FEBS Letters 580 (2006) 1502–1508

Characterization of the cysteine protease domain of Semliki Forest virus replicase protein nsP2 by in vitro mutagenesis

Andrey Golubtsov, Leevi Kääriäinen, Javier Caldentey*

Institute of Biotechnology, Biocenter Viikki, P.O. Box 56, 00014 University of Helsinki, Finland

Received 6 December 2005; revised 15 January 2006; accepted 23 January 2006

Available online 31 January 2006

Edited by Judit Ovádi

Abstract The function of Semliki Forest Virus nsP2 protease was investigated by site-directed mutagenesis. Mutations were introduced in its protease domain, Pro39, and the mutated proteins were expressed in *Escherichia coli*, purified and their activity in vitro was compared to that of the wild type Pro39. Mutations M781T, A662T and G577R, found in temperature-sensitive virus strains, rendered the enzyme temperature-sensitive in vitro as well. Five conserved residues were required for the proteolytic activity of Pro39. Changes affecting Cys^{478}, His^{548}, and Trp^{549} resulted in complete inactivation of the enzyme, whereas the replacements N600D and N605D significantly impaired its activity. The importance of Trp^{549} for the proteolytic cleavage specificity is discussed and a new structural motif involved in substrate recognition by cysteine proteases is proposed.

Keywords: Alphavirus; Semliki Forest virus; nsP2 protease; Temperature-sensitive mutants; Glycine specificity motif; Cysteine protease

1. Introduction

Semliki Forest virus (SFV) is an enveloped positive-strand RNA virus belonging to the Alphavirus genus of the Togaviridae family. The structure and replication of alphaviruses have been studied in detail (reviewed in [1,2]). The virus has been used as an important tool in studies of protein folding [3,4], intracellular membrane transport and endocytosis [5,6] and viral pathogenesis [7]. The SFV based replicons have been used as expression vectors for the production of recombinant proteins in eukaryotic cells [8]. Attempts to use them for the production of vaccines and in gene and cancer therapies have also been reported [9].

Upon infection of the host cell, the 5′ two-thirds of the SFV 42S RNA genome is translated into a 2432 amino acid-long polyprotein, designated P1234, which is autocatalytically processed to yield non-structural proteins nsP1–nsP4. All these function as virus-specific components of the membrane-associated RNA polymerase [2,10]. The processing intermediates P123 plus nsP4 are needed for the synthesis of complementary 42S RNA early in infection [11,12], while the complete cleavage products are responsible for the synthesis of positive sense 42S RNA genomes and the subgenomic 26S mRNA [2,11,12].

The protease activity responsible for the non-structural polyprotein cleavage resides in the C-terminal domain of nsP2 protein [13–15]. It belongs to the papain-like peptidase type (C9 family of CA clan in the MEROPS database [16]). The papain-related proteases have little sequence similarity but they share some biochemical and structural properties [17]. We have recently isolated the protease domain of SFV nsP2 (Pro39) and tested its activity using model substrates containing short sequences from each of the P1234 polyprotein processing sites fused to thioredoxin protein (Trx12, Trx23 and Trx34) [15].

Several temperature sensitive (*ts*) mutants of SFV have previously been characterized in our laboratory [18,19]. Temperature-shift experiments performed with these mutants have provided invaluable information on the biology of SFV, particularly on RNA synthesis and the function of the individual non-structural proteins [20–23]. Sequence analysis of the cDNA derived from the genome of those *ts* mutants has shown that *ts4*, *ts6*, and *ts11* mutations are the result of single aminoacid changes in the protease domain of nsP2, M781T, A662T and G577R, respectively [23,24]. Despite these substitutions being located in poorly conserved regions of the nsP2 protease, they cause functional defects in virus propagation. The *ts4* mutation has been shown to result in the halt of viral RNA synthesis and processing of the non-structural polyprotein at the restrictive temperature of 39 °C [23,25,26]. The *ts6* and *ts11* mutants also displayed an RNA negative phenotype, failing to synthesize any viral RNA at 39 °C [18].

To obtain an insight on the mechanisms responsible for the virus *ts* phenotype as well as the role played by other residues of the SFV protease in the proteolytic reaction, we produced Pro39 variants with substitutions at key positions and analyzed their activity with Trx34 model substrate.

2. Materials and methods

2.1. Pro39 expression plasmids

The plasmid construct for the expression of *wt* Pro39 (residues 459–799 of SFV nsP2), tagged with the peptide LEHHHHHH at its C-terminus, has been described previously [15]. This plasmid was used as a template to obtain the various mutated protein forms. The point mutations were introduced using the Quick Change XL Site-Directed Mutagenesis kit (Stratagene) and verified by DNA sequencing.

*Corresponding author. Present address: European Science Foundation, COST Office, 149 Avenue Louise, 1050 Brussels, Belgium. Fax: +32 2 5333890.
E-mail address: jcaldentey@cost.esf.org (J. Caldentey).

0014-5793/$32.00
doi:10.1016/j.febslet.2006.01.071

2.2. Expression and purification of Pro39

Wt and mutant Pro39 were expressed in *E. coli* and purified essentially as described previously [15] with some modifications in the isolation procedure. Briefly, the cell lysates in 20 mM sodium phosphate buffer, pH 7.4, 200 mM NaCl, 0.1% Tween 20 (buffer A), supplemented with 0.1 mM EDTA and 1 mM PMSF, were cleared by centrifugation at 15000 × g for 30 min at 4 °C. The supernatant was supplemented with 150 mM imidazole and loaded to a Hi-Trap Chelating HP column (Amersham Biosciences) charged with $NiSO_4$. After extensive washing, Pro39 was eluted with buffer A containing 500 mM imidazole. The resulting protein solution was supplemented with 2 mM EDTA and the buffer was changed to 20 mM HEPES, pH 7.4, containing 200 mM NaCl, 20% glycerol, 0.1% Tween 20, 1 mM DTT, and 1 mM EDTA using a PD10 gel filtration column (Amersham Biosciences). The final preparation was divided into aliquots, frozen and stored at −70 °C. Under these storage conditions, Pro39 retained its activity for at least one year.

2.3. Protease assay

An 18 kDa substrate, comprising 19 (nsP3) and 18 (nsP4) residues of the SFV nsP3–nsP4 junction fused to 115 residues of thioredoxin, was used to assess the activity of the purified proteins [15]. The protease activity of the various Pro39 proteins was assayed in 50 mM HEPES, pH 7.4, containing 50 mM NaCl, 1 mM DTT, 2 mM spermidine and 10% glycerol, at a substrate concentration of 0.8 mg/ml. For time-course experiments, enzyme and substrate were mixed in pre-warmed buffer. At the indicated time points, 5 μl aliquots were withdrawn, mixed with electrophoresis sample buffer and boiled immediately. In experiments aimed at determining and comparing enzymatic activities, Pro39 was pre-incubated in buffer for 2 min at the desired temperature prior to the addition of the substrate. The proteolytic products of the reaction were analyzed by electrophoresis in 17% polyacrylamide gels.

2.4. General analytical methods

SDS–PAGE was carried out according to Laemmli [27]. Gels were stained with Coomassie Brilliant Blue and destained in 10% acetic acid before being dried. For the quantification of the proteolytic activity of the proteins, gels were scanned and analyzed by densitometry using the Tina 2.0 program. Care was taken not to overload the gels and the linearity in the signal detection was checked in gels containing known amounts of Pro39 or Trx34. Protein concentrations were determined with the Bradford assay [28] using bovine serum albumin as the standard.

3. Results

3.1. Expression and purification of mutant Pro39

In order to study the role of individual residues in activity of SFV protease, point mutations were introduced in the protease domain of nsP2 by site-directed mutagenesis to obtain eight Pro39 mutants (M781T (*ts4*), A662T (*ts6*), G577R (*ts11*), C478A, H548A, W549A, N600D, N605D). The proteins were expressed in *E. coli* and purified by metal affinity chromatography (Fig. 1). In all cases, the conditions for expression and purification were reproduced as precisely as possible to facilitate comparison of the properties of the mutant enzymes. The expression of the N600D and N605D mutants had a strong negative effect of the host cell, resulting in a slower growth rate and thus smaller bacterial mass as well as in a reduced expression level of the corresponding soluble Pro39. This, in turn, resulted in preparates with lower degree of purity, as can be seen in Fig. 1. Since the two mutant proteins showed a tendency to aggregate, particularly when submitted to the changes of ionic strength associated to many chromatography techniques, further purification steps were not successful and the shown preparates were used for the experiments reported here.

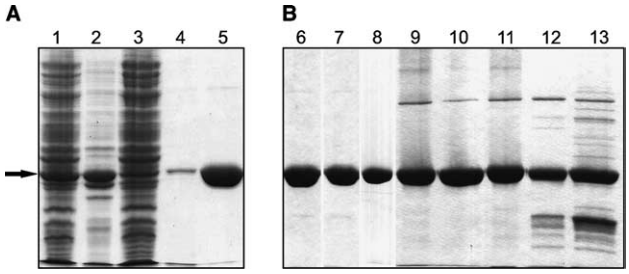

Fig. 1. Purification of Pro39 by immobilized Ni affinity chromatography. Panel A: Purification of *wt* Pro39. Lanes (1) soluble protein fraction after cell lysis and clearing at 15000 × g; (2) insoluble protein fraction; (3) flow through of the chromatographic run; (4) 150 mM imidazole column wash; (5) final Pro39 preparation. Panel B: Samples of final preparations of purified Pro39 point mutants. Lanes (6) *ts4* (M781T); (7) *ts6* (A662T); (8) *ts11* (G577R); (9) C478A; (10) H548A; (11) W549A; (12) N600D; (13) N605D. The numbering of the residues corresponds to their position in SFV nsP2. The gel has been overloaded to give an impression of the purity of the preparations used in this study.

3.2. Effect of temperature on the activity of Pro39

It has been shown previously that Pro39 readily cleaves the site between nsP3 and nsP4 in the synthetic substrate Trx34 [15]. The proteolytic reaction results in the appearance of a large fragment, L, of 14 kDa that can easily be detected and quantified after electrophoresis (Fig. 2). A smaller fragment, S, with a molecular mass of 4.1 kDa often appears as a broad and rather diffuse band at the bottom of the gel.

To determine if the mutations causing the *ts* phenotype in the virus would also render the protease temperature sensitive in vitro, *wt* and *ts* Pro39 were mixed with the Trx34 substrate

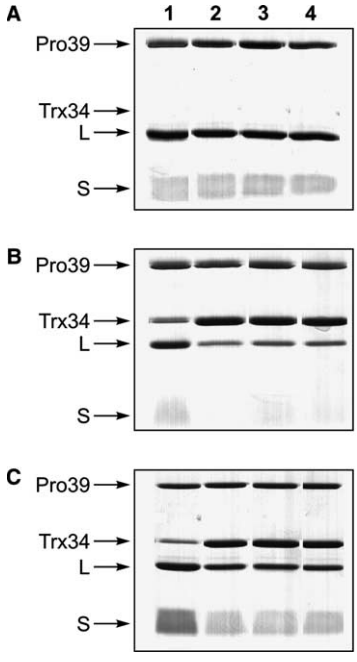

Fig. 2. Effect of temperature on Pro39 activity. The proteolytic activity of the *wt* enzyme (lane 1) and the *ts4* (lane 2), *ts6* (lane 3), and *ts11* (lane 4) mutants was assayed as described in Section 2 at 28 °C (panel A), 39 °C (B) and 28 °C after a 10 min pre-incubation at 39 °C (C). The positions of Pro39, Trx34, and the cleavage products L and S are indicated with arrows.

in buffer pre-warmed at either 28 or 39 °C, and incubated further for 60 min at the same temperature. As can be seen in Fig. 2A, *wt* as well as the mutant Pro39 carrying *ts* mutations were able to cleave the substrate completely at 28 °C. However, performing the reaction at the restrictive temperature of 39 °C clearly resulted in an impairment of the enzymatic activity of the mutant Pro39, whereas the *wt* enzyme appeared to be only slightly affected (Fig. 2B). 10 min pre-incubation of the samples at 39 °C, followed by further 60 min incubation at 28 °C, resulted in an intermediate pattern of activity in the case of the *ts* proteins (Fig. 2C). Thus, the properties of the mutant Pro39 in vitro correlated with the temperature-sensitive phenotype of the viruses observed in vivo.

Initial experiments were carried out at an enzyme:substrate molar ratios of 1:8. Under those conditions, the reaction was quite fast, reaching a plateau approximately after 10 min (Fig. 3, panels A and B). For a quantitative comparison of the enzymatic activity of the *ts* and the *wt* Pro39, we lowered the rate of the reaction by decreasing the amount of enzyme in the samples. At a Pro39:Trx34 molar ratio of 1:32 the reaction proceeded linearly for a period of up to 10 min (Fig. 3C), allowing accurate determination of the reaction rate. Thus, these conditions were used to compare the activity of the different proteases at both the permissive and restrictive temperatures. For this, Pro39 was preincubated in buffer for 2 min in the absence of Trx34 in order to avoid any possible contribution of the substrate on the stability of the enzyme. Upon substrate addition, samples were withdrawn at time points 1, 2 and 5 min, and analyzed. We found the activity of the *wt* and mutant Pro39 to be practically identical when the reaction was performed at 28 °C. However, at 39 °C the activity of the *ts* enzymes clearly decreased to levels of 55 ± 8% (mean of 3 experiments), 55 ± 4% (2), and 65 ± 3% (3) for *ts4*, *ts6*, and *ts11*, respectively, whereas the activity of the *wt* Pro39 was much less affected by the rise in temperature (82 ± 12%, mean of 5 experiments). From the results of the experiments carried out with the *wt* enzyme at different enzyme:substrate ratios, an apparent K_M of approximately 250 μM and a V_{max} close to 160 pmol/min was calculated for Pro39 at 28 °C (not shown).

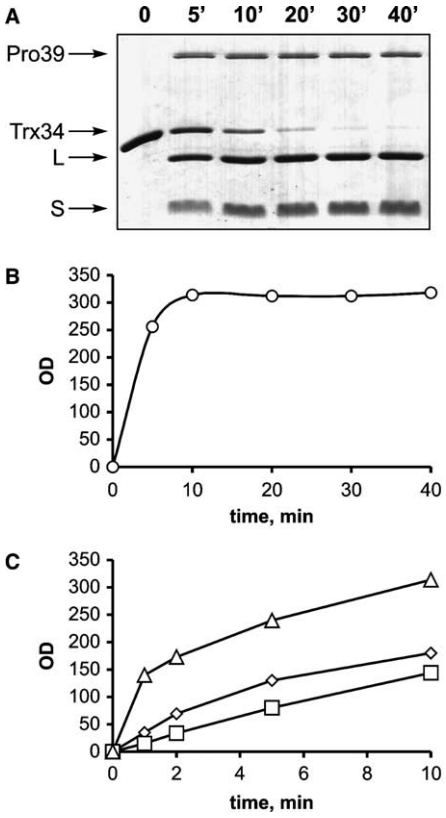

Fig. 3. Kinetics of Trx34 cleavage by *wt* Pro39. Panel A: The cleavage reaction was performed at 28 °C at a Pro39:Trx34 molar ratio of 1:8. At the indicated time points, aliquots were withdrawn for analysis. Time 0 represents the amount of substrate present in the sample prior to the addition of the enzyme. Panel B: Quantitation of the proteolytic reaction shown in panel A. The reaction product L was quantified by densitometry and the amounts, expressed as arbitrary OD units, plotted against the reaction time. Panel C: Time course of Trx34 cleavage by *wt* Pro39 at 28 °C at enzyme:substrate ratios of 1:8 (triangles), 1:16 (rhombs), and 1:32 (squares).

3.3. Mutagenesis at the putative active site

Previously, the protease activity of closely related Sindbis virus (SIN) was assayed via the autocatalytic processing of the non-structural polyprotein synthesized by in vitro translation. Cys^{481}, His^{558} and Trp^{559} were shown to be critical for the proteolytic activity [29]. We mutagenized the homologous residues in SFV Pro39 and analyzed the activity of the purified proteases using Trx34 as the substrate. Mutations C478A and H548A resulted in completely inactive Pro39 supporting the view that they represent the conserved catalytic dyad of the papain-like proteases, being Cys^{481}/His^{558} the equivalent residues in SIN. Substitution of the homologous Trp^{549} with Ala also rendered Pro39 inactive (Fig. 4).

The work of Strauss and coworkers [29] also revealed two Asn residues conserved in alphaviruses important for the SIN protease activity. N609D mutation inhibited processing of the polyprotein but allowed viral replication, whereas N614D mutation had the opposite effect – it enhanced proteolysis but was lethal for the virus [12,29]. In order to investigate the effect of these mutations on the catalytic activity of SFV protease directly, Asn residues at the corresponding positions of Pro39 were replaced by Asp. Both N600D and N605D Pro39 mutants showed significantly reduced efficiency of the in vitro Trx34 substrate cleavage compared to the *wt* enzyme (Fig. 4). Because of the lower purity of these samples, it could be argued that the detected activity, or at least part of it, could be due to the presence of contaminant bacterial proteins able to act upon the same bond on Trx34. It should be noted, however, that we were not able to detect any endogeneous protease activity capable of specific cleavage of Trx34 neither in *E. coli*

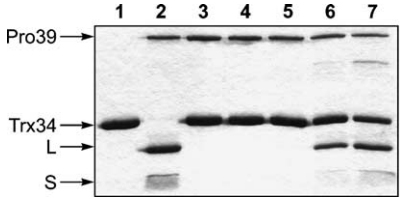

Fig. 4. Effect of point mutations on the activity of Pro39. The gel shows the proteolytic activity of the *wt* (lane 2), C478A (lane 3), H548A (lane 4), W549A (lane 5), N600D (lane 6), and N605D (lane 7) enzymes, determined at Pro39:Trx34 ratio of 1:8 at 28 °C. A control sample containing only the substrate is shown in lane 1. The positions of Pro39, substrate, and large fragment are indicated by arrows.

BL21 extracts nor in the purified mutant Pro39 preparations. In the latter case, prolonged incubation of these samples at room temperature resulted in aggregation and precipitation of the mutant enzyme, whereas the bulk of contaminating proteins remained soluble. After separation of the precipitated Pro39 by centrifugation, the resulting supernatant did not show any detectable activity against Trx34 (not shown), suggesting that the observed reduced enzymatic activity of the preparates was due to the mutant Pro39 and not to the presence of the contaminating proteins.

3.4. Role of Trp^{549} in the cleavage specificity

Trp^{549} (Trp^{559} in SIN), the residue following the catalytic His, is conserved in the alphavirus cysteine proteases. It is also critical for the protease activity of SFV Pro39 (Fig. 4). To obtain further information on the role of this essential Trp, we scanned the MEROPS database to check if the "HW" motif is present in other cysteine proteases. It appeared in eight families of peptidases and in six of them, this motif was strictly conserved. Analysis of the experimental data published for these proteases revealed that all of them process polypeptides with Gly residue at P2 position of the cleavage site, as is the case of all three processing sites of alphavirus non-structural polyprotein [30], or at P1 position or at both. Moreover, we found that this conclusion could be extended to other cysteine proteinases where the catalytic His is followed by Tyr or Phe residue (Table 1).

The proteases listed in Table 1, denoted here as GSM (glycine specificity motif) proteases, are mainly representatives of clans CA and CE, where papain and adenain, respectively, are considered the prototypes. The group includes several families of viral proteases, ubiquitin and SUMO hydrolases, pseudomurein endoisopeptidases and bacteriocin processing enzymes. Representatives of five families have had their tertiary structures solved and deposited in the Protein Data Bank [31]. Despite the lack of sequence similarity, the architecture of the catalytic center is very similar, as illustrated in Fig. 5. What is particularly important is that the aromatic residue (Trp, Tyr or Phe) appears to occupy a virtually identical position in all the structures. Moreover, in the structures where the substrate molecule is present, the contact of the aromatic residue with the penultimate Gly of the substrate can be directly visualized.

Another distinctive feature unifying the GSM proteases is their resistance to E-64, known to be a specific inhibitor of papain-like peptidases [17]. Although the majority of GSM enzymes belonging to the papain type, none of them are sensitive to E-64. However, where this has been checked directly, E-64 always failed to block the proteolytic activity of the enzymes listed in Table 1.

Table 1
List of cysteine protease families containing the GSM motifs and their cleavage site specificity[a]

MEROPS clan	MEROPS family	MEROPS family name	GSM sequence[b]	Cleavage site[b]	E-64 inhibition
CA	C9	SIN-type nsP2 peptidase	HW	G[AGC]↓	No
CA	C33	Equine Arterivirus Nsp2-type cysteine peptidase	HW	G↓G[c]	
CA	C39	Bacteriocin-processing endopeptidase	H[WYF]	GG↓of bacteriocin leader peptide	
CE	C5	Adenain	HW	GG↓	No
CE	C48	Ulp1 endopeptidase	HW	GG↓ of SUMO	
CE	C57	Vaccinia virus I7L processing peptidase	HW	AG↓A	
CE	C63	African Swine Fever virus processing peptidase	HW	GG↓	No
Unassigned	C53	Pestivirus Npro endopeptidase	H[QEKNDHPW]		
CA	C12	Ubiquitinyl hydrolase-L1	H[YF]	GG↓ of ubiquitin	
CA	C16	Coronavirus papain-like endopeptidase:	H[SCY]		
		Murine Hepatitis coronavirus $PL^{pro}1$	HS	RG↓, CG↓	Yes
		Murine Hepatitis coronavirus $PL^{pro}2$	HY	GG↓A	No
		Avian Infectious Bronchitis coronavirus $PL^{pro}2$	HC	AG↓G	
		Human coronavirus 229E $PL^{pro}1$ and $PL^{pro}2$	HY	CG↓, AG↓G, GA↓G[c]	
		SARS coronavirus	HY	GG↓A	No
CA	C19	Ubiquitin-specific peptidase 14	H[YF]	GG↓ of ubiquitin	
CA	C21	Tymovirus endopeptidase	H[YF]	GA↓[d]	
CA	C23	Carlavirus endopeptidase	H[ILMAF]		
CA	C27	Rubella virus endopeptidase	HF	RG↓G	No
CA	C36	Beet Necrotic Yellow Vein Furovirus-type papain-like endopeptidase	HF		
CA	C54	ATG4 peptidase:	H[TEDVIHYF]		
		Saccharomyces cerevisiae ATG4 peptidase	HI	FG↓	
		Human autophagin-3	HY	FG↓[c]	
CA	C64	Cezanne deubiquitinating peptidase	HF	GG↓ of ubiquitin	
CA	C65	Otubain	HY	GG↓ of ubiquitin	
CA	C67	CylD protein	HY	GG↓ of ubiquitin	
CA	C71	Pseudomurein endoisopeptidase Pei	HY	Eγ-A↓εK	
CE	C55	YopJ endopeptidase	H[HIRF]	GG↓ of SUMO[c]	

[a]An extended version of Table 1 with references to original reports is available as Supplementary Material.
[b]If a residue is not conserved the variety of aminoacids at the position is given in square brackets.
[c]Provisional assignment, cleavage site has not been determined experimentally.
[d]Turnip Yellow Mosaic virus endopeptidase containing "HF" GSM is the only protease in the family, which site specificity has been shown experimentally.

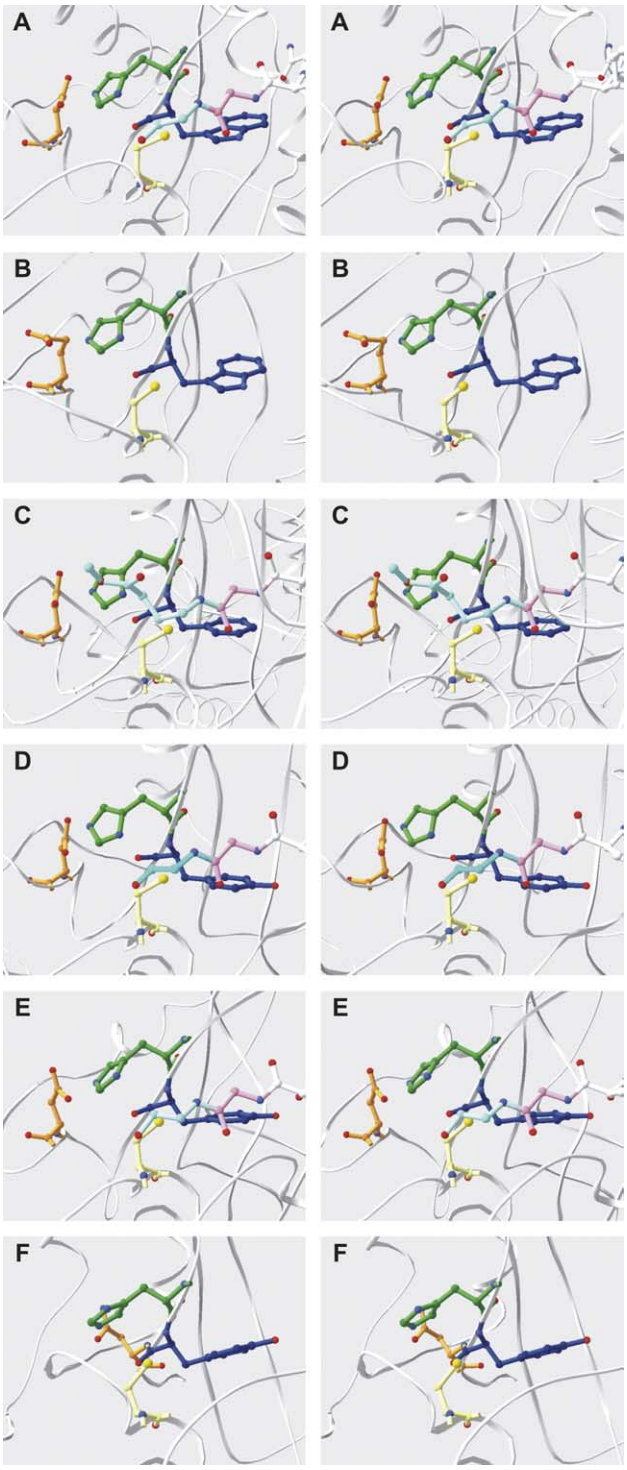

Fig. 5. Architecture of the active site of the GSM proteases. The structrures of six GSM-containing proteases were aligned by catalytic site residues with Deep View software [43] and are shown as ribbon models. The catalytic residues shown as ball-stick models: Cys – in yellow, His – in green, the third member of the catalytic triad (Asp, Glu, Asn) – in orange; the aromatic residue of GSM motif (Trp, Phe, Tyr) – in dark blue. The last and the penultimate Gly residues of the substrates are shown in turquoise and pink, respectively. The structures presented: Ulp1–SUMO complex, PDB code 1EUV (A); adenain, 1AVP (B); UCH L3-ubiquitin vinylmethylester complex, 1XD3 (C); Yuh1–ubiquitin aldehyde complex, 1CMX (D); HAUSP–ubiquitin aldehyde complex, 1NBF (E); otubain, 1TFF (F).

4. Discussion

The alphavirus protein nsP2 plays an essential role in virus propagation. It is a multifunctional enzyme involved in many replication processes. Its protease activity, residing at the C-terminus of the protein, is responsible for the maturation of the non-structural polyprotein [32,33]. The N-terminal half of nsP2 possesses helicase [34,35] and RNA triphosphatase activities [36]. The protein plays a central role in regulation of the 26S subgenomic RNA transcription [19,23,26]. In addition, it has been reported to be involved in virus–host interactions [37,38].

The variety of nsP2 functions often makes it difficult to analyze nsP2 function in vivo and in complex systems in vitro where multiple secondary effects are possible. Even the protease activity itself shows different behavior, acting upon three different cleavage sites in the non-structural polyprotein [39]. The cleavage between nsP1 and nsP2 occurs only in *cis*, whereas the nsP2/nsP3 site is processed in *trans* and requires the presence of a free nsP2 terminus. However, the site between nsP3 and nsP4 requires only the protease activity and a short specific polypeptide stretch as a substrate for efficient cleavage [15]. Taking these facts into account, we exploited the benefits of a simplified experimental system consisting of the basal protease Pro39 and the model substrate Trx34 [15] to characterize the SFV protease by in vitro mutagenesis.

The first question we addressed was the proteolytic activity of purified Pro39 carrying mutations derived from SFV temperature-sensitive mutants *ts4* (M781T), *ts6* (A662T) and *ts11*(G577R) [23,24]. The activities of the mutant enzymes were assayed at 28 and 39 °C, which are the permissive and restrictive temperatures for the *ts* viruses, respectively. The *wt* and the mutant Pro39 cleaved the Trx34 substrate with practically the same efficiency at 28 °C. However, a marked reduction in the activity of the mutated proteins was observed at the non-permissive temperature of 39 °C. Thus, Pro39 proteins carrying the *ts* mutations display temperature sensitivity in vitro that correlates with the effects of higher temperatures observed in vivo, suggesting that an impaired protease domain is most likely responsible for the viral *ts* phenotypes. It is important to note that after a short exposure to 39 °C the cleavage efficiency of Pro39 did not revert to the initial 28 °C level (Fig. 2C), indicating that there was a loss in the amount of active enzyme. Therefore, the *ts* mutations affected rather protein stability than the catalytic activity itself, although we cannot exclude that both change at higher temperatures.

The nsP2 protease of alphaviruses belongs to a papain-like family of cysteine proteases [40]. In the case of the Sindbis enzyme, site-directed mutagenesis of the non-structural polyprotein strongly suggested that Cys^{481} and His^{558} form the catalytic dyad of the enzyme [29]. Our present results, based on in vitro analysis of the purified protease domain activity, confirmed that, in the case of SFV, Cys^{478} and His^{548} are essential for the proteolytic activity. Similarly, the mutation N609D in the SIN polyprotein resulted in impaired polyprotein processing [29], which correlates with the reduced proteolytic activity of the corresponding N605D mutant of SFV Pro39. However, an interesting difference was noticed in the case of the Pro39 N605D mutation. The corresponding change in SIN virus resulted in hyperactive polyprotein processing that was lethal for the virus [12,29], whereas the activity of the

N609D mutant of SFV Pro39 was reduced. This discrepancy may reflect differences in the organization of the active site of these two similar proteases. At the same time, it cannot be excluded that the hyperactive phenotype observed in polyprotein processing is caused at the level of regulation of the nsP2 activity rather than at the level of the catalytic activity itself.

The role of $Trp^{549/559}$ in SFV and SIN nsP2, respectively, has been poorly understood. It has been suggested that it might act as a third member of the catalytic triad providing correct orientation for the catalytic histidine [29]. Here we suggest that this Trp residue is crucial because of its role in recognition of the penultimate Gly residue in the processing sites of the non-structural polyprotein. Indeed, the papain-like proteinases are known to be very specific for the amino acid residue at the P2 position of the substrate, usually a bulky or hydrophobic residue [17]. However, the non-structural polyproteins of alphaviruses have Gly preserved at this position [30]. This difference may provide the most straightforward explanation for the resistance of the SFV protease to the E-64 inhibitor [15]. The E-64 leucine moiety mimics the most common occupant of the substrate pocket of the papain-like peptidases (see [41] for papain family and [42] for calpains), but, most probably, it cannot fit the pocket arranged for a tiny Gly residue.

The putative role of Trp^{549} can be inferred from the finding that all the cystein proteases listed in MEROPS, which have the catalytic His followed by either Trp, Tyr or Phe, processed the substrates with a Gly residue in the cleavage site. The only exception was the pseudomurein endoisopeptidases. In this case, however, the Glu-Ala peptide, linked together via the glutamate side chain carboxyl group, is very similar to the conventional Gly-Ala and can fit also in a small substrate pocket. The structures available for these peptidases showed the same active site architecture and the same position of the aromatic residue of the motif directly interacting with the substrate P2 Gly. There may be some variations in the active site organization since two groups of GSM proteases cleave the substrates with Gly at P1 position instead of P2 (see Table 1). We assume that in those cases, either another residue is allowed at P2 or the Gly at P1 plays a similar role in the recognition event. It should be noted, however, that the Rubella virus protease has been shown to be resistant to E-64, a property seemingly common to all the peptidases in the GSM group.

An especially illustrative case is provided by Murine Hepatitis coronavirus, which has two very similar orthologous papain-like peptidases $PL^{pro}1$ and $PL^{pro}2$. The difference in the catalytic site of these two enzymes, "HS" and "HY", respectively, correlates with the change at the P2 substrate position and with the resistance to E-64 (Table 1).

Based on empiric bioinformatics data, structural and biochemical information, we suggest that alphavirus proteinases, as well as a number of other viral and cellular cystein proteinases, utilize recognition of the Gly residue at the P2 position of the substrate. The glycine specificity motif – the catalytic His followed by an aromatic residue – can be considered a hallmark of this type of peptidases and can be used for predicting the specificity of newly discovered proteinases.

Acknowledgements: The technical assistance of Ms. Airi Sinkko is gratefully acknowledged. Drs. Lidia Vasiljeva and Denis Kainov are thanked for fruitful discussions. Thanks are also due to Dr. John Williams for useful advice and critically reading the manuscript. This work was supported by the EU Consortium Agreement QLK3-CT-2002-02052 (SFVECTORS) and the Sigrid Juselius Foundation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2006.01.071.

References

[1] Strauss, J.H. and Strauss, E.G. (1994) The alphaviruses: gene expression, replication and evolution. Microbiol. Rev. 58, 491–562.
[2] Kääriäinen, L. and Ahola, T. (2002) Functions of alphavirus nonstructural proteins in RNA replication. Prog. Nucleic Acid Res. Mol. Biol. 71, 187–222.
[3] Nicola, A.V., Chen, W. and Helenius, A. (1999) Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. Nat. Cell. Biol. 1, 341–345.
[4] Kowarik, M., Kung, S., Martoglio, B. and Helenius, A. (2002) Protein folding during cotranslational translocation in the endoplasmic reticulum. Mol. Cell 10, 769–778.
[5] Pelkmans, L. and Helenius, A. (2003) Insider information: what viruses tell us about endocytosis. Curr. Opin. Cell. Biol. 15, 414–422.
[6] Vonderheit, A. and Helenius, A. (2005) Rab7 associates with early endosomes to mediate sorting and transport of Semliki Forest Virus to late endosomes. PLoS Biol. 21, e233.
[7] Fazakerley, J.K. (2004) Semliki Forest virus infection of laboratory mice: a model to study the pathogenesis of viral encephalitis. Arch. Virol. Suppl. 18, 179–190.
[8] Karlsson, G.B. and Liljeström, P. (2004) Delivery and expression of heterologous genes in mammalian cells using self-replicating alphavirus vectors. Methods Mol. Biol. 246, 543–557.
[9] Atkins, G.J., Smyth, J.W., Fleeton, M.N., Galbraith, S.E. and Sheahan, B.J. (2004) Alphaviruses and their derived vectors as anti-tumor agents. Curr. Cancer Drug Targets 4, 597–607.
[10] Salonen, A., Ahola, T. and Kääriäinen, L. (2004) Viral RNA replication in association with cellular membranes. Curr. Top. Microbiol. Immunol. 285, 139–173.
[11] Kim, K.H., Rümenapf, T., Strauss, E.G. and Strauss, J.H. (2004) Regulation of Semliki Forest virus RNA replication: a model for the control of alphavirus apthogenesis in vertebrate hosts. Virology 323, 153–163.
[12] Lemm, J.A., Bergqvist, A., Read, C.M. and Rice, C.M. (1998) Template-dependent initiation of Sindbis virus RNA replication in vitro. J. Virol. 72, 6546–6553.
[13] Takkinen, K., Peränen, J. and Kääriäinen, L. (1991) Proteolytic processing of Semliki Forest virus-specific non-structural polyprotein. J. Gen. Virol. 72, 1627–1633.
[14] Merits, A., Vasiljeva, L., Ahola, T., Kääriäinen, L. and Auvinen, P. (2001) Proteolytic processing of Semliki Forest virus-specific non-structural polyprotein by nsP2 protease. J. Gen. Virol. 82, 765–773.
[15] Vasiljeva, L., Valmu, L., Kääriäinen, L. and Merits, A. (2001) Site-specific protease activity of the carboxyl-terminal domain of Semliki Forest virus replicase protein nsP2. J. Biol. Chem. 276, 30786–30793.
[16] Rawlings, N.D., Tolle, D.P. and Barrett, A.J. (2004) MEROPS: the peptidase database. Nucleic Acids Res. 32, D160–D164.
[17] Barrett, A.J. and Rawlings, N.D. (2001) Evolutionary lines of cysteine peptidases. Biol. Chem. 382, 727–733.
[18] Keränen, S. and Kääriäinen, L. (1974) Isolation and basic characterization of temperature-sensitive mutants from Semliki Forest virus. Acta Pathol. Microbiol. Scand. [B] Microbiol. Immunol. 82, 810–820.
[19] Keränen, S. and Kääriäinen, L. (1979) Functional defects of RNA-negative temperature-sensitive mutants of Sindbis and Semliki Forest viruses. J. Virol. 32, 19–29.
[20] Lachmi, B.E. and Kääriäinen, L. (1976) Sequential translation of non-structural proteins in cells infected with a Semliki Forest virus mutant. Proc. Natl. Acad. Sci. 73, 1936–1940.
[21] Kääriäinen, L., Sawicki, D. and Gomatos, P.J. (1978) Cleavage defect in the non-structural polyprotein of Semliki Forest virus

has two separate effects on virus RNA synthesis. J. Gen. Virol. 39, 463–473.
[22] Keränen, S. and Ruohonen, L. (1983) Non-structural proteins of Semliki Forest virus: synthesis, processing, and stability in infected cells. J. Virol. 47, 505–515.
[23] Suopanki, J., Sawicki, D.L., Sawicki, S.G. and Kääriäinen, L. (1998) Regulation of alphavirus 26S mRNA transcription by replicase component nsP2. J. Gen. Virol. 79, 309–319.
[24] Lulla, V., Merits, A., Sarin, P., Kääriäinen, L., Keränen, S., Ahola., T. (in press). Identification of mutations causing temperature-sensitive defects in Semliki Forest virus RNA synthesis. J. Virol.
[25] Saraste, J., Kääriäinen, L., Söderlund, H. and Keränen, S. (1977) RNA synthesis directed by a temperature-sensitive mutant of Semliki Forest virus. J. Gen. Virol. 37, 399–406.
[26] Sawicki, D.L. and Sawicki, S.G. (1985) Functional analysis of the A complementation group mutants of Sindbis HR virus. Virology 144, 20–34.
[27] Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.
[28] Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 7, 248–254.
[29] Strauss, E.G., De Groot, R.J., Levinson, R. and Strauss, J.H. (1992) Identification of the active site residues in the nsP2 proteinase of Sindbis virus. Virology 191, 932–940.
[30] Strauss, J.H. and Strauss, E.G. (2004) Sindbis virus nsP2 endopeptidase in: Handbook of Proteolytic Enzymes (Barrett, N.D., Rawlings, N.D. and Woessner, J.F., Eds.), 2nd edn, pp. 1269–1271, Elsevier, London.
[31] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. Nucleic Acids Res. 28, 235–242.
[32] Ding, M.X. and Schlesinger, M.J. (1989) Evidence that Sindbis virus NSP2 is an autoprotease which processes the virus non-structural polyprotein. Virology 171, 280–284.
[33] Hardy, W.R. and Strauss, J.H. (1989) Processing the non-structural polyproteins of sindbis virus: non-structural proteinase is in the C-terminal half of nsP2 and functions both in *cis* and in *trans*. J. Virol. 63, 4653–4664.
[34] Rikkonen, M., Peränen, J. and Kääriäinen, L. (1994) ATPase and GTPase activities associated with Semliki Forest virus non-structural protein nsP2. J. Virol. 68, 5804–5810.
[35] Gómez de Cedrón, M., Ehsani, N., Mikkola, M.L., García, J.A. and Kääriäinen, L. (1999) RNA helicase activity of Semliki Forest virus replicase protein NSP2. FEBS Lett. 448, 19–22.
[36] Vasiljeva, L., Merits, A., Auvinen, P. and Kääriäinen, L. (2000) Identifiction of a novel function of the alphavirus capping apparatus. RNA 5′-triphosphatase of nsP2. J. Biol. Chem. 275, 17281–17287.
[37] Rikkonen, M. (1996) Functional significance of the nuclear-targeting and NTP-binding motifs of Semliki Forest virus non-structural protein nsP2. Virology 218, 352–361.
[38] Frolova, E.I., Fayzulin, R.Z., Cook, S.H., Griffin, D.E., Rice, C.M. and Frolov, I. (2002) Roles of non-structural protein nsP2 and Alpha/Beta interferons in determining the outcome of Sindbis virus infection. J. Virol. 76, 11254–11264.
[39] Vasiljeva, L., Merits, A., Golubtsov, A., Sizemskaja, V., Kääriäinen, L. and Ahola, T. (2003) Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. J. Biol. Chem. 278, 41636–41645.
[40] Gorbalenya, A.E., Koonin, E.V. and Lai, M.M. (1991) Putative papain-related thiol proteases of positive-strand RNA viruses. Identification of rubi- and aphthovirus proteases and delineation of a novel conserved domain associated with proteases of rubi-, alpha- and coronaviruses. FEBS Lett. 288, 201–205.
[41] Matsumoto, K., Mizoue, K., Kitamura, K., Tse, W.C., Huber, C.P. and Ishida, T. (1999) Structural basis of inhibition of cysteine proteases by E-64 and its derivatives. Biopolymers 51, 99–107.
[42] Moldoveanu, T., Campbell, R.L., Cuerrier, D. and Davies, P.L. (2004) Crystal structures of calpain-E64 and -leupeptin inhibitor complexes reveal mobile loops gating the active site. J. Mol. Biol. 343, 1313–1326.
[43] Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723.