Macromolecular Assembly-Driven Processing of the 2/3 Cleavage Site in the Alphavirus Replicase Polyprotein

Aleksei Lulla, Valeria Lulla, and Andres Merits

Institute of Technology, University of Tartu, Nooruse 1, Tartu 50411, Estonia

Semliki Forest virus (SFV) is a member of the *Alphavirus* genus, which produces its replicase proteins in the form of a nonstructural (ns) polyprotein precursor P1234. The maturation of the replicase occurs in a temporally controlled manner by protease activity of nsP2. The template preference and enzymatic capabilities of the alphaviral replication complex have a very important connection with its composition, which is irreversibly altered by proteolysis. The final cleavage of the 2/3 site in the ns polyprotein apparently leads to significant rearrangements within the replication complex and thus denotes the “point of no return” for viral replication progression. Numerous studies have devised rules for when and how ns protease acts, but how the alphaviral 2/3 site is recognized remained largely unexplained. In contrast to the other two cleavage sites within the ns polyprotein, the 2/3 site evidently lacks primary sequence elements in the vicinity of the scissile bond sufficient for specific protease recognition. In this study, we sought to investigate the molecular details of the regulation of the 2/3 site processing in the SFV ns polyprotein. We present evidence that correct macromolecular assembly, presumably strengthened by exosite interactions rather than the functionality of the individual nsP2 protease, is the driving force for specific substrate targeting. We conclude that structural elements within the macrodomain of nsP3 are used for precise positioning of a substrate recognition sequence at the catalytic center of the protease and that this process is coordinated by the exact N-terminal end of nsP2, thus representing a unique regulation mechanism used by alphaviruses.

Semliki Forest virus (SFV) and Sindbis virus (SINV) are the best-studied members of the genus *Alphavirus* (family *Togaviridae*), which consists of positive-strand RNA viruses that infect vertebrate hosts and insect vectors (69). Their RNA replication occurs on cellular membranes and involves two steps: the synthesis of negative strands followed by the synthesis of two classes of positive strands, new genomes and subgenomic (SG) mRNAs, which express structural proteins. Synthesis of the viral RNA is not only highly asymmetric (positive strands are made in great excess), but it is also temporally regulated: the synthesis of negative-strand RNA ceases 3 to 4 h postinfection, whereas the production of positive strands continues throughout the infectious cycle (61).

Alphavirus replication strategy and its regulation relies on the production of replicase proteins in the form of a nonstructural (ns) polyprotein precursor called P1234 (or P123 and P1234 polyproteins for SINV), which is then co- and posttranslationally processed (69). These events are directly related to the assembly of replicase complexes: if the replicase proteins (nsPs) are coexpressed as separate proteins, they fail to interact with each other, are mislocalized, and cannot support RNA replication (37, 60). Processing of the ns polyprotein is performed by the papain-like protease activity residing in nsP2 and is highly regulated (8, 28, 29, 73). It is generally accepted that during the course of viral infection, sequential processing of the ns polyprotein leads to rearrangements of the viral replication complex (reviewed in reference 31). Indeed, it has been shown that unprocessed P1234 is incapable of supporting alphavirus replication and that the short-lived negative-strand polymerase consists of P123 and nsP4, which is released by the first processing event; the stable positive-strand replicase consists of fully processed nsPs (34, 38, 39, 65). An intermediate form of the replicase, consisting of nsP1, P23, and nsP4, has been investigated with the use of mutant viruses unable to process the cleavage site between nsP2 and nsP3. This replicase can synthesize both positive and negative strands, but it prefers the

promoter for genomic positive-sense RNA to that of SG RNA (23, 34); replication complexes of such mutant viruses are unstable (47). The role of the intermediate form of the replicase, if any, in the replication of alphaviruses is unclear, because in cells that are infected with wild-type (wt) alphaviruses, the P23 cannot be detected, even with very short pulses (45) or at the earliest time points after infection (8, 30), indicating that this processing intermediate has an extremely short half-life.

The nsP2 is the sole protease responsible for the processing of SFV P1234 (49). It recognizes and cleaves all three cleavage sites, located between nsP1 and nsP2 (here referred to as the 1/2 site), nsP2 and nsP3 (the 2/3 site), and nsP3 and nsP4 (the 3/4 site). The sequences of these cleavage sites are generally not highly conserved within the same virus or between the different alphaviruses (45). The presence of a Gly residue at the P2 position of the processing site (nomenclature according to Schechter and Berger [63]) is the only absolutely conserved feature (21). Clear rules for the processing of P1234 of SFV have been identified. The 3/4 site is processed relatively independently from the others; short substrates containing this site can be processed efficiently by nsP2 or its protease domain alone in a cell-free *trans*-cleavage reaction (45, 72, 73). During the formation of the replicase complex of SINV, this cleavage is also the first to take place (64). The second cleavage is an *in cis* cleavage of the 1/2 site, which is rapidly followed by an *in trans* cleavage of the 2/3 site (73). It is very likely that the order of these cleavages is similar for SINV P1234 as well (8, 64), al-

Received 25 May 2011 Accepted 19 October 2011

Published ahead of print 26 October 2011

Address correspondence to Aleksei Lulla, aleksei.lulla@ut.ee.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.05195-11

though the processing of the 2/3 site prior to the 1/2 site has been reported in early studies (29). Detailed studies of the *in trans* processing of the SFV 3/4 site revealed that the main determinants of the cleavage efficiency are located in the region preceding the scissile bond and that the protease recognizes at least the residues P4 to P1'; the sequence of this cleavage site most likely reflects a compromise between efficiency of protease recognition and other requirements of the viral life cycle (45). However, even for alphaviruses, whose three processing sites are substantially different, the existence of the fixed processing order cannot be fully explained by the differences in the primary sequence of the processing sites. For example, it was found that the amino acid (aa) residues in positions P4, P3, P2, and P1 of the 3/4 site of SFV are rather reluctant to accept substitutions; i.e., changes in these positions resulted in a significant reduction or a complete block (P2 position) of *trans*-cleavage. However, the same mutations in the P4 or P3 positions had little to no effect on the infectivity of the viral genome (45), indicating that either very small amounts of free nsP4 were sufficient for viral replication and/or that during formation of the replicase, the 3/4 site is primarily processed in *cis* and a different set of sequence requirements exists for this mode of 3/4 site cleavage.

In contrast to the 1/2 and 3/4 sites, the 2/3 site is obligatorily processed in *trans* (73). This property was elucidated by resolving the three-dimensional (3-D) structure of the protease domain of a related alphavirus, which revealed that the C terminus of nsP2, and thus the scissile peptide bond of the 2/3 site, is located too far from the protease active site to be processed in an *in cis* reaction (59). Another specific feature of the 2/3 site cleavage is the requirement for the full-length nsP2 protease, indicating that sequences at or near the N terminus of nsP2 are involved (45, 72, 73). Furthermore, the free N-terminal region of nsP2 is needed, because the nsP2 included in the P123 precursor cannot process this site, whereas it can do so in the form of the P23 precursor (8). Additionally, it has been shown that a short, 17-aa extension at the N terminus of nsP2 also blocks its ability to process 2/3 site-containing substrates (73). Finally, in contrast to the situation with other cleavage sites, neither the protease domain of nsP2 nor full-length nsP2 is capable of recognizing and cleaving a short substrate corresponding to the 2/3 site (45, 72), indicating that its recognition requires additional factor(s).

Despite the finding that alphavirus mutants that are unable to process the 2/3 site are nonetheless viable in cell culture (23, 34, 65), this processing event is extremely important for infection. First, it represents the "point of no return" in replicase formation: once the 2/3 site cleavage is performed, the ability of the replicase to initiate the synthesis of negative-strand RNAs is abolished (38, 39, 65). Second, the 2/3 site cleavage releases mature nsP2. Because only about 25% of nsP2, which is produced during infection, is included in the replicase complexes (54), part of nsP2 is present in the form of free protein. Two-thirds of this free nsP2 is transported into the nucleus, where it causes cytotoxic effects at least in vertebrate cells infected by Old World alphaviruses (19, 20) and counteracts the activation of antiviral responses (23). Third, the rest of the released nsP2 is dispersed throughout the cytoplasm of infected cells, where it changes the processing pattern of newly made P1234 polyproteins presumably by rapidly cleaving them at the 2/3 site (30), thereby abolishing the formation of new negative-strand replicase complexes and subsequently preventing the formation of positive-strand replication complexes (73).

Thus, it is the processing of the 2/3 site that triggers the cessation of negative-strand RNA synthesis, which is a hallmark feature of alphavirus infection (61, 62). Fourth, although it has never been directly demonstrated, several studies have suggested that, for alphaviruses, the phenomenon of superinfection exclusion is dependent on free cytoplasmic nsP2 protease, which supposedly blocks the formation of the early replicase (unprocessed P123 and nsP4) of superinfecting viruses by processing of P123 or P1234 polyproteins (32, 34).

All of these findings suggest that the processing of the 2/3 site is one of the key events in the alphavirus infection cycle. Nevertheless, the conditions that are required for this cleavage have so far not been determined. Here, we present evidence that the recognition of the 2/3 site is not based on the recognition of the short substrate sequence by the protease domain of nsP2 but that it requires the presence of the conserved macrodomain of nsP3 in the substrate and the free N terminus of the full-length nsP2 in the enzyme. Furthermore, we conclude that amino acids residues, located distantly from the scissile bond at the end of the macrodomain (or perhaps between the macrodomain and the second domain of nsP3), play a crucial role in 2/3 site processing. The analysis of the corresponding mutant viral genomes confirmed these findings and revealed that the sequences that are involved in 2/3 site processing also have other important functions in viral infection. Our findings support a model in which the processing of the 2/3 site is intrinsically linked to the macromolecular assembly of alphavirus ns proteins into the viral replication complex.

MATERIALS AND METHODS

Protein expression and purification. The expression constructs for SFV nsP2 protease and its variants were prepared using derivatives of the pBAT-1 vector (55). The N terminus of nsP2 was fused to the thioredoxin domain, the GS-spacer, and the sequence preceding the efficiently processed 3/4 site (thioredoxin-GSGSGS-SGITFGDDVLRIGRAGA). This design was chosen to generate the methionine-free N terminus of the nsP2 variants through autoproteolysis inside the cells. The absence of the tag and the authenticity of the expected N terminus were verified by Edman sequencing of the final nsP2 preparations. The C terminus of nsP2 was fused to the sequence LEHHHHHH for purification purposes. The penultimate nsP2 Gly residue was replaced with Glu to prevent potential autocleavage at the C terminus. The construct for the 174-aa-residue-long N-terminal domain of SFV nsP2 (N174 or NTD) was designed in the same way. In this case, the N-terminal thioredoxin carrier was removed during purification by the addition of purified SFV Pro39 protease, which was subsequently removed by chromatography. The construct for SINV nsP2 expression was designed similarly, except that the SINV 1/2 site upstream sequence (IEAAAEVVCVEGLQADIGA) was used within the N-terminal tag. The constructs for the 2/3 site substrates were based on the pET32b vector (Novagen). The indicated upstream and downstream regions of the SFV and/or SINV 2/3 sites were fused to the thioredoxin-GSGSGS carrier at the N terminus and to the NSSSVDKLAALHHH HHHH sequence at the C terminus.

Both the proteases and the substrates were produced in the Rosetta2(DE3) strain of *Escherichia coli* (Novagen). After overnight expression at 17°C, cells were harvested, washed, and broken with a French press in a buffer containing 20 mM HEPES (pH 8.0), 350 mM NaCl, and 10 mM MgCl₂ and supplemented with complete EDTA-free protease inhibitor cocktail (Roche) and benzonase nuclease (Novagen). The lysates were clarified by centrifugation, supplemented with 25 mM imidazole, and loaded onto a column with Ni-nitrilotriacetic acid (NTA) Superflow resin (Qiagen). After being washed extensively, the captured proteins were eluted by elevated imidazole concentrations (up to 200 mM imidazole). For substrate preparations, the eluted proteins were transferred to a

FIG 1 Mapping of the sequence elements involved in the processing of the 2/3 site in the SFV ns polyprotein. (A) Schematic representation of the P1234 polyprotein. Arrows above the bar indicate the processing sites, numbers indicate the order of processing events during replicase maturation, *cis* or *trans* indicate the processing mode, and the star symbol indicates the protease active site. Lower bars show the basic organization of the main constructs used in the current study. NTD and Pro, N-terminal (N174) and protease domains of nsP2, respectively; macro, macrodomain of nsP3. (B) Processing of the substrates containing the 2/3 site with nsP2 protease. Substrates were *in vitro* translated in the presence of [³⁵S]methionine; recombinant nsP2 protease was added for processing (+ lanes). Samples were separated by SDS-PAGE; radioactive products were visualized using a Typhoon imager. (C) Processing of purified recombinant proteins containing the 2/3 site by recombinant nsP2. Substrates at a 5 μ M concentration were incubated with nsP2 protease in a 0.2 μ M concentration; components and products of cleavage reactions were subjected to SDS-PAGE and visualized with Coomassie blue staining. Substrate names are given at the top of the lines, and the incubation times are shown below the panel. The positions of the components and products of the cleavage reactions are indicated at the right of the panel, macro and Trx-20 cleavage products, corresponding to the P' side and the P side of substrate, respectively. (D) Processing of the P1234 polyprotein in cells infected with SFV4 (wt) and SFVΔ25-49. BHK-21 cells were infected with 100 PFU/cell and labeled at 3 h postinfection by [³⁵S]methionine for 15 min; then, they were either not chased (pulse) or chased with unlabeled methionine for 10 or 60 min. Viral proteins were captured by immunoprecipitation with anti-nsP1 and anti-nsP2 antibodies (lines marked "1+2") or anti-nsP3 and anti-nsP4 antibodies ("3+4"), resolved by SDS-PAGE, and visualized with a Typhoon imager. The position of the polyproteins, processing intermediates, and mature SFV nsPs are indicated on the right side of the gel.

buffer containing 20 mM Tris-HCl (pH 8.0) and 50 mM NaCl using PD10 desalting columns (GE Lifesciences), loaded onto a 1-ml Resource Q column (GE Lifesciences), and purified with a 50 to 500 mM NaCl gradient. For proteases, the Ni-NTA eluate was purified further by size exclusion chromatography on a HiLoad 16/60 Superdex 200 pg column (GE Lifesciences) in a buffer containing 20 mM HEPES (pH 8.0), 150 mM NaCl, and 1 mM dithiothreitol. The peak fractions were concentrated on a 1-ml Resource S column (GE Lifesciences) by a 150 to 500 mM NaCl gradient. The concentration of the proteins in the preparations was measured by UV spectroscopy using extinction coefficients calculated based on their amino acid compositions (53).

Proteolytic assay with purified proteins. To assess the proteolytic activity of the purified proteins, 0.2 μ M protease and 5 μ M substrate were combined in a buffer containing 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 2 mM EDTA, and 1 mM dithiothreitol. After 6 or 60 min of incubation at 28°C, the reaction was stopped by the addition of an equal volume of 2 $\times$ Laemmli sample buffer. The components of the reaction mixture were separated by SDS-PAGE (12% Laemmli gels) and visualized by Coomassie R-250 (Sigma) staining; a volume corresponding to 5 μ l of the original mixture was loaded per gel lane. The basic design of the analysis was modified in some experiments, as indicated in Results.

***In vitro* translation and protease assay.** The region encoding the 2/3 site in the SFV ns polyprotein and sequences around it were amplified by PCR and cloned into the pTM1 vector (50) under the control of the T7 RNA polymerase promoter and an encephalomyocarditis virus internal ribosome entry site. The constructs bearing 19 aa residues upstream of the cleavage position (Fig. 1B) were designed to take advantage of the natural Met residue at the P19 position of the substrate. The polypeptides were expressed in the T7 TNT rabbit reticulocyte lysate system (Promega) according to the manufacturer's protocol and in the presence of [³⁵S]methionine. After 60 min of synthesis, translation was stopped by the addition of 1 mM cycloheximide; 0.5 μ g of nsP2 per 10 μ l of the lysate was added, and the incubation was continued for 60 min at 28°C. No protease was added to the corresponding control reactions. Aliquots of the proteolytic reactions (1 μ l) were separated by SDS-PAGE, and the radioactive products were visualized using a Typhoon imager (GE Life Sciences).

The mutations E154R, D157R, R159E, E163R, R159E+E163R, or R159A+E163A in the nsP3 were introduced into the coding sequence of SFV P123 in the pTM1 vector by PCR-based mutagenesis. To obtain the P123^{S4} expression construct of SINV, the stop codon at the end of nsP3 was changed to a Ser codon; to obtain the P12^{CA}S4 construct, the codon

corresponding to the Cys residue in the nsP2 active site of P123^{S4} was replaced with an Ala codon. To obtain the SFV P12^{CA}34 construct, a G798I mutation was introduced into the nsP2 region of P12^{CA}34. All of these sequences, as well as sequences encoding the P1234 and P12^{CA}34 polyproteins of SFV (49), were cloned into a pTM1 vector and were expressed and analyzed as described above.

Analysis of virus replication and polyprotein processing in infected cells. Deletions of the nsP3 25 to 49 or 23 to 68 aa regions or point mutations of nsP3 residues (R159E, E163R, their combination, and a combination of R159A+E163A mutations) were introduced into the infectious cDNA clone pSFV4 (41) using PCR-based mutagenesis and subcloning. A similar approach was used to introduce the G798I mutation into the nsP2 region of pSFV4 (the resulting construct was designated pSFV4-2'3). Capped RNA transcripts of the SpeI-linearized templates were electroporated into BHK-21 cells to assay the efficiency of plaque formation (25) and to generate virus stocks. The virus titers, plaque sizes, and induction of cytopathic effects (CPEs) were all assayed using BHK-21 cells. The analysis of the sequences of rescued viruses was carried out as follows. For each mutant, up to eight different isolates were obtained by plaque purification. Then the viral RNA was extracted from each isolate using TRIzol reagent (Invitrogen), reverse transcribed using the first-strand cDNA synthesis kit (Fermentas), and PCR amplified using SFV-specific primers. PCR fragments corresponding to the mutated region were sequenced for all plaque-purified isolates. To identify possible second-site mutations, a set of PCR fragments covering the SFV genome (excluding its structural part) was sequenced for at least two individual plaque-purified isolates.

Analysis of the processing of P1234 in infected cells was performed as described previously (45). BHK-21 cells were infected with SFV4, SFVΔ25–49, or plaque-purified isolates of SFV4-R159E, SFV4-E163R, SFV4-R159A+E163A, or SFV4-2'3 with 100 PFU per cell. At 3 h postinfection, the cells were labeled by [³⁵S]methionine for 15 min and chased with unlabeled methionine for 10 or 60 min. Cells were collected and lysed at these time points, viral proteins were captured by immunoprecipitation with antibodies against the ns proteins, resolved by SDS-PAGE, and visualized with a Typhoon imager.

RESULTS

The presence of at least 165 N-terminal aa residues of nsP3 is required for efficient cleavage of a substrate representing the 2/3 site. To map sequences involved in the recognition of the 2/3 site, serial deletions were introduced into the SFV P2^{CA}3 expression construct (CA denotes Cys to Ala mutation in the active site of nsP2 and abolishes the self-processing of the molecule), and the *in vitro*-produced substrates were subjected to cleavage by nsP2. Deletions of the substrate from the nsP2 side were found to have no effect on cleavage efficiency, and the substrate [20:nsP3], which contained 20 aa residues from the C terminus of nsP2 and the full sequence of nsP3 (similar designations of substrates are used here [Fig. 1A]), was cleaved as efficiently as P2^{CA}3. Similarly, deletions leaving 180 or more aa residues from the nsP3 side in the substrate did not affect the cleavage. In contrast, a deletion that left 120 aa residues of nsP3 abolished cleavage completely (data not shown). Thus, the minimal efficiently cleaved substrate, as revealed by this analysis, was [20:180]. The size of the mapped P' region corresponded closely with that of the conserved macrodomain (or "X-domain") of nsP3 (22), which is composed of the first 160 aa residues of nsP3. Subsequent analysis revealed that substrate [19:160], the P' side of which corresponded to the crystallized macrodomain of Chikungunya virus (48), could not be cleaved by the nsP2 of SFV, whereas the slightly larger [19:165], [19:167], and [19:170] substrates were efficiently processed (Fig. 1B).

To map the requirements for processing of the 2/3 site in

clearly defined reactions, the set of recombinant substrates was expressed in *E. coli* and purified to homogeneity. Consistent with previous analyses, the assay using purified proteins revealed that the [20:165] and [20:170] substrates were efficiently processed; in contrast, the processing of the [20:160] substrate was very inefficient and was detectable only in the case of prolonged incubation with the enzyme (Fig. 1C). The truncation of the P side of the substrate to six SFV-specific amino acid residues had no adverse effect on the processing; furthermore, the substrate [2:170] was also processed efficiently (Fig. 1C). Combined, these data indicate that the efficient processing of the 2/3 site absolutely requires the presence of the conserved N-terminal one-third of nsP3 in the substrate, whereas the nonconserved residues, located on the P side, have little impact on the processing.

Next, we found that large deletions inside the macrodomain of the [20:170] substrate abolished its processing by nsP2 (data not shown). This effect may result from the removal of elements required for the binding of the nsP2 protease and/or because the 3-D structure of the substrate was destroyed. Indeed, in the 3-D structure of this domain sequence, elements located more than 150 aa residues downstream of the scissile bond of the 2/3 site are positioned close to it in space (48). To reveal whether or not all sequences of the macrodomain are required for 2/3 site processing, two deletions based on the 3-D structure of nsP3 were made: the Δ25–49 deletion, which removed the region corresponding to the α1 helix and loops surrounding this element, and the Δ23–68 deletion, which removed the α1 helix and the region corresponding to the β3 and β4 sheets. The design of these deletions was anticipated to preserve overall fold while removing structural elements that are located on the opposite side (in respect to the protease cleavage site) of the domain. We found that the shorter deletion had no detectable effect on the processing of the corresponding substrate and that even the larger deletion caused only a slight reduction of the cleavage efficiency (Fig. 1C), indicating that not all sequences in the macrodomain are required for processing of the 2/3 site in cell-free reactions. To verify that this is also the case in virus-infected cells, both deletions were introduced into the infectious cDNA clone pSFV4. The transfection of BHK-21 cells with *in vitro*-synthesized transcripts revealed that the mutant harboring the Δ23–68 deletion was not infectious, indicating that although the deleted region was dispensable for 2/3 site processing, it nonetheless had other crucial function(s) in SFV infection. The mutant containing the Δ25–49 deletion was viable, although the infectivity of the corresponding RNA was reduced in comparison to that of the original SFV4 (5×10^3 and 6×10^5 PFU/μg of RNA, respectively). SFVΔ25–49 also formed smaller plaques on the BHK-21 monolayers but was capable of growth to virus titers comparable to those of SFV4. The analysis of processing of P1234 in infected cells, carried out using the pulse-chase method, failed to reveal any difference between SFV4 and SFVΔ25–49 (Fig. 1D), indicating that the region of the macrodomain located between aa residues 25 to 49 is dispensable for 2/3 site processing and is also not strictly required for other crucial aspects of viral infection. Importantly, as the results obtained in the cell-free reaction system (Fig. 1C) correspond perfectly with those obtained using infected cells (Fig. 1D), we can conclude that the cell-free assay represents an adequate model to study the effects of different mutations on the processing of the 2/3 site, including lethal mutations that are impossible to analyze in the context of virus infection (exemplified by the Δ23–68 mutation).

FIG 2 Effects of point mutations of amino acid residues located in the C-terminal region of the minimal 2/3 substrate on the P123 processing, infectivity of the corresponding mutant genomes, and the properties of rescued viruses. (A) Processing of *in vitro*-translated P123 polyproteins with mutations E154R, D157R, R159E, E163R, or R159A + E163A in the nsP3 region. Substrates (indicated at the top of the drawing) were *in vitro* translated in the presence of [³⁵S]methionine. Reaction mixtures (left) or probes obtained by immunoprecipitation using anti-nsP3 antibodies (right) were separated by SDS-PAGE, and radioactive products were visualized by a Typhoon imager. The positions of polyproteins and the products of the cleavage reactions are shown between the panels. (B) Infectivities of *in vitro*-transcribed RNAs, plaque sizes of rescued viruses, and CPEs in the infected cell cultures analyzed in BHK-21 cells. Results of the sequence analysis of mutated regions are shown. (C) Properties of the plaque-purified viruses containing point mutations in the regions involved in the 2/3 site processing. Plaque-purified virus isolates were propagated in BHK-21 cells, and the final titers (PFU/ml) of these virus stocks (second passage) and plaque sizes formed on the monolayers of BHK-21 cells are shown. Mutations in the original site and revealed second-site mutations are shown.

Critical determinants for 2/3 site processing are located in the C-terminal part of the minimal substrate. The removal of five amino acid residues from the C terminus of [19:165] and [20:165] substrates abolishes their cleavage by nsP2 (Fig. 1B and C), suggesting that some crucial elements required for processing are located in the C-terminal region of the minimal 2/3 substrate. In this region, four charged amino acid residues, which are presumably clustered on one solvent-exposed face of the last alpha helix of the macrodomain, could potentially be involved in electrostatic interactions with each other or with the protease domain of nsP2 and/or its cofactors. Therefore, the E154R, D157R, R159E, and E163R mutations, which change the polarity of the charge, or the R159A + E163A mutation were introduced into P123 of SFV,

and their effects on the processing were studied using *in vitro*-translated polyproteins. This assay revealed that the E154R and D157R mutations had no effect on P123 processing (Fig. 2A), indicating that these amino acid residues are not involved in the regulation of 2/3 site processing. In contrast, the R159E, E163R, and R159A + E163A mutations all lead to the accumulation of unprocessed P123 polyprotein (Fig. 2A, left). Two processes are responsible for the turnover of P123: self-*cis*-cleavage of the 1/2 site (results in nsP1 and P23) and *trans*-cleavage of the 2/3 site (which results in P12 and nsP3) either by released nsP2 or by P23 processing intermediates; the first occurs in a constant (and relatively slow) manner, whereas the speed of the second (and, accordingly, its contribution to the P123 turnover) increases over time as the

FIG 3 Effects of point mutations of amino acid residues located in the C-terminal region of the minimal 2/3 site substrate on the P1234 processing in infected cells. Processing of the P1234 polyprotein in cells infected with SFV4 (wt) or plaque-purified isolates of SFV4-2*3 (one clone), SFV4-R159E→K (two clones, both with the same pseudoreversion), SFV4-E163R (two clones), and SFV4-R159A+E163A (two clones) is shown. Designations of constructs and clones are the same as those described for Fig. 2C. A pulse-chase experiment was performed as described in the legend to Fig. 1D. The positions of the polyproteins, processing intermediates, and mature SFV nsPs are indicated on the right side of the gel. Note the lack of the processing of P23 for SFV4-2*3, large delay in P23 processing (and, consequently, delay in the accumulation of mature nsP2) for both clones of SFV4-E163R or SFV4-R159A+E163A, and the nearly wild-type processing pattern for pseudoreverted SFV4-R159E→K viruses.

released nsP2 and P23 intermediate accumulate in the reaction mixture. To distinguish between comigrating P12 and P23 polyproteins, the reaction products were subjected to immunoprecipitation with antibodies against nsP3. This analysis revealed that in the case of P123, the P23 polyprotein was barely detectable, a finding which is consistent with previously published data (49, 73). This was also the case for polyproteins with E154R or D157R mutations (Fig. 2A, right). In contrast, the R159E, E163R, and tandem R159A+E163A mutations resulted in elevated levels of unprocessed P23 (Fig. 2B), thus clearly demonstrating that the charged amino acid residues Arg-159 and Glu-163 of nsP3 have an impact on 2/3 site processing. Out of these two residues, the Glu-163 apparently has a more significant role in the processing, a result which is consistent with the finding that substrate [20:165] is efficiently processed, whereas substrate [20:160] is not (Fig. 1C).

Next, the impact of selected mutations on the infectivity of SFV RNA was analyzed. To differentiate between effects originating from defects in 2/3 site processing and those originating from other defects in the assembly of the replicase complex, a mutant designated SFV4-2*3, containing a Gly-to-Ile substitution in the P2 position of the 2/3 site (the GGG codon of aa residue 798 of nsP2 was changed to ATA; note that reversion back to the Gly codon requires two nucleotide changes), was constructed and analyzed. The mutation did result in a severe, more than 10,000-fold reduction of the infectivity of *in vitro*-synthesized viral RNAs; however, the infectivity was rescued and the resulting virus replicated to a high titer. SFV4-2*3 failed to produce CPEs by 24 h postinfection but was able to form plaques that were only slightly smaller than those formed by wt SFV4 (Fig. 2B). Sequencing of the region of the 2/3 site in the genomes of the plaque-purified-rescued SFV4-2*3 viruses confirmed that the introduced mutation was preserved (Fig. 2C). Consistently, the analysis of expressed

proteins revealed the lack of processing at the 2/3 site (Fig. 3), indicating that the growth properties of SFV4-2*3 were likely restored by the appearance of second-site compensatory mutations (one possible candidate mutation was found in nsP3 [Fig. 2C]), similarly to the previously described SINV/1V2V/green fluorescent protein (GFP) mutant (23). Thus, it was confirmed that the cleavage of the 2/3 site is not a prerequisite for virus replication, although its absence severely compromised the infectivity of the corresponding RNA genome.

Similar to SFV4-2*3, recombinant genomes containing mutations in nsP3 at positions 159 and/or 163 were found to be infectious; however, their phenotypes clearly differed from each other and from that of SFV4-2*3. The genome of SFV4-R159E had strongly reduced infectivity, but the rescued virus replicated to a high titer, produced detectable CPEs by 24 h postinfection, and formed large plaques on the BHK-21 monolayers (Fig. 2B). Sequence analysis revealed that the resulting virus originates from pseudoreversion: the introduced acidic Glu residue in position 159 was changed to a positively charged Lys residue (note that the introduced GAG Glu codon cannot revert to any of the Arg codons but can pseudorevert to an AAG Lys codon through a single-nucleotide change); no common mutations were found anywhere in the genome, indicating that this pseudoreversion was sufficient to restore infectivity of the virus by restoring the efficient processing of the 2/3 site in infected cells (Fig. 3). The genome of SFV4-E163R also had relatively low infectivity; in contrast to the previous case, the rescued virus made very small plaques and failed to induce CPEs. As expected, the sequence analysis failed to detect any reversion or pseudoreversion in the mutated position of the rescued genomes (the introduced CGC Arg codon cannot be changed to any codon for Asp or Glu by a single-nucleotide change). Instead, numerous second-site changes, likely responsi-

ble for the rescue of infectivity, were identified in nsP3, nsP4, and, surprisingly, the 3' untranslated region of the viral genome (Fig. 2C). Pulse-chase experiments, performed with these plaque-purified mutant viruses, revealed that the processing of the 2/3 site remained inefficient (Fig. 3); thus, these changes had most likely no effect on the 2/3 site processing, and the analysis of their functional significance remains a topic for future studies. Thus, the presence of a basic residue in position 163 of nsP3 is not optimal for infectivity of SFV; however, in the absence of the possibility of generating a negative charge by a single-nucleotide change, the infectivity of mutant genomes, but not processing of the 2/3 site itself, can be restored by various second-site mutations.

Finally, we studied whether positions 159 and 163 function together or have separate roles in the virus life cycle, including processing of the 2/3 site. For these investigations, two mutant genomes were constructed and assayed. The genome of SFV4-R159E+E163R, which contains both R159E and E163R mutations, displayed extremely low infectivity, and the rescued virus made small plaques (Fig. 2B). Analysis of the sequences of rescued genomes was consistent with our previous findings: the Glu residue at position 159 was invariably pseudoreverted to a Lys residue, whereas the Arg residue was maintained at position 163. The search for second-site mutations, which are most likely required to compensate the reduced efficiency of 2/3 site processing, was not performed for this mutant. These data indicate that charged residues at the end of the minimal 2/3 substrate region have independent roles and likely do not interact with each other. In contrast, the genome of SFV4-R159A+E163A, containing R159A and E163A mutations, was found to be 100 times more infectious than that of SFV4-R159E+E163R; however, although the resulting virus did replicate to a high titer, it formed smaller plaques than SFV4 and failed to cause CPEs (Fig. 2B). Sequence analysis of the rescued virus revealed that the introduced mutations did not revert or pseudorevert, despite the fact that the introduced GCA Ala codon can revert to a Glu codon through a single-nucleotide change. Instead, various second-site changes, possibly responsible for the rescue and increased infectivity of this mutant, were found in nsP3 (Fig. 2C). Again, it was found that these changes do not restore the processing of the 2/3 site in virus-infected cells (Fig. 3); the analysis of the roles of these changes is under way in our laboratory. Taken together, our results confirm that the region around aa position 160 of nsP3 is crucial for processing of the 2/3 site and also has other possibly overlapping functions in viral infection.

Processing of the 2/3 site requires perfect positioning of the scissile bond with respect to the nsP3 macrodomain. We have shown that the P2 Gly residue and the N-terminal one-third of nsP3 (roughly corresponding to the macrodomain with the charged Arg-159 and Glu-163 residues at its C-terminal part) represent two crucial elements required for processing of the 2/3 substrate. To investigate how these elements should be positioned with respect to each other, a set of mutant substrates was prepared and analyzed. Two substrates were designed to shift the cleavage position toward the nsP3 side of the 2/3 site, thus decreasing the distance between the scissile bond and the body of the macrodomain: [20:170] Δ AP and [20:170] Δ APSY, where two or four N-terminal amino acid residues of the nsP3 were deleted, respectively. In [20:170]2 $\times$ GC, the first two amino acid residues of nsP3 were substituted with the last two amino acid residues of nsP2, thus generating an additional potential cleavage position; in [20:

FIG 4 The role of the nsP3 macrodomain in positioning of the scissile bond of the 2/3 substrate for the nsP2 protease. (A) Deletions, insertions, or substitutions were introduced into the original [20:170] substrate in the vicinity of the scissile bond. The black arrow indicates the original cleavage position, the open arrows indicate the potential cleavage positions generated by mutations, and the crossed arrows indicate cleavage positions that were not used. All reaction products were verified by N-terminal sequencing. (B) Efficiency of the modified [20:170] substrate cleavage by nsP2 was assayed as described in the legend to Fig. 1C.

170]2 $\times$ AP, the first two N-terminal amino acid residues of nsP3 were duplicated in order to increase the distance between the macrodomain and the scissile bond (Fig. 4A) while preserving the correct amino acid environment. The results obtained using the cell-free cleavage assay clearly showed that shifting the scissile bond closer to the macrodomain resulted in noncleavable substrates (Fig. 4B). The inability of the protease to process these substrates cannot be explained by the unfavorable nature of the new P1' residues, since the nsP2 protease prefers Ser or Arg residues over the original Ala residue (45). In contrast, the [20:170]2 $\times$ GC substrate was efficiently processed. N-terminal sequencing revealed that only the bond in the position corresponding to the original scissile bond, but not the bond at the newly generated position, was processed. In the efficiently processed [2:170] substrate, only the native scissile bond was cleaved; no cleavage of upstream (linker-derived) sites with favorable P1'-Gly residues was detected (Fig. 4A). Inefficient cleavage was observed for [20:170]2 $\times$ AP, for which sequencing of processing products revealed that the native scissile bond, which was shifted away from the macrodomain, was processed (Fig. 4). Combined, these results indicate the following: first, consistent with the results of previous studies (21, 45), the invariant P2-Gly residue is absolutely required for the processing of the 2/3 site; second, the precise distance between the scissile bond and crucial elements located in the macrodomain is required. Thus, both elements required for 2/3 site processing must not only be present in the

FIG 5 The role of the N terminus of nsP2 in the 2/3 site processing. wt, wild type; Pro, protease domain of nsP2; NTD, N-terminal domain (N174) of nsP2. (A) Sequences of the N termini of the modified enzymes; (B) proteolytic activity of the modified enzymes assayed as described in Fig. 1C. Substrates were assayed, and the results are presented as described for Fig. 1C.

substrate, but they must also be located correctly with respect to each other, and even very small changes in this arrangement can abolish or drastically reduce the processing efficiency.

The N terminus of nsP2 protease is required for efficient 2/3 site cleavage. It has been shown previously that the addition of 17 or 21 extra aa residues to the N terminus of nsP2 or its truncation by 8 aa residues inhibits the ability of the enzyme to process *in vitro*-translated polyproteins at the 2/3 site (73); however, a careful analysis of the significance of the native N terminus of nsP2 using a well-defined system has not been performed. Therefore, a set of modified enzymes was designed, expressed, and purified in a way that excludes the presence of an extra Met residue at their N termini. The mutant enzymes included nsP2(AL), in which two N-terminal amino acid residues of nsP2 of SFV (Gly and Val) were substituted with those of nsP2 of SINV (Ala and Leu), nsP2(+1) with an extra Ala residue at the N terminus, and nsP2(Δ 1) and nsP2(Δ 2), in which one or two N-terminal amino acid residues were removed, respectively (Fig. 5A). When these enzymes were used in our assay, it was found that nsP2(AL) processed the [20:170] substrate as efficiently as wt nsP2 (Fig. 5B), indicating that the two N-terminal amino acid residues of nsP2 of SFV and SINV are functionally equivalent. In contrast, addition or deletion of a single amino acid residue at the N terminus of nsP2 or deletion of two amino acid residues strongly reduced or almost completely abolished its ability to cleave the [20:170] substrate, respectively (Fig. 5B). Thus, not only is the N-terminal region of nsP2 required for 2/3 site processing, but so is the precise N terminus of nsP2.

The exact functions of the N-terminal region of alphavirus nsP2, which has no homology to known cellular or viral proteins outside the genus *Alphavirus*, are poorly understood. It has been shown that SFV P123 polyproteins containing deletions of nsP2 aa residues 20 to 457 or 60 to 457 are incapable of 2/3 site processing, whereas polyproteins with smaller deletions (aa residues 121 to 457) can process this site (73). Although it cannot be excluded that the defective processing observed in former mutants may origi-

nate from their incorrect folding or introduced steric problems, these data suggest that some crucial determinants required for 2/3 site processing may be located between aa residues 61 and 120 of nsP2.

To further study the role of the nsP2 N-terminal domain in the processing of the 2/3 site, we expressed and purified the 174-aa-long N-terminal domain (NTD) of nsP2, the protease domain of nsP2 (Pro), and the protease domain together with the first 170 aa residues of nsP3 (Pro170) as recombinant proteins (Fig. 1A). Consistent with previous reports (45, 72, 73), we found that the protease domain alone was incapable of processing the [20:170] substrate (Fig. 5B). Addition of the recombinant NTD to the reaction did not result in cleavage regardless of whether the 2/3 site was present in a separate [20:170] substrate or on the same molecule as the protease (data not shown). As no processing was observed despite the presence of all revealed determinants required for 2/3 site processing, it can be concluded that these elements failed to form the correct functional assembly needed for 2/3 site processing. The most straightforward explanation for this is that separated components (or some of them) failed to interact with each other; this assumption is supported by experiments using size exclusion chromatography. In these experiments, no complex formation between the NTD, Pro, and the [20:170] substrate was detected (data not shown). Thus, either the NTD or Pro must be present in one and the same molecule (as they are in native full-length nsP2), or they must be brought together in the correct assembly by an additional cofactor(s). Notably, it was demonstrated that temperature-sensitive (ts) mutation ts9 (G389R) within the helicase part of SFV nsP2 (missing in the constructs used in the above-mentioned experiments) is responsible for the specific defect in the 2/3 site processing, suggesting that the helicase region of the nsP2 may be required for the proper spatial orientation of the elements involved in the 2/3 site cleavage (1).

NSP2 of SINV is capable of processing the 2/3 site of P1234 of SFV. *Trans*-cleavage of the P1234 polyprotein by free nsP2 protease is dominant in the late stage of infection of alphaviruses (8, 28, 64) and is suggested to be the event that causes the cessation of negative-strand RNA synthesis (62) and superinfection exclusion (32). Interestingly, superinfection exclusion has also been observed for different yet related alphaviruses (12), suggesting that a protease from one virus may recognize and process the cleavage site(s) in the P1234 polyprotein of other viruses (32). As the 2/3 site is the only site in P1234 that is always processed *in trans*, and this study concludes that its recognition is based on the presence of the invariant P2 Gly residue and the well-conserved structure (and/or sequences) of the nsP3 macrodomain, it could be considered the most likely target in the P1234 polyprotein of a superinfecting virus. However, it has also been reported that alphavirus proteases have a poor cross-processing ability and that the protease domain of nsP2 of SINV could not cleave short substrates corresponding to cleavage sites of SFV and vice versa (79). As the system used in a study conducted by Zhang et al. (79) did not fulfill the requirements for 2/3 site processing, the ability of SFV and SINV proteases to cross-process the P1234 polyproteins was analyzed in our system. We found that the *in vitro*-translated P12^{CA34} of SFV and P12^{CA3S4} of SINV can be cleaved efficiently by purified nsP2 proteases from the corresponding viruses. For SFV, this cleavage resulted in the formation of all four mature nsPs, whereas for SINV, only mature nsP1 and nsP2 and unprocessed P34 were formed (Fig. 6A); this finding is consistent with a previous report

FIG 6 Processing of native, nonnative, and chimeric 2/3 sites by SFV and SINV nsP2 proteases. (A) Processing of P12^{CA}34 polyproteins of SFV and SINV by purified nsP2 of each virus. Substrates (indicated in the top row of the drawing) were *in vitro* translated in the presence of [³⁵S]methionine, and recombinant nsP2 protease of SFV or SINV was added as indicated above the panel. Samples were separated by SDS-PAGE; radioactive products were visualized by a Typhoon imager. The positions of polyproteins and the products of the cleavage reactions are shown at the right. P12 and P34, resulting from the *trans*-cleavage of P12^{CA}34 of SFV by nsP2 of SINV at the 2/3 site, are indicated by asterisks. (B) Processing of the recombinant substrates by SFV and SINV proteases. The names of sequences and components originating from SINV are given in white text on a black background, whereas SFV components are indicated with black text. The combinations of half-sites used in the chimeric substrates and the enzymes used for their processing are indicated at the top of each lane. The positions of components and products of the cleavage reactions shown at the right are designated in Fig. 1C.

(8). The SFV protease failed to process the polyprotein of the SINV, whereas P12^{CA}34 of SFV was efficiently processed by nsP2 of SINV (Fig. 6A). Importantly, in this case, only the 2/3 site was processed, as is evident from the sizes of the processing products (which correspond to P12 and P34), from the lack of products corresponding to mature nsPs, and from the inability of nsP2 of SINV to process P12^{CA}3^{CA}4 of SFV (Fig. 6A). Thus, nsP2 of SINV was able to process the 2/3 site in the polyproteins of both viruses, a finding that correlates with reports that SINV infection can block superinfection by SFV (12, 32). In contrast, nsP2 of SFV was unable to process P12^{CA}3^{CA}4 of SINV, which again is consistent with the observation that SFV-infected cells can be superinfected by SINV (our unpublished data).

To map the determinants responsible for the observed substrate specificity of SFV and SINV proteases, the cross-processing experiments were performed using recombinant proteases and [20:170] substrates. This assay confirmed that the nsP2 of SINV is capable of processing both native and SFV-derived [20:170] sub-

strates, whereas the nsP2 of SFV was able to process only its native substrate (Fig. 6B). The inability of the nsP2 of SFV to process the SINV-specific [20:170] substrate could, at least in part, be attributed to the P side of the substrate. Specifically, the P3 position in all three sites of SFV is occupied by an Ala residue, and it has been shown that only small amino acid residues are permitted in this position (45). In contrast, the P3 position in the processing sites of SINV is always occupied by large, hydrophobic Ile (1/2 site) or Val (2/3 and 3/4 sites) residues, which may not be suitable for the substrate recognition pocket of the SFV nsP2. This hypothesis was experimentally confirmed through the substitution of P3 Val for an Ala residue in [20:170] of SINV, which allowed the cleavage of this substrate by the nsP2 of SFV (Fig. 6B).

To further study the role of specificity determinant(s) located on the P and P' sides of the substrate, experiments with shuffling of the half-sites were performed. These experiments revealed that the SFV protease cleaved both chimeric substrates well, albeit slightly less efficiently than its own substrate (Fig. 6B). This result indicates that the nsP2 of SFV is capable of recognizing and using the P and P' sides of both substrates; thus, the P and P' sides of the SINV substrate represent a combination that is uncleavable by this enzyme only when they are together, and only in that context is the P3-Val residue not acceptable for nsP2 of SFV. The SINV protease was also capable of processing both chimeric substrates. Notably, in this case, the processing of the substrate containing the P side from SFV and the P' side from SINV was rather inefficient, even compared to the processing of the SFV-only substrate (Fig. 6B). Thus, correct interactions are required not only between the substrate and the different domains of the enzyme but also inside the 2/3 substrate, where the P and P' sides (and, accordingly, crucial elements located in these regions) act cooperatively, modulating each other's recognition and/or usage.

DISCUSSION

The progress of maturation of the viral replication complex must be properly controlled in order to prevent internal conflicts, such as a premature switch to generation of progeny virions when deployment of sufficient numbers of new replication complexes has not yet been achieved. To fulfill the requirements for gradual progression of infection, positive-strand RNA viruses (2, 78, 80), retroviruses (56), and several DNA viruses (e.g., African swine fever virus [66]) employ a polyprotein expression strategy in which proteolytic maturation plays a pivotal role (67). In the case of positive-strand RNA viruses, polyprotein precursors produced directly from genomic RNA are processed in an ordered cascade of cleavages. Viral polyprotein processing may release up to 16 mature proteins and a corresponding number of processing intermediates (68), some of which have functions different from those of the mature proteins and are therefore biologically significant. Host-encoded proteases are involved mostly in the processing of virus-encoded structural polyproteins or the structural parts of the single polyprotein (43), whereas the crucial cleavages within the replicase polyproteins are carried out by the virus's own enzymes (10).

It is important to emphasize that since the viral protease and its substrate(s) intimately coevolved in the single body of polyprotein, their relationships, through interacting and even interpenetrating surfaces, are evolutionarily justified to be considerably deeper than what is normally observed between an enzyme active site and its corresponding substrate cleavage sequence. It is note-

worthy, therefore, that such a multipoint contact strategy may alleviate the requirement for high-affinity binding of a substrate recognition sequence, which would otherwise lead to competitive inhibition of protease activity, and, most importantly, it potentially provides vast possibilities for regulation. Macromolecular assemblies are, undoubtedly, greater than the sum of their parts due to the tremendous role of cooperative effects in complex multicomponent biological systems, as they can enable a plethora of potentially synergistic or antagonistic combinations by using just a few existing basic biological principles and simple molecular building blocks in order to create sophisticated regulated networks (76, 77).

The intrinsic specificity of a protease is inevitably defined by its subsite selectivity for optimal protein motif recognition (3, 42). It is important, however, that the manifestation of cooperativity in the regulation of protease activity is apparent in many instances of enhancement of the proteolytic reactions through the use of exosite interactions (26). Exosites, as secondary binding platforms distant from the enzyme active site, act by increasing the surface area for protein-protein interactions and thus provide higher selectivity, affinity, and additional possibilities for regulation (44), being the main driving force for specific and sequential targeting. As more data become available, it appears that exosite interactions represent an indispensable principle of regulation in various areas of biology (5, 18, 35, 36, 51, 52, 57, 74, 75) and therefore emerge as potential targets in inhibitor design (11, 14). Among rare examples from the viral world, two different picornaviral proteinases, papain-like Lb^{pro} of the foot-and-mouth disease virus and the chymotrypsin-like 2A^{pro} of human rhinovirus 2, were reported to employ exosite-driven interactions to specifically target cellular protein eIF4G1 (15, 16). However, due to obvious difficulties with identification of potential protein cleavage determinants dispersed across the large protein surface, the involvement of exosites can be suspected mostly in cases of weak recognition of cleavage peptide derived from otherwise efficiently processed protein, as it was the case for the processing of the 2/3 site in the alphaviral ns polypeptide, which, to the best of our knowledge, represents the first example of the use of an exosite inside a viral proteome.

Historically, alphaviruses have represented one of the most valuable viral models for proteolytic processing studies (37, 38). Each cleavage within the ns polypeptide has a specific importance and unique characteristics. The clear order and precise timing of processing events certainly result from the activity of a well-tuned complex machinery of replication rather than that of an independent enzyme. Nevertheless, despite its significance, the functional requirements for the 2/3 site, processing of which is obligatorily performed *in trans* (59, 73) and certainly requires additional factors located outside the immediate cleavage sequence (45), remained poorly understood.

Normally, proteases tend to recognize a specific amino acid composition of the cleavage peptide and cleave substrates in solvent-exposed flexible loops by accommodating peptides in the recognition pocket in an extended beta-strand conformation (46, 71). The 1/2 and 3/4 sites in the alphaviral ns polypeptide are preceded by the regions with the predicted intrinsic disorder that apparently assist in the structural availability of these sites. This accessibility and the sufficient affinity of the recognized peptide, which is enhanced by an intramolecular mode of cleavage, which ensures intrinsically high local substrate concentrations, easily explain the cleavability of the 1/2 and 3/4 sites. In turn, the 2/3 site is

certainly different in its properties, as it is located between two well-structured domains, the nsP2 protease domain (59) and the nsP3 macrodomain (48), presumably in a kink of a beta-hairpin, and amino acid residues in the vicinity of the cleavage position evidently do not provide sufficient affinity for protease binding, thus making it, in general terms, a poor substrate (45, 79). These conditions apparently exist to prevent the occasional premature cleavage based on the short recognition sequence. Indeed, here we conclude that macromolecular assembly of nsP2 with the macrodomain region of nsP3 is a prerequisite for the precise positioning of the 2/3 site cleavage peptide at the protease active site.

Specifically, we found that in the 2/3 site, a conserved P2 Gly residue is absolutely required, as could be expected if the same protease subsites are involved in the accommodation of the P-side residues to bring the scissile bond to the catalytic center. The minor influence of the P6 to P3 region of the 2/3 site on the processing efficiency suggests that the substrate has to just fit into the pocket of the active site (Fig. 1C). Accordingly, the main power for forced targeting of the substrate to the active site should be provided by secondary-site interactions imposed by higher-order complex formation. The main determinant of 2/3 site processing appears to be contained within the carboxy-terminal region of the macrodomain of nsP3, which is spatially located in proximity to the 2/3 scissile bond and the P-side residues (48). Nonetheless, it is unlikely that the P side of the cleavage site and macro C-terminal elements simply represent two halves of a single site brought together by a macrofold. Evidently, the crucial elements recognized by the nsP2 protease are independent and positively cooperative. This finding is consistent with the primary (cleavage)- and secondary (exosite)-site binding model. A changing disposition of the recognition elements (by internal deletions or insertions) was found to be deleterious for processing (Fig. 4), suggesting that their mutual correct spatial arrangement is crucial. Suboptimal recognition at the primary site containing the scissile bond is thus compensated for by additional conserved contacts (most probably of an ionic nature) provided by the exosite residues at the carboxy-terminal part of the macrodomain. Correct spatial positioning of flanking regions of the macrodomain involved in proteolytic recognition is thus a novel function of the macrodomain fold in alphaviruses. Occasionally, the influence of the macrodomain on proteolytic processing has been also observed in case of other viruses (rubella virus [7, 40], mouse hepatitis virus [6, 70]) and thus may be a common theme for proteases of positive-strand RNA viruses that possess this domain. Nonetheless, it is premature to draw any parallels to the significance and mode of recognition of these domains in other viruses. Additionally, the effect of internal deletions in the macrodomain suggests that it has an independent and vital role in viral replication.

An essential player in 2/3 site processing is the N-terminal domain of nsP2, especially a subset of its N-terminal-most residues (Fig. 5B). This domain is dispensable for 3/4 site processing (45); however, it is important for the next two cleavage events. The *in cis* processing of the 1/2 site ensures that the released nsP2 has its own N terminus prepositioned in the region of the active site of the protease. Therefore, it is tempting to conclude that because the addition or removal of just a few amino acid residues at the native N terminus of nsP2 causes a dramatic and specific effect, it is the N terminus of nsP2 that is required for the recognition of the determinants located in the P' region of the [20:165] substrate. However, we were not able to demonstrate direct physical interactions

between these components, nor could we reconstitute the processing reaction using the purified [20:165] substrate and the protease- and N-terminal domains of nsP2; also, the introduction of charge-reversion substitutions at the C terminus of the macrodomain did not lead to the appearance of any potential compensatory mutation in the N-terminal region of nsP2 (Fig. 2C). Therefore, a more realistic explanation is that the mode of interaction is more complicated and that the internal architecture of nsP2, and not its extreme elements, ensures correct conformations and protein-protein associations. Hypothetically, the N-terminal peptide of nsP2, which is quite conserved and is predicted to form a beta-strand, once liberated, may contribute to the macrodomain structure, thus bringing essential protease-substrate complex elements into place. Notably, protein-protein interactions through beta-strand addition are common (58) and were found in many cases of cofactor activation of viral proteases (e.g., flaviviral [13] and adenoviral [9] proteases) and were in-depth studied in case of the hepatitis C virus NS4A cofactor which is known to intercalate with the beta-sheet within the core of the NS3 protease (33). Clearly, direct structural studies are necessary to disclose the actual mode of molecular recognition.

Results of this and previous studies unequivocally demonstrate the importance of correct macromolecular assembly for 2/3 site processing. As the *in trans* mode of cleavage implies, at least two molecules, namely, nsP2 (or P23) as an enzyme and P23, P123, or P1234 as a substrate, are required for this intermolecular reaction. Due to the specific cleavage requirements (extended binding surface with multiple distributed contacts), the processing of the 2/3 site is by far the most efficient cleavage within the alphavirus ns polypeptide. Accordingly, only traces of P23 can be detected in *in vitro* translation/processing experiments (Fig. 2A, right) unless the incubation time of reaction is reduced and/or the translation mixture is strongly diluted (73). The same is true for infected cells; in contrast to easily detectable P123, P12, P34, and P1234 (Fig. 1D), P23 has been detected only in certain ts mutants (ts17, ts18, and ts24) of SINV at restrictive temperatures (30). In their classical study, Hardy et al. (30) demonstrated that the ts mutations in nsP2 did not block the protease activity but specifically inhibited cleavage of the 2/3 site. These data strongly support the assembly-dependent mechanism of 2/3 site processing, as the same ts mutations also inhibited viral RNA synthesis (27)—another process that requires correct macromolecular assembly rather than functionality of the individual nsP2 protein.

Why is the 2/3 site cleavage so efficient when, even in its absence, the virus can survive (23, 47)? It has been proposed that the cleavage of the 2/3 site in P1234 by free nsP2 limits the number of replicase complexes per infected cell (62); however, again, the number of replicase complexes is also controlled even in the case of mutant viruses with an uncleavable 2/3 site (47). Similarly, the processing of the 2/3 site is not strictly required for formation of spherular structures containing viral replication complexes (17); so, instead of being required for functional replicase assembly, the 2/3 site processing rather serves to sense that correct assembly has been achieved and makes it essentially irreversible. In any case, maturation of nsP2 into its free form is found to have other functional consequences, such as the recognition of the SG promoter (34) or suppression of cellular antiviral responses (23) and, possibly, gaining the ability to process ns polypeptides of superinfecting viruses (reference 32 and this study). Similarly, cleavage of the 2/3 bond triggers the release of nsP3 into the cell, which may also

have specific effects on the host (24). All these events are clearly important for the infectious cycle of the virus and thus deserve special studies.

Commonly enough, the sequential mode of viral polypeptide processing is attempted to be explained by apparent differences in the cleavage efficiencies of short peptides representing cleavage sites. However, it appears unlikely that such vital decisions rely on the stochastic nature of affinity matching in the search for potential substrates. Instead, the ns protease in the context of viral replicase can be viewed as a watchdog awaiting a command to bite. In line with this analogy, the protease, which is embedded in a maturing polypeptide, which, in turn, is almost always membrane bound, is thus kept “on a chain” due to the restriction of its opportunity for free movement. The replication complex is a dynamic system with many moving parts, so it can be envisioned that the regulatory role of the protease is to monitor the succession and completion of the events of viral infection and to respond with a cleavage once respective conformational changes in the complex allow the presentation of the scissile bond and other essential determinants. In this case, correct completion of the translation of the viral polypeptide, capture of viral mRNA and its reassignment for genome replication, proper assembly of the components of the ribonucleoprotein complex, association with cellular membranes and subsequent enwrapping of replication complexes into spherular structures, completion of replicative intermediate RNA synthesis, and other similar events, which are most certainly accompanied by conformational transitions, may thus serve as valid signals for the protease to perform cleavage and to shift the whole system to a new level. Curiously, analogous ratcheting of a substrate from a zymogen to a proteinase conformation was found to allow sequential presentation of the cleavage sites, leading to the ordered action of prothrombinase on prothrombin, and this sequence of events was proposed as a broadly applicable mechanism underlying the ordered action of proteinases at multiple sites in their polypeptide substrates (4). We believe that in the viral world the use of secondary binding sites is a common but, so far, largely overlooked property of viral proteases, which can and must be exploited for restriction of viral replication and infection.

ACKNOWLEDGMENTS

We thank Andrey Golubtsov and Tero Ahola for their involvement in this study.

This work was supported by the Wellcome Trust grant 067575, Estonian Science Foundation grants 7501 and 7407, target financing project SF0180087s08, and the European Union through the European Regional Development Fund via the Center of Excellence in Chemical Biology.

REFERENCES

1. Balistreri G, Caldentey J, Kääriäinen L, Ahola T. 2007. Enzymatic defects of the nsP2 proteins of Semliki Forest virus temperature-sensitive mutants. *J. Virol.* 81:2849–2860.
2. Bedard KM, Semler BL. 2004. Regulation of picornavirus gene expression. *Microbes Infect.* 6:702–713.
3. Bianchi E, Pessi A. 2002. Inhibiting viral proteases: challenges and opportunities. *Biopolymers* 66:101–114.
4. Bianchini EP, Orcutt SJ, Panizzi P, Bock PE, Krishnaswamy S. 2005. Ratcheting of the substrate from the zymogen to proteinase conformations directs the sequential cleavage of prothrombin by prothrombinase. *Proc. Natl. Acad. Sci. U. S. A.* 102:10099–10104.
5. Bock PE, Panizzi P, Verhamme IM. 2007. Exosites in the substrate specificity of blood coagulation reactions. *J. Thromb. Haemost.* 5(Suppl. 1):81–94.
6. Bonilla PJ, Hughes SA, Piñón JD, Weiss SR. 1995. Characterization of

- the leader papain-like proteinase of MHV-A59: identification of a new in vitro cleavage site. *Virology* 209:489–497.
7. Chen HH, Stark CJ, Atreya CD. 2006. The rubella virus nonstructural protease recognizes itself via an internal sequence present upstream of the cleavage site for trans-activity. *Arch. Virol.* 151:1841–1851.
 8. De Groot RJ, Hardy WR, Shirako Y, Strauss JH. 1990. Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo. *EMBO J.* 9:2631–2638.
 9. Ding J, McGrath WJ, Sweet RM, Mangel WF. 1996. Crystal structure of the human adenovirus proteinase with its 11 amino acid cofactor. *EMBO J.* 15:1778–1783.
 10. Dougherty WG, Semler BL. 1993. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiol. Rev.* 57:781–822.
 11. Drag M, Salvesen GS. 2010. Emerging principles in protease-based drug discovery. *Nat. Rev. Drug Discov.* 9:690–701.
 12. Eaton BT. 1979. Heterologous interference in *Aedes albopictus* cells infected with alphaviruses. *J. Virol.* 30:45–55.
 13. Erbel P, et al. 2006. Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat. Struct. Mol. Biol.* 13: 372–373.
 14. Farady CJ, Craik CS. 2010. Mechanisms of macromolecular protease inhibitors. *ChemBiochem* 11:2341–2346.
 15. Foeger N, Kuehnle E, Cencic R, Skern T. 2005. The binding of foot-and-mouth disease virus leader proteinase to eIF4G1 involves conserved ionic interactions. *FEBS J.* 272:2602–2611.
 16. Foeger N, Schmid EM, Skern T. 2003. Human rhinovirus 2 APro recognition of eukaryotic initiation factor 4G1. Involvement of an exosite. *J. Biol. Chem.* 278:33200–33207.
 17. Frolova EI, Gorchakov R, Pereboeva L, Atasheva S, Frolov I. 2010. Functional Sindbis virus replicative complexes are formed at the plasma membrane. *J. Virol.* 84:11679–11695.
 18. Fuentes-Prior P, Salvesen GS. 2004. The protein structures that shape caspase activity, specificity, activation and inhibition. *Biochem. J.* 384: 201–232.
 19. Garmashova N, Gorchakov R, Frolova E, Frolov I. 2006. Sindbis virus nonstructural protein nsP2 is cytotoxic and inhibits cellular transcription. *J. Virol.* 80:5686–5696.
 20. Garmashova N, et al. 2007. The Old World and New World alphaviruses use different virus-specific proteins for induction of transcriptional shut-off. *J. Virol.* 81:2472–2484.
 21. Golubtsov A, Kääriäinen L, Caldentey J. 2006. Characterization of the cysteine protease domain of Semliki Forest virus replicase protein nsP2 by in vitro mutagenesis. *FEBS Lett.* 580:1502–1508.
 22. Gorbalenya AE, Koonin EV, Lai MM. 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.
 23. Gorchakov R, et al. 2008. A new role for ns polyprotein cleavage in Sindbis virus replication. *J. Virol.* 82:6218–6231.
 24. Gorchakov R, Garmashova N, Frolova E, Frolov I. 2008. Different types of nsP3-containing protein complexes in Sindbis virus-infected cells. *J. Virol.* 82:10088–10101.
 25. Gorchakov R, Frolova E, Williams BR, Rice CM, Frolov I. 2004. PKR-dependent and -independent mechanisms are involved in translational shutoff during Sindbis virus infection. *J. Virol.* 78:8455–8467.
 26. Hacisalihoglu A, Panizzi P, Bock PE, Camire RM, Krishnaswamy S. 2007. Restricted active site docking by enzyme-bound substrate enforces the ordered cleavage of prothrombin by prothrombinase. *J. Biol. Chem.* 282:32974–32982.
 27. Hahn YS, Strauss EG, Strauss JH. 1989. Mapping of RNA⁺ temperature-sensitive mutants of Sindbis virus: assignment of complementation groups A, B, and G to nonstructural proteins. *J. Virol.* 63:3142–3150.
 28. Hardy WR, Strauss JH. 1988. Processing of the nonstructural polyproteins of Sindbis virus: study of the kinetics in vivo using monoclonal antibodies. *J. Virol.* 62:998–1007.
 29. Hardy WR, Strauss JH. 1989. Processing the nonstructural polyproteins of Sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in *cis* and in *trans*. *J. Virol.* 63:4653–4664.
 30. Hardy WR, Hahn YS, de Groot RJ, Strauss EG, Strauss JH. 1990. Synthesis and processing of the nonstructural polyproteins of several temperature-sensitive mutants of Sindbis virus. *Virology* 177:199–208.
 31. Kääriäinen L, Ahola T. 2002. Functions of alphavirus nonstructural proteins in RNA replication. *Prog. Nucleic Acid Res. Mol. Biol.* 71:187–222.
 32. Karpf AR, Lenches E, Strauss EG, Strauss JH. 1997. Superinfection exclusion of alphavirus in three mosquito cell lines persistently infected with Sindbis virus. *J. Virol.* 71:7119–7123.
 33. Kim JL, et al. 1996. Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* 87: 343–355.
 34. Kim KH, Rümenapf T, Strauss EG, Strauss JH. 2004. Regulation of Semliki Forest virus RNA replication: a model for the control of alphavirus pathogenesis in invertebrate hosts. *Virology* 323:153–163.
 35. Kleanthous C, Walker D. 2001. Immunity proteins: enzyme inhibitors that avoid the active site. *Trends Biochem. Sci.* 26:624–631.
 36. Krishnaswamy S. 2005. Exosite-driven substrate specificity and function in coagulation. *J. Thromb. Haemost.* 3:54–67.
 37. Lemm JA, Rice CM. 1993. Roles of nonstructural polyproteins and cleavage products in regulating Sindbis virus RNA replication and transcription. *J. Virol.* 67:1916–1926.
 38. Lemm JA, Rümenapf T, Strauss EG, Strauss JH, Rice CM. 1994. Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *EMBO J.* 13:2925–2934.
 39. Lemm JA, Bergqvist A, Read CM, Rice CM. 1998. Template-dependent initiation of Sindbis virus RNA replication in vitro. *J. Virol.* 72: 6546–6553.
 40. Liang Y, Yao J, Gillam S. 2000. Rubella virus nonstructural protein protease domains involved in *trans*- and *cis*-cleavage activities. *J. Virol.* 74:5412–5423.
 41. Liljeström P, Lusa S, Huylebroeck D, Garoff H. 1991. In vitro mutagenesis of a full-length cDNA clone of Semliki Forest virus: the small 6,000-molecular-weight membrane protein modulates virus release. *J. Virol.* 65: 4107–4113.
 42. Lim MD, Craik CS. 2009. Using specificity to strategically target proteases. *Bioorg. Med. Chem.* 17:1094–1100.
 43. Lindenbach BD, Thiel HJ, Rice CM. 2007. Flaviviridae: the viruses and their replication, p. 1101–1152. *In* D. M. Knipe and P. M. Howley (ed.), *Fields virology*, Lippincott Williams & Wilkins, Philadelphia, PA.
 44. López-Otín C, Overall CM. 2002. Protease degradomics: a new challenge for proteomics. *Nat. Rev. Mol. Cell Biol.* 3:509–519.
 45. Lulla A, Lulla V, Tints K, Ahola T, Merits A. 2006. Molecular determinants of substrate specificity for Semliki Forest virus nonstructural protease. *J. Virol.* 80:5413–5422.
 46. Madala PK, Tyndall JD, Nall T, Fairlie DP. 2010. Update 1 of: Proteases universally recognize beta strands in their active sites. *Chem. Rev.* 110: PR1–P31.
 47. Mai J, Sawicki SG, Sawicki DL. 2009. Fate of minus-strand templates and replication complexes produced by a p23-cleavage-defective mutant of Sindbis virus. *J. Virol.* 83:8553–8564.
 48. Malet H, et al. 2009. The crystal structure of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. *J. Virol.* 83:6534–6545.
 49. Merits A, Vasiljeva L, Ahola T, Kääriäinen L, Auvinen P. 2001. Proteolytic processing of Semliki Forest virus-specific non-structural polyprotein by nsP2 protease. *J. Gen. Virol.* 82:765–773.
 50. Moss B, Elroy-Stein O, Mizukami T, Alexander WA, Fuerst TR. 1990. Product review. New mammalian expression vectors. *Nature* 348:91–92.
 51. Nagase H, Fushimi K. 2008. Elucidating the function of non catalytic domains of collagenases and aggrecanases. *Connect Tissue Res.* 49: 169–174.
 52. Overall CM. 2002. Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites. *Mol. Biotechnol.* 22:51–86.
 53. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 4:2411–2423.
 54. Peränen J, Rikonen M, Liljestöm P, Kääriäinen L. 1990. Nuclear localization of Semliki Forest virus-specific nonstructural protein nsP2. *J. Virol.* 64:1888–1896.
 55. Peränen J, Rikonen M, Hyvönen M, Kääriäinen L. 1996. T7 vectors with modified T7lac promoter for expression of proteins in *Escherichia coli*. *Anal. Biochem.* 236:371–373.

56. Pettit SC, Clemente JC, Jeung JA, Dunn BM, Kaplan AH. 2005. Ordered processing of the human immunodeficiency virus type 1 GagPol precursor is influenced by the context of the embedded viral protease. *Society* 79: 10601–10607.
57. Prudent R, Sautel CF, Cochet C. 2010. Structure-based discovery of small molecules targeting different surfaces of protein-kinase CK2. *Bioch. Bioph. Acta* 1804:493–498.
58. Remaut H, Waksman G. 2006. Protein-protein interaction through beta-strand addition. *Trends Biochem. Sci.* 8:436–444.
59. Russo AT, White MA, Watowich SJ. 2006. The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. *Structure* 14: 1449–1458.
60. Salonen A, et al. 2003. Properly folded nonstructural polyprotein directs the Semliki Forest virus replication complex to the endosomal compartment. *J. Virol.* 77:1691–1702.
61. Sawicki DL, Sawicki SG. 1980. Short-lived minus-strand polymerase for Semliki Forest virus. *J. Virol.* 34:108–118.
62. Sawicki DL, Perri S, Polo JM, Sawicki SG. 2006. Role for nsP2 proteins in the cessation of alphavirus minus-strand synthesis by host cells. *J. Virol.* 80:360–371.
63. Schechter I, Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27:157–162.
64. Shirako Y, Strauss JH. 1990. Cleavage between nsP1 and nsP2 initiates the processing pathway of Sindbis virus nonstructural polyprotein P123. *Virology* 177:54–64.
65. Shirako Y, Strauss JH. 1994. Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus-strand RNA synthesis. *J. Virol.* 68:1874–1885.
66. Simón-Mateo C, Andrés G, Viñuela E. 1993. Polyprotein processing in African swine fever virus: a novel gene expression strategy for a DNA virus. *EMBO J.* 12:2977–2987.
67. Spall VE, Shanks M, Lomonosoff GP. 1997. Polyprotein processing as a strategy for gene expression in RNA viruses. *Seminars Virol.* 8:15–23.
68. Steuber H, Hilgenfeld R. 2010. Recent advances in targeting viral proteases for the discovery of novel antivirals. *Curr. Top. Med. Chem.* 10: 323–345.
69. Strauss JH, Strauss EG. 1994. The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* 58:491–562.
70. Teng H, Piñón JD, Weiss SR. 1999. Expression of murine coronavirus recombinant papain-like proteinase: efficient cleavage is dependent on the lengths of both the substrate and the proteinase polypeptides. *J. Virol.* 73:2658–2666.
71. Tyndall JD, Nall T, Fairlie DP. 2005. Proteases universally recognize beta strands in their active sites. *Chem. Rev.* 105:973–999.
72. Vasiljeva L, Valmu L, Kääriäinen L, 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.
73. Vasiljeva L, et al. 2003. Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. *J. Biol. Chem.* 278:41636–41645.
74. Vauquelin G, Packeu A. 2009. Ligands, their receptors and plasma membranes. *Mol. Cell Endocrinol.* 311:1–10.
75. Vogel A, Schilling O, Späth B, Marchfelder A. 2005. The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties. *Biol. Chem.* 386:1253–1264.
76. Whitty A. 2008. Cooperativity and biological complexity. *Science* 4:435–439.
77. Williamson JR. 2008. Cooperativity in macromolecular assembly. *Nat. Chem. Biol.* 4:458–465.
78. Yao N, Reichert P, Taremi SS, Prosise WW, Weber PC. 1999. Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase. *Structure* 7:1353–1363.
79. Zhang D, Tözsér J, Waugh DS. 2009. Molecular cloning, overproduction, purification and biochemical characterization of the p39 nsp2 protease domains encoded by three alphaviruses. *Protein Expr. Purif.* 64: 89–97.
80. Ziebuhr J, Snijder EJ, Gorbalenya AE. 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* 81:853–879.