Functional Characterization of the Cleavage Specificity of the Sapovirus Chymotrypsin-Like Protease

Ivonne Robel,1 Julia Gebhardt,1 Jeroen R. Mesters,2 Alexander Gorbalenya,3 Bruno Coutard,4 Bruno Canard,4 Rolf Hilgenfeld,2 and Jacques Rohayem1†

The Calicilab, Institute of Virology, Dresden University of Technology, Dresden, Germany1; Institute of Biochemistry, Center for Structural and Cell Biology in Medicine, University of Lübeck, Lübeck, Germany2; Leiden Medical Center, Department of Microbiology, Leiden, The Netherlands3; and AFMB, CNRS, Université Aix-Marseille, Marseilles, Marseilles, France4

Sapovirus is a positive-stranded RNA virus with a translational strategy based on processing of a polyprotein precursor by a chymotrypsin-like protease. So far, the molecular mechanisms regulating cleavage specificity of the viral protease are poorly understood. In this study, the catalytic activities and substrate specificities of the predicted forms of the viral protease, the 3C-like protease (NS6) and the 3CD-like protease-polymerase (NS6-7), were examined in vitro. The purified NS6 and NS6-7 were able to cleave synthetic peptides (15 to 17 residues) displaying the cleavage sites of the sapovirus polyprotein, both NS6 and NS6-7 proteins being active forms of the viral protease. High-performance liquid chromatography and subsequent mass spectrometry analysis of digested products showed a specific trans cleavage of peptides bearing Gln-Gly, Gln-Ala, Glu-Gly, Glu-Pro, or Glu-Lys at the scissile bond. In contrast, peptides bearing Glu-Ala or Gln-Asp at the scissile bond (NS4-N5 and NS5-N6, or NS6-N7 junctions, respectively) were resistant to trans cleavage by NS6 or NS6-7 proteins, whereas cis cleavage of the Glu-Ala scissile bond of the NS5-N6 junction was evidenced. Interestingly, the presence of a Phe at position P4 overruled the resistance to trans cleavage of the Glu-Ala junction (NS5-N6), whereas substitutions at the P1 and P2’ positions altered the cleavage efficiency. The differential cleavage observed is supported by a model of the substrate-binding site of the sapovirus protease, indicating that the P4, P1, and P2’ positions in the substrate modulate the cleavage specificity and efficiency of the sapovirus chymotrypsin-like protease.

Sapovirus is a RNA virus belonging to Caliciviridae, where it segregates in a separate genus including human pathogenic and nonhuman pathogenic strains. Because of the lack of a cell culture system to propagate human pathogenic sapovirus strains, the translation strategy remains poorly understood. In particular, so far, the active forms of the protease as well as the molecular mechanisms of cleavage by the viral protease remain unclear.

The sapovirus single-stranded positive-oriented RNA genome (about 7.6 kb) encodes structural and nonstructural proteins located in two open reading frames (ORF1 and ORF2). Like other RNA viruses with a positive-oriented genome (i.e., Flaviviridae and Picornaviridae), sapovirus uses a translational strategy based on cotranslational processing of a polyprotein precursor encoded in ORF1. This cotranslational processing is mediated by one of the key players in viral replication, the viral chymotrypsin-like protease. The sapovirus protease (3C-like, NS6, 16 kDa) is predicted to be located in the C-terminal half of the polyprotein precursor, upstream of the viral RNA-dependent RNA polymerase (3D-like, NS7, 57 kDa).

In this study, it was postulated that (i) both NS6 and NS6-7 proteins are active forms of the viral protease, (ii) differential cis/trans processing of the cleavage sites in the polyprotein occurs, and (iii) the specificity of cleavage is regulated by distinct positions in the substrate. Our results show that both NS6 and NS6-7 proteins are able to cleave nearly all the scissile bonds in the polyprotein precursor in trans. Moreover, cleavage specificity and efficiency seem to be modulated by the positions P4, P1, and P2’ of the substrate. Cleavage of peptides displaying the scissile bonds of the sapovirus protease reveals different cleavage efficiencies by NS6 and NS6-7, suggesting a differential release of the nonstructural proteins from the polyprotein precursor.

MATERIALS AND METHODS

Prediction of sapovirus NS6 protein boundaries. The boundaries of the sapovirus NS6 protein were based on the previous predictions performed on the active domain of the sapovirus NS7 RNA-dependent RNA polymerase (3), indicating a putative cleavage of the protease-polymerase precursor at a Glu-Pro, Asp-Pro scissile bond in the polyprotein precursor of the sapovirus pG-Sap 1 clone (GenBank accession no. AY96184).

Expression and purification of recombinant sapovirus NS6 protein. Production and purification of recombinant sapovirus NS6 were performed as described previously (3, 13–15), with slight modifications. Briefly, a cDNA fragment (459 bp) encompassing the sapovirus 3C-like protease gene and encoding the NS5-NS6 and NS6-NS7 cleavage sites at its 5’ and 3’ ends, respectively, was generated by PCR from sapovirus clone pG-Sap1 (GenBank accession no. AY96184; Fig. 1) using primers VZ-SapNS6-H6C-for (5’-GG GACAAGTTTGTACAAAAAAGCAGGCTTCAAGAGGAGTAGCCCATGAAAAGCCGCCGACACATTGGGGAGTTCACGCAG-3’) and VZ-SapNS6-H6C-rev (5’-GGGAGACCCATTTGTAACAGAAAGCTGGGCTTATTAGTGATGGTGATGGGAGTGATGTTGTGATGGGAGTGATGGTGATGGGAGTTCACGCAG-3’). cDNA was cloned into the pDEST14 vector (Gateway; Invitrogen), resulting in expression vector pVZ-SapNS6-H6C, which was sequenced and used to transform Escherichia coli Rosetta(DE3) cells. After induction of cells grown at 37°C in 2YT medium, cultures were incubated...
FIG. 1. Organization of the polyprotein precursor encoded in ORF1 of sapovirus clone pJG-Sapl (GenBank accession no. [GB acc. No.] AY694184). (A) ORF1 encoding the nonstructural proteins (5,176 nucleotides in length) is shown. Putative cleavage sites are indicated. Both NS6 and NS6-7 proteins and their active-site mutant were examined in *Escherichia coli* Rosetta(DE3) cells. The recombinant proteins bear a His tag at the C terminus. The nonstructural proteins NS1 to NS7 corresponding to the proteins p11, p28, 2C-like NTase, p32, VPg, 3C-like protease, and 3D-like polymerase, respectively, are shown. (B) The proteins HA-[NS5/NS6]-NS6-7-His and HA-[NS4/NS5]-NS6-7-His bearing the authentic cleavage sites of the NS5-NS6 and the NS4-NS5 junctions, respectively, were expressed in *Escherichia coli* Rosetta(DE3) cells. Those constructs bear a His tag at the C terminus and an HA tag at the N terminus. The HA tag is underlined.

Expression of mutated sapovirus NS6 gene. Point mutations in the putative active site of sapovirus NS6 protein (GDC143GC) were generated with the Quick Change site-directed mutagenesis kit (Stratagene) according to the manufacturer’s instructions. pVZ-SapNS6-H6C or pVZ-SapNS6-7H6C templates and primers 84-Sap-mNS6-5’-GCCGACACCACTTTGTACAAGAAAAGCAGGCTCGAAG-3’ and 85-Sap-mNS6-3’-AGATGCGCCACACGAGCTCGTGGCTTTGACACGGGAATGCAG-3’ were used. The HA tag was inserted in the protein’s predicted active site to alanine. Mutant NS6 gene expression was performed as described above, and the resulting proteins were termed mNS6 and mNS6-7 (Fig. 1).

Synthesis and purification of peptides. Peptides were synthesized by GL Biochem (Shanghai, China) and purified by reversed-phase high-performance liquid chromatography (RP-HPLC) on an RP column. The lyophilized products were characterized by HPLC and fast atom bombardment mass spectrometry (MS).

Analysis of the trans proteolytic activity of sapovirus NS6 and NS6-7 proteins. Twenty-one different peptides (S1 to S21, 15 to 17 residues) displaying the cleavage sites in the sapovirus ORF1 polyprotein precursor as well as mutations in the P4 to P1 and P1 to P1’ positions were used as substrates (Table 1). Lyophilized peptides were dissolved in Tris buffer (40 mM Tris, 150 mM NaCl [pH 7.2]) at a final concentration of 1 mg/ml. Proteolytic cleavage was examined by HPLC as described previously (15). The cleavage efficiencies of the peptides by the viral protease were reported as percentages of the fraction of substrate that is converted to product based on the integrated peak area (12, 15, 17). Comparative analysis of cleavage efficiencies was performed using a nonparametric Mann-Whitney U test using GraphPad InStat, version 3.0a, for Macintosh (GraphPad Software, San Diego, CA). Statistical significance was set at \( P < 0.05 \).
Molecular characterization of the cleavage products. MS analysis of HPLC fractions containing the products of limited proteolysis was performed as described previously (15) at the Mass Spectrometry Facility of the Max Planck Institute of Molecular Cell Biology and Genetics (Dresden, Germany). Matrix-assisted laser desorption ionization analysis was carried out on a Reflex IV matrix-assisted laser desorption ionization-time-of-flight mass spectrometer (Bruker Daltonics, Bremen, Germany) in reflective mode using α-cyano-4-hydroxy-cinnamic acid as the matrix (17). Spectra were externally calibrated using masses of abundant autolysis products of trypsin. Nano-electrospray ionization-MS and tandem-MS (MS/MS) analyses were performed on a hybrid quadrupole time-of-flight mass spectrometer QSTAR Pulsar i (MDS Sciex, Concord, Canada) equipped with an automated chip-based nanoflow ion source (Advion, Ithaca, NY). Spectra were processed with BioAnalyzer QS v.1 software (MDS Sciex, Concord, Canada).

Analysis of the cis proteolytic activity of sapovirus protease. A time course experiment was performed to investigate the cis cleavage activity of sapovirus NS6-7. Therefore, two constructs bearing the sciase bonds corresponding to the NS4/NS5 or NS5/NS6 cleavage sites fused to a hemagglutinin (HA) tag at the N terminus were designed, resulting in expression vectors pIR-SAP-NS6/NS5-NS6-7 and pIR-SAP-ΔtNS5/NS5/NS6-7. Expression in Escherichia coli Rosetta(DE3) and purification were performed as described above, yielding proteins HA-[NS5/NS6]-NS6-7-His and HA-[NS4/NS5]-NS6-7-His. The cultures were harvested 8 h after induction, and the pellets were diluted 1:25, 1:50, 1:75, and 1:100. The pellets were suspended in 0.5 ml 1× sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) loading solution, separated by 12% SDS-PAGE, and transferred onto Hybond nitrocellulose filters (Amersham Biosciences). Filters were blocked overnight in 1% milk powder dissolved in phosphate-buffered saline-0.1% Tween 20. Blotted proteins were detected with the mouse monoclonal anti-His primary antibody (Qiagen) or the mouse monoclonal anti-HA antibody (Sigma) at a 1:500 dilution. A phosphatase-labeled antimouse secondary antibody (Promega) was used at a 1:10,000 dilution. After each incubation step, filters were washed four times with phosphate-buffered saline-0.1% Tween 20. Detection was performed as described previously (15).

Homology model building. A homology model for the sapovirus 3C-like protease was manually constructed on the crystal structure of the norovirus 3C-like protease (Protein Data Bank [PDB] identification no. 2FYQ) (19). Sequence alignments and structure modeling were aided by DeepView/SwissPDBViewer, version 3.96 (Gloor-SmithKline, 2006).

RESULTS

Expression and purification of recombinant SV NS6 and NS6-7 proteins. Expression of recombinant SV NS6 and NS6-7 genes as well as their respective active-site mutants was performed in E. coli Rosetta(DE3) cells. After separation of the protein fractions by Ni-NTA affinity chromatography, the eluted protein was further purified by gel filtration, and the fractions containing the highest level of purity were used for further experiments (Fig. 2).

FIG. 2. Expression and purification of the sapovirus NS6 and NS6-7 proteins and their respective active-site mutants. The cDNA clones pIG-SAP-NS6, pIG-SAP-μNS6, pIG-SAP-NS6-7, and pIG-SAP-μNS6-7 were used for expression in Escherichia coli Rosetta(DE3) cells. The NS6 and NS6-7 proteins as well as their active-site mutants (μNS6 and μNS6-7) were purified by affinity chromatography on a Ni-NTA column followed by gel chromatography. Coomassie staining after SDS-PAGE of eluted fractions is shown.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Cleavage site</th>
<th>Mutation</th>
<th>Sequence (NH₂-COOH)</th>
<th>Mean ± SE % cleavage efficiency (n=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV1</td>
<td>NS1-NS2</td>
<td></td>
<td>D G Y T F V E E G L L D M F G T</td>
<td>100.00</td>
</tr>
<tr>
<td>SV2</td>
<td>NS2-NS3</td>
<td></td>
<td>V K T T F T A Q G P T D L G W A</td>
<td>13.2 ± 1.8</td>
</tr>
<tr>
<td>SV3</td>
<td>NS3-NS4</td>
<td></td>
<td>M E R R F K E Q A G P L Q N L</td>
<td>43.2 ± 0.6</td>
</tr>
<tr>
<td>SV4</td>
<td>NS4-NS5</td>
<td></td>
<td>T R G S T V E E A K G P T D L G W A</td>
<td>0.0</td>
</tr>
<tr>
<td>SV5</td>
<td>NS5-NS6</td>
<td></td>
<td>P Q O H Y E E E A P T A I V E F T</td>
<td>0.0</td>
</tr>
<tr>
<td>SV6</td>
<td>NS6-NS7</td>
<td></td>
<td>V K E V T T Q D E F Q W K G L</td>
<td>0.0</td>
</tr>
<tr>
<td>SV7</td>
<td>NS7-VP1</td>
<td></td>
<td>V K I V F E M E G N S N P E P K</td>
<td>94.03 ± 0.5</td>
</tr>
<tr>
<td>SV8</td>
<td>NS1-NS2</td>
<td></td>
<td>D G Y T F V E E G L L D M F G T</td>
<td>0.0</td>
</tr>
<tr>
<td>SV9</td>
<td>NS2-NS3</td>
<td></td>
<td>V K T T F T A G G P T D L G W A</td>
<td>0.0</td>
</tr>
<tr>
<td>SV10</td>
<td>NS4-NS5</td>
<td></td>
<td>T R G S F T E E A K G K T K H G</td>
<td>0.0</td>
</tr>
<tr>
<td>SV11</td>
<td>NS4-NS5</td>
<td></td>
<td>T R G S R K E E A K G K T K H G</td>
<td>0.0</td>
</tr>
<tr>
<td>SV12</td>
<td>NS4-NS5</td>
<td></td>
<td>R Q R F T S S K</td>
<td>0.0</td>
</tr>
<tr>
<td>SV13</td>
<td>NS5-NS6</td>
<td></td>
<td>P Q O H Y K E A A P T A I V E F T</td>
<td>0.0</td>
</tr>
<tr>
<td>SV14</td>
<td>NS5-NS6</td>
<td></td>
<td>Y 1003F</td>
<td>19.27 ± 1.77</td>
</tr>
<tr>
<td>SV15</td>
<td>NS5-NS6</td>
<td></td>
<td>E 1051V</td>
<td>0.0</td>
</tr>
<tr>
<td>SV16</td>
<td>NS5-NS6</td>
<td></td>
<td>T 1055D</td>
<td>0.0</td>
</tr>
<tr>
<td>SV17</td>
<td>NS5-NS6</td>
<td></td>
<td>E 1051V</td>
<td>0.0</td>
</tr>
<tr>
<td>SV18</td>
<td>NS2-NS3</td>
<td></td>
<td>A 1051E</td>
<td>9.48 ± 0.48</td>
</tr>
<tr>
<td>SV19</td>
<td>NS2-NS3</td>
<td></td>
<td>G 1051E</td>
<td>33.64 ± 1.14</td>
</tr>
<tr>
<td>SV20</td>
<td>NS1-NS2</td>
<td></td>
<td>G 1051E</td>
<td>97.95 ± 2.05</td>
</tr>
<tr>
<td>SV21</td>
<td>NS5-NS6</td>
<td></td>
<td>Y 1003F, P 1004L</td>
<td>0.0</td>
</tr>
</tbody>
</table>

* The peptides reflect the authentic or mutated sites in the sapovirus polyprotein.

* The numbering corresponds to the polyprotein of clone pJG-SapI (GenBank accession no. AY694184).

* Key positions from P5 to P1 and P1’ to P5’ are noted. The mutated positions are highlighted in boldface.

Table 1. Sequence and trans cleavage efficiency of peptides by sapovirus protease
SV NS6 and NS6-7 display cleavage activities in trans. For characterization of the trans cleavage activities of NS6 and NS6-7, a cell-free assay was designed as previously described (15). This assay examines the trans cleavage activity of the purified NS6 and NS6-7 proteins on peptides bearing the scissile bonds from the SV polyprotein. Seven peptides corresponding to the cleavage sites within the sapovirus polyprotein precursor were designed (SV1 to SV7; Table 1). Overnight incubation of the peptides with NS6 or NS6-7 protein followed by analysis of the reaction by RP-HPLC provided evidence of proteolytic cleavage of peptides NS1-NS2, NS2-NS3, NS3-NS4, and NS7-virion protein 1 (VP1). A representative example is shown in Fig. 3, where incubation of peptide NS1-NS2 with NS6 or NS6-7 protein generated cleavage products, as indicated by two RP-HPLC peaks in addition to the substrate peak, presumably resulting from proteolytic processing of this peptide at the Gln-Gly scissile bond. When the active-site mutants were incubated with the peptide, no cleavage was evidenced, indicating that the cleavage is specific for NS6 and NS6-7 proteins. Similarly, incubation of peptides NS2-NS3, NS3-NS4, and NS7-VP1 with purified NS6 or NS6-7 proteins generated two cleavage products, presumably resulting from proteolytic processing of peptides at Glu-Gly, or Glu-Ala or Gln-Ala scissile bonds, respectively (Table 1). In contrast, no trans cleavage products were observed when peptide NS5-NS6, NS6-NS7, or NS4-NS5 was incubated with NS6 or NS6-7 protein, suggesting resistance of these peptides to proteolytic cleavage in trans.

To further characterize the specificities of NS6 and NS6-7 for the predicted cleavage sites, the RP-HPLC fractions containing the products of peptide proteolysis were subjected to MS analysis. Analyzed fractions contained the expected products of proteolysis, allowing the precise identification of the scissile bonds for peptides NS1-NS2, NS2-NS3, NS3-NS4, and NS7-VP1 (Table 1).

Sapovirus peptides NS1-NS2, NS2-NS3, NS3-NS4, NS4-NS5, NS5-NS6, NS6-NS7, and NS7-VP1 are differentially cleaved by NS6 and NS6-7 proteins. The activity of the sapovirus protease NS6 and NS6-7 was examined in the peptidolytic assay under standard conditions. As shown in Fig. 4 and Table 1, sapovirus NS6 and NS6-7 proteins displayed similar cleavage efficiencies in trans. However, both proteins cleaved peptides NS1-NS2 and NS7-VP1 more efficiently than NS2-NS3 and NS3-NS4 ($P < 0.05$ by nonparametric Mann-Whitney U test). Moreover, peptides NS4-NS5, NS5-NS6, and NS6-NS7 were resistant to cleavage in trans. These data further suggest...
a differential cleavage pattern of the sapovirus polyprotein by NS6 and NS6-7, possibly modulated by the amino acid sequence pattern around the scissile bond.

**Investigation of cis cleavage at the NS4-NS5, NS5-NS6, and NS6-NS7 scissile bonds.** To investigate whether a cis cleavage occurs at the NS4-NS5 and NS5-NS6 junction, two NS6-7 proteins bearing the NS4-NS5 or NS5-NS6 scissile bond fused to an HA tag at the N terminus and a His tag at the C terminus located downstream of the NS6-NS7 cleavage site were expressed in *E. coli*. The cultures were harvested 8 h after induction, and the pellets were diluted 1:25, 1:50, 1:75, and 1:100, analyzed by SDS-PAGE, and visualized by Western blotting using anti-HA or anti-His antibodies. The NS6-7 protein was not able to cleave the His tag at the C terminus downstream of the NS6-NS7 cleavage site, indicating no cleavage in cis between NS6 and NS7 (Fig. 5A). The NS6-7 protein was able to cleave the HA tag at the N terminus upstream of the NS6-NS7 cleavage site, indicating that the scissile bond is located downstream of the NS6-NS7 cleavage site (Fig. 5B) up to a 1:100 dilution, indicating that cis cleavage of the NS5-NS6 scissile bond occurs. In contrast, the NS6-7 protein could not cleave the HA tag at the N terminus upstream of the NS4-NS5 scissile bond site cis.

The amino acid sequence around the scissile bond modulates the trans cleavage by sapovirus protease. To examine whether the P1 to P4 and P1′ to P2′ positions in the peptide sequence modulate trans cleavage, 14 additional peptides bearing amino acid substitutions at either side of the scissile bond were designed and synthesized. Gln or Gla is present at the P1 position of all cleavage sites of the polyprotein precursor (Table 1). In order to assess the specificity of the protease for either amino acid at this position, substitution of Gla for Gly or Gln for Gly was performed in peptides NS1-NS2 (SV8) or NS2-NS3 (SV9), respectively. This completely abolished processing of the peptides. In contrast, at the P1′ position, no effect on cleavage efficiency was observed when Gly was substituted for with a Pro or Lys (SV19 and SV20, respectively), indicating that at the scissile bond (P1′-P1 position), the P1 position modulates the sensitivity to trans cleavage by the sapovirus protease.

At the P2 position, Gla, Ala, Thr, or Met is present. Mutation of Ala to Gln in this position did not result in an increase of cleavage efficiency (SV2 versus SV17). However, by replacing a Pro at the P2′ position with an Asp, cleavage efficiency increased significantly (SV2 versus SV18; *P < 0.05*). This indicates that the P2′ position also modulates cleavage efficiency in trans by the sapovirus 3C-like protease. This is also observed in peptide SV21, where mutation at the P2′ position (Pro to Leu) completely abolished cleavage of the peptide. The P3 position does not seem to influence cleavage efficiency, as observed in peptides SV11 to SV13, SV15, and SV16. In contrast, position P4 positively modulated cleavage, as observed in peptide SV14 in comparison to peptide SV5. There, replacing Tyr by Phe resulted in cleavage of the peptide NS5-NS6 that
involved are Asn$_{99}$ (side-chain amide), Thr$_{7/7me}$, Ile$_{10/6}$, and Val$_{50}$. The latter residue is Ile$_{10/6}$ in norovirus 3C-like protease, but the side chain has to be shorter in the sapovirus protease, otherwise the above-mentioned Arg$_{10/6}$ of the S2 pocket would not fit. The S1’ pocket in all 3C and 3C-like proteases is small in size, and the sapovirus NS6 protein is no exception. Thus, the most preferred P1’ residue is glycine. The residues involved are the carbonyl of Gly$_{113}$, Gly$_{114}$, His$_{115}$ of the catalytic triad, and Val$_{32}$. The main-chain amides of Gly$_{114}$ and Cys$_{116}$ are involved in stabilizing the oxanion. The S2’ subsite cannot be modeled reliably, but we note that the P2’ residue will be partly exposed to solvent. Beyond P2’, the substrate residues are not in contact with the enzyme.

**DISCUSSION**

The sapovirus protease is predicted to play an essential role in the replication strategy of the virus. The molecular mechanisms underlying the catalytic activity and specificity of the sapovirus chymotrypsin-like protease have so far remained enigmatic. Understanding the regulation of polyprotein processing in sapovirus is an important step toward elucidating the sapovirus replication strategy. Such knowledge is also an important prerequisite for the development of antiviral drugs to control sapovirus infection.

In this study, the cleavage efficiency and substrate specificity of the catalytic in-trans and cis activities of sapovirus NS6 and NS6-7 proteins were examined in vitro. The NS6 and NS6-7 proteins were able to cleave peptides containing Gln-Gly, Glu-Gly, or Gln-Ala bonds. However, NS6 and NS6-7 were not able to cleave the NS6-NS7 scissile bond, neither in trans nor in cis, indicating that the NS6-7 protein is the active form of the sapovirus protease. These results are in accordance with previous reports on the processing of sapovirus polyprotein in a cell-free system (9, 10). Indeed, production of the sapovirus ORF1 polyprotein in vitro leads to cotranslational release of the NS1, NS2, NS3, NS4-NS5, and NS6-7 proteins. To characterize the differential cis/trans cleavage pattern of the polyprotein precursor, two constructs featuring the authentic cleavage sites of the NS4-NS5 junction or the NS5-NS6 junction at the N terminus were designed. Expression of the constructs revealed that the NS5-NS6 junction was cleaved in cis, whereas the NS4-NS5 junction was resistant to cis cleavage. Cleavage in cis by the 3C-like protease has already been discussed in the cases of norovirus (16) and poliovirus (6). Our data on the autocatalytic cleavage of the norovirus 3C-like protease from its protease-polymerase precursor (15) support these observations. In our study, the peptide bearing the cleavage site between the NS4 and NS5 proteins was resistant to cleavage in trans by both NS6 and NS6-7 proteins. Discordant results on processing of the sapovirus NS4-NS5 boundary have been reported so far. In the study of Oka et al. (8), processing of the ORF1 polyprotein of strain Mc10 leads to the detection of the NS4 and NS5 proteins putatively released from an N-terminally glutathione S-transferase-labeled NS4-NS5 precursor.

The cleavage between the NS4 and NS5 proteins was also observed in cell-free system, where cleavage of the whole polyprotein was investigated (9, 10). However, in a further study using the same system, the NS4-NS5 precursor was also detected (11), indicating no cleavage of the scissile bond, as was resistant to cleavage in trans. To sum up, the positions P4, P1, and P2’ seem to play an important role in modulating cleavage efficiency of the scissile bonds of the polyprotein precursor by the sapovirus protease.

**Modeling of sapovirus NS6 protease active site.** In order to generate a reliable model of the active site of the sapovirus protease, automatic sequence alignment of the sapovirus and norovirus 3C-like protease was used, yielding a sequence identity of 32%. However, as is often the case, this preliminary alignment gave an overestimated value for the sequence identity, and our structure-based alignment yielded a more realistic value of approximately 20%. According to our model (Fig. 6) of the sapovirus protease active site, the walls of the S1 specific pocket are formed by Ala$_{132mc}$ ("mc" represents main-chain atoms), Gly$_{113}$, Thr$_{134}$, Thr$_{111}$, and Gly$_{114}$. The two threonine residues are at the far end of the pocket (see Fig. 6, upper left). At the bottom, there is His$_{115}$ residue highly conserved in 3C and 3C-like proteases. As in most of these enzymes, the histidine accepts a hydrogen bond from the hydroxyl of a tyrosine residue: in this case, Tyr$_{120}$. The structural position of the tyrosine OH group is similar to the one in the coronavirus main proteases (1, 2), although the residue is on a different strand in the β-barrel of domain II. In the S2 subsite, Asn$_{109}$ at the tip of the βII-cII hairpin forms part of the S2 pocket, and so does the Gly$_{122}$-His$_{115}$ pair of the catalytic triad. Interestingly, the guanidinium group of Arg$_{104}$ is also part of the S2 pocket, according to our model. As to the S3 subsite, and similar to most 3C or 3C-like proteases, there is no evident S3 pocket and a specificity for P3 of the substrate is not recognized, except that the side chain tends to be hydrophilic since the subsite is open toward bulk solvent. In contrast, there is pronounced specificity for P4 being hydrophilic. Since a part of the S4 pocket involves a segment of the protease that is highly unrelated to norovirus 3C-like protease, we had difficulty completing the model for this subsite. Residues likely
observed in our study. In order to investigate the reason for this discordance, we have performed a control experiment in which the processing of a so-far cleavage-resistant peptide (SV4 reflecting the putative NS4-NS5 scissile bond) was investigated in the presence of reticulocyte lysates (see the supplemental material). Interestingly, the peptide was completely cleaved in the presence of reticulocyte lysates, whether sapovirus NS6-7, NS6, or their respective active-site mutants were added to the reaction, indicating that a nonspecific cleavage of the NS4-NS5 site occurs in the translation-transcription assay used. This nonspecific cleavage is not due to a cofactor of the NS6 nor NS6-7 present in the reticulocyte lysates, as would have been expected if the active protease incubated with the lysates cleaves the peptide, in comparison to the lack of cleavage observed when the lysates are absent. Those results indicate that processing of the NS4-NS5 scissile bond strongly depends on the expression system used.

Moreover, the cleavage specificity and efficiency of the scissile bonds seem to depend on the sequence of the peptide, more precisely the P4, P1, and P2' positions. The sapovirus NS6 and NS6-7 proteins displayed a higher cleavage efficiency of NS1-NS2 and NS7-VP1 sites than of the NS2-NS3 and NS3-NS4 sites, indicating that Gln in P1 decreases the cleavage efficiency compared to Glu. Mutation of the P1 Gln to Gly or Glu to Gln abolished completely processing of the peptides. In contrast, mutation at the P1' position did not affect the cleavage efficiency. These data indicate that at the scissile bond (P1-P1' position), the P1 position modulates the sensitivity to trans cleavage by the sapovirus protease. The NS5-NS6 scissile bond, consisting of Glu-Ala, was not cleaved by NS6 nor by NS6-7 proteins in trans. Peptide NS6-NS7 that was not cleaved in trans displayed a Gln-Asp scissile bond, thus differing in the P1' position from the above-mentioned cleavage sites. However, the Glu-Ala scissile bond of the NS5-NS6 junction was cleaved in cis, indicating that this pair of residues may play a role in regulating cis versus trans cleavage. Interestingly, the P2' position seems to modulate cleavage efficiency in trans by the sapovirus 3C-like protease, as observed in peptide SV21 (NS5-NS6). Moreover, by replacing at the P2' position of the NS2-NS3 peptide a Pro with an Asn, cleavage efficiency increased significantly (SV2 versus SV18, P < 0.05). Similarly, position P4 positively modulates cleavage, where replacing Tyr with Phe results in trans cleavage of the peptide NS5-NS6. The relevance of a Phe at position P4 has also been highlighted in the cases of norovirus (4) and feline calcivirus (11). In summary, the positions P4, P1, and P2' seem to play an important role in modulating cleavage efficiency of the scissile bonds of the polyprotein precursor by the sapovirus protease.

The differential cleavage observed in this study is supported by our model of the substrate-binding site of the sapovirus protease. In the absence of a crystal structure for an enzyme, three-dimensional homology models have their merits when it comes to interpreting experimental data on substrate specificity (such as in the present case) and even for preliminary steps in drug discovery (5). Of course, the quality of a homology model depends on the overall degree of sequence identity between the template(s) and the model, as well as the quality of the template and the number of homologous sequences available. In this study, we only focus on regions of high similarity between the sapovirus and norovirus 3C-like proteases that allowed reliable construction of a three-dimensional model.

In the SV NS6 3C-like protease, the glutamate residue of the catalytic triad Cys116-His131-Glu152 is in contact with another histidine residue, His102. This residue is not conserved in the norovirus 3C-like protease, where it is replaced by Val214. The reason for the presence of the second histidine in the immediate catalytic site will be discussed below. In order to interpret the cleavage specificity data described above, we will focus on the substrate-binding subsites of the enzyme as revealed by the homology model. At the bottom of the S1 pocket, His131 is absolutely conserved in all calcivirus, picornavirus, and coronaviruses 3C-like proteases. It has been shown to interact with the P1 side chain of the substrate in several crystal structures of complexes between substrate-like inhibitors and these enzymes (1, 2, 7, 18). Most of these proteases are absolutely specific for glutamine in the P1 position. In order to reject the binding of glutamate residues in the S1 pocket, the conserved histidine should be uncharged over a wide range of pHs. In the coronaviruses 3C-like proteases (main proteases), this is ensured by a tyrosine residue, which donates a hydrogen bond from its hydroxyl group to the histidine (1). A tyrosine is also present in the picornavirus 3C proteases, even though at a very different position in the amino acid sequence (residue 138 in the 3Cpros from human rhinovirus 2, poliovirus, and foot-and-mouth disease virus; as an exception, hepatitis A virus 3Cpro does not have this tyrosine). Interestingly, the hydroxyl group of Tyr138 in the picornavirus 3Cpro occupies exactly the same position as the hydroxyl of Tyr160/161 (numbering for severe acute respiratory syndrome-coronavirus Mr160 in brackets) in the coronaviral enzymes and can thus donate an ideal hydrogen bond to the histidine at the bottom of the S1 pocket (1). Sapoviruses and norovirus 3C proteases also feature the conserved tyrosine, and yet they accept not only glutamine, but also glutamate in the P1 position of the substrate. As far as the tyrosine is concerned, the calcivirus 3C-like proteases seem to be more similar to the coronavirus main proteases than to the picornavirus 3C enzymes. According to our model, the presence of a guanidinium group of Arg104 in the S2 pocket could explain the observed specificity of the sapovirus protease for a glutamate in P2. In norovirus 3C-like protease, this residue is replaced by Gly116, which obviously cannot interact with a P2-glutamate. Methionine and alanine also occur (once each) at the P2 position of sapovirus cleavage sites. Our model would be compatible with both: methionine is sufficiently flexible to adapt its conformation to the amphiphilic S2 pocket, and alanine does not pose a steric problem due to its short side chain. The presence of Arg104 at the bottom of the S2 pocket jeopardizes the integrity of the catalytic triad, because according to our model, the guanidinium group of this residue is placed within 4.5 Å from Glu152, which might reorient toward the arginine since the latter is in a largely hydrophobic environment. To prevent this, His102 plays an important role by holding Glu152 in its proper position in the triad. Thus, the triad in fact becomes a Cys116-His131-Glu152-His102 tetrad (Fig. 6). This feature does not exist in norovirus 3C-like protease (which lacks the arginine in the S2 subsite).

Based on our data, we postulated that differences in susceptibility of the predicted cleavage sites to proteolysis by sapovirus NS6-7 protein may result in a sequential processing of the
scissile bonds. According to the observed relative cleavage efficiencies of the peptides by NS6-7 (Fig. 4), a sequential map for processing of the sapovirus polyprotein is proposed (Fig. 7) based on a temporal and/or spatial ranking for release of the active nonstructural proteins from the polyprotein precursor. This ranking may be based on the half-life of the precursor proteins, as observed during the processing of norovirus and poliovirus polyproteins (12, 15), in which a similar mechanism regulating availability of the nonstructural proteins of the replication complex was postulated. Therefore, it is conceivable that proteolytic processing by sapovirus NS6-7 regulates the formation of the replication complex through a ranking order of stability of the nonstructural proteins. It should be emphasized, however, that the proposed model of polyprotein processing cascade is based on data generated in vitro, i.e., in the absence of cellular factors that may play an essential role in modulating the replication of the sapovirus. In this sense, our in vitro model will have to be completed if and when an efficient culture system for sapovirus is available.

ACKNOWLEDGMENT

This study was supported by the European project “VIZIER” (Comparative Structural Genomics of Virus-encoded Proteases and Protease Inhibitors) funded by the 6th Framework Program of the European Commission under reference LSHG-CT-2004-511960.

REFERENCES


