



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

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

Crystal structure of the serine protease domain of *Sesbania mosaic virus* polyprotein and mutational analysis of residues forming the S1-binding pocket

P. Gayathri^a, P.S. Satheshkumar^b, K. Prasad^a, Smita Nair^b, H.S. Savithri^b, M.R.N. Murthy^{a,*}

^a Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

^b Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, India

Received 25 August 2005; returned to author for revision 28 October 2005; accepted 7 November 2005

Available online 13 December 2005

Abstract

Sesbania mosaic virus (SeMV) polyprotein is processed by its N-terminal serine protease domain. The crystal structure of the protease domain was determined to a resolution of 2.4 Å using multiple isomorphous replacement and anomalous scattering. The SeMV protease domain exhibited the characteristic trypsin fold and was found to be closer to cellular serine proteases than to other viral proteases. The residues of the S1-binding pocket, H298, T279 and N308 were mutated to alanine in the ΔN70-Protease–VPg polyprotein, and the *cis*-cleavage activity was examined. The H298A and T279A mutants were inactive, while the N308A mutant was partially active, suggesting that the interactions of H298 and T279 with P1-glutamate are crucial for the E–T/S cleavage. A region of exposed aromatic amino acids, probably essential for interaction with VPg, was identified on the protease domain, and this interaction could play a major role in modulating the function of the protease.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Serine protease; *Sesbania mosaic virus*; Crystal structure; MIRAS; Trypsin fold; Glutamate-specific protease; Polyprotein processing; S1-binding pocket

Introduction

Sesbania mosaic virus (SeMV) is an ss-RNA plant sobemovirus found infecting *Sesbania grandiflora* in India. Its genome of 4149 nucleotides codes for four open reading frames (ORFs) (Lokesh et al., 2001). ORF1 codes for an 18-kDa protein predicted to be a movement protein (Sivakumaran et al., 1998). The ORF2 codes for a polyprotein, which encompasses more than one functional domain, while ORF3 is present as an internal ORF in ORF2 and is expressed via ribosomal frame shifting mechanism (Lokesh et al., 2001). The ORF4 is expressed from a sub-genomic RNA and codes for the coat protein of the virus.

Abbreviations: SeMV, *Sesbania mosaic virus*; ss-RNA, single-stranded RNA; ORF, open reading frame; VPg, viral protein genome linked; RdRP, RNA-dependent RNA polymerase; MIRAS, multiple isomorphous replacement and anomalous scattering; IPTG, isopropyl thiogalactoside; PCMBs, *p*-chloro mercury benzene sulphonate.

* Corresponding author. Fax: +91 80 23600535.

E-mail address: mrm@mbu.iisc.ernet.in (M.R.N. Murthy).

Polyprotein processing is one of the major strategies employed by both animal and plant viruses to generate more than one functional protein from the same polypeptide chain (Wellink and van Kammen, 1988). To accomplish cleavage at specific sites, viruses employ one or more proteases with unique cleavage specificities. In an earlier study, we have shown that in SeMV, the processing is mediated by the N-terminal protease domain coded by ORF2 (Satheshkumar et al., 2004). It is a serine protease, similar to cellular proteases like trypsin and chymotrypsin (Gorbalenya et al., 1988). The catalytic residues are H181, D216 and S284. The protease cleaves the ORF2 polyprotein at three different positions, at E325–T326, E402–T403 and E498–S499 to release four different domains—protease, VPg (viral protein genome linked), p10 and RdRP (RNA-dependent RNA polymerase) (Satheshkumar et al., 2004). In most other viruses that have VPg at the 5' end of their genome, the domain arrangement is VPg–protease, whereas, in sobemoviruses, it is protease–VPg.

Recent biochemical studies on the protease–VPg domains of SeMV have shown that the interaction of VPg with the

Table 1A
Data collection statistics for native and derivative datasets

	Native	HgCl ₂	Thiomersal	HgI ₂	PCMBs
Crystal	ΔN92-Pro	ΔN92-Pro	ΔN92-Pro	ΔN70-Pro	ΔN70-Pro
Concentration of heavy atom solution	–	2.5 mM	25 mM	2.5 mM	2.5 mM
Time of soak	–	3 h	30 min	2 h	48 h
Unit cell (Å) ^{a,c}	74.07, 68.75	74.11, 68.70	73.42, 68.82	75.05, 68.88	74.47, 68.61
Space group	P3 ₁ 21	P3 ₁ 21	P3 ₁ 21	P3 ₁ 21	P3 ₁ 21
Resolution range	50–2.3 Å	50–2.8 Å	50–2.8 Å	50–3.5 Å	50–3.2 Å
Overall completion ^b	92.3 (56.5)	99.4 (98.7)	99.9 (99.6)	97.4 (90.1)	97.8 (85.7)
R_{merge} ^{a,b}	5.5 (26.6)	10.8 (34.5)	10.9 (47.0)	9.0 (25.6)	6.7 (17.6)
$\langle I \rangle / \langle \sigma I \rangle$ ^b	32.1 (3.8)	16.0 (4.8)	18.72 (3.4)	12.2 (3.5)	20.2 (5.6)
Wilson B	45.606	63.655	69.435	49.981	87.152

^a $R_{\text{merge}} = \Sigma(I_{hj} - \langle I_h \rangle) / \Sigma I_{hj}$, where $\langle I_h \rangle$ is the average of the intensities of symmetry related reflections.

^b Values in parenthesis refer to the highest resolution bin.

protease domain modulates the protease activity (Satheskumar et al., 2005). The serine protease domain lacking the 70 amino acids (ΔN70-Pro) was found to be inactive in *trans*. However, the presence of VPg at the C-terminus of ΔN70-Pro rendered the polyprotein active in *cis* and *trans*. By mutational analysis, it was demonstrated that interaction of W43 in the VPg domain with the protease was responsible for both *cis* and *trans* proteolytic activities and the associated conformational changes. It was suggested that the natively unfolded VPg is an activator of the protease and could regulate the polyprotein processing.

In the present study, the crystal structure of SeMV protease domain was determined to a resolution of 2.4 Å by multiple isomorphous replacement coupled with anomalous scattering, with a view to identify the residues involved in substrate binding as well as protease–VPg interactions. Based on structural alignment of the S1-binding pocket with proteases of Glu/Gln specificity, H298, T279 and N308 were identified as glutamate-binding residues, and their role has been established by mutational analysis. A region of exposed aromatic residues, probably essential for interaction with the VPg domain, has been identified on the protease domain. This is the first report of the structure of a non-structural protein from the genus *Sobemovirus* and provides the framework for understanding polyprotein processing in the genus.

Results and discussion

Protein crystallization and data collection

Cloning and expression of the full-length protease domain (residues 1–325 of the polyprotein) in *Escherichia coli* resulted in insoluble aggregates. Examination of the amino acid sequence of the protease domain suggested that its amino terminal residues might correspond to a transmembrane region. It was found that the solubility and the activity of the protease were enhanced by the deletion of 70 (ΔN70-Pro) and 92 (ΔN92-Pro) residues. The ΔN70-Pro and ΔN92-Pro domains of the polyprotein were cloned, expressed and purified as described earlier (Satheskumar et al., 2004). Both ΔN70-Pro and ΔN92-Pro domains were crystallized using microbatch

method in the presence of 0.2 M Tris, pH 8.0, 0.2 M ammonium sulphate, 0.6 M 1,6-hexane diol, 5 mM β-mercaptoethanol and 4% glycerol. The protein crystals were shown to belong to the space group P3₁21. The cell parameters were determined to be $a = b = 74.07$ Å, $c = 68.75$ Å, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$. The asymmetric unit of the crystal was compatible with a monomer of the protease with a Matthews's coefficient of 2.5 Å³/Da and a solvent content of 52% (Matthews, 1968).

A very mosaic diffraction pattern of limited resolution was observed with many crystals. The quality of diffraction of such crystals could be substantially improved by repeated soaking of the crystals in the cryoprotectant solution (crystallization buffer with 20% ethylene glycol) followed by flash freezing in liquid nitrogen. Interestingly, the data from the annealed and non-annealed crystals scaled well. Hence, both types of datasets were combined and used for structure solution. The use of a similar annealing procedure to improve diffraction has been reported earlier (Kriminski et al., 2002).

Surprisingly, both the ΔN70-Pro and ΔN92-Pro crystals had identical cell parameters even though ΔN70-Pro was 22 amino acids longer than ΔN92-Pro. On mass spectroscopic analysis, it was observed that both had the same mass of ~20,000 Da suggesting an internal cleavage. N-terminal sequencing of the proteins suggested a possible cleavage between the residues A134 and V135 (data not shown) in both the mutants. The N-termini of both ΔN70-Pro and ΔN92-Pro therefore correspond to the residue V135. Henceforth, the ΔN70-Pro and ΔN92-Pro

Table 1B
MIRAS phasing statistics

	HgCl ₂	Thiomersal	HgI ₂	PCMBs
No. of heavy atom sites	2	7	2	2
R_{Cullis} (acentric) ^a	0.65	0.67	0.73	0.69
R_{Cullis} (centric)	0.75	0.78	0.77	0.77
Phasing power (centric) ^b	1.51	1.42	1.40	1.42
Phasing power (acentric)	1.24	1.19	1.12	1.16
R_{Cullis} anomalous	0.91	0.91	0.93	0.91

^a $R_{\text{Cullis}} = \Sigma(|F_{\text{PH}} \pm F_{\text{P}}| - |F_{\text{H}}|_{\text{calc}}) / \Sigma|F_{\text{PH}} \pm F_{\text{P}}|$.

^b Phasing Power = $[\Sigma F_{\text{H}}^2 / \Sigma(|F_{\text{PH}}|_{\text{obs}} - |F_{\text{P}}|_{\text{calc}})^2]^{1/2}$, where F_{PH} and F_{H} are the derivative and calculated structure factors respectively.

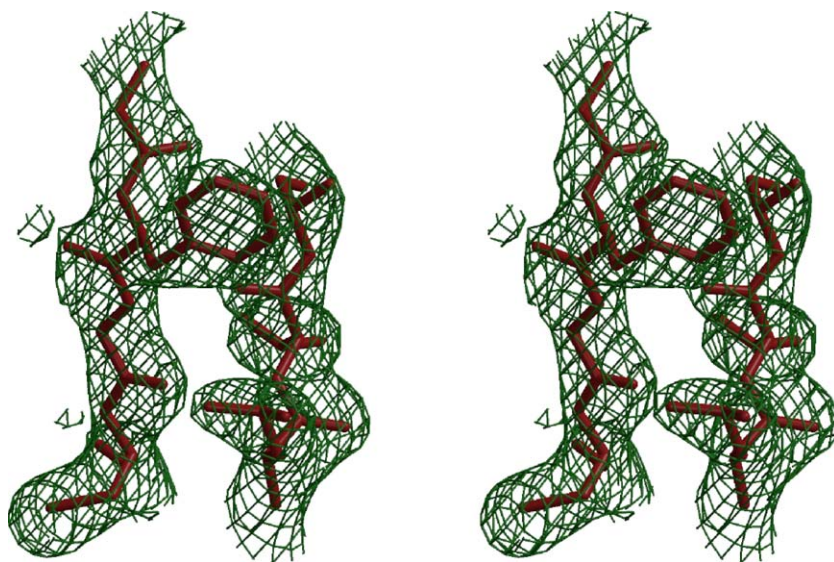


Fig. 1. Quality of electron density map. Stereo view of a representative section of the electron density ($2F_o - F_c$) contoured at 1.2σ is shown.

were considered indistinguishable, and data from both constructs were combined for phasing. SeMV protease does not have any significant sequence similarity to any of the cellular or viral serine proteases. Hence, molecular replacement method could not be used for phase determination, and data from heavy atom derivatives and anomalous scattering were used to determine the structure. The data collection statistics for the native and derivative datasets and phasing statistics are shown in Tables 1A and 1B, respectively.

Quality of the model

Most of the residues in the final map are in good electron density, except for a few residues at the protein surface. The side chains of a few lysines, arginines and aspartates with poor density have been truncated according to the extent of density observed for them. A section of the electron density map is shown in Fig. 1. Electron density is absent for the first three residues at the N-terminus and two residues at the C-terminus. Hence, the final model includes residues 138 to 323. The electron density for the stretch of residues from 171 to 173 is poor, and these residues are omitted in the model. The loop region comprising of residues 251 to 254 is partially disordered, and there is a break in the electron density at S252. A total of 45 water molecules have been added. The structure has been refined to a resolution of 2.4 Å though the data extended to 2.3 Å, due to the presence of very few test reflections in the last resolution bin. Table 1C lists the refinement statistics.

Polypeptide fold

SeMV protease belongs to the trypsin-like family of serine proteases. The overall fold exhibits the characteristic features of the trypsin fold. Fig. 2 shows the overall fold of the protease and a topology diagram with the secondary structure elements labeled according to the convention followed for

trypsin. It consists of two β barrels (domains I and II) connected by a long inter-domain loop. Both the domains belong to the all β class of proteins. The active site and the substrate-binding cleft occur in between the two domains and are fairly exposed to the solvent. There are only three helices in SeMV protease.

The barrel formed by the β -strands aI, bI, cI, dI, eI and fI constitutes the domain I. The first β -strand aI of domain I begins at L152. The active site residue H181 occurs in the small helix in the segment connecting cI and dI strands (Fig. 2). Strand eI extends from residues 197 to 210 and consists of two strands eIa and eIb connected by the stretch of residues from 201 to 205 in an extended but irregular conformation. D216 of the catalytic triad occurs at the end of the loop eI–fI. The loop connecting the two domains is a long stretch of residues extending from 222 to 243. This region contains a small helix, formed by the residues 223 to 231.

Table 1C
Refinement statistics

<i>R</i> factor (%)	25.04 (35.3)
<i>R</i> _{free} (%)	28.46 (35.1)
No. of atoms	
Protein atoms	1349
Solvent atoms	45
Ligand atoms	6
<i>Model quality</i>	
Average B-factors (Å ²)	
Protein atoms	40.800
Water	44.096
Ligand atoms	54.538
RMS deviation from ideal values	
Bond length (Å)	0.008
Bond angle (degree)	1.208
Dihedral angles (degree)	6.605
Residues in Ramachandran plot (%)	
Most allowed region	81.7%
Allowed region	18.3%
Generously allowed region	0

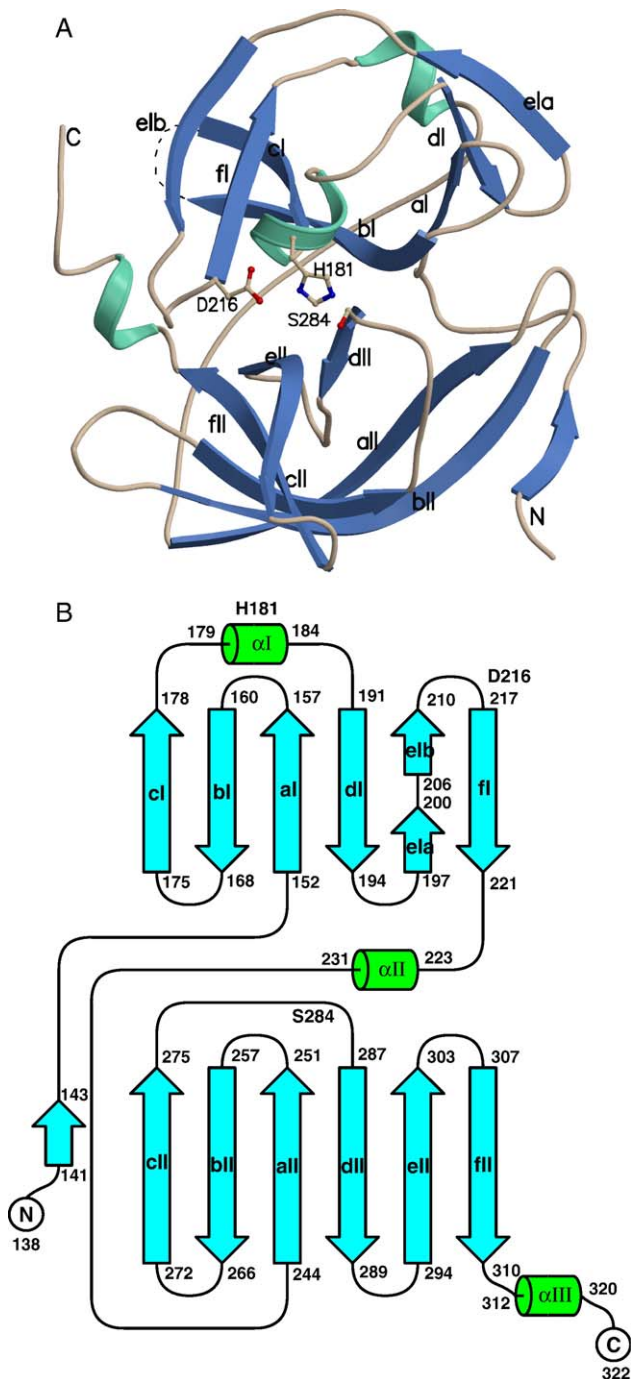


Fig. 2. Polypeptide fold of SeMV protease. (A) Overall fold of the SeMV serine protease. The active site residues H181, D216 and S284 are shown in ball-and-stick representation. The secondary structure elements are numbered according to the convention followed for trypsin proteases. The stretch of residues 171 to 173, which is missing in the model, is shown as broken line. (B) Topology diagram for SeMV protease.

The domain II consists of the strands aII, bII, cII, dII, eII, fII and the C-terminal helix. The strand cII forms part of the wall of the S1-binding pocket. There is an intramolecular disulfide bond connecting strands aII and cII. The loop connecting cII and dII forms the oxyanion hole and contains the active site S284. The polypeptide chain ends in the C-terminal helix formed by residues ranging from 312 to 320.

Intermolecular disulfide bond

The asymmetric unit consists of a monomer of SeMV protease. However, it was observed that there is an intermolecular disulfide bond between C256 of 2-fold symmetry related molecules. This bond was formed despite the presence of 5 mM β -mercaptoethanol in the crystallization buffer. Fig. 3 shows the two molecules connected by the disulfide. This disulfide is likely to be an artifact of crystal packing as gel filtration studies have shown that the protease is a monomer in solution. This is further reflected in the area of the interface formed between the molecules (470.1 \AA^2), which is not very extensive. Hence, the protease can be considered as a monomer in the crystal also, consistent with its behavior in solution. Similar intermolecular disulfide bonds have been observed in the *tobacco etch virus* (TEV) NIa protease structure (Phan et al., 2002) and in a few other cases.

Comparison with other members of the protease family

A comparison of the three-dimensional structure of SeMV protease domain with all the available entries in the Protein Data Bank was carried out using the DALI server (Holm and Sander, 1993). The DALI server identifies significant similarities in the three-dimensional structures of polypeptides irrespective of the sequence similarities between them. The highest z score (17.1) was observed for the heparin-binding protein (1a7s), a serine protease homolog. The list of the top 20 structures from the DALI server output is shown in Table 2A. The z scores and RMSD values suggest that the SeMV protease is closer to the non-viral proteases than to the viral proteases. The structures listed in Table 2A and other viral proteases of known three-dimensional structure (Table 2B) were compared to the SeMV protease by pair wise alignment. The salient features of such a comparison are described below.

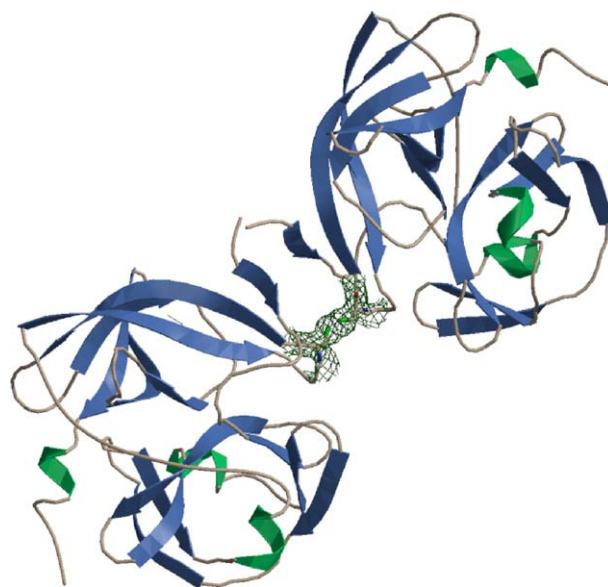


Fig. 3. Intermolecular disulfide bond. The relative orientation of the two symmetry related molecules illustrating the sparse contacts and the disulfide bond between them. The $2F_o - F_c$ electron density contoured at 1.2σ is shown around the disulfide bond.

Table 2A

List of proteins superposed by DALI server and superposition statistics

Protein name	PDB code	Z score	RMSD (Å)	Length aligned	Length of sequence	% identity
Heparin binding protein	1a7s	17.1	2.7	169	221	12
Bovine (<i>Bos taurus</i>) beta trypsin	5ptp	16.9	2.7	165	223	14
<i>Bacillus intermedius</i> glutamyl endopeptidase	1p3c	16.7	2.9	174	215	18
<i>Staphylococcus aureus</i> epidermolytic toxin A	1agj	15.2	3.0	172	242	16
<i>Achromobacter</i> protease	1arb	15.1	3.1	173	263	17
HTRA2 serine protease (Homo sapiens)	1lcy	14.9	3.0	158	296	16
Clr complement serine protease fragment	1md7	14.1	3.0	166	304	10
Complement factor b fragment (<i>Homo sapiens</i>)	1dle	13.1	3.0	164	288	15
Protease DEGP HTRA	1ky9	12.7	3.1	158	311	10
Equine arteritis virus NSP4 protease	1mbm	12.5	3.1	141	198	18
Tobacco etch virus NIa protease	1lvb	12.4	3.7	154	208	14
Hepatitis C virus serine protease	1jxp	12.3	3.0	136	177	11
Hepatitis A virus 3C protease	1hav	12.2	3.2	157	216	15
Rhinovirus 3C protease	1cqq	12.2	3.2	150	180	11
Foot-and-mouth disease 3C protease	2bhg	11.6	3.6	156	184	10
<i>Lysobacter enzymogenes</i> alpha-lytic protease	1qq4	11.4	3.2	149	198	10
Sindbis virus capsid protein	1svp	10.1	3.0	132	160	16
Coronavirus main proteinase	1lvo	10.0	3.1	141	299	17
Dengue virus NS3 serine protease	1bef	10.0	3.0	132	177	13
Human rhinovirus 2 2A cysteine protease	2hrv	7.1	2.2	66	139	14

A helix is present at the N-terminus in all the viral proteases except the capsid-forming proteases of Sindbis, Semilki forest and Venezeulan equine encephalitis viruses (1svp, 1vcp and 1ep5 respectively). This N-terminal helix is not part of the canonical trypsin fold and is absent in the case of SeMV protease. The N-terminal residues of the SeMV protease pack against the bII strand of domain II and have characteristics of β -strand-forming residues. Main chain O and N of S141 form hydrogen bonds with the backbone atoms of the bII strand residue S259. This conformation of the N-terminus agrees with those of the cellular and bacterial proteases namely heparin-binding protein (1a7s), *Bacillus intermedius* glutamyl endopeptidase (1p3c), bovine beta trypsin (5ptp) and *Staphylococcus aureus* epidermolytic toxin A (1agj).

The rhino and polio virus proteases 1cqq and 1l1n harbor the conserved RNA-binding motif KFRDIR (Mosimann et al., 1997) in the helical segment occurring in the inter-domain loop. Although this motif is absent in SeMV, the residues H225, K229 and K233 constitute a positively charged patch in

this region. However, the RNA-binding properties of SeMV protease have not been investigated.

Another region, which shows considerable variability among the different families of serine proteases, is the loop connecting bII and cII. In many of the viral serine-like cysteine proteases of picornaviruses (1cqq, 1l1n, 1hav and 2hrv), this loop has a long insertion that forms a β -hairpin, which extends to the substrate-binding site. It contributes to the peptide binding and shields the active site from the solvent. The peptide substrate binds as a bridging strand between eII and the β -hairpin. SeMV protease does not have this insertion, and this leads to an exposed active site. This β -hairpin is not present in most of the non-viral proteases. The loop is also not found in the viral serine proteases equine arteritis virus NSP4 protease, hepatitis C virus protease (1mbm and 1jxp, respectively) and the capsid-forming proteases (1ep5, 1vcp and 1svp), suggesting that an exposed active site is a characteristic of the viral serine proteases.

The active site

The active site residue H181 forms hydrogen bonds with both S284 and D216, the other two residues of the catalytic triad. The active site and the hydrogen bonds involved are shown in Fig. 4A. The side chain of D216 is stabilized by two hydrogen bonds—between its O δ 1 and the main chain N of H181, and its O δ 2 and N δ 1 of H181. The N ϵ 2 of H181 is at a hydrogen bonding distance from O γ of S284, although the geometry for the hydrogen bond is not optimal. The electron density for the imidazole ring of the H181 is not very clear, implying conformational flexibility. This is a feature found in the active site histidines of other serine proteases also. In the HRV protease, the active site H40 exists in alternate conformations (Petersen et al., 1999), while in the polio viral 3C protease, the electron density for the side chain is weak, and the mean temperature factor of the atoms is above the average (Mosimann et al., 1997). The oxyanion hole, a highly

Table 2B

List of viral protease structures of trypsin fold

Protein name	PDB code	Active site residue
Human rhinovirus 2 2A cysteine protease	2hrv	Cys
Tobacco etch virus NIa protease	1lvb	Cys
Hepatitis A virus 3C protease	1hav	Cys
Rhinovirus 3C protease	1cqq	Cys
Foot-and-mouth disease 3C protease	2bhg	Cys
Gastroenteritis virus main proteinase	1lvo	Cys
Human corona virus main proteinase	1p9s	Cys
SARS corona virus	1q2w	Cys
Poliovirus 3C protease	1l1n	Cys
Equine arteritis virus NSP4 protease	1mbm	Ser
Semilki Forest Virus	1vcp	Ser
Sindbis virus capsid protein	1svp	Ser
Dengue virus NS3 serine protease	1bef	Ser
Venezeulan equine encephalitis virus	1ep5	Ser
Hepatitis C virus serine protease	1jxp	Ser

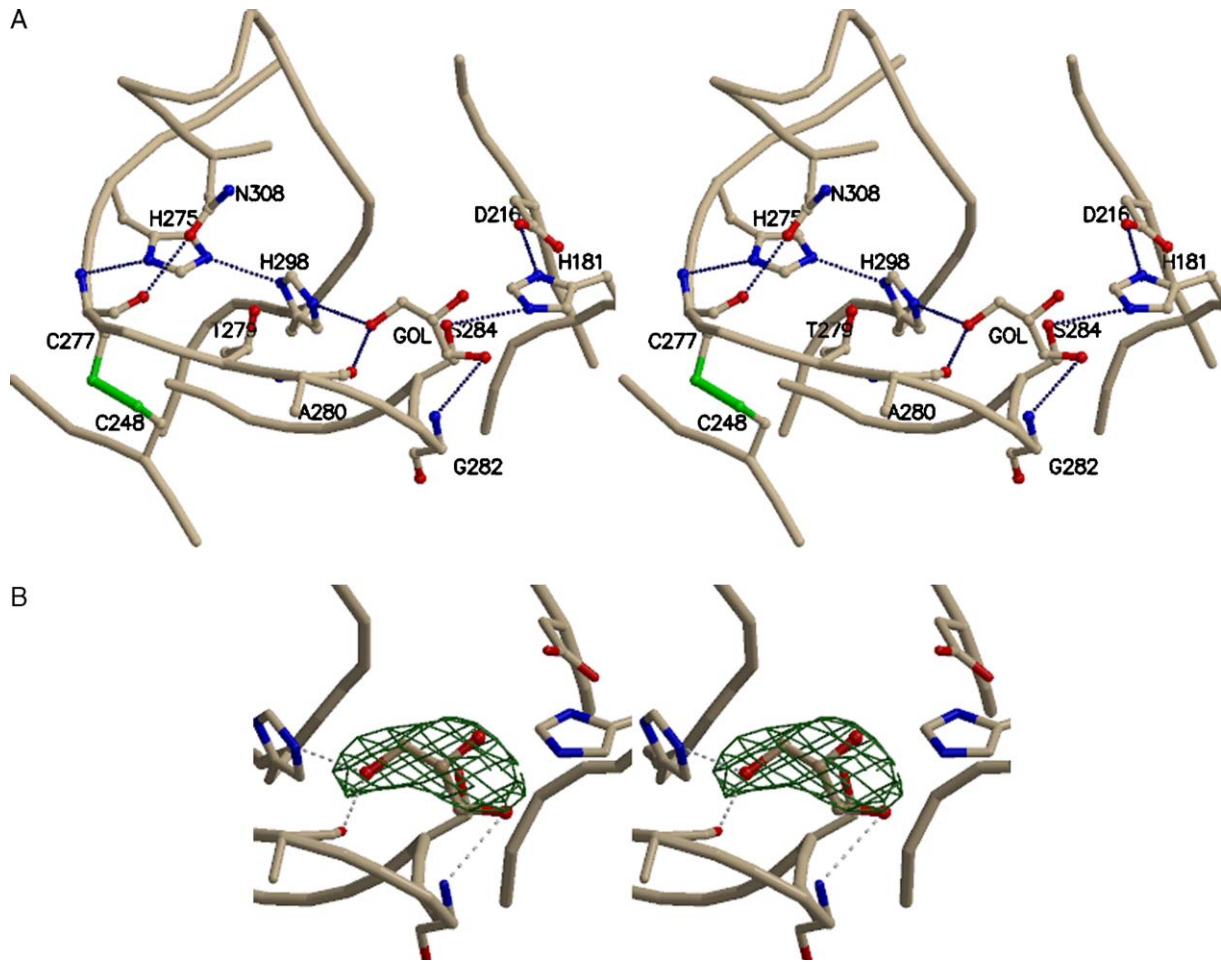


Fig. 4. The active site and the S1-binding pocket of SeMV protease. (A) A stereo view of the active site and the S1 region of the substrate-binding site. The bound glycerol is labeled as GOL. The hydrogen bonds are shown as broken lines. (B) The 2Fo–Fc electron density contoured at 1.0σ for the bound glycerol is shown, along with the hydrogen bonds.

conserved region of the active site in terms of sequence (GXSG) and structure, is important for the stabilization of the oxyanion intermediate formed during the reaction. In SeMV protease, the main chain amide nitrogens of G282 and S284 are involved in the oxyanion hole formation. The structure is in conformity with the earlier mutational studies on the active site residues of SeMV protease (Satheshkumar et al., 2004).

Substrate specificity

Viral proteases involved in polyprotein processing are known to have very stringent substrate specificities. Mutational analyses have demonstrated that the serine protease domain of SeMV cleaves at E–T occurring between protease–VPg, and VPg–p10 domains, and also at E–S between p10 and RdRP (Satheshkumar et al., 2004). In order to understand the structural basis of the stringent cleavage specificity of the enzyme, a comparison was carried out between SeMV protease and the structures of other proteases that display glutamate specificity (Table 2C). It was observed that the histidine that binds to the carbonyl oxygen of Glu/Gln is conserved in all Glu/Gln-specific proteases except the *S. griseus* proteases (2sga and 2sgp) where a positively charged arginine at the base of the substrate-binding

pocket stabilizes the glutamate (Read et al., 1983; Read and James, 1988). The corresponding residue in SeMV protease is H298, and it is directed towards the active site. N δ 1 of H298 forms a hydrogen bond with N ϵ 2 of H275 (Fig. 4A). The conformation of H275 side chain is, in turn, stabilized by a hydrogen bond from N δ 1 to the main chain N of the disulfide-bonded C277. In the glutamate-specific *S. griseus* protease (1hpg), the role of the histidine triad, H213, H199 and H228 (1hpg numbering), in stabilizing the charge on glutamic acid has been emphasized (Nienaber et al., 1993). In SeMV protease, a corresponding histidine triad does not exist, and H275 hydrogen

Table 2C
List of Glutamic acid specific protease structures

Protein name	PDB code
<i>Streptomyces griseus</i> protease	1hpg
<i>Bacillus intermedius</i> glutamyl endopeptidase	1p3c
<i>Staphylococcus aureus</i> epidermolytic toxin A	1agj
<i>Streptomyces griseus</i> protease A	2sga
<i>Streptomyces griseus</i> protease B	2sgp
<i>Staphylococcus aureus</i> V8 protease	1wcz
Equine arteritis virus NSP4 protease	1mbm
Foot-and-mouth disease 3C protease	2bhg

bonded to the conserved H298 is not in the same position as that of H199 in the *S. griseus* protease. However, the chain of hydrogen bonds is maintained. This can possibly help in the stabilization of the charge of the glutamic acid side chain.

The residues present in the proposed glutamate-binding site or the S1 pocket are T279, A280, H298, F301 and N308 (Fig. 4A). Out of these residues, only H298 and T279 are conserved across most of the Glu/Gln-binding proteins. In a few cases, Thr is replaced by Ser, whose O γ substitutes for the Thr O γ . His and Thr have been implicated to have a major role in Glu/Gln binding, and the mutations of these have resulted in inactive enzymes in poliovirus 3C protease (Ivanoff et al., 1986). In the crystal structure of TEV complexed with substrate, it has been observed that the carboxamide group

of glutamine forms hydrogen bonds with both His and Thr (Phan et al., 2002).

In the present structure, electron density observed near the oxyanion hole has been interpreted as a glycerol molecule. Fig. 4B shows the glycerol bound near the S1-binding pocket with the 2F_o–F_c electron density around it contoured at 1.0 σ level. It is possible that the orientation of the glycerol is related to the mode of glutamate binding. The main chain carboxyl group of Glu should be directed towards the amide nitrogens of Gly and Ser at the oxyanion hole (Perona and Craik, 1995). Accordingly, the O3 of glycerol forms hydrogen bond with the main chain N of G282 (Fig. 4). At the other end, O1 forms hydrogen bonds with both N ϵ 2 of H298 and main chain O of A280. A water molecule, Wat28, is present between H298 and the

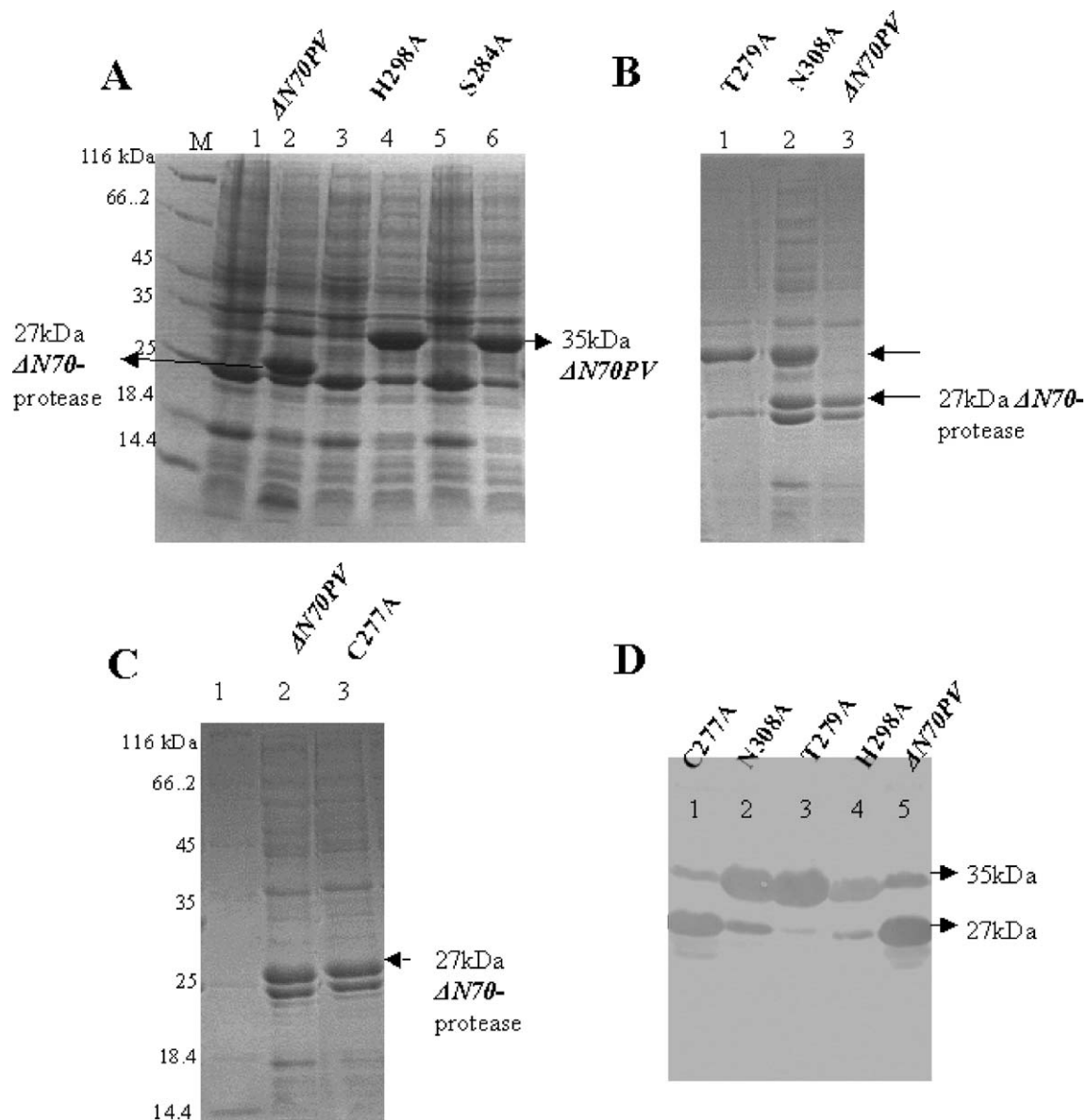


Fig. 5. Mutational analysis of $\Delta N70PV$ fusion protein. The expressed proteins upon induction with 0.3mM IPTG were analyzed on 12% SDS-PAGE. (A) Lanes 1 and 2—uninduced and induced $\Delta N70PV$, lanes 3 and 4—uninduced and induced $\Delta N70PV$ -H298A, lanes 5 and 6—uninduced and induced $\Delta N70PV$ -S284A. (B) Lane 1—induced $\Delta N70PV$ -T279A, lane 2—induced $\Delta N70PV$ -N308A, lane 3—induced $\Delta N70PV$. (C) Lane 1—induced $\Delta N70PV$ -C277A, lane 2—induced $\Delta N70PV$. (D) Western blot analysis of $\Delta N70PV$ mutants using anti-protease antibodies. Lane 1— $\Delta N70PV$ -C277A, lane 2— $\Delta N70PV$ -N308A, lane 3— $\Delta N70PV$ -T279A, lane 4— $\Delta N70PV$ -H298A, lane 5— $\Delta N70PV$.

glycerol. This water molecule is close to the aromatic ring of F301 and comes within hydrogen bonding distance from H298 and T279 side chain atoms. However, F301 aromatic ring and the water molecule are not in well-defined density.

F301 is in a position equivalent to S170 of TEV protease and S137 of equine arteritis virus NSP4 protease. These serine residues are involved in substrate binding at the P1 site (Phan et al., 2002; Barrette-Ng et al., 2002). Further, all Glu/Gln-binding proteases, except SeMV protease, have either a Ser or Gly at this position. The absence of well-defined density for F301 in SeMV protease suggests that its side chain might undergo substantial displacement on binding of the substrate or on conformational changes induced by the interaction of the protease domain with VPg. TEV and equine arteritis virus proteases, in which a serine residue occurs at the position corresponding to F301, are active in *trans*. Mutation of this F301 to serine, the corresponding residue in other Glu/Gln-binding proteases, might therefore bestow activity in *trans* to SeMV protease, even in the absence of VPg domain.

An extended C-terminus results in a buried cleft in the case of TEV protease. The absence of this extension makes SeMV substrate-binding site significantly different from that of TEV protease. The residues in this region constitute the specificity pocket in TEV protease. In SeMV protease, the cleft is solvent exposed, and the loop 212–215 appears to be closest to the substrate-binding region. The other side of the substrate-binding cleft is similar in TEV and SeMV proteases and is lined by the β strand eII. This is a common feature in most of the proteases and mainly involves main chain interactions.

Mutational analysis

As mentioned earlier, the Δ N70-Pro or Δ N92-Pro was not active in *trans*. Hence, in order to confirm the role of the proposed glutamate-binding residues (Fig. 4A) in protease activity, the residues H298, T279 and N308 were mutated to Ala in Δ N70-protease–VPg (Δ N70PV) fusion protein, and the *cis*-cleavage of the expressed protein into protease and VPg was monitored as described in the Materials and methods section. The results obtained were confirmed by a Western blot analysis using anti-protease antibodies (Fig. 5D).

Expression of Δ N70PV showed a band of size 27 kDa corresponding to the protease domain (Fig. 5A, lane 2; Fig. 5D, lane 5) confirming that nearly complete cleavage of Δ N70PV (35 kDa) had occurred. The active site mutant Δ N70PV-S284A (Satheshkumar et al., 2004) showed a prominent band corresponding to 35 kDa on expression (Fig. 5A, lane 6). The H298A mutant gave a band of size 35 kDa corresponding to the Δ N70PV (Fig. 5A, lane 4; Fig. 5D, lane 4), confirming that the mutation had affected the cleavage activity of Δ N70PV. Similarly T279A mutant also did not show any *cis*-cleavage activity (Fig. 5B, lane 1; Fig. 5D, lane 3). However, N308 to A mutation in the Δ N70PV resulted only in a partial loss of the protease activity (Fig. 5B, lane 2; Fig. 5D, lane 2). The mutant Δ N70PV-N308A partially cleaved the Δ N70PV giving a 35-kDa Δ N70PV band and a 27-kDa Δ N70-protease band, implying that N308 is not absolutely essential for the substrate

binding, but contributes to the proper binding of the glutamate residue and hence to the optimal activity.

Role of the intramolecular disulfide bond

Disulfides are generally implicated in the stability of the three-dimensional structures of serine proteases (Wang et al., 1997). Most of the prokaryotic and eukaryotic trypsins have highly conserved disulfide bonds. On the contrary, disulfide bonds have not been observed in any of the viral protease structures reported so far. In the poliovirus and hepatitis 3C virus proteases (111n and 1jxp), binding of a metal ion contributes to the structural integrity (De Francesco et al., 1996).

In SeMV protease, C248 forms a disulfide bond with C277. This disulfide bond connects the β strands aII and cII in domain II (Fig. 4A). It holds the walls of the S1 specificity pocket and could have a role in the maintenance of its rigid conformation. A disulfide bond is responsible for the rigidity of S1-binding pocket in many of the mammalian trypsins. Though the disulfide bond in SeMV protease does not correspond to any of the conserved disulfide bonds in trypsin, it involves the cII strand of domain II. The same strand in trypsin has the conserved disulfide between the residues 198 and 220 (Wang et al., 1997).

In order to assess the role of the unique disulfide bond in SeMV protease in the maintenance of the S1-binding pocket, C277 was mutated to alanine in Δ N70PV, and the effect of the mutation on *cis*-cleavage was examined. As shown in Fig. 5C, lane 2, and Fig. 5D, lane 1, the 35-kDa Δ N70PV-C277A mutant was completely cleaved in *cis* to 27-kDa band like the wild-type Δ N70PV (Fig. 5A, lane 1; Fig. 5D, lane 5). Hence, it can be concluded that this disulfide bond is not essential for the maintenance of the rigidity of the S1-binding pocket.

Presence of exposed aromatic residues

An interesting observation in the crystal structure of SeMV protease is the occurrence of a stretch of aromatic residues exposed to the surface (Fig. 6). These aromatic residues F269, W271, Y315 and Y319 are not consecutive in sequence but form a stack near the C-terminus of the protein. Two of these residues form part of the C-terminal helix. These residues may be of functional significance to the polyprotein. SeMV residues F269 and W271 are conserved across genomes of sobemoviruses. The presence of exposed aromatic residues is believed to indicate a protein–protein interaction interface. It has been demonstrated that there are extensive interactions between protease and VPg, and the conformational changes that accompany these interactions enhance and regulate protease activity (Satheshkumar et al., 2005). W43 of VPg was shown to mediate aromatic interactions with the protease, which results in a positive peak at 230 nm in the CD spectrum of protease–VPg fusion protein, but the interacting partner in protease was not identified. The presence of exposed aromatic residues in the present structure strongly suggests that this might be the site of protease–VPg interaction. It is probable that W271 of protease interacts with W43 of VPg domain. The positive peak observed at 230 nm in the CD spectrum of protease–VPg fusion protein might be the

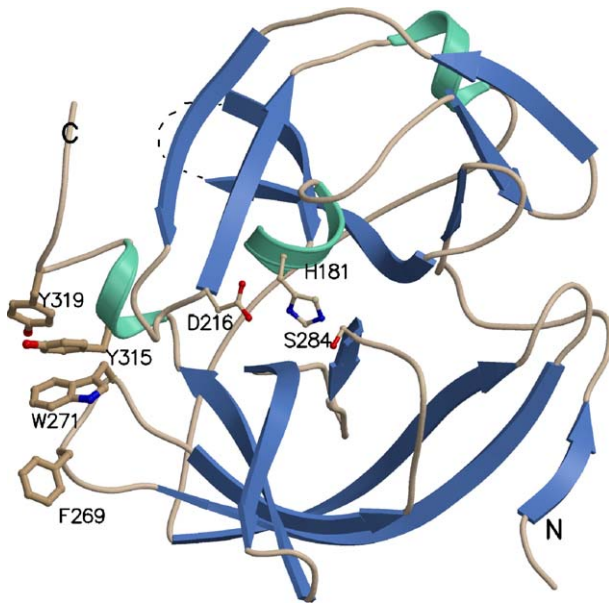


Fig. 6. Exposed aromatic residues on the surface of SeMV protease. The residues involved are labeled. The missing residues in the model are shown as broken line.

result of this interaction. Further mutational studies are required to delineate the significance of the aromatic stretch of residues in protease activity.

Polyprotein processing

The crystal structure of SeMV protease provides insights on the possibilities of *cis* (intramolecular)- or *trans* (intermolecular)-cleavages. The C-terminal helix ends at residue 320, and this is followed by a stretch of five residues. The helix is a stable structure and is closely packed against the rest of the protein. The presence of a C-terminal helix is a characteristic of many serine proteases. The disordered short segment at the C-terminus is not long enough to reach the active site suggesting the possibility of intermolecular cleavage between protease and VPg (*trans*-cleavage). However, it is possible that there could be a major conformational change because of the presence of the natively unfolded VPg at the C-terminus, and this may position the residues for *cis*-cleavage at the active site.

Biochemical evidences (data not shown) as well as the present structure suggest that a cleavage occurs between A134 and V135 at the N-terminus. This is an unexpected finding, as this site does not correspond to the canonical site for substrate cleavage. The relative positions of the N-terminus and the active site suggest that an intramolecular cleavage is indeed possible. The N-terminus in the crystal structure is disordered, and clear electron density is observed only from S138. A further extension of approximately four residues is long enough to approach the active site through the cleft below the bII-cII loop of domain II. This gains additional support by the observation that the N-terminus of SeMV protease superposes very well with the glutamyl endopeptidase from *Bacillus intermedius* (1p3c), as shown in Fig. 7. In the case of 1p3c, the N-terminus extends to the active site, and this extension is essential for zymogen activation and charge compensation for

glutamate specificity (Meijers et al., 2004). Therefore, based on the relative positions of the N- and C-termini and the active site in the crystal structure, it can be concluded that the cleavage between protease and VPg could occur in *cis* or *trans*, while the N-terminal cleavage could be *cis*.

The non-specific cleavage at Ala-Val could be due to the conformation of the polypeptide, which positions the A134–V135 peptide bond optimally for the cleavage. The residues Ala and Val are small and could be accommodated in the active site without steric hindrance. These observations also suggest that the specificity of the protease depends not only on sequence but also on the conformation of the polypeptide. A non-specific cleavage on the polypeptide chain has been observed at the C-terminus of TEV protease also (Nunn et al., 2005). There have also been reports in the literature on other viral proteases such as rhinovirus HRV2-2A protease, where the specificity requirements for *cis*-cleavage are less stringent than for *trans*-cleavage. The entropy term for substrate binding is favorable in *cis*-cleavage, and hence, the specificity requirements are less stringent (Petersen et al., 1999). In spite of this, most viral proteases are extremely specific, and they function mainly in the processing of the respective polyproteins. Their activities are modulated by the presence of VPg domains either at the N- or C-terminus. Determination of the structure of the protease with the VPg domain could shed light on the molecular mechanism of such regulations.

Materials and methods

Protein purification and crystallization

The ΔN70-protease (ΔN70-Pro) and ΔN92-protease (ΔN92-Pro) domains of the polyprotein were cloned in pRSET C vector

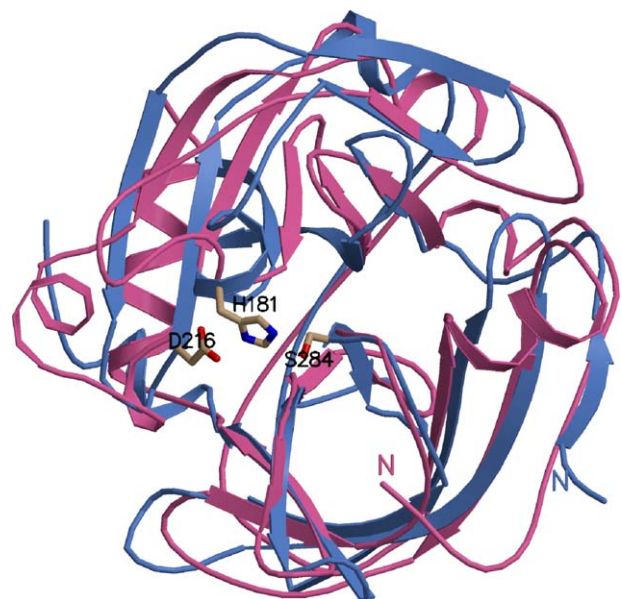


Fig. 7. Superposition of SeMV protease (blue) and *Bacillus intermedius* glutamyl endopeptidase (pink). The superposition was generated using the output from DALI server.

at NheI and BamHI sites. The cloning strategy resulted in the addition of 11 amino acids from the vector sequence at the N-terminus including the hexa-histidine residues inserted for affinity purification. The proteins were expressed in *E. coli* BL21 (DE3) pLysS cells by the addition of 0.3 mM IPTG at a culture density 0.6 OD at 660 nm and purified using Ni-NTA affinity chromatography (Novagen) as previously described (Satheshkumar et al., 2004). The final protein preparation was in 50 mM Tris, pH 8.0, buffer containing 200 mM NaCl.

Both Δ N70-Pro and Δ N92-Pro domains formed small crystals in the presence of 0.2 M Tris, pH 8.0, 0.2 M ammonium sulphate, 0.6 M 1,6-hexane diol, 5 mM β -mercaptoethanol and 4% glycerol (crystallization buffer). Larger crystals could be obtained using the microbatch method when a mixture of silicone and mineral oils in the ratio 2:3 was layered over the crystallization droplets (5 μ l of 30 mg/ml protein and 3 μ l of crystallization buffer). The crystals appeared from a layer of precipitate within 2 weeks and grew to a size of approximately $0.3 \times 0.3 \times 0.2$ mm³ during a period of 1 month.

Data collection and processing

The crystals were transferred to the crystallization buffer containing 20% ethylene glycol as the cryoprotectant for a few seconds and then mounted in a cryo-loop. The crystals were exposed to Cu-K α radiation at liquid nitrogen temperature (100 K). X-ray diffraction data were collected using a rotating anode X-ray generator and a MAR Research image plate detector system. The datasets were processed using *DENZO*, and the resulting intensities were scaled using *SCALEPACK* (Otwinowsky, 1997). The frames were processed in the space group P3₁21. Systematic absences indicated the presence of a 3₁ or 3₂ screw along the c axis.

Structure determination

The structure solution was attempted by multiple isomorphous replacement (MIR). Four isomorphous mercury derivatives were obtained by soaking the crystals in *p*-chloro mercury benzene sulfonate (PCMBs), mercuric iodide, mercuric chloride and thiomersal (ethyl mercury thio salicylic acid or EMTS). The native and derivative datasets were scaled using the *SCALEIT* program in the CCP4 program suite version 4.2.1 (CCP4, 1994). Both isomorphous and anomalous difference Patterson maps

were calculated for these derivatives using CCP4 program suite and Harker sections for the space group P3₁21 were plotted.

Initially, one site in the PCMBs derivative was obtained using the isomorphous differences in the program suite *SOLVE* version 2.03 (Terwilliger, 2004). The rest of the sites in all the derivatives were identified from the difference Fourier maps using *MLPHARE* in CCP4 program suite (1994). The fractional coordinates of all the sites, their occupancies and anomalous occupancies were refined in *MLPHARE* for all the derivatives to a resolution of 2.8 Å. The B-factors of the sites were not refined and were fixed at the values obtained from the Wilson plots. The space group was confirmed to be P3₁21, as this choice resulted in an interpretable map with the correct handedness.

The experimental phases obtained from isomorphous and anomalous signals were used in the program *RESOLVE* (Terwilliger, 2004) to perform phase extension to 2.3 Å by density modification followed by automated model building. This resulted in the placement of 34 residues with side chains and 123 without side chains into the electron density map. The coordinates of the partial model and the phases from *RESOLVE* were used as input for *ARP/wARP* version 6.1 (Morris et al., 2003) for building the rest of the model. 80% of the residues were built by the program, which included residues 140–161, 174–194, 205–250, 257–303 and 306–323. The rest of the residues were built manually in subsequent cycles of refinement using *COOT* (Emsley and Cowtan, 2004) and *REFMAC* 5.2 (Murshudov et al., 1997). The progress of the refinement was monitored by using 5% (459) of the total 9272 independent reflection measurements for the calculation of the free R-factor. The Hendrickson–Lattman coefficients were used as restraints during the refinement (Skubak et al., 2004). Individual B-factors of non-hydrogen atoms were refined.

All structural alignments were done using the output from DALI server (Holm and Sander, 1993). The figures were prepared using *MOLSCRIPT* (Kraulis, 1991) and *BOBSCRIPT* (Esnouf, 1999) and rendered using *Raster3D* (Merritt and Murphy, 1994). The topology diagram for the protease was prepared using *TOPDRAW* (Bond, 2003).

Mutational analysis of Δ N70-protease–VPg (Δ N70PV) fusion protein

Site directed mutagenesis was performed by PCR-based approach (Weiner et al., 1994). The sense and antisense primers were designed with desired changes in the nucleotides (Table 3),

Table 3
List of primers used for cloning and mutagenesis

Name	Sequence	Description
H298A-s	AGAGGCGTGGT <u>CGAC</u> ATGGCCGTCGGTTTTGGAG	Primers used for mutation of H298 to A. The created SalI site is underlined
H298A-a	CTCAAAACCGACGGCCATGTCGACCACGCCTCT	
T279A-s	CTTGAAAATTGACCCACACCTGTCCTGCAGCTGCC	Primers used for mutation of T279 to A. The abolished HinCII site is underlined
T279A-a	GGCAGCTGCAGGACAGGTGTGGGTCAATTTCCAG	
N308A-s	GAAATAGGAAA <u>ACTCGCCCGGGT</u> GTAACATG	Primers used for mutation of N308 to A. The SmaI site created is underlined
N308A-a	CATGTTTACACCCGGGCGAGTTTTCCTATTTC	
C277A-s	GGAAGTTGACCCACACGGCGCTACAGCTGCC	Primers used for mutation of C277 to A. The Ehel site created is underlined
C277A-a	GGCAGCTGTAGGCGCCGTGTGGGTCAACTTCC	

and restriction sites were incorporated in the primers to enable easy screening. The PCR was carried out using Deep Vent polymerase (New England Biolabs) according to the manufacturer's instructions. The PCR product was treated with 0.5 µl DpnI enzyme at 37 °C for 45 min to remove the template DNA and transformed into 2 aliquots of DH5α competent cells. The cells were plated after transformation in antibiotic containing plates. The colonies obtained were inoculated separately. Plasmid isolation was carried out and digested with the appropriate restriction enzyme. The mutants screened by restriction enzyme digestion were further confirmed for the presence of mutation by DNA sequencing.

E. coli BL21 (DE3) pLysS cells were transformed with the recombinant clones, and the proteins were expressed by induction with 0.3 mM IPTG for 4–5 h at 30 °C. The cell pellet was resuspended in buffer A (50 mM Tris–HCl, pH 8.0, 300 mM NaCl, 0.2% Triton X100, 5% Glycerol), sonicated and the expression was checked by SDS polyacrylamide gel electrophoresis. The cleavage activity was monitored by the appearance of bands of expected size on the SDS polyacrylamide gel. The results were confirmed by Western blot analysis as described in Satheshkumar et al. (2004).

Protein data bank accession number

The coordinates and structure factors for SeMV protease have been submitted to the Protein Data Bank, and the structure has been assigned the accession code 1ZYO.

Acknowledgments

MRNM and HSS thank the Department of Science and Technology (DST) and the Department of Biotechnology (DBT) of the Government of India for financial support. Diffraction data were collected at the X-ray facility for Structural Biology at Molecular Biophysics Unit, Indian Institute of Science (IISc), supported by DST and DBT. We thank the staff in the X-ray laboratory and Supercomputer Education and Research Centre of IISc for their co-operation during the course of these investigations. PG and PSS acknowledge the Council for Scientific and Industrial Research, Government of India, for the fellowships.

References

- Barrette-Ng, I.H., Ng, K.K., Mark, B.L., Van Aken, D., Cherney, M.M., Garen, C., Kolodenco, Y., Gorbalenya, A.E., Snijder, E.J., James, M.N., 1994. Structure of arterivirus nsp4. The smallest chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. *J. Biol. Chem.* 277, 39960–39966.
- Bond, C.S., 2003. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* 19 (2), 311–312.
- CCP4 (Collaborative Computational Project, Number 4), 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., D: Biol. Crystallogr.* 50, 760–763.
- De Francesco, R., Urbani, A., Nardi, M.C., Tomei, L., Steinkuhler, C., Tramontano, A., 1996. A zinc binding site in viral serine proteinases. *Biochemistry* 35 (41), 13282–13287.
- Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60 (Pt. 12 Pt. 1), 2126–2132.
- Esnouf, R.M., 1999. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 55 (Pt. 4), 938–940.
- Gorbalenya, A.E., Koonin, E.V., Blinov, V.M., Donchenko, A.P., 1988. Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Lett.* 236 (2), 287–290.
- Holm, L., Sander, C., 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233 (1), 123–138.
- Ivanoff, L.A., Towatari, T., Ray, J., Korant, B.D., Petteway Jr., S.R., 1986. Expression and site-specific mutagenesis of the poliovirus 3C protease in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 83 (15), 5392–5396.
- Kraulis, P.J., 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950.
- Kriminski, S., Caylor, C.L., Nonato, M.C., Finkelstein, K.D., Thorne, R.E., 2002. Flash-cooling and annealing of protein crystals. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 58 (Pt. 3), 459–471.
- Lokesh, G.L., Gopinath, K., Satheshkumar, P.S., Savithri, H.S., 2001. Complete nucleotide sequence of *Sesbania mosaic virus*: a new virus species of the genus Sobemovirus. *Arch. Virol.* 146 (2), 209–223.
- Matthews, B.W., 1968. Solvent content of protein crystals. *J. Mol. Biol.* 33 (2), 491–497.
- Meijers, R., Blagova, E.V., Levnikov, V.M., Rudenskaya, G.N., Chestukhina, G.G., Akimkina, T.V., Kostrov, S.V., Lamzin, V.S., Kuranova, I.P., 2004. The crystal structure of glutamyl endopeptidase from *Bacillus intermedius* reveals a structural link between zymogen activation and charge compensation. *Biochemistry* 43 (10), 2784–2791.
- Merritt, E.A., Murphy, M.E., 1994. Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 50 (Pt. 6), 869–873.
- Morris, R.J., Perakis, A., Lamzin, V.S., 2003. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* 374, 229–244.
- Mosimann, S.C., Cherney, M.M., Sia, S., Plotch, S., James, M.N., 1997. Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* 273 (5), 1032–1047.
- Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 53 (Pt. 3), 240–255.
- Nienaber, V.L., Breddam, K., Birktoft, J.J., 1993. A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. *Biochemistry* 32 (43), 11469–11475.
- Nunn, C.M., Jeeves, M., Cliff, M.J., Urquhart, G.T., George, R.R., Chao, L.H., Tsuchia, Y., Djordjevic, S., 2005. Crystal structure of tobacco etch virus protease shows the protein C terminus bound within the active site. *J. Mol. Biol.* 350, 145–155.
- Otwinowsky, Z.M.W., 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326.
- Perona, J.J., Craik, C.S., 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci.* 4 (3), 337–360.
- Petersen, J.F., Cherney, M.M., Liebig, H.D., Skern, T., Kuechler, E., James, M.N., 1999. The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *EMBO J.* 18 (20), 5463–5475.
- Phan, J., Zdanov, A., Evdokimov, A.G., Tropea, J.E., Peters III, H.K., Kapust, R.B., Li, M., Wlodawer, A., Waugh, D.S., 2002. Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* 277 (52), 50564–50572.
- Read, R.J., Fujinaga, M., Sielecki, A.R., James, M.N., 1983. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8-Å resolution. *Biochemistry* 22 (19), 4420–4433.
- Read, R.J., James, M.N., 1988. Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution. *J. Mol. Biol.* 200 (3), 523–551.
- Satheshkumar, P.S., Lokesh, G.L., Savithri, H.S., 2004. Polyprotein processing:

- cis* and *trans* proteolytic activities of *Sesbania mosaic virus* serine protease. *Virology* 318 (1), 429–438.
- Satheshkumar, P.S., Gayathri, P., Prasad, K., Savithri, H.S., 2005. “Natively Unfolded” VPg is essential for *Sesbania mosaic virus* serine protease activity. *J. Biol. Chem.* 280 (34), 30291–30300.
- Sivakumaran, K., Fowler, B.C., Hacker, D.L., 1998. Identification of viral genes required for cell-to-cell movement of southern bean mosaic virus. *Virology* 252 (2), 376–386.
- Skubak, P., Murshudov, G.N., Pannu, N.S., 2004. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60 (Pt. 12 Pt. 1), 2196–2201.
- Terwilliger, T., 2004. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* 11 (Pt. 1), 49–52.
- Wang, E.C., Hung, S.H., Cahoon, M., Hedstrom, L., 1997. The role of the Cys191-Cys220 disulfide bond in trypsin: new targets for engineering substrate specificity. *Protein Eng.* 10 (4), 405–411.
- Weiner, M.P., Costa, G.L., Schoettlin, W., Cline, J., Mathur, E., Bauer, J.C., 1994. Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene* 151 (1–2), 119–123.
- Wellink, J., van Kammen, A., 1988. Proteases involved in the processing of viral polyproteins. Brief review. *Arch. Virol.* 98 (1–2), 1–26.