Site-specific Protease Activity of the Carboxyl-terminal Domain of Semliki Forest Virus Replicase Protein nsP2

Received for publication, May 25, 2001
Published, JBC Papers in Press, June 15, 2001, DOI 10.1074/jbc.M104786200

Lidia Vasiljeva, Leena Valmu, Leevi Kääriäinen, and Andres Merits

From the Program in Cellular Biotechnology, Institute of Biotechnology, Biocenter Viikki, P. O. Box 56, University of Helsinki, FIN-00014 Helsinki, Finland

The virus-specific components (nsP1–nsP4) of Semliki Forest virus RNA polymerase are synthesized as a large polypeptide (P1234), which is cleaved by a virus-encoded protease. Based on mutagenesis studies, nsP2 has been implicated as the protease moiety of P1234. Here, we show that purified nsP2 (799 amino acids) and its C-terminal domain Pro39 (amino acids 459–799) specifically process P1234 and its cleavage intermediates. Analysis of cleavage products of in vitro synthesized P12, P23, and P34 revealed cleavages at sites 1/2, 2/3, and 3/4. The cleavage regions of P1/2, P2/3, and P3/4 were expressed as thioredoxin fusion proteins (Trx12, Trx23, and Trx34), containing 20 amino acids on each side of the cleavage sites. After exposure of these purified fusion proteins to nsP2 or Pro39, the reaction products were analyzed by SDS-polyacrylamide gel electrophoresis, mass spectrometry, and amino-terminal sequencing. The expected amino termini of nsP2, nsP3, and nsP4 were detected. The cleavage at 3/4 site was most efficient, whereas cleavage at 1/2 site required 5000-fold more of Pro39, and 2/3 site was almost resistant to cleavage. The activity of Pro39 was inhibited by N-ethylmaleimide, Zn2+, and Cu2+, but not by EDTA, phenylmethylsulfonyl fluoride, or pepstatin, in accordance with the thiol proteinase nature of nsP2.

The genomes of many positive strand RNA viruses are expressed as polypeptides in order to achieve the expression of multiple proteins from a single message, unlike the mRNAs of their eukaryotic host cells, which mostly code for single proteins. Thus, proteolyses of the polypeptide precursors are essential events in the regulation of the replication and morphogenesis of these RNA viruses. In picornaviruses and flaviviruses, the entire RNA genome is translated as a single polypeptide, from which the structural and nonstructural proteins are processed by proteolysis. In picornavirus-infected cells, the processing is carried out by virus-encoded proteases within the polypeptide, whereas the processing of flavivirus polyprotein is assisted by host proteases (1, 2). The large RNA genomes of coronaviruses (approximately 30 kilobases) and arteriviruses (12.7–15.7 kilobases), together classified as Nidovirales, use in addition to the polypeptide strategy also a set of subgenomic mRNAs (3). Alphaviruses and rubella virus, members of the Togaviridae family, express two polypeptides. The nonstructural polypeptide is expressed directly from the RNA genome, whereas the structural polypeptide is synthesized from a subgenomic mRNA (4–6).

Semliki Forest virus (SFV) is a typical alphavirus with a lipoprotein envelope surrounding the nucleocapsid. The 5′ two-thirds of the 11.5-kilobase 42 S RNA genome codes for the nonstructural polyprotein (P1234) of 2432 aa, which is autocatalytically cleaved to finally yield the virus-specific components of the RNA polymerase complex, nsP1–nsP4 (4, 5, 7). The processing of the nonstructural polyproteins P1234 and P123 of Sindbis virus (SIN), another alphavirus, has been studied in detail (4, 8, 9). Using mostly in vitro translation and site-directed mutagenesis as tools, autocatalytic protease activity was detected in the polyprotein and its cleavage intermediates. The protease activity was localized to nsP2, and more precisely, to its carboxyl-terminal part (10, 11). Cysteine 481 and histidine 558 were identified as essential residues for the autoprotease activity (12), supporting the view that the protease is a thiol proteinase belonging to the papain superfamily. The same conclusion was reached by sequence comparisons (13). The respective mutation of Cys-478 in SFV nsP2 also inactivates the autocatalytic processing of P1234, P123, and P23 (14).

The amino-terminal domain of SFV nsP2 (residues 1–470) has been shown to house several enzymatic activities including RNA triphosphatase (15), nucleoside triphosphatase (16), and RNA helicase (17). Interestingly, a significant amount of nsP2, synthesized during infection, is transported into the nucleus (18, 19). The core of the nuclear localization signal was mapped to a short sequence P647RRR (20). nsP2 mutant PRDR is not lethal for the virus in cell culture, but SFV carrying this mutation is apathogenic for mouse (21). In addition, the carboxyl-terminal domain of SFV nsP2 has been implicated in the regulation of the synthesis of the subgenomic mRNA (22–24). Taken together, these data suggest that alphavirus nsP2 consists of two structurally independent domains, each possessing a distinct set of biological activities. However, direct proof that purified nsP2 or its carboxyl-terminal part has protease activity has been lacking. To this end, recombinant nsP2 and a set of its amino-terminally truncated variants were expressed in Escherichia coli, purified by metal-chelate chromatography, and assayed for the presence of protease activity. Both full-length nsP2 and its soluble carboxyl-terminal fragment Pro39 (aa 458–799) catalyzed site-specific proteolysis of

* This work was supported by Academy of Finland Grant 8397 and by grants from the Technology Development Center and Center for International Mobility. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

* Biocentrum Helsinki fellow. To whom correspondence should be addressed. Tel.: 358-9-191-59400; Fax: 358-9-191-59560; E-mail: leevi.kaariainen@helsinki.fi.
SFV P1234 in vitro. Furthermore, both nsP2 and Pro39 could specifically cleave purified recombinant fusion proteins containing short (~40 aa) SFV-specific peptides, which span protease cleavage sites. Using this newly devised in vitro assay, we show that Pro39 was inactivated by N-ethylmaleimide, in accordance with the catalytic mechanism of cysteine proteases.

**Fig. 1.** Schematic presentation and properties of nsP2 deletion mutants. **Filled boxes** represent the deleted regions. **C478** represents catalytic residue of the active site; +, protease activity; −, no protease activity; nd, activity could not be determined.

**Table I**

| Name          | Polarity | Sequence (5' → 3') |
|---------------|----------|--------------------|
| Mut1F         | Sense    | GAGGCGAGGGTCGTGGAACACCC |
| Mut1R         | Anti-sense | TGGGTGACACCTCTCCCTCAAA |
| Mut2F         | Sense    | GAGGCGAGGGTCGTGGAAGTT |
| Mut2R         | Anti-sense | GAGGCGAGGGACCATGCCCTTCCCTCC |
| Mut3F         | Sense    | GAGGCGAGGGTCGTGGAACACCC |
| Mut3R         | Anti-sense | TGGGTGACACCTCTCCCTCAAA |
| Δ210          | Sense    | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ350          | Sense    | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ400          | Sense    | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ458          | Sense    | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ470          | Sense    | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ80           | Anti-sense | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ120          | Anti-sense | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S1/2up        | Sense    | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S1/2down      | Anti-sense | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S2/3up        | Sense    | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S2/3down      | Anti-sense | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S3/4up        | Sense    | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S3/4down      | Anti-sense | ATACCGGACGACCTCGGCTGCTGCTGATG |

**Table II**

| Oligonucleotides used for site-directed mutagenesis and PCR amplification | Restriction sites used in cloning are underlined. |
|---------------------------------------------------------------|-------------------------------------------------|
| Mut1F Sense                                                   | GAGGCGAGGGTCGTGGAACACCC |
| Mut1R Anti-sense                                             | TGGGTGACACCTCTCCCTCAAA |
| Mut2F Sense                                                  | GAGGCGAGGGTCGTGGAAGTT |
| Mut2R Anti-sense                                             | GAGGCGAGGGACCATGCCCTTCCCTCC |
| Mut3F Sense                                                  | GAGGCGAGGGTCGTGGAACACCC |
| Mut3R Anti-sense                                             | TGGGTGACACCTCTCCCTCAAA |
| Δ210 Sense                                                   | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ350 Sense                                                   | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ400 Sense                                                   | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ458 Sense                                                   | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ470 Sense                                                   | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ80 Anti-sense                                               | ATACCGGACGACCTCGGCTGCTGCTGATG |
| Δ120 Anti-sense                                              | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S1/2up Sense                                                 | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S1/2down Anti-sense                                          | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S2/3up Sense                                                 | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S2/3down Anti-sense                                          | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S3/4up Sense                                                 | ATACCGGACGACCTCGGCTGCTGCTGATG |
| S3/4down Anti-sense                                          | ATACCGGACGACCTCGGCTGCTGCTGATG |

**Expression**, Solubility, and Proteolytic activity for the mutant constructs of nsP2.

| Mutant Construct of nsP2 | Expression | Solubility | Proteolytic activity |
|--------------------------|------------|------------|----------------------|
| N-terminal deletion      |            |            |                      |
| P2NΔ60                   | low        | soluble    | +                    |
| P2NΔ120                  | low        | soluble    | +                    |
| P2NΔ180                  | low        | soluble    | +                    |
| P2NΔ240                  | low        | soluble    | +                    |
| P2NΔ300                  | low        | soluble    | +                    |
| P2NΔ350                  | low        | soluble    | +                    |
| P2NΔ400                  | high       | soluble    | +                    |
| P2NΔ458                  | high       | soluble    | +                    |
| P2NΔ470                  | low        | soluble    | +                    |
| C- and N-terminal deletions |            |            |                      |
| P2NΔ300CΔ40              | high       | insoluble  | nd                   |
| P2NΔ300CΔ80              | high       | insoluble  | nd                   |
| P2NΔ300CΔ120             | medium     | 50% insoluble | -                  |
EXPERIMENTAL PROCEDURES

nsP2 Expression Plasmids—Full-length nsP2 of SFV was produced and purified as described previously (15). To obtain a set of nsP2 mutants, containing progressive amino-terminal truncations, nine separate PCR amplifications were carried out using 

\[ Pfu \text{ Turbo DNA polymerase} (\text{Stratagene}), \text{SFV infectious cDNA as a template and} \]

\[ 3' /H11032 \text{-Mut} (\text{Table I}) \text{as common} \]

\[ 3' /H11032 \text{primer}. \text{This oligonucleotide was designed to introduce a G} \rightarrow \text{E mutation in the} \]

\[ 2/3 \text{cleavage site to prevent possible self-proteolysis of the expressed protein at the nsP2-His}6 \text{tag boundary.} \]

\[ \text{Oligonucleotides} N/H9004_{30}, N/H9004_{120}, N/H9004_{180}, N/H9004_{240}, N/H9004_{300}, N/H9004_{350}, N/H9004_{400}, N/H9004_{458}, \text{and} \]

\[ N/H9004_{470} (\text{Table I}) \text{were used as} \]

\[ 5' /H11032 \text{primers. Similarly, a set of bidirectional truncations of nsP2 was also prepared, using} \]

\[ N/H9004_{300} \text{as a common} \]

\[ 5' /H11032 \text{primer and} C/H9004_{40}, C/H9004_{80}, \text{or} C/H9004_{120} (\text{Table I}) \text{as} \]

\[ 3' /H11032 \text{primers.} \]

\[ \text{All PCR products were cleaved with} \]

\[ NcoI-XhoI \text{and cloned into} \]

\[ \text{FIG. 2. Protease domain Pro39 of nsP2. A, a schematic presentation of nsP2. The carboxyl-terminal protease domain of 341 aa (Pro39) with the proposed active site residue Cys-478 is indicated. B, purification of Pro39 by metal affinity chromatography monitored by SDS-PAGE in 10% gel. Lane 1, total extract from the isopropyl-1-thio-\text{-}\beta\text{-}d\text{-galactopyranoside-induced bacterial cells; lane 2, supernatant after centrifugation at} 100,000 \times g; lane 3, purified Pro39 eluted from the nickel-chelate column. Arrow indicates the position of Pro39. Molecular mass markers in kDa are shown in lane M. C, determination of molecular mass of Pro39 by electrospray mass spectrometry. The peak of 39.3 kDa is the calculated molecular mass of Pro39.} \]

\[ \text{FIG. 3. Protease activity of purified nsP2 and Pro39 assayed on in vitro synthesized SFV polyprotein substrates with active site mutation C478A. A, polyproteins P12C3, P2C3, and P34 as substrates. Identification of the reaction products of Pro39 with pertinent antiserum is shown in lanes 4, 5, 9, 10, 14, and 15. In vitro synthesized nsP1 and polyproteins containing nsP1 show regularly a truncated form ΔP1, due to internal initiation of translation. B, polyproteins 12C3 and 12C34 as substrates. Immunoprecipitation analyses are shown in lanes 4–6 and 10–13. The proteins were synthesized using a cell-free in vitro system in the presence of [\text{35S}]methionine for 1 h at 30 °C. The synthesis was stopped by adding cycloheximide. The enzyme preparations (100 ng) were added to the reaction mixture followed by incubation for 1 h at 30 °C. The reaction products were analyzed by SDS-PAGE in 10% gels before and after immunoprecipitation, followed by visualization by autoradiography. Positions of the polyproteins and their cleavage products are indicated with the arrows.} \]
pET21d vector (Novagen) to yield pP2NΔ60, pP2NΔ120, pP2NΔ180, pP2NΔ240, pP2NΔ300, pP2NΔ350, pP2NΔ400, pP2NΔ458, pP2NΔ470, pP2NΔ900Δ340, pP2NΔ900Δ380, and pP2NΔ900Δ1200 expression plasmids, respectively. To obtain plasmid pP2NΔ58^3^, encoding inactive protease, PCR amplification with NΔ58 and 3'-Mut primers was carried out on the SFV CA template (14), and the obtained plasmids, designated as pTrx1^2^, pTrx2^3^, and pTrx3^4^ (their cleavage sites) and pTrx1^2^, pTrx2^3^, and pTrx3^4^ (their non-cleavable analogs), were used to produce respective thioredoxin fusion proteins.

**Protein Expression and Purification**—Recombinant proteins were expressed in *E. coli* BL21(DE3) (Stratagene) and purified using Ni^2+^-affinity chromatography as described previously (16, 15). Protein purification was monitored by 10–17% SDS-PAGE (25), and protein concentration was determined with the Bradford assay (26) using Bio-Rad reagents. Purified proteins (1–5 mg/ml) were stored in a buffer, containing 50 mM Tris-HCl, pH 7.0, 200 mM NaCl, and 1 mM dithiothreitol at 4 °C.

**In Vitro Synthesis of the Nonstructural Polyprotein Substrates**—Coupled transcription-translation of the constructs encoding SFV P12^2^, P2^2^, P12^3^, P12^4^, and P34 was carried out in the T7 TNT rabbit reticulocyte lysate system (Promega) according to the manufacturer’s instructions. Reaction mixtures (10 μl) supplemented with 10 μCi of [35S]methionine (>1000 Ci/mmol; Amersham Pharmacia Biotech) and 1 μg of plasmid DNAs were incubated for 1 h at 30 °C, and the protein synthesis was stopped by adding 1 mM cycloheximide. For protease assay with [35S]labeled nonstructural polyprotein substrates, typically 0.1–0.5 μg of the isolated protease was added to the cyclization-arrested translation mixtures (see above), and the mixture was incubated for 1 h at 30 °C. Reaction products were separated by SDS-PAGE. Gels were dried, and radioactive protein bands were visualized using phosphorimager (BAS-1500, Fuji).

**Protease Assay with Purified Thioredoxin Fusion Protein Substrates**—Protease activity of nsP2 and Pro39 was assayed using thioredoxin (Trx) substrates (see above). Purified substrates and enzymes were incubated for 1 h at 30 °C, and the expected cleavage sites are shown.

A schematic presentation of Trx12, Trx23, and Trx34 substrates. The amino-terminal sequence of thioredoxin is followed by 42–37 aa from the cleavage regions of the 1/2, 2/3, and 3/4 sites; 13 aa from pET32C; and a His11 tag. Six residues on both sides of the expected cleavage sites are shown. The uncleavable fusion proteins Trx1^2^ and Trx2^3^ differ only in replacements of the penultimate glycine (bold) to glutamic acid at the cleavage site. In Trx3^4^, an additional mutation of alanine (bold) to valine was engineered. Arrowheads indicate the predicted cleavage positions. B, proteolysis of Trx34 by nsP2 and Pro39 (lanes 5 and 6); D, proteolysis of Trx12 (lane 4); E, proteolysis of Trx23. The products were analyzed by SDS-PAGE in 17% gels, which were stained by Coomassie Blue 250. The position of Pro39, uncleaved substrates, and the large cleavage product (L) are indicated by arrows. Molecular mass markers in B, D, and E are given in lanes 1, C, separation of large (L; 14 kDa) and small (S; 4 kDa) proteolytic products of Trx34 by Pro39 using Tricine-SDS electrophoresis (28).
Alphavirus Nonstructural Protease

RESULTS

Expression and Purification of Recombinant Proteins—Full-length SFV nsP2 was expressed in E. coli and purified using metal-chelate chromatography as described (15). The same expression strategy was utilized to prepare a set of amino-terminally truncated variant proteins of nsP2: P2N/H9004/30790/P2N/H9004, terminally truncated variant proteins of nsP2: P2N/H9004.

Expression strategy was utilized to prepare a set of amino-terminally truncated variant proteins of nsP2: P2N/H9004/30790/P2N/H9004, terminally truncated variant proteins of nsP2: P2N/H9004.

were mixed in the buffer containing 50 mM HEPES-NaOH, pH 7.2, 20–100 mM NaCl, and 1 mM dithiothreitol, and the mixtures were incubated for 1 h at 30 °C. Reaction products were analyzed by SDS-PAGE and/or mass spectrometry after HPLC purification. In the protease inhibition assays, enzyme was pre-incubated with the inhibitor for 10 min.

Reversed Phase Chromatography—Substrates and the reaction products were separated on a 0.1 x 15-cm Vydac C8 column (300 Å, 5 μm, LC-Packings) using a SMART™ system (Amersham Pharmacia Bio-tech, Uppsala, Sweden). Elution was performed using linear gradients of acetonitrile (0–60% in 100 min) in 0.1% trifluoroacetic acid. Chromatography was monitored for absorbancy at 214 nm, and the peptide-containing fractions were collected automatically.

Mass Spectrometry and NH2-terminal Sequence Analysis—MALDI-TOF mass spectrometry was performed on a Biflex III flight hybrid mass spectrometer (Micromass, Manchester, United Kingdom). Pro39 was dissolved in a mixture of 0.1% trifluoroacetic acid and 50% acetonitrile (1:2) as a matrix. Samples were prepared by mixing 1 μl of reversed phase HPLC eluate with 1 μl of sinapic acid matrix on the target plate and dried under a gentle stream of warm air. All mass spectra were calibrated externally with either cytochrome C or myoglobin as standards. Electrospray ionization mass spectra were obtained using a Micromass Q-TOF quadrupole/time-of-flight mass spectrometer with a syringe pump at a flow rate of 30 nA/h. The mass spectrometer was calibrated using sodium trifluoroacetate as reference standard. The mass spectrometer was calibrated using sodium trifluoroacetate as reference standard. The mass spectrometer was calibrated using sodium trifluoroacetate as reference standard.

FIG. 5. Dependence of Trx12 and Trx34 proteolysis on Pro39 concentration. A, Trx34 (1 μg) was treated by different nanogram amounts of Pro39 as indicated for lanes 1–5. B, Trx12 (1 μg) was treated with microgram amounts of Pro39 as indicated for lanes 1–5. The reactions were carried out for 1 h at 30 °C. The proteolysis products were separated by SDS-PAGE, and the gel was stained with Coomassie Blue R-250. Positions of Pro39, Trx12, Trx34, and large (L; ~14 kDa) products are indicated with the arrows.

FIG. 6. HPLC separation and MALDI-TOF mass spectra of Trx12 and its cleavage products by Pro39. A, reversed phase HPLC purification of Trx12 and MALDI-TOF mass spectrum of the indicated peak in the inset. B, HPLC separation of the reaction products and MALDI-TOF mass spectrometric analyses of S- and l-fragment-containing fractions (filled arrows) in the insets. The peaks corresponding to Trx12 are shown with empty arrows in A and B. P2NΔ458, and P2NΔ470; as well as three bidirectional deletion mutants P2NΔ300CΔ40, P2NΔ300CΔ80, and P2NΔ300CΔ120. All amino-terminally truncated mutants were at least partially soluble. On the other hand, deletions from the carboxyl terminus rendered nsP2 fragments insoluble, except for P2NΔ300CΔ120, which was partly soluble but had no protease activity (Fig. 1). Interestingly, the P2NΔ458 mutant showed solubility comparable to, or even higher than, that of full-length nsP2. Since this 341-aa-long fragment coincides with the putative nsP2 protease domain, it will be hereafter referred to as Pro39 (Fig. 2A). This soluble carboxyl-terminal fragment was purified to near homogeneity from cleared bacterial lysates using metal-chelate chromatography, as shown in Fig. 2B. Pro39 migrated in SDS-PAGE as a 39-kDa protein. Its molecular mass was determined by electrospray mass spectrometry to be 39.3 kDa (Fig. 2C).

Both nsP2 and Pro39 Are Proteolytically Active in Vitro—To assay protease activities of the isolated proteins, several SFV nonstructural polypeptide substrates containing [35S]methionine were synthesized in a cell-free transcription-translation system. These were P12CA, P2CA3, P12CA3, P12CA34 (with a protease-inactivating mutation of C478A in the nsP2 moiety), as well as P34. Both nsP2 and its amino-terminally truncated fragments were active as proteases; however, Pro39 showed the...
highest specific activity (data not shown). From the polyprotein pairs, both preparations cleaved readily at sites 1/2 (Fig. 3A, lanes 2–5) and 3/4 (Fig. 3A, lanes 12–15), whereas cleavage at site 2/3 was much less efficient, but detectable (Fig. 3A, lanes 9 and 10). The two larger polyproteins were cleaved also at sites 1/2 and 3/4 and to some extent at site 2/3, as revealed by immunoprecipitation of the cleavage products (Fig. 3B, lanes 4–6 and 10–13). The cleavage at site 3/4 was more efficient than that at site 1/2, as seen after reduction of enzyme concentration or time of incubation (data not shown). As expected, the C478A mutation inactivated both nsP2 and Pro39 completely (data not shown). Overall, these data demonstrate site-specific protease activity in vitro of purified nsP2 and Pro39. The high solubility and specific activity of Pro39 suggest that this fragment represents a structurally compact protease domain of nsP2.

Protease Activity of nsP2 and Pro39 Assayed with Purified Substrates—In the following experiments, protease activity of nsP2 and Pro39 was assayed with a set of purified recombinant Trx fusion proteins as substrates (Fig. 4A). These fusion proteins (~18 kDa) contained approximately 40-aa sequences spanning the SFV P1234 polyprotein cleavage sites (Trx12, Trx23, and Trx34). As controls we used non-cleavable analogs with point mutations (glycine to glutamic acid) in the penultimate amino acid of the predicted cleavage site (Trx1^2, Trx2^3) as well as an additional mutation (Ala → Val) for 3/4 cleavage site yielding Trx3^4 (Fig. 4). Both nsP2 and Pro39 cleaved Trx34 with the highest efficiency. As expected, two proteolytic products were formed, one larger (L; ~14 kDa) and one smaller (S; ~4 kDa). The L-fragment (thioredoxin moiety plus 19 carboxyl-terminal residues from nsP3) could be detected in standard SDS-PAGE gels (Fig. 4B). The S-fragment migrated in this gel system in the front (data not shown).

**TABLE II**

| Substrate | Full-length mass (Da) | L-fragment mass (Da) | S-fragment mass (Da) | S-fragment N-terminal sequence |
|-----------|----------------------|---------------------|---------------------|-------------------------------|
|           | Predicted | Mass spectrometry | Predicted | Mass spectrometry | Predicted | Mass spectrometry | Predicted | Mass spectrometry |
| Trx12     | 18,560    | 18,568             | 14,261    | 14,248             | 4,317    | 4,310             | GVVETPR. . . |
| Trx23     | 18,514    | 18,509             | 14,173    | ND                 | 4,359    | 4,353             | APXYRVK. . . |
| Trx34     | 18,158    | 18,073             | 14,057    | 14,042             | 4,119    | 4,116             | YIFSSDT. . . |

*a* Determined by Edman degradation method.  
*ND,* not determined.

**Fig. 7.** Kinetics of the proteolytic cleavage of Trx34 catalyzed by Pro39. A 50-μl reaction mixture contained 20 μg of Trx34 and 100 ng of the enzyme. Aliquots of 5 μl were taken during 120 min of incubation at 30 °C, and then reaction was stopped by addition of sample buffer and analyzed by SDS-PAGE in 17% gel. The gel was stained with Coomassie Blue R-250. Arrows show the positions of Pro39, Trx34, and 14-kDa large proteolytic product (L).

**Fig. 8.** Effects of inhibitors (A) and divalent cations (B) on proteolytic activity of Pro39. Pro39 (10 ng) was incubated with the following inhibitors: 10 mM EDTA; 2.5 mM NEM; 100 μM E-64; and 2 mM Zn2+, Co2+, Mg2+, Mn2+, Ca2+, and Ni2+ for 20 min prior to addition of 5 μg of Trx34, followed by 60-min incubation at 30 °C. Lanes 2 contain Trx34 incubated alone. Lanes 3 show the result of pre-incubation of Pro39 with the reaction buffer (Tris-HCl, pH 7.5, 25 mM NaCl) prior to substrate addition. Molecular mass markers (kDa) are indicated in lanes 1. Arrows show the positions of Pro39, Trx34, and the large proteolytic product of 14 kDa (L).
However, Tricine SDS-PAGE method (28) resolved this fragment as a defined band (Fig. 4C). Pro39 also cleaved Trx12, as judged by the appearance of the l-fragment (Fig. 4D). Under the same conditions, the cleavage products obtained with nsP2 as protease and Trx12 as substrate were barely detectable (data not shown). No cleavage of Trx23 could be detected with either nsP2 or Pro39 when analyzed by SDS-PAGE (Fig. 4E). Neither nsP2 nor Pro39 could cleave Trx12, Trx23, or Trx34, which served as controls (Fig. 4, B, D, and E). The relative cleavage efficiency by Pro39 of sites 1/2 and 3/4 was further analyzed using serial dilutions of the enzyme (Fig. 5). From these results it could be estimated that complete cleavage of Trx12 required ~5000-fold more enzyme than cleavage of Trx34.

HPLC Purification and MALDI-TOF Mass Spectrometry of the Proteolytic Products—In addition to the SDS-PAGE analysis, the proteolytic products of Trx12, Trx23, and Trx34 were also separated by reversed phase HPLC (see “Experimental Procedures” for details). Fractions containing S- and L-fragments, as well as non-cleaved substrates, were analyzed by MALDI-TOF mass spectrometry as shown in detail for Trx12 (Fig. 6). Amino-terminal sequences of the S-fragments were also determined. The results of these experiments confirmed that Pro39 cleaves Trx-fusion protein substrates exactly at the predicted positions of P1234, and were supported by previous NH4-terminal radiosequence analyses of in vivo labeled nsP2, nsP3, and nsP4 (29, 30) (Table II). Importantly, this approach was sensitive enough to detect the cleavage of the Trx23 substrate. Thus, 2/3 site can be hydrolyzed correctly in vitro by Pro39 but with a very poor efficiency (Table II).

Characterization of the Pro39-catalyzed Reaction—Effects of several reaction parameters on the protease activity of Pro39 were studied systematically. The enzyme was active over a broad range of pH values (pH 6.8–9.5) and different ionic strengths (0–500 mM NaCl), and the optimal reaction temperature was 30 °C (data not shown). Omission of reducing reagents from the reaction mixture had no detectable effect on Pro39 activity. The time course of the proteolysis under optimized conditions was also studied. In this experiment, Pro39 demonstrated a very high specific activity, hydrolyzing 50% of 400-fold molar excess of Trx34 substrate in 5 min (Fig. 7).

The effect of different protease inhibitors on the enzymatic activity of Pro39 was also tested. The enzyme is completely resistant to the inhibitors of serine proteases (PMSF), metalloproteases (EDTA), aspartic proteases (pepstatin), and some cysteine protease inhibitors (leupeptin and E-64) (Fig. 8A). Cleavage of both Trx12 and Trx34 was completely inhibited by 2.5 mM N-ethylmaleimide (NEM), a typical cysteine protease inhibitor. Surprisingly, the protease was also insensitive to some divalent cations in the reaction mixture. Addition of 2 mM Zn2+ or Cu2+ resulted in total inactivation of the protease activity, and Co2+ and Ni2+ caused partial inhibition (Fig. 8B). On the other hand, the same concentrations of Ca2+, Mg2+, and Mn2+ had no effect on Pro39 activity.

**DISCUSSION**

Previous work on Semliki Forest virus showed that early in infection the synthesis of the negative strand RNA was strictly dependent on protein synthesis and ceased in about 15 min after addition of cycloheximide (31), whereas late in infection the synthesis of positive strand RNAs continued for several hours in the absence of protein synthesis. Solution to this dilemma came from findings with Sindbis virus, another alphavirus, where the processing intermediate P123 of the nonstructural polyprotein together with nsP4 was shown to be responsible for the synthesis of the negative strand RNA (32–34). Cleavage of P123 is essential for the synthesis of the positive RNA strands. Thus, the regulated processing of the nonstructural polyprotein controls the early events of virus infection. An overactive processing mutant of Sindbis virus nsP2 (N614D) cannot replicate, as the P123 intermediate is too short-lived to enable the necessary synthesis of the complementary RNA (35).

Our knowledge of the processing of alphavirus nonstructural polyprotein(s) is based mostly on experiments of *in vitro* translation of Sindbis virus RNA. Ingenious constructions by which the cleavage sites were mutated alone and in different combinations, together with constructs coding for enzymatically inactive polyprotein as substrates, have been used to analyze this complex process (4, 8, 9). These experiments showed that the polyprotein P1234 itself and all its cleavage intermediates containing nsP2 to be active proteases. The cleavability of the different sites varied and was dependent on the order of removal of different nsPs from the polyprotein substrate. Particularly interesting was the finding where the cleavage of site 2/3 in P1234 or P123 was only possible after the cleavage of nsP1 (36). Evidently, conformational changes in P1234 and its cleavage products affect the interactions between the cleavage sites and the protease domain of nsP2 in a complex manner. To understand the processes better, we have characterized purified SFV nsP2 and its carboxyl-terminal fragment Pro39 as proteolytic enzymes. As substrates we used *in vitro* synthesized polyproteins P1234, P123, P23, and P34, as well as recombinant thioredoxin fusion proteins, which contain short SFV-specific fragments, spanning the polyprotein processing sites 1/2, 2/3, and 3/4.

We show for the first time that purified nsP2 has proteolytic activity, which cleaves readily the 3/4 site of P1234 and P34. Deletion series of nsP2 resulted in a soluble, active carboxyl-terminal fragment consisting of amino acid residues 459–799, which was designated as Pro39. It was purified to near homogeneity by metal-affinity chromatography. The specificities of Pro39 and nsP2 were identical, indicating that the amino-terminal half of nsP2 does not affect the fidelity of the protease. According to sequence alignments with the thiol protease superfamily, Pro39 contains a conserved protease domain (459–600), but also almost 200 carboxyl-terminal extra amino acids (9, 13). Our attempts to delete 40–120 amino acids from the carboxyl terminus of nsP2 resulted in insoluble or inactive proteins (Fig. 1). Experiments with temperature-sensitive mutants of SIN and SFV nsP2 have shown that amino acid replacements N700K in SIN ts133, K736S in SIN ts24, and M781T in SFV ts4 result in inhibition of protease activity at 39 °C, suggesting that the extreme carboxyl terminus of nsP2 participates somehow in the protease function (24).

Establishment of a biochemical assay system consisting of isolated Pro39 and thioredoxin attached cleavage regions of the SFV nonstructural polyprotein allowed characterization of the viral protease under defined experimental conditions. Pro39 was inactivated by N-ethylmaleimide but not with pepstatin, EDTA, or PMSF. These are properties, which are in accordance with its classification as a thiol protease of the papain superfamily. However, Pro39 is not inhibited by E-64, which is a typical inhibitor of cysteine proteases (37). Sensitivity for NEM and resistance for E-64 have been previously reported for poliovirus 3C thiol proteinase (38, 39). Another interesting feature of Pro39 is the inhibition by zinc ions (Fig. 8).

When Pro39 (or nsP2) was added to the reaction mixture, after *in vitro* translation of P12 or P123, almost a quantitative release of nsP1 was observed. Similarly, when P34 or P1234 were used as substrates, quantitative release of nsP4 was seen, whereas only a small amount of nsP3 was released from P23, P123, or P1234 (Fig. 3). These results suggested that sites 1/2 and 3/4 were exposed to the added protease, whereas site 2/3
was not. To study this phenomenon under controlled conditions, in which the large protein domains would not interfere sterically with the proteolysis, we constructed fusion proteins with a different number of amino acid residues around the cleavage sites.

Constructs with less than 10 amino acids on both sides of the cleavage site were not digested by Pro39 or nsP2 (data not shown). We ended up using thioredoxin fusion proteins with about 40 residues of each cleavage region (Trx12, Trx23, and Trx34). Isolation of the cleavage products and their mass spectrometric analysis, as well as amino-terminal sequencing showed that Pro39 cleavage products were derived exactly from the predicted cleavage sites, determined previously by radiosequence analysis (29, 30) (Table II). As controls we used thioredoxin fusion proteins with mutations close to the cleavage site, which were not digested by Pro39 or nsP2 (Fig. 4). Thus, we conclude that both proteases recognize specifically the three cleavage sites of the SFV nonstructural polyprotein. However, there were large differences in the sensitivity of the different cleavage sites, the 3/4 junction being most sensitive. Roughly 5000-fold more Pro39 was needed for complete cleavage of site 1/2 (Fig. 5). Under the same conditions, only a small amount of Trx23 was cleaved.

The different sensitivities of the three cleavage sites may well reflect the different specificities of the protease associated with the polyproteins, in which the cleavage at site 2/3 of P123 or P1234 does not take place unless preceded by cleavage of nsP1. The fact that Pro39 can catalyze cleavage at 1/2 site of in vitro translated SFV P12, P123, and P1234, which normally undergoes cleavage in cis (14), is better understood when real-time reading of the manuscript.

Acknowledgments—We thank Airi Sinkko for excellent technical assistance. We are grateful to Dr. Nisse Kalkkinen for valuable advice and discussions. We thank Dr. Marja Makarov and Dr. Tero Ahola for critical reading of the manuscript.

REFERENCES

1. Ryan, D. M., and Flint, M. (1997) J. Gen. Virol. 78, 699–723
2. Ryan, D. M., Monaghan, S., and Flint, M. (1998) J. Gen. Virol. 79, 947–959
3. Ziebuhr, J., Snijder, J. E., and Gorbalenya, A. E. (2000) J. Gen. Virol. 81, 853–879
4. Strauss, J. H., and Strauss, E. G. (1994) Microbiol. Rev. 58, 491–562
5. Schlesinger, M. J. (1999) in Encyclopedia of Virology (Granoff, A., and Webster, R. G., eds) pp. 1566–1663, 2nd Ed., Oxford University Press, Oxford
6. Frey, T. K., and Wolinsky, S. J. (1999) in Encyclopedia of Virology (Granoff, A., and Webster, R. G., eds) pp. 1592–1601, 2nd Ed., Oxford University Press, Oxford
7. Kaariainen, L., Takkinen, K., Keranen, S., and Soderlund, H. (1987) J. Cell Sci. Suppl. 7, 231–250
8. Strauss, J. H., and Strauss, E. G. (1990) Semin. Virol. 1, 347–356
9. ten Dam, E., Flint, M., and Ryan, M. D. (1999) J. Gen. Virol. 80, 1879–1888
10. Hardy, W. R., and Strauss, J. H. (1989) J. Virol. 63, 4653–4664
11. Ding, M. X., and Schlesinger, M. J. (1989) Virology 171, 260–264
12. Strauss, E. G., de Greot, R. J., Levinson, R., and Strauss, J. H. (1992) Virology 191, 932–940
13. Gorbalenya, A. E., Koonin, E. V., and Lai, M.-C. (1991) FEBS Lett. 288, 201–203
14. Merits, A., Vasiljeva, L., Ahola, K., Kaariainen, L., and Auvinen, P. (2001) J. Gen. Virol. 82, 765–773
15. Vasiljeva, L., Merits, A., Auvinen, P., and Kaariainen, L. (2000) J. Biol. Chem. 275, 17281–17287
16. Rikkenen, M., Peranen, J., and Kaariainen, L. (1994) J. Virol. 68, 5804–5810
17. Gomez de Cedro, M., Ehsani, N., Mikkola, M. L., Garcia, J. A., and Kaariainen, L. (1999) FEBS Lett. 445, 19–22
18. Peranen, J., Rikkenen, M., Liljestrom, P., and Kaariainen, L. (1990) J. Virol. 64, 1888–1896
19. Peranen, J., Rikkenen, M., and Kaariainen, L. (1993) J. Histochem. Cytochem. 41, 447–454
20. Rikkenen, M., Peranen, J., and Kaariainen, L. (1992) Virology 189, 462–473
21. Rikkenen, M. (1996) Virology 218, 352–361
22. Keranen, S., and Kaariainen, L. (1979) J. Virol. 47, 565–515
23. Sawicki, D. L., and Sawicki, S. G. (1985) Virology 144, 20–34
24. Suopanki, J., Sawicki, D. L., Sawicki, S. G., and Kaariainen, L. (1988) J. Gen. Virol. 79, 309–319
25. Cleveland, D. W., Fischer, S. G., Kirschner, M. W., and Laemmli, U. K. (1977) J. Biol. Chem. 252, 1102–1106
26. Bradford, M. M. (1976) Anal. Biochem. 72, 248–254
27. Moini, M., Jones, B. L., Rogers, R. M., and Jiang, L. (1998) J. Am. Soc. Mass Spectrom. 9, 977–980
28. Schagger, H., and von Jagow, G. (1987) Anal. Biochem. 166, 368–379
29. Kalkkinen, N., Laaksonen, M., Soderlund, H., and Jorvall, H. (1981) Virology 113, 188–195
30. Kalkkinen, N. (1986) in Advanced Methods in Protein Microsequence Analysis (Wittmann-Liebold, B., Salnikow, J., and Erdmann, V. A., eds) pp. 194–206, Springer-Verlag, Heidelberg
31. Sawicki, D. L., and Sawicki, S. G. (1980) J. Virol. 34, 108–118
32. Lemm, J. A., and Rice, C. M. (1993) J. Virol. 67, 1905–1915
33. Lemm, J. A., and Rice, C. M. (1993) J. Virol. 67, 1916–1926
34. Shirako, Y., and Strauss, J. H. (1994) J. Virol. 68, 1874–1885
35. Lemm, J. A., Runenapf, T., Strauss, E. G., Strauss, J. H., and Rice, C. M. (1994) EMBO J. 13, 2925–2934
36. Shirako, Y., and Strauss, J. H. (1990) Virology 177, 54–64
37. Beynon, R. J., and Bond, J. S. (1989) in Proteolytic Enzymes: A Practical Approach (Beynon, R. J., and Bond, J. S., eds) pp. 83–104, Oxford University Press, Oxford, United Kingdom
38. Lawson, M. A., and Semler, B. L. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 9912–9923
39. Yu, P. S., and Lloyd, R. R. (1991) Virology 182, 615–625
40. Failla, C., Tomoi, L., and De Francesco, R. (1994) J. Virol. 68, 3753–3760
41. Lin, C., and Rice, C. M. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 7622–7626
42. Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R., and Thomson, J. A. (1996) Cell 87, 345–355
43. Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T., and Hostomsky, Z. (1996) Cell 87, 331–342
44. Shimizu, Y., Yamaji, K., Masuho, Y., Yokota, T., Inoue, H., Sudo, K., Satoh, S., and Shimotono, K. (1996) J. Virol. 70, 127–132
45. Kujala, P., Ikaiheimonen, A., Ehsani, N., Viihinen, H., Auvinen, P., and Kaariainen, L. (2001) J. Virol. 75, 3873–3884