacts with main-chain atoms of residues Gly97–Gly98, and its amino group interacts through a hydrogen bond with the main-chain carbonyl group of Glu93 and through weak ionic interactions with the side-chain carboxylates of Glu93 and Glu96. Arg198' makes van der Waals contacts with Gly98, Pro99, Leu143 and extensively with Gln146. Its guanidinium group also hydrogen-bonds to the amide side chain of Gln146. Residue Gln197' (P<sub>5</sub>) has its side chain exposed to the solvent, but still contacts through its main chain Pro99, thus making a very open subsite S<sub>5</sub>. Finally, Val196' is buried in a hydrophobic cavity (subsite S<sub>6</sub> formed by residues Pro99, Ala101, Val127 and Leu178), located on the interdomain cleft, just underneath subsite S<sub>2</sub>.

### **Biological implications of the structure**

**Self-cleavage at the C-terminus.** The presence of the CTE in the active site of adjacent molecules argues for intermolecular self-processing. However, although in the crystal structure of Lb<sup>P<sup>ro</sup></sup> the CTE projects away towards neighbouring molecules, instead of folding back into its own substrate-binding cleft, several structural features suggest that self-processing *in cis* is possible and might

even be favoured. First, the interface between the globular domains exchanging their CTEs is composed of weak interactions, indicating that this region is not designed to promote an intermolecular reaction. Secondly, residues located immediately after Tyr183, at which point the polypeptide chain leaves the globular region to begin the CTE, favour a turn. Notably, Asp184 and Glu186 are conserved in all serotypes of FMDV and in ERV1, enabling the placement of a conserved hydrophobic residue (Leu188) into a shallow hydrophobic pocket formed by residues Ala118, Pro121, Thr130, Met132 and Cβ of Asp136. This interaction leads the CTE polypeptide in a direction compatible with both *cis* and *trans* self-processing. A polar, highly flexible stretch (residues Asn189–Glu191; modelled in Figure 7) should overcome the distance between this pocket and the above-described subsites in the substrate-binding cleft. A tryptophan residue at position 192 (or the aromatic residue found in ERV1 Tyr201) would enhance self-processing *in cis* by stacking its aromatic ring with the exposed and conserved Trp105 (ERV1 Tyr103) of the globular domain. Thus, the CTE would reach subsite S<sub>6</sub> (interaction with Val196) with minor rearrangements of the main chain (Figure 7); the electrostatic and van der Waals interactions of the CTE with the substrate-binding cleft would maintain the correct orientation for the self-processing *in cis*.

Why are intramolecular CTE interactions not observed in the Lb<sup>P<sup>ro</sup></sup> crystal structure? First, the path of the CTE to its 'own' active site is partially blocked by the intermolecular disulfide bridge between neighbouring molecules. This bond is not believed to be present in the

reducing environment inside the cell. Secondly, crystal packing requirements may have also favoured the observed CTE interactions. Finally, the ability to cleave eIF4G requires the CTE to be flexible and not remain in the active site. Evidence for the flexible nature of the CTE is provided by the lack of density in the sLb<sup>pro</sup> form and the differences in the positions of certain residues in the disordered form of the CTE. The disorder in the CTE also appears to be energetically favoured, as freezing the polar CTE in a fixed conformation would impose a high entropic penalty.

**Cleavage of eIF4G.** The tendency of the CTE to leave the active site of the same polypeptide chain after *cis* cleavage is of significance for the recognition and cleavage of eIF4G by Lb<sup>pro</sup>. Thus, the enzyme cannot be inhibited by binding to its own C-terminus, which served as the recognition site for the cleavage on the viral polyprotein. However, as indicated above, this implies that the S site does not provide sufficient interactions to maintain the substrate in the active site; indeed, the absence of significant intramolecular product inhibition is probably a direct result of this inability. Thus, to bind to its cleavage site on eIF4G, which lacks a basic P<sub>1</sub> residue, the enzyme appears to employ the acidic patch of the S' site (Figure 5) to provide an ionic interaction with the P'<sub>1</sub> Arg residue of the cleavage site on eIF4G. Indeed, it is noteworthy that all intermolecular substrates of Lb<sup>pro</sup> identified so far *in vitro* contain basic residues at P'<sub>1</sub> or P'<sub>2</sub> or both, even when the P<sub>1</sub> residue is basic (data not shown). Taken together, these observations suggest that a substrate containing basic residues at both P<sub>1</sub> and P'<sub>1</sub> should be an optimized substrate for the Lb<sup>pro</sup>. As, however, no data are available on the sequence preference of Lb<sup>pro</sup> on peptide substrates, experiments are underway to investigate this notion.

The acidic patch at the S' site of Lb<sup>pro</sup>, coupled with the narrow cleft traversing the active site (Figures 5 and 6A), also appear to be the reasons why the Lb<sup>pro</sup> is clearly much more specific than papain, although the two enzymes possess an S<sub>2</sub> pocket almost identical in composition and topology. Thus, Lb<sup>pro</sup> does not cleave an immunoglobulin molecule, a classic substrate of papain. Furthermore, although eIF4G is an efficient substrate for papain, the cleavage products are not the same as those of Lb<sup>pro</sup> (B.Hampoelez and T.Skern, unpublished).

## Summary

The FMDV Lb<sup>pro</sup> is the first crystal structure determined of a viral papain-like cysteine protease. The structure shows the interactions involved in C-terminal processing which enable cleavage between a lysine and a glycine residue to take place. In addition, examination of the substrate-binding cleft indicates how the enzyme can carry out the specific cleavage of the host protein eIF4G which, in contrast, requires cleavage between a glycine and an arginine residue.

## Materials and methods

### Protein expression, purification and crystallization

The Cys51Ala mutant of the FMDV serotype O<sub>1k</sub> Lb<sup>pro</sup> was expressed in *Escherichia coli* BL21 (DE3) pLysS and purified as described (Kirchweger *et al.*, 1994). A variant of the Cys51Ala mutant lacking six amino acids at

the C-terminus (sLb<sup>pro</sup>) was expressed and purified similarly. Lb<sup>pro</sup> crystals, belonging to space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with cell dimensions  $a = 65.4 \text{ \AA}$ ,  $b = 101.6 \text{ \AA}$  and  $c = 277.0 \text{ \AA}$ , were obtained by vapour diffusion against solutions containing 10% PEG 6000, 0.8 M MgCl<sub>2</sub>, 0.1 M Tris-HCl, pH 8.5. There are eight molecules in the asymmetric unit and an estimated solvent content of 59%. sLb<sup>pro</sup> crystals, belonging to space group C222<sub>1</sub> with unit cell dimensions of  $a = 51.0 \text{ \AA}$ ,  $b = 130.0 \text{ \AA}$  and  $c = 126.2 \text{ \AA}$ , were grown from 8% PEG 4000, 0.2 M MgCl<sub>2</sub>, 0.1 M Tris-HCl, pH 8.5. This crystal form has two molecules in the asymmetric unit and an estimated solvent content of 55%. sLb<sup>pro</sup> crystals were harvested in solutions containing 10% PEG 4000, 0.2 M MgCl<sub>2</sub>, 0.1 M Tris-HCl, pH 8.5, prior to data collection or heavy atom screening.

The two heavy atom derivatives of the sLb<sup>pro</sup> crystal used for phase determination were prepared by soaking at room temperature for 48 h in the harvesting solution containing 0.1 mM HgCl<sub>2</sub> or for 24 h in the harvesting solution which had been adjusted to 10 mM K<sub>2</sub>PtCl<sub>4</sub>.

### Data collection

For cryogenic X-ray data collection, native and derivative crystals were soaked in harvesting solutions made up to 25% ethylene glycol and flash-frozen under a stream of boiled-off nitrogen at 100 K (Oxford CryoSystems). X-ray diffraction data sets were collected on MarResearch image plate detector systems using a Rigaku RU-200B rotating anode and synchrotron facilities at the X11 EMBL outstation (DESY, Hamburg). Data were indexed, reduced, scaled and merged with DENZO and SCALEPACK (Otwinowski and Minor, 1997) (Table I). Most subsequent calculations were performed with the CCP4 Program Suite (Collaborative Computational Project Number 4, 1994).

### Structure determination

The structure of the sLb<sup>pro</sup> crystal form was determined by multiple isomorphous replacement. Heavy atom sites were located by Patterson methods and confirmed using cross-phased difference maps. Refinement of heavy atom parameters and phase calculation were performed with SHARP (de la Fortelle and Bricogne, 1997). Initial phases were calculated at 3.0 Å (Table I), and were significantly improved using the density modification procedures in SOLOMON. An electron density map calculated using these phases showed clear molecular boundaries and allowed the identification of several secondary structure elements to which polyalanine chains were fitted. These elements and the positions of the heavy atom substitutions were used to determine the position of the local 2-fold axis relating the two molecules in the asymmetric unit. Masks covering the monomer were created from skeletonized electron density and partial models using MAMA (Kleywegt and Jones, 1994), edited interactively with program O (Jones *et al.*, 1991), and used for local averaging and solvent flattening in further cycles of density modification performed with DM (Cowtan, 1994). The final figure of merit for data in the range 25.0–3.0 Å was 0.68. The resulting electron density map was used for model building with program O. The initial model comprised 138 amino acid residues for each of the two copies in the asymmetric unit and had a crystallographic *R*-factor of 0.47 for all reflections in the resolution range 10.0–3.0 Å.

Initial phases for the Lb<sup>pro</sup> (residues 29–201) crystal form were obtained by molecular replacement with the sLb<sup>pro</sup> coordinates as a model. Rotation/translation parameters were calculated with 95% of data between 15 and 3.5 Å using the AMoRe package (Navaza, 1994), and the information about the presence of a translational symmetry derived from the strong peak found in the native Patterson at position  $u = 0.5$ ,  $v = 0.5$ ,  $w = 0.225$  (Navaza *et al.*, 1998). Four independent solutions were found that generate the second four by translation. The molecular replacement solution gave a final correlation factor of 0.69 and an *R*-factor of 0.40 for data in the same resolution range.

### Refinement

Refinement was done following standard protocols using iteratively program X-PLOR (Brünger, 1992) alternating with manual rebuilding in the interactive graphics program O. For both crystal forms, bulk solvent, overall anisotropic *B*-factor corrections and tight non-crystallographic restraints were introduced based on the behaviour of the *R*<sub>free</sub> index (Table I). The non-crystallographic restraints for the Lb crystal form were applied considering two groups of 4-fold symmetrically related molecules.

The refined atomic model for the sLb<sup>pro</sup> form comprises residues 29–187 for both copies in the asymmetric unit and 20 ordered solvent molecules, and has an *R*-factor of 23.5% (*R*<sub>free</sub> = 28.8%) for all data between 20.0 and 3.0 Å. Of the non-glycine residues, 82.1% fall within the 'most favoured regions' of the Ramachandran plot, on the higher side of acceptable for a 3.0 Å structure as defined by the program PROCHECK (Laskow-

ski *et al.*, 1994). The rest are inside the 'additional allowed regions', except for Asp164 which presents good electron density and is located at position  $i + 1$  of a type II'  $\beta$  turn. No electron density is observed for residues 184–187, which form part of the CTE and project away from the globular catalytic domain. The remainder of the residues, apart from a few solvent-exposed side chains, present well-defined electron density for both molecules in the asymmetric unit.

The refined atomic model for the Lb<sup>pro</sup> form comprises residues 29–201 for four copies in the asymmetric unit while the other four do not include residues 186–192. The N-terminal residue appears to be modified by a well-defined acetyl group in the eight independent subunits. The present model has an *R*-factor of 25.8% ( $R_{\text{free}} = 29.6\%$ ) for all data between 20.0 and 3.0 Å. No electron density is observed for residues 186–192 in copies with the disordered conformation of the CTE; however, the remaining residues, apart from a few solvent-exposed side chains, present well-defined electron density for all molecules in the asymmetric unit.

## Acknowledgements

We thank F.Torrents and A.Marina for assistance, J.Navaza for his valuable advice, and J.Bravo, F.X.Gomis-Rüth and J.Seipelt for critical reading of the manuscript. This work was supported by the Austrian Science Foundation (P-11222 to T.S.) and DGICYT [PB95-0218 (to I.F.) and P96-0271 (to J.T.)]. Data collection in Hamburg was supported by the Human Capital Mobility Project, contract CHGE-CT93-0040. A.G. is the recipient of a fellowship from the Ministerio de Educacion y Cultura (Spain).

## References

- Allaire, M., Chernaia, M.M., Malcolm, B.A. and James, M.N.G. (1994) Picornaviral 3C cysteine proteases have a fold similar to chymotrypsin-like serine proteases. *Nature*, **369**, 72–76.
- Babé, L.M. and Craik, C.S. (1997) Viral proteases: evolution of diverse structural motifs to optimize function. *Cell*, **91**, 427–430.
- Belsham, G.J., Brangwyn, J.K., Ryan, M.D., Abrams, C.C. and King, A.M.Q. (1990) Intracellular expression and processing of foot-and-mouth disease virus capsid precursors using vaccinia virus vectors: influence of the L protease. *Virology*, **176**, 524–530.
- Berti, P.J. and Storer, A.C. (1995) Alignment/phylogeny of the papain superfamily of cysteine proteases. *J. Mol. Biol.*, **246**, 273–283.
- Brocklehurst, K., Watts, A.B., Patel, M., Verma, C. and Thomas, E.W. (1998) Cysteine proteinases. In Sinnott, M. (ed.), *Comprehensive Biological Catalysis*. Vol. I. Academic Press, San Diego, CA, pp. 381–423.
- Brünger, A.T. (1992) *XPLOR Version 3.1: A System for X-Ray Crystallography and NMR*. Yale University Press, New Haven, CT.
- Chen, P. *et al.* (1996) Structure of the human cytomegalovirus protease catalytic domain reveals a novel serine protease fold and catalytic triad. *Cell*, **86**, 835–843.
- Collaborative Computational Project, Number 4 (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr.*, **D50**, 760–763.
- Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J.S. and Cygler, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.*, **15**, 5492–5503.
- Cowan, K. (1994) DM, an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newslett. Protein Crystallogr.*, **31**, 24–28.
- de la Fortelle, E. and Brice, G. (1997) Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.*, **276**, 472–494.
- Ding, J., McGrath, W.J., Sweet, R.M. and Mangel, W.F. (1996) Crystal structure of the human adenovirus protease with its 11 amino acid cofactor. *EMBO J.*, **15**, 1778–1783.
- Esnouf, R.M. (1997) An extensively modified version of Molscript that includes greatly enhanced colouring capabilities. *J. Mol. Graph.*, **15**, 133–138.
- Gorbalenya, A.E., Koonin, E.V. and Lai, M.M.-C. (1991) Putative papain-related thiol proteases of positive-strand RNA viruses. *FEBS Lett.*, **288**, 201–205.
- Guarné, A., Kirchweger, R., Verdaguer, N., Liebig, H.D., Blaas, D., Skern, T. and Fita, I. (1996) Crystallization and preliminary X-ray diffraction studies of the Lb protease from foot-and-mouth disease virus. *Protein Sci.*, **5**, 1931–1933.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in this model. *Acta Crystallogr.*, **A47**, 110–119.
- Kamphuis, I.G., Kalk, K.H., Swarte, M.B.A. and Drenth, J. (1984) Structure of papain refined at 1.65 Å resolution. *J. Mol. Biol.*, **179**, 233–256.
- Kamphuis, I.G., Drenth, J. and Baker, E.N. (1985) Thiol proteases: comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information from cathepsins B and H, and stem bromelain. *J. Mol. Biol.*, **182**, 317–329.
- Kim, J.L. *et al.* (1996) Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell*, **87**, 345–355.
- Kirchweger, R. *et al.* (1994) Foot-and-mouth disease virus leader protease: purification of the Lb form and determination of its cleavage site on eIF-4γ. *J. Virol.*, **68**, 5677–5684.
- Kleywegt, G.J. and Jones, T.A. (1994) Halloween masks and bones. In Bailey, S., Hubbard, R. and Waller, D. (eds), *From the First Map to Final Model*. SERC Daresbury Laboratory, Warrington, UK, pp. 59–66.
- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Laskowski, R.A., MacArthur, M.W., Smith, D.K., Jones, D.T., Hutchinson, E.G., Morris, A.L., Naylor, D., Moss, D. and Thornton, J.M. (1994) *PROCHECK Manual. Version 3.0*. Oxford Molecular Ltd, Oxford, UK.
- Love, R.A., Parge, H.E., Wickersham, J.A., Hostomsky, Z., Habuka, N., Moomaw, E.W., Adachi, T. and Hostomska, S. (1996) The crystal structure of hepatitis C virus NS3 protease reveals a trypsin-like fold and a structural zinc binding site. *Cell*, **87**, 331–342.
- Matthews, D.A. *et al.* (1994) Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell*, **77**, 761–771.
- Medina, M., Domingo, E., Brangwyn, J.K. and Belsham, G.J. (1993) The two species of the foot-and-mouth disease virus leader protein, expressed individually, exhibit the same activities. *Virology*, **194**, 355–359.
- Merrit, E.A. and Murphy, M.E.P. (1994) Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr.*, **D50**, 869–873.
- Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.
- Navaza, J., Panepucci, E.H. and Martin, C. (1998) On the use of strong Patterson function signals in many-body molecular replacement. *Acta Crystallogr.*, **D54**, 817–812.
- Nicholls, A., Sharp, K. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
- Otwinski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Piccione, M.E., Zellner, M., Kumosinski, T.F., Mason, P.W. and Grubman, M.J. (1995a) Identification of the active-site of the L protease of foot-and-mouth disease virus. *J. Virol.*, **69**, 4950–4956.
- Piccione, M.E., Sira, S., Zellner, M. and Grubman, M.J. (1995b) Expression in *Escherichia coli* and purification of biologically active L proteinase of foot-and-mouth disease virus. *Virus Res.*, **35**, 263–275.
- Roberts, P.J. and Belsham, G. (1995) Identification of critical amino acids with the foot-and-mouth disease Leader protein, a cysteine protease. *Virology*, **213**, 140–146.
- Ryan, M.D. and Flint, M. (1997) Virus-encoded proteases of the picornavirus super-group. *J. Gen. Virol.*, **78**, 699–723.
- Skern, T., Fita, I. and Guarné, A. (1998) A structural model of picornavirus leader proteases based on papain and bleomycin hydrolase. *J. Gen. Virol.*, **79**, 301–307.
- Stuart, D.I., Levine, M., Muirhead, H. and Stammers, D.K. (1979) The crystal structure of a cat pyruvate kinase at a resolution of 2.6 Å. *J. Mol. Biol.*, **134**, 109–142.
- Tong, L., Qian, C., Massariol, M.-J., Bonneau, P.R., Cordingley, M.G. and Lagacé, L. (1996) A new serine-protease fold revealed by the crystal structure of human cytomegalovirus protease. *Nature*, **383**, 272–282.
- Yamamoto, D., Matsumoto, K., Ohishi, H., Ishida, T., Inoue, M., Kitamura, K. and Mizuno, H. (1991) Refined X-ray structure of papain-E-64-c complex at 2.1 Å resolution. *J. Biol. Chem.*, **266**, 14771–14777.
- Yamamoto, A. *et al.* (1992) Crystal structure of papain-succinyl-Gln-Val-Val-Ala-Ala-p-nitroanilide complex at 1.7 Å resolution: noncovalent binding mode of a common sequence of endogenous thiol protease inhibitors. *Biochemistry*, **31**, 11305–11309.
- Ziegler, E., Borman, A.M., Kirchweger, R., Skern, T. and Kean, K.M. (1995) Foot-and-mouth disease virus Lb proteinase can stimulate rhinovirus and enterovirus IRES-driven translation and cleave several proteins of cellular and viral origin. *J. Virol.*, **69**, 3465–3474.

Received August 13, 1998; revised October 14, 1998;  
accepted October 15, 1998