JOURNAL OF VIROLOGY, Apr. 2006, p. 4157–4167 Vol. 80, No. 8
0022-538X/06/$08.00+0 doi:10.1128/JVI.80.8.4157–4167.2006

Characterization of a Torovirus Main Proteinase

Saskia L. Smits,[1] Eric J. Snijder,[2] and Raoul J. de Groot[1]*

Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands,[1] and Molecular Virology Laboratory, Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands[2]

Received 12 October 2005/Accepted 2 February 2006

Viruses of the order *Nidovirales* encode huge replicase polyproteins. These are processed primarily by the chymotrypsin-like main proteinases (M^{pro}s). So far, M^{pro}s have been studied only for corona-, arteri-, and roniviruses. Here, we report the characterization of the M^{pro} of toroviruses, the fourth main Nidovirus branch. Comparative sequence analysis of polyprotein 1a of equine torovirus (EToV) strain Berne, identified a serine proteinase domain, flanked by hydrophobic regions. Heterologous expression of this domain resulted in autoprocessing at flanking cleavage sites. N-terminal sequence analysis of cleavage products tentatively identified FxxQ ↓ (S, A) as the substrate consensus sequence. EToV M^{pro} combines several traits of its closest relatives. It has a predicted three-domain structure, with two catalytic β-barrel domains and an additional C-terminal domain of unknown function. With respect to substrate specificity, the EToV M^{pro} resembles its coronavirus homologue in its preference for P1-Gln, but its substrate-binding subsite, S1, more closely resembles that of arteri- and ronivirus M^{pro}s, which prefer P1-Glu. Surprisingly, in contrast to the M^{pro}s of corona- and roniviruses, but like that of arterivirus, the torovirus M^{pro} uses serine instead of cysteine as its principal nucleophile. Under the premise that the M^{pro}s of corona- and toroviruses are more closely related to each other than to those of arteri- and roniviruses, the transition from serine- to cysteine-based proteolytic catalysis (or vice versa) must have happened more than once in the course of nidovirus evolution. In this respect, it is of interest that a mutant EToV M^{pro} with a Ser^{165}→Cys substitution retained partial enzymatic activity.

Many positive-strand RNA viruses have adopted a "polyprotein" strategy to express (part of) their genome. Translation of viral RNA results in the synthesis of large multidomain protein precursors, which are proteolytically cleaved to yield a collection of intermediate and mature products. As a rule, polyproteins that are to yield secretory or membrane (glyco)proteins are processed by host-encoded proteinases within the lumen of the compartments of the exocytotic pathway (in particular the endoplasmic reticulum and the *trans*-Golgi network). Cytosolic processing, however, involves co- and posttranslational cleavage by viral proteinases that are present as distinct domains within the polyprotein itself (51, 72). Processing is often a prerequisite for the assembly of the viral RNA replicase/transcriptase. To add to the complexity, intermediate and mature products may have distinct functions in replication and may be required during different stages of the infection. Hence, the properties of the viral proteinases must be carefully tailored to ensure accurate and efficient cleavage in a substrate-specific and temporally well-coordinated fashion.

Many RNA virus proteinases resemble, in their mechanism of peptide hydrolysis and in their three-dimensional structure, either one of the prototypical cellular proteases papain and chymotrypsin. Papain and related proteinases are characterized by a two-domain α+β-fold with typically a cysteine and downstream histidine as the catalytic residues. Chymotrypsin has a distinctive two-β-barrel structure in which the principal nucleophile, a serine residue, is part of a catalytic His-Asp-Ser triad (20, 49). Chymotrypsin-like serine proteinase domains have been identified within the polyproteins of among others flavi-, alpha-, luteo-, astro-, and sobemoviruses (13, 17, 29, 38, 42, 53, 66). Intriguingly, however, many viruses, for instance picorna-, como-, poty-, and caliciviruses, encode proteinases with a (predicted) chymotrypsin-like fold in which, instead of serine, a cysteine acts as the main catalytic residue (1, 5, 6, 27, 28, 31, 52, 55).

The order *Nidovirales* comprises a group of evolutionarily related enveloped (positive) stranded RNA viruses (corona-, toro-, arteri-, and roniviruses) (10, 15, 19, 26, 56, 58, 62), which have taken the polyprotein strategy to the extreme. The 5′-most two-thirds of the nidovirus genome is occupied by two large overlapping open reading frames (ORFs), ORF1a and ORF1b, the most downstream of which is only translated upon ribosomal frameshifting (9, 15, 18, 58). These ORFs specify two huge polyproteins, pp1a and the C-terminally extended frameshift protein pp1ab. In the case of the coronaviruses, pp1ab measures up to 7,200 amino acids in length, representing the largest polyprotein known to date. Among nidoviruses, the largest degree of sequence conservation is found within the ORF1b-encoded part of pp1ab, which contains the chief replicative domains (15, 19, 30, 57), i.e., the RNA dependent RNA polymerase (RdRp) and the helicase, as well as up to three additional domains, ExoN, NendoU, and 2′-*O*-methyltransferase (2′-*O*-MT), which display sequence identity to cellular RNA-modifying enzymes (8, 37, 57). Of the latter, the ExoN and 2′-*O*-MT domains are lacking in the arterivirus pp1ab.

Processing of the arteriviral and coronaviral polyproteins gives rise to 12 to 13 and 15 to 16 mature products, respectively, and an as yet unknown number of functional intermediates (57, 76). The N termini of the replicase polyproteins are

* Corresponding author. Mailing address: Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, The Netherlands. Phone: 31 30 2531463/2485. Fax: 31 30 2536723. E-mail: R.Groot@vet.uu.nl.

processed by "accessory" papain-like proteinases (31). The vast majority of cleavages, however, are carried out by a single chymotrypsin-like main proteinase (M^{pro}), which is located in the C-terminal half of pp1a, flanked by hydrophobic regions that are thought to be involved in targeting the replicase to intracellular membranes (30, 54, 67, 71). Recent X-ray crystallographic analysis of the structures of corona- and arterivirus M^{pro}s confirmed the predicted chymotrypsin-like two-β-barrel fold (2, 4, 73). Interestingly, the corona- and arteriviral enzymes both possess an additional C-terminal domain, a feature not commonly seen in other viral and cellular chymotrypsin-like proteinases (2, 4). Also for the ronivirus M^{pro}, such an additional C-terminal domain was predicted (74).

One might surmise that the nidoviral M^{pro}s, given their pivotal role in viral replication, are well conserved. Conversely, however, the primary sequences of the M^{pro}s of corona-, roni-, and arteriviruses have diverged almost beyond recognition, and sequence identity is restricted to the regions immediately surrounding the main catalytic residues. Moreover, whereas the arteriviral M^{pro} is a serine proteinase with a canonical His-Asp-Ser catalytic triad (4, 64), the M^{pro}s of the corona- and roniviruses are cysteine proteinases (28, 30, 43, 74). In the latter, the acidic catalytic residues have been lost and, apparently, these enzymes solely rely on the remaining catalytic His-Cys dyad (2, 73, 74).

Toroviruses have been poorly studied, but because of their relatively close phylogenetic relationship to coronaviruses they may in many respects represent "missing links" in comparative nidovirus studies. Here, we are the first to report the biochemical characterization of a torovirus M^{pro}. Upon analysis of the C-terminal half of pp1a of equine torovirus (EToV) strain Berne, we identified a chymotrypsin-like proteinase domain, which, like the M^{pro}s of the arteri-, corona-, and roniviruses, is flanked by hydrophobic regions. When expressed in prokaryotic and mammalian cells, the proteinase domain autocatalytically released itself from its surrounding sequences. Surprisingly, the toroviral M^{pro} seems no more related to that of coronaviruses than to the M^{pro}s of arteri- and roniviruses. Most saliently, in contrast to the M^{pro}s of corona- and roniviruses, but like that of arteriviruses, the torovirus M^{pro} is a serine proteinase. A tentative processing scheme for the torovirus pp1a/pp1ab polyproteins is presented, and the implications of our findings for nidovirus evolution are discussed.

MATERIALS AND METHODS

Cells and virus. Equine dermis (Ederm) cells (American Type Tissue Culture Collection) and OST7-1 cells (23) were maintained as monolayers in Dulbecco's modified Eagle's medium (DMEM; BioWhitaker) supplemented with penicillin (100 IU/ml) and streptomycin (100 μg/ml) and 20% (DMEM20) or 10% (DMEM10) fetal calf serum, respectively. EToV strain Berne was propagated in Ederm cells as described previously (68). Recombinant vaccinia virus vTF7-3 (24) was obtained from B. Moss (NIH, Bethesda, MD).

Isolation of intracellular EToV RNA, cDNA synthesis, and PCR amplification. Total cytoplasmic RNA, isolated from the equivalent of 4×10^6 EToV-infected Ederm cells (68), was subjected to reverse transcription-PCR (RT-PCR) with RNase H-free Moloney murine leukemia virus reverse transcriptase (Superscript II; Invitrogen) and with the Expand long template PCR system (Roche Diagnostics GmbH), according to the instructions of the manufacturers. Oligonucleotide primers 1071 and 1072 (Table 1) were designed after sequences located at the 5′ end and 3′ end of EToV ORF1A, respectively (59, 61). Primers 1288 and 1325 were designed after sequences from cDNA clones p133 and p135. These clones had been identified previously as EToV specific after screening of a genomic cDNA library (61) and had been tentatively mapped in ORF1a (E. J. Snijder, unpublished). Amplicons of approximately 5.5 kb were obtained, which were gel purified, blunt ended using large fragment DNA polymerase I (LFDP) according to the instructions of the manufacturer (Invitrogen), and cloned into an EcoRV-digested vector pWSK29 (70). Sequence analysis was performed commercially (BaseClear Labservices) using the ABI PRISM BigDye Terminators v3.0 cycle sequencing kit (Applied Biosystems) on at least two independent clones in both orientations.

Preparation of EToV M^{pro}-specific antisera. The coding region for EToV M^{pro} amino acids (aa) 5 to 196 (as counted from the M^{pro} N terminus) was RT-PCR amplified with oligonucleotide primers 1548 and 1549 (Table 1). The resulting PCR product was digested with BamHI and EcoRV and inserted into HindIII (blunt ended)-BamHI-digested pQE9 (QIAGEN), yielding plasmid pQE-SS-M^{pro}(5–196). This expression vector, encoding a 22-kDa M^{pro} fragment fused at the N terminus to a His_6 tag, was introduced in *Escherichia coli* M15/pREP4 cells (QIAGEN). Expression of the recombinant protein was induced by adding IPTG (isopropyl-β-D-thiogalactopyranoside) to a final concentration of 1 mM to bacterial cultures during exponential growth (optical density [OD] of 0.6). Bacteria were lysed 5 h after induction, and the His-tagged expression product was purified by Ni-nitrilotriacetic acid affinity chromatography using the XPRESS System version 2.0 (Invitrogen) according to the instructions of the manufacturer. The recombinant protein was eluted with a mixture of 20 mM sodium phosphate, 8 M urea, and 500 mM sodium chloride (pH 6.0) containing 500 mM imidazole. The eluate was concentrated on a Centricon concentrator (30-kDa cutoff; Amicon), dialyzed for 6 h against phosphate-buffered saline (PBS)-0.05% sodium dodecyl sulfate (SDS) and, finally, dialyzed for 16 h against PBS-0.02% SDS. New Zealand White rabbits were immunized subcutaneously with 100 μg of His-M^{pro} 1548/1549 fusion protein as described previously (12); the resulting antiserum was designated K158.

Antisera specific for the N terminus of EToV M^{pro} (K270), the C terminus of the M^{pro} (K272), or the N terminus of pp1a/pp1ab (K274) were prepared by repeated subcutaneous immunization of New Zealand White rabbits with 200 μg of the bovine serum albumin-coupled synthetic peptides NH_2-SVFSKATSPFT LHARPPK-COOH, NH_2-KPLQYFHVPSFWQPFKKQ-COOH, and NH_2-MF RILKNNTRETEQHLSSSK-COOH, respectively, as described previously (63).

Construction of prokaryotic and eukaryotic EToV M^{pro} expression vectors. For prokaryotic expression of EToV M^{pro}, residues 10376 to 11472 of the EToV genome were RT-PCR amplified with oligonucleotide primers 1606 and 1607 (Table 1). A product of the anticipated size of 1,112 bp was obtained, which was gel purified, inserted into plasmid pQE9, and cloned in *E. coli* M15/pREP4 cells (QIAGEN) as described above. The resulting expression vector, pQE-SS-M^{pro}, codes for a 42-kDa polypeptide, encompassing the EToV M^{pro} domain and flanking sequences.

The expression construct pQE9-SS-M^{pro}-GST is a pQE-SS M^{pro} derivative, which consists of nucleotides (nt) 10376 to 11473 of EToV ORF1a and nt 258 to 935 of the glutathione *S*-transferase (GST) gene from plasmid pGEX2T (Amersham Pharmacia Biotech). This expression cassette was produced via PCR with oligonucleotides 2124 and 2125 and conventional cloning procedures. It codes for a 60-kDa fusion protein consisting of the 366-amino-acid-residue M^{pro} domain with flanking sequences linked by a proline residue to amino acid residues 1 to 226 of the GST protein.

For mammalian expression, we employed the recombinant vaccinia virus-based vTF7-3 system. To generate a suitable expression vector, the 1606/1607 PCR amplicon, digested with BamHI and EcoRV, was cloned into SacI/BamHI-digested pBluescript K/S+ (Strategene) together with a double-stranded linker, consisting of oligonucleotides 1721 and 1722. The resulting plasmid, pBS-SS-M^{pro}, contains residues 10376 to 11472 of the EToV genome, placed under the control of the bacteriophage T7 promoter and provided in frame with an initiation codon in an optimal context for translation initiation (40).

Site-directed mutagenesis of (putative) active-site residues was performed by splicing overlap extension-PCR mutagenesis (35) and conventional cloning techniques; mutations introduced in M^{pro} and the primers used are listed in Table 1.

The dual-expression construct pBS-SS-M^{pro}-dual, is a pBS-SS-M^{pro} derivative in which, downstream of the M^{pro} sequences, a second expression cassette is inserted, which is placed under the translational control of the encephalomyocarditis virus (EMCV) internal ribosomal entry site (IRES; taken from plasmid pTN2) (69) and which consists of nt 821 to 1128 of EToV ORF1a fused in frame to nt 10.378 to 11.472 of the EToV genome (encoding M^{pro} and surrounding sequences) and nt 3 to 483 of the EToV N gene. This chimeric expression cassette was produced via a series of splicing overlap extension-PCRs with oligonucleotides 1638, 1640, 1902, and 1914 through 1918 (Table 1) and conventional cloning procedures. It codes for a 66-kDa fusion protein, consisting as listed from the N terminus to C terminus of (i) residues 1 to 103 of pp1a; (ii) the 65 aa residues immediately upstream of M^{pro}; (iii) the 292-residue M^{pro} domain; (iv) the 8 residues downstream of M^{pro}, including the cleavage site; and (v)

TABLE 1. Oligonucleotides used in this study for amplification or mutagenesis of EToV sequences

Primer	Oligonucleotide sequence (5′ to 3′)[a]	Nucleotide no.[b]	Polarity	Use[c]
1072	GTCGTCTACGTGGTTTAAGGC	13521–13541	−	PCR
1288	GCGACACTCTCCTTCTGC	6918–6935	−	PCR
1325	TTAGTTCAATGGTTTGTTTTGAA	7973–7995	+	PCR
1548	TTGGATCCAAAGCAACATCGCCATTTACA	10586–10606	+	PCR
1549	AAGATATCTCCTTCAAAATCAGACGCC	11143–11161	−	PCR
1606	TTGGATCCGAGCAGATTTCAAAGTGTGTG	10376–10396	+	PCR
1607	AAGATATCATGACACATTATTGGAAACAG	11451–11471	−	PCR
1638	ATCACCACCTTCAGTATTATA	11045–11065	−	PCR
1639	ACTGAAGGTGGTGATTGTGGTGCGCCATTAGTATGT	11051–11086	+	CM
1640	ACTGAAGGTGGTGATGCTGGTGCGCCATTAGTATGT	11051–11086	+	CM
1694	TTTTGCTGTTACAATAGTATT	10709–10729	−	PCR
1695	ATTGTAACAGCAAAAGGTCTTTTTGAGTATGATGAT	10715–10750	+	CM
1696	ATTGTAACAGCAAAACGTCTTTTTGAGTATGATGAT	10715–10750	+	CM
1697	AGGTTCCTTCCAAATTATGGA	10811–10831	−	PCR
1698	ATTTGGAAGGAACCTAATGTAAAAGGTTGGACATTT	10817–10852	+	CM
1699	ATTTGGAAGGAACCTGAAGTAAAAGGTTGGACATTT	10817–10852	+	CM
1724	AACCTGAATGTAAGCATTTTCACC	10856–10879	−	PCR
1725	AATGCTTACATTCAGGTTGCAAATTTGAAAGATTTTTAT	10862–10900	+	CM
1726	AATGCTTACATTCAGGTTGATAATTTGAAAGATTTTTAT	10862–10900	+	CM
1727	AATGCTTACATTCAGGTTCAAAATTTGAAAGATTTTTAT	10862–10900	+	CM
1728	ACCTTTAAATGTCCAACCTTT	10838–10858	−	PCR
1729	TGGACATTTAAAGGTGCAAATGCTTACATTCAGGTT	10844–10879	+	CM
1730	TGGACATTTAAAGGTGATAATGCTTACATTCAGGTT	10844–10879	+	CM
1731	TGGACATTTAAAGGTCAAAATGCTTACATTCAGGTT	10844–10879	+	CM
1732	TTTCAAATTTTCAACCTGAAT	10871–10891	−	PCR
1733	GTTGAAAATTTGAAAGCTTTTTATATTGAGGATTTT	10877–10912	+	CM
1734	GTTGAAAATTTGAAAGAATTTTATATTGAGGATTTT	10877–10912	+	CM
1735	GTTGAAAATTTGAAAAATTTTTATATTGAGGATTTT	10877–10912	+	CM
1902	GGATCCATGAGTACCAGTTCCAGTAT	821–840	+	PCR
1914	TGAAATCTGAGTAGACCCGGTAGTCCAT	1111–1128	−	PCR
1915	CCGGGTCTACGCAGATTTCAAAGTGTGTGT	10378–10397	+	PCR
1916	TAAAGCCGGCAAAGCTAAAATTAGATGGTTG	10550–10570	−	PCR
1917	ATTTGCCGGCGTGTTTTCAAAAGCAACATCG	10577–10597	+	PCR
1918	CGGGATCCTTAATTCATAACCACCTTAATAG	27773–27795	−	PCR
2124	AAATGGCCAATGTCCCCTATACTAGGTTAT	1–21[d]	+	PCR
2125	AACTGCAGTCAGTCACGATGAATTCCCGG	679–699[d]	−	PCR

[a] The underlined sequences were required for cloning and mutagenesis of EToV sequences.
[b] The nucleotide numbers correspond to the determined sequence of equine torovirus strain Berne.
[c] Oligonucleotides were used for PCR or codon mutagenesis (CM).
[d] Position as counted from the initiation codon of the GST gene of vector pGEX2T.

residues 2 to 160 of the EToV N protein. In this fusion protein, the N-terminal M^{pro} cleavage site is abolished (QS→AG) and the M^{pro} proteinase is inactivated by an active site Ser-to-Ala substitution.

Purification and N-terminal sequence analysis of M^{pro} cleavage products. EToV M^{pro} and M^{pro}-GST were expressed from plasmids pQE-SS-M^{pro} and pQE9-SS-M^{pro}-GST, respectively, in *E. coli* M15/pREP4 as described above. Of the 1-ml bacterial culture, the cells were collected by centrifugation at 14,000 rpm for 30 s, resuspended in 100 μl Laemmli sample buffer, and incubated for 5 min at 95°C. Proteins were separated in 15% SDS-polyacrylamide gel electrophoresis (PAGE) gels (20 μl cleared supernatant/lane), transferred to polyvinylidene difluoride membranes (Bio-Rad Laboratories), and subsequently stained with Coomassie brilliant blue R-250. The regions containing the cleavage products of interest were excised. N-terminal sequence analysis was performed by subjecting the proteins to 5 to 8 rounds of Edman degradation using standard procedures with a pulse-liquid protein sequencer (model 476A; Applied Biosystems).

Infection/transfection of mammalian cells and metabolic labeling. Monolayers of Ederm cells (~1×10^6 cells), grown in 35-mm wells, were washed once with PBS containing 50 μg/ml DEAE-dextran and then infected in PBS-DEAE with EToV at a multiplicity of infection of 5 PFU/cell. At 1 h postinfection (p.i.), the inoculum was replaced by 1 ml DMEM10 and incubation at 37°C continued.

Subconfluent monolayers of OST7-1 cells (~2×10^6 cells), grown in 35-mm wells, were infected with recombinant vaccinia virus vTF7-3 and, at 1 h p.i., transfected with 2 μg plasmid DNA as described previously (46).

Prior to metabolic labeling, the cells were depleted for methionine and cysteine for 30 min by replacing the tissue culture supernatant with 600 μl of prewarmed starvation medium (DMEM without L-cysteine and L-methionine [Invitrogen] supplemented with 5% FCS, 10 mM hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 7.4], 1 mM Glutamax). Then 100 μCi of Tran^{35}S-label (ICN) was added and the cells were labeled for the indicated length of time at 37°C.

After labeling, the cells were washed once with ice-cold PBS and then lysed on ice in 600 μl ice-cold lysis buffer (20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl, 1% Triton X-100, and 1 μg/ml aprotinin, leupeptin, and pepstatin A each). Nuclei and cell debris were pelleted by centrifugation for 5 min at 14,000 rpm and 4°C, and the supernatant was subjected to radioimmunoprecipitation assay (RIPA). To this end, the lysates were adjusted to 2% SDS, diluted 10 times with detergent solution (50 mM Tris-HCl, pH 8.0, 62.5 mM EDTA, 0.4% Na deoxycholate, 1% NP-40, and 1 μg/ml of aprotinin, leupeptin, and pepstatin A each). RIPA was performed in a final volume of 1 ml with 3 μl of antiserum K158, K270, K272, K274 (see below), or R-αEToV N (14). The immune complexes were adsorbed to Pansorbin cells (Calbiochem) for 30 min at 4°C, collected by centrifugation for 1 min at 14,000 rpm, and washed three times with detergent solution containing 0.25% SDS. The pellets were then resuspended in 25 μl Laemmli sample buffer and heated for 5 min at 95°C. Proteins were separated in 15% SDS-PAGE gels, which were fixed for 30 min in 50% methanol-10% acetic acid and dried for 60 min at 80°C. Labeled proteins were visualized by fluorography using Kodak intensifying screens.

RESULTS

Comparative sequence analysis of EToV M^{pro}. The EToV Berne ORF1a was sequenced completely (GenBank accession no. DQ310701) and encodes a product, pp1a, 4,569 amino

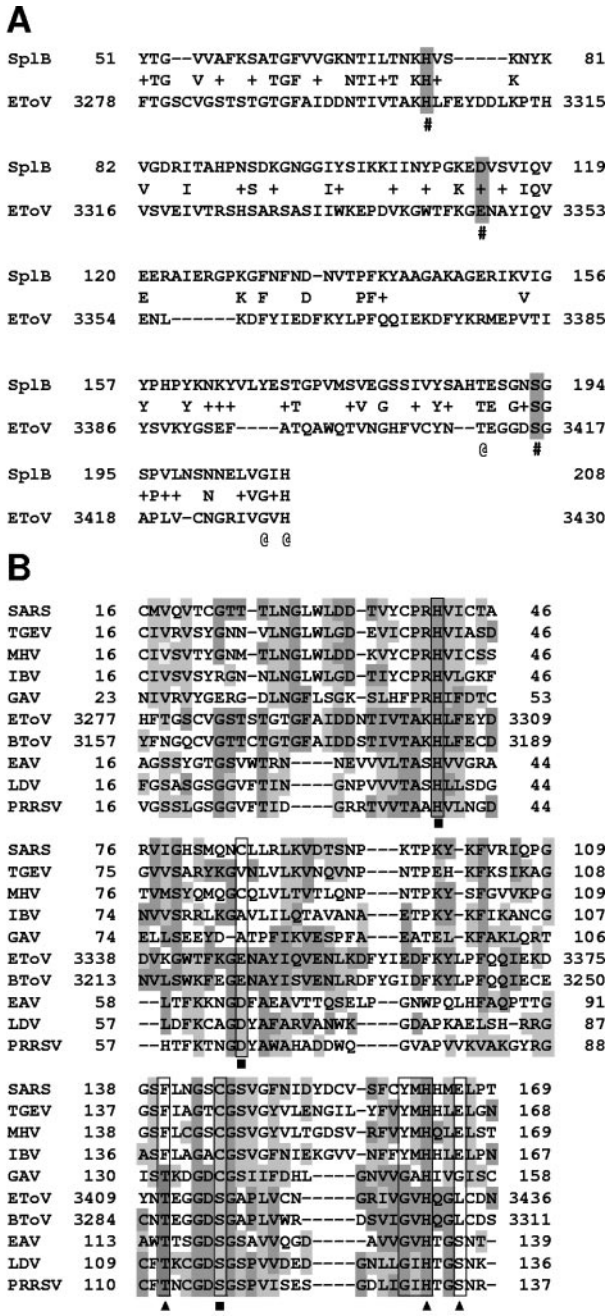

FIG. 1. Identification and sequence comparison of the toroviral M^{pro}. (A) Alignment of EToV strain Berne M^{pro} with SplB of *Staphylococcus aureus*. Standard and PHI and PSI BLAST searches with pp1a residues 2385 to 4570 as a query revealed sequence identity across an extensive region (residues 3278 to 3430) with serine proteinase SplB of *Staphylococcus aureus*. Active-site residues predicted for SplB and corresponding residues in the EToV sequence are boxed and indicated by #; predicted key residues of the S1 pocket are shown in bold and indicated by @. Residues of EToV M^{pro} are presented as numbered from the initiating Met^1 of pp1a. (B) Multiple sequence alignment of corona-, arteri-, roni-, and torovirus M^{pro} domains. Clustal X-based alignments of coronavirus, arterivirus, and gill-associated okavirus (GAV) M^{pro}s, generated previously (3, 4, 74), were used to align the EToV and BToV sequences. Abbreviations of virus names and GenBank accession numbers for the sequences are as follows: TGEV, porcine transmissible gastroenteritis virus (group 1; CAB91144); MHV, mouse hepatitis virus strain A59 (group 2; NC_001846); SARS, SARS coronavirus (group II; NC_004718); IBV, avian infectious bronchitis virus (group 3; AY692454); EToV, equine torovirus strain Berne; BToV, bovine torovirus strain Breda 1 (AY427798); GAV, gill-associated okavirus (AF227196); EAV, equine arteritis virus (NC_002532); LDV; lactate dehydrogenase-elevating virus neurovirulent type C (NC_02534); and PRRSV, porcine reproductive and respiratory syndrome virus strain Lelystad (AY588319). Residues shared by a torovirus M^{pro} and any of the other sequences are indicated by dark shading; residues conserved between M^{pro}s of different nidoviruses are indicated by light shading. Residue numbers for the mature M^{pro}s of arteri-, corona-, and roniviruses are given; residues of EToV M^{pro} are numbered from the initiating Met^1 of pp1a. Residues of the catalytic center and the S1 specificity pocket are boxed. Squares indicate the catalytic His and Ser/Cys residues, as well as the catalytic Asp residues of arterivirus M^{pro}s; the latter are aligned with SARS coronavirus and TGEV residues, which in the tertiary structure of the coronavirus M^{pro} occupy an equivalent main chain position. Triangles indicate key residues involved in P1 substrate binding.

acids long; a ribosomal frameshift at the pseudoknot at the 3′ end of ORF1a (58) would produce a C-terminally extended product, pp1ab, of 6,857 amino acids. During the preparation of our article, Draker et al. (22) reported the completion of the sequence of the bovine torovirus (BToV) strain Breda genome. These authors identified a BToV pp1a domain with sequence similarity to the adenosine diphosphate-ribose 1″-phosphatase processing enzyme family (22), a homologue of which also occurs at a similar topological position in pp1a of coronaviruses, but not in those of roni -and arteriviruses (57). Moreover, two proteinase domains were found: a cysteine proteinase within the N-terminal half of pp1a and a more C-terminally located serine proteinase (22). The pp1a polyproteins of EToV and BToV share 65% sequence identity and the three domains described by Draker and colleagures (22) are conserved in pp1a of EToV. They were in fact identified independently using standard protein-protein and PSI and PHI BLAST searches of the NCBI database with EToV pp1a sequences as a query.

The EToV serine proteinase domain, comprising residues 3278 to 3430, displayed sequence similarity with the SpIB serine protease of *Staphylococcus aureus* and with the pp1a serine proteinase of turkey astrovirus. Similarity was particularly evident in the areas surrounding the predicted active-site residues, with His^{75}, Asp^{113}, and Ser^{193} of *S. aureus* SplB (50); GenBank accession no. AF271715) aligning with His^{3304}, Glu^{3347}, and Ser^{3416} in the EToV sequence (Fig. 1A). The putative EToV proteinase domain resides within pp1a/pp1ab at a relative position similar to that of the M^{pro}s of other nidoviruses and, likewise, is flanked by hydrophobic regions (HR-1 and -2; Fig. 2A). Based upon these observations and upon the fact that no other chymotrypsin-like proteinase motifs were evident in the EToV pp1a/pp1ab sequence, we postulated the identified domain to represent the toroviral M^{pro}.

Although sequence conservation between the putative torovirus M^{pro} and the M^{pro}s of other nidoviruses is very weak, its sequence could be fitted to existing M^{pro} alignments (3, 4, 74) in the regions containing the active-site His and Cys/Ser residues. The putative EToV M^{pro} is predicted to adopt a two-β-barrel structure with His^{3304} and Ser^{3416} as the main active-site residues and, again, Glu^{3347} as the most likely candidate for the role of third residue within the catalytic triad (Fig. 1B). From

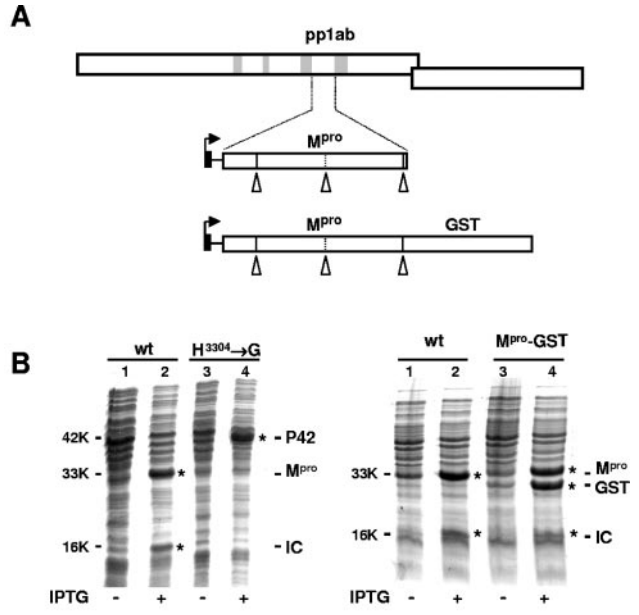

FIG. 2. Autocatalytic processing of torovirus M^{pro} in *E. coli*. (A) Schematic representation of the EToV polymerase polyprotein pp1ab with the ORF1a- and ORF1b-encoded polypeptides shown as boxes (top). Also shown is the domain organization of the C-terminal half of pp1a with hydrophobic regions depicted as gray boxes. The M^{pro} domain and flanking hydrophilic sequences (indicated as a white box) were cloned in pQE9 and placed under the control of an IPTG-inducible bacteriophage T5 promoter (indicated by the black box and arrow). Furthermore, the structure of a construct encoding M^{pro} fused to GST is shown. White arrowheads indicate (predicted) M^{pro} cleavage sites (bottom). (B) Prokaryotic expression of the EToV M^{pro} domain. (Left panel) *E. coli* M15/pREP4 cells, transformed with pQE-SS-M^{pro} (lanes 1 and 2) or pQE-SS-M^{pro}/Gly^{3304} (lanes 3 and 4), were grown in the presence (+) or absence (−) of 1 mM IPTG for 5 h. Total cell lysates were analyzed in 15% SDS-polyacrylamide gels. Proteins were stained with Coomassie brilliant blue. Positions of the 42-kDa precursor (P42) and the 33-kDa M^{pro} and 16-kDa internal cleavage products (IC) are indicated. (Right panel) Expression of an EToV M^{pro}-GST chimera in *E. coli*. *E. coli* M15/pREP4 cells, transformed with pQE-SS-M^{pro} (lanes 1 and 2) or pQE9-SS-M^{pro}-GST (lanes 3 and 4), were grown in the presence (+) or absence (−) of 1 mM IPTG for 5 h. Total cell lysates were analyzed in 15% SDS-polyacrylamide gels. Proteins were stained with Coomassie brilliant blue. Positions of the 33-kDa M^{pro} and 16-kDa internal cleavage products are indicated as well as the 26-kDa GST product. wt, wild type.

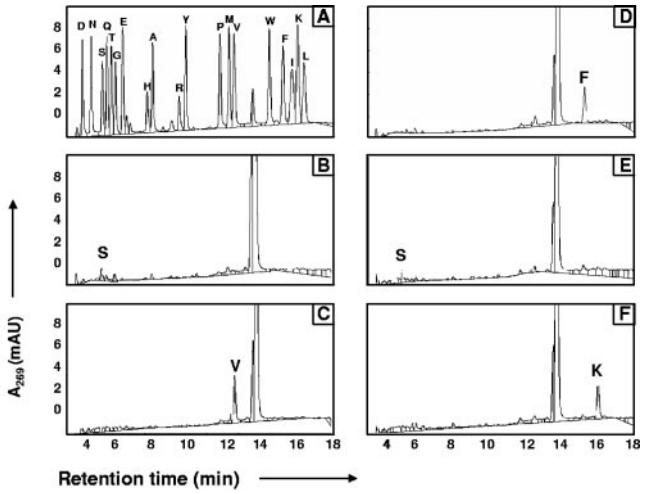

FIG. 3. Characterization of the N-terminal EToV M^{pro} autoprocessing site by protein sequencing. The 33-kDa cleavage product, purified from bacterial lysates (Fig. 2B, lane 2), was subjected to Edman degradation. Phenylthiohydantoin (PTH)-amino acids generated during each reaction cycle were detected by their absorbance at 269 nm (expressed as milliabsorption units [mAU]; *y* axis) and identified by their characteristic retention times (in minutes; *x* axis) on a reversed-phase high-pressure liquid chromatography support. (A) Chromatogram of PTH-amino acid standards. (B to F) Chromatograms of PTH-amino acids from reaction cycles 1 to 5. Specific peaks are indicated by the single-letter code.

our data, the torovirus M^{pro} appears to be no more closely related to the coronavirus M^{pro} than to its roni- and arterivirus homologues. Perhaps most surprising, the torovirus M^{pro} differs from the M^{pro}s of corona- and roniviruses in that it is not a cysteine proteinase, but rather, like its arterivirus homologue, uses serine as the principal nucleophile.

Proteolytic activity of EToV M^{pro}. The bioinformatical analyses of the putative EToV M^{pro} provided a theoretical framework for subsequent experimental studies. The hydrophilic region between HR-1 and -2, which spans the predicted proteinase domain, was expressed in *E. coli* (Fig. 2). A construct in which the predicted active-site His residue was replaced by Gly was used as a control (Fig. 2B). IPTG-induced expression of these constructs should yield a 42-kDa primary translation product. Indeed, a protein of this size was detected in cells expressing the His-to-Ala mutant. However, in cells expressing the intact proteinase domain, two products with apparent molecular masses of 33 and 16 kDa were found instead. These results suggested that the EToV pp1a region comprising residues 3186 to 3550 indeed contains a functional proteinase domain mediating autocatalytic processing.

The 33- and 16-kDa products were purified from *E. coli* and N-terminally sequenced by performing 5 or 8 cycles of Edman degradation, respectively. Analysis of the 33-kDa product indicated that cleavage had occurred at the sequence ^{3246}SNFS FQ ↓ SVFSKAT^{3258}, thus identifying Ser^{3252} as the N-terminal residue of EToV M^{pro} (Fig. 3) (Note that, throughout this article, amino acid residues flanking the scissile bond, indicated by ↓, are given from N to C terminus in the single-letter code, where "x" indicates any residue.) The 16-kDa product proved to be a mixture of two protein species. For one of these, the N terminus was identical to that of the 33-kDa product. The other product apparently arose from proteolytic cleavage at an internal sequence, ^{3392}SEFATQ ↓ AWQTVN^{3403} (data not shown). Indeed, cleavage of M^{pro} at this site would cut the proteinase effectively in half, thus giving rise to two equal-size (~16.5 kDa) products. These results tentatively identified FxxQ ↓ (S, A) as the EToV M^{pro} substrate consensus sequence (residues are given in parentheses to indicate the variation occurring at a particular position within the cleavage site).

Analysis of the M^{pro} flanking sequences identified a putative additional cleavage site downstream of the proteinase domain, ^{3540}FKKQ ↓ S^{3544}. In the full-length precursor product encoded by plasmid pQE-SS-M^{pro}, this sequence is located only 5 residues upstream of the C terminus. Hence, cleavage at this site could not be confirmed. As the hydrophobic region HR-2, C-terminally flanking M^{pro}, might interfere with efficient ex-

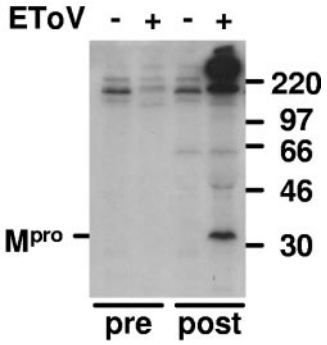

FIG. 4. Detection of M^{pro} in EToV-infected cells. Ederm cells, infected (+) or mock-infected (−) with EToV, were metabolically labeled with [^{35}S]Met plus Cys from 6 to 9 h p.i. Cell lysates were supplemented with SDS to a final concentration of 2% in order to dissociate protein-protein complexes. The lysates were then diluted 10-fold with detergent solution and subjected to RIPA with polyclonal antiserum K274, directed against the C-terminal 17 residues of M^{pro} (post) or with preserum (pre). Similar results were obtained with polyclonal antisera K158 and K272, directed against a peptide comprising residues 5 through 196 or the N-terminal 17 residues of EToV M^{pro}, respectively (not shown). The samples were analyzed in 15% SDS-polyacrylamide gels. Molecular size markers are given in kilodaltons. The 33-kDa M^{pro} species is indicated. High-molecular-mass products—presumably M^{pro} precursors—are evident at the top of the gel.

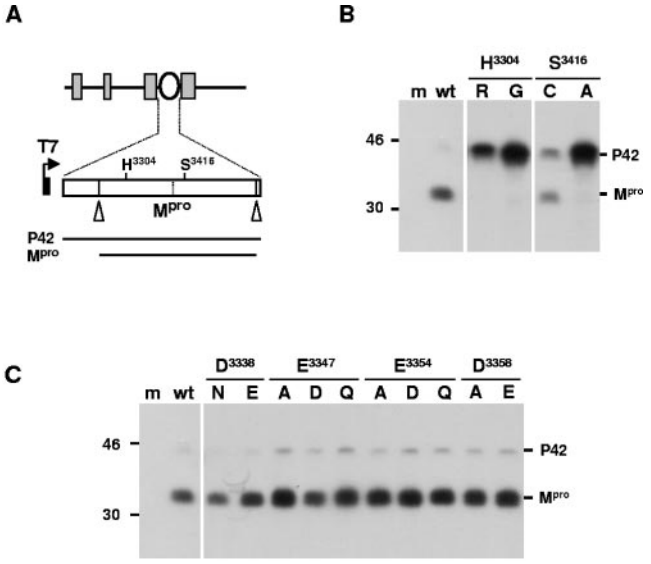

FIG. 5. Mutagenesis of predicted active-site residues of EToV M^{pro}. (A) Mammalian expression constructs for the EToV M^{pro} domain. The upper panel shows the organization of the C-terminal half of pp1a with hydrophobic regions depicted as gray boxes and the M^{pro} domain as an oval. The M^{pro} domain and flanking hydrophilic sequences (represented by a white box) were cloned in pBluescript and placed under the control of the bacteriophage T7 promoter (depicted by the black box and arrow). The relative locations of the predicted active-site His^{3304} and Ser^{3416} residues are indicated. White arrowheads indicate M^{pro} cleavage sites. Also shown schematically are the structure of the P42 precursor and that of the mature 33-kDa M^{pro} species. (B) vTF7-3 expression of EToV M^{pro}. OST7-1 cells were infected with recombinant vaccinia virus vTF7-3 and transfected with pBluescript-based expression vectors encoding either the wild-type EToV M^{pro} (wt) or mutants in which the predicted active-site His^{3304} and Ser^{3416} residues had been replaced by Arg or Gly or Cys or Ala, respectively. Mock-transfected cells were taken along as a negative control (m). Cells were metabolically labeled from 5 to 6 h p.i. Cell lysates were subjected to RIPA with a polyclonal antibody directed against M^{pro} (K158). Precipitates were analyzed in 15% SDS-polyacrylamide gels, and labeled proteins were visualized by fluorography. The positions of the 42-kDa precursor (P42) and of the 33-kDa mature M^{pro} species are indicated. Molecular masses of marker proteins are given in kilodaltons. (C) Mutagenesis of putative active-site acidic residues. The effect of conservative and nonconservative substitutions of Asp^{3338}, Glu^{3347}, Glu^{3354}, and Asp^{3358} was assessed in the vTF7-3 system as described above. Amino acid residues are indicated in the single-letter code.

pression in *E. coli*, we did not extend the expression product with EToV sequences, but instead designed an M^{pro}-fusion protein, which contained the predicted cleavage site including the downstream seven residues, linked via a newly introduced proline residue to residues 1 to 226 of GST (Fig. 2A). Expression in *E. coli* yielded, as anticipated, two major products of 33 kDa and 26 kDa (Fig. 2B), apparently corresponding to mature M^{pro} and the cleaved-off GST moiety. N-terminal sequence analysis of the 26-kDa product confirmed this notion and indicated that proteolytic cleavage indeed had occurred at the predicted site ^{3543}Gln ↓ Ser^{3544} (data not shown).

Cleavage at the sites upstream and downstream of the M^{pro} domain should yield a 33-kDa protein. Accordingly, in lysates from metabolically labeled EToV-infected cells, a product of this size, comigrating exactly with the bacterial cleavage product, was detected by RIPA with M^{pro}-specific antisera (Fig. 4) (data not shown). The 16-kDa products, however, were not detected in EToV-infected cells.

Mutagenesis of the predicted EToV M^{pro} active-site residues. Further analysis of the EToV M^{pro} was performed by employing the vaccinia virus-based vTF7-3 mammalian expression system (24) (Fig. 5). To this end, residues 10376 to 11472 of the EToV genome were provided with an AUG codon in an optimal context for translation initiation (40) and the resulting ORF was placed under the control of the bacteriophage T7 promoter (Fig. 5A). Expression of this construct predominantly yielded the 33-kDa M^{pro} species. To assess the importance of the predicted catalytic residues, single-amino-acid substitutions were introduced, and their effects were studied by analyzing the processing activities of the mutant proteinases. Consistent with their proposed catalytic function, replacement of His^{3304} by either Gly or Arg and of Ser^{3416} by Ala completely abolished proteolytic activity. Interestingly, however, an M^{pro} mutant with a Ser^{3416}→Cys substitution retained partial activity (Fig. 5B).

EToV M^{pro} displays the distinctive traits of a classical serine proteinase, and hence the existence of a third acidic catalytic residue was surmised. On the basis of the alignments (Fig. 1A and B), Glu^{3347} was considered the most likely candidate to occupy a main chain position equivalent to that of the catalytic Asp residue in chymotrypsin. In addition, Asp^{3383}, Glu^{3354}, and Asp^{3358} were selected for mutagenesis. However, the substitution for neither of these residues by Asn or Ala affected proteolytic processing of the 42-kDa M^{pro} precursor (Fig. 5C). One explanation for this result would be that yet another acidic residue is part of the catalytic triad. Still, the fact that our assay relied on autocatalytic cleavages that may occur in *cis* posed a caveat. Mutagenesis studies of other serine proteases have

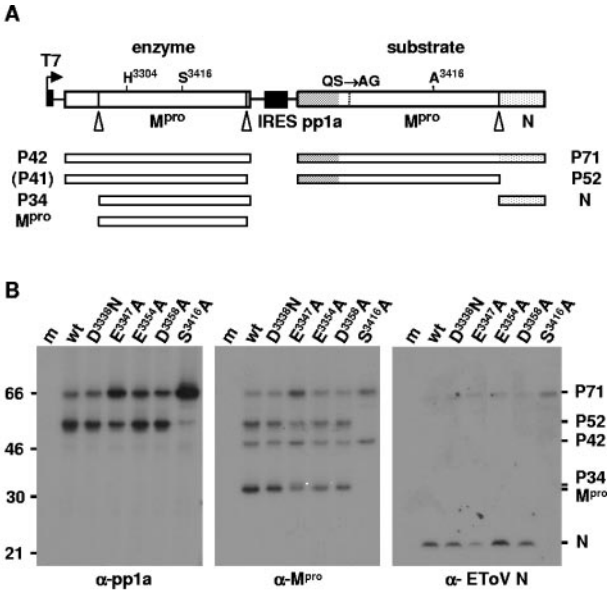

FIG. 6. Effect of substitutions of possible active-site acidic residues as determined in a *trans*-cleavage assay. (A) Schematic outline of the *trans*-cleavage assay. The upper panel shows the structure of the expression constructs with the ORFs encoding the M^{pro} enzyme and substrate depicted as boxes. The expression of the ORFs was under the control of the T7 promoter (black box and arrow). Translation of the downstream ORF for the substrate was driven by the EMCV IRES (indicated by a black box). For the M^{pro} enzyme, the relative locations of the active-site His and Ser residues are indicated. Cleavage sites in the M^{pro} precursor and in the substrate are indicated by white arrowheads. Substrate sequences derived from ORF1a/pp1a and from EToV N are indicated by hatching or dots, respectively. In the substrate, the N-terminal cleavage site was abolished (QS→AG) and the enzyme was inactivated by a Ser^{3416}→Ala substitution. Shown below are the structures of the various precursor and cleavage products. (B) Mutagenesis of possible active-site acidic residues. The effect of conservative and nonconservative substitutions of Asp^{3338}, Glu^{3347}, Glu^{3354}, and Asp^{3358} was assessed in the vTF7-3 system essentially as described in Fig. 5. Substitutions are indicated in the single-letter code. Metabolic labeling was from 3.5 to 4.5 h p.i., and RIPA was performed with antisera K270 (α-pp1a), K158 (α-M^{pro}), and rabbit anti-EToV N (α-EToV N). The positions of the various precursors and mature products are indicated on the right. Molecular masses of marker proteins are given in kilodaltons on the left. Amino acid residues are indicated in the single-letter code. m, mock-transfected negative control; wt, wild type.

shown that the acidic active-site residue is not absolutely required for peptide hydrolysis, although its substitution may cause the enzymatic activity to drop by several orders of magnitude (11, 16, 33, 41, 64, 65). Yet, in an experimental setup in which processing occurs by highly efficient cleavage in *cis*, even such considerable differences in enzymatic activity may go unnoticed. We therefore designed a more sensitive *trans*-cleavage assay, outlined in Fig. 6A, which entailed coexpression of enzymatically active M^{pro} and a substrate protein. Enzyme and substrate were expressed from the same transcript with translation of the latter being driven by the EMCV IRES.

The substrate was a 71-kDa chimeric protein consisting of (i) residues 1 to 103 of pp1a/pp1ab, (ii) a EToV M^{pro}-containing segment consisting of pp1a/pp1ab residues 3187 to 3551, and (iii) residues 2 to 160 of the EToV N protein, as listed from N to C terminus. In this fusion protein, the N-terminal M^{pro} cleavage site was abolished by a QS→AG substitution, while the M^{pro} was inactivated by changing the active site Ser^{3416} to Ala. However, the cleavage site C terminal of M^{pro} was left intact and hence remained available for proteolytic cleavage in *trans*. This should give rise to products of 52 kDa (the pp1a-M^{pro} fusion protein) and of 19 kDa (the cleaved-off N moiety), which are readily distinguishable from the enzymatically active M^{pro} and its precursors. Processing was monitored by RIPA with antisera against the N terminus of pp1a, M^{pro}, and the N protein and the effect of substitutions of the selected acidic residues was assessed (Fig. 6B). A construct in which the M^{pro} enzyme was inactivated by a Ser^{3416}→Ala substitution was taken along as a negative control.

M^{pro} derivatives, in which Asp^{3338}, Glu^{3354}, and Asp^{3358} had been altered, behaved like the wild-type enzyme in the *trans*-cleavage assay. Substitution of Glu^{3347}, however, reduced proteolytic activity as indicated by accumulation of the 71-kDa noncleaved substrate precursor and of the M^{pro} precursors of 42 and 34 kDa; the latter protein species apparently represented an intermediate generated by processing of the 42-kDa M^{pro} precursor at the N-terminal but not at the C-terminal cleavage site (Fig. 6B).

DISCUSSION

The main proteinases, which mediate most of the processing steps of nidovirus replicase polyproteins, have been studied so far only for corona-, arteri-, and roniviruses. Here, we report the biochemical characterization of the fourth member, the torovirus M^{pro}. The nidoviral M^{pro}s all belong to a family of viral and cellular proteinases, resembling picornavirus 3C (3C-like proteinases [3CLPs]). Moreover, they are located at cognate positions within the C-terminal half of pp1a, typically flanked by hydrophobic domains at either side (74, 76). Hence, the null hypothesis is that the proteinases of extant nidoviruses have all evolved from a single ancestral proto-nidovirus 3CLP. However, if so, they have diverged to an extreme extent. As noted previously (74, 76), primary sequence diversity of the corona-, roni-, and arterivirus M^{pro}s is so pronounced that even the most sensitive computer-aided methods do not reveal any specific relationship among them that would justify their grouping and separation from other cellular or viral 3CLPs.

Within the order *Nidovirales*, the toroviruses seem to be most closely related to the coronaviruses (10, 25, 58, 60). We therefore anticipated at the onset of this study that the torovirus M^{pro} might bridge the large evolutionary gap between the main proteinase of the coronaviruses and those of the other nidoviruses (76). Surprisingly, however, from its primary sequence, the M^{pro} of EToV Berne seems no more related to the coronaviral proteinase than to those of the arteri- and roniviruses (Fig. 1B). The characteristics of the torovirus M^{pro} in relation to those of the other nidoviruses are discussed in the paragraphs below.

M^{pro}-mediated processing and cleavage site specificity. Nidoviral M^{pro}s and related proteinases in other RNA viruses are released from adjoining polyprotein domains by autocatalytic processing (21, 76). Likewise, when pp1a segments, containing the predicted EToV M^{pro} domain, were heterologously expressed in prokaryotic and in mammalian cells, the proteinase liberated itself by cleaving both at an N-terminal site and at a C-terminal site. This yielded a mature M^{pro} of 33 kDa, a

protein of which size was also detected in EToV-infected cells by RIPA with M^{pro}-specific antisera. Interestingly, in EToV-infected cells, autocatalytic release of M^{pro} might be inefficient as during a 3-h continuous metabolic labeling period a considerable proportion of the proteinase seemed to remain part of a large precursor (Fig. 4). Similar observations were reported for the arteri- and coronavirus M^{pro}s (32, 54, 63, 71). It has been suggested, that the long-lived M^{pro} intermediates are membrane associated via the flanking hydrophobic domains. Anchoring of proteinase precursors to intracellular membranes might well be a central phenomenon in nidovirus polyprotein processing and/or replication (39, 54, 71).

N-terminal sequence analysis of mature EToV M^{pro} showed that autocatalytic cleavage occurred at the location ^{3246}SNFS FQ ↓ SVFS^{3255}. The sequence ^{3538}QPFKKQ ↓ SVSN^{3547} was identified as the C-terminal cleavage site. Unexpectedly, in prokaryotic cells, processing occurred also at a site internal to M^{pro}, ^{3392}SEFATQ ↓ AWQT^{3401}, effectively cleaving the proteinase in half. It is tempting to speculate that, during EToV infection, M^{pro} might inactivate itself as part of a feedback autoregulatory mechanism or, otherwise, that processing at the internal site gives rise to an alternative cleavage pathway (71). However, we have no evidence that internal processing actually occurs in mammalian cells. The 16-kDa cleavage products were detected neither in the vTF7-3 expression system nor in EToV-infected cells. Hence, it cannot be excluded that internal cleavage, as observed in *E. coli*, is an artifact resulting from overexpression and possibly aided by protein misfolding, which might have exposed cleavage sites that are normally inaccessible. Intriguingly, however, the internal cleavage site is also conserved in BToV M^{pro} (22).

Catalytic center and substrate specificity. Although sequence identity between EToV M^{pro} and its nidovirus homologues or other viral and cellular chymotrypsin-like enzymes is limited, sequence alignments clearly implicated Ser^{3416} and His^{3304} as the main catalytic residues. Consistent with this notion, their replacement by Ala and by Gly or Arg, respectively, abrogated enzymatic activity. The M^{pro}s of corona- and roniviruses (2, 73, 74), like the hepatitis A virus 3C proteinase (1, 7), employ a catalytic Cys-His dyad within the context of the chymotrypsin-fold. Under the assumption that the EToV M^{pro} is a classical serine proteinase, a third acidic catalytic residue was expected (20). Although sequence alignments of EToV M^{pro} with other proteinases were far from robust, particularly in the region that should contain this third residue, our analyses suggested Glu^{3347} as the most likely candidate. However, neither substitution mutagenesis of Glu^{3347} nor that of the proximal Asp^{3338}, Glu^{3354}, and Asp^{3358} noticeably affected autocatalytic release of M^{pro} in heterologous expression studies. Yet, in an assay that tested for processing in *trans*, the efficacy of M^{pro} cleavage of an artificial substrate was clearly reduced by substitution of Glu^{3347}, but not by that of the other acidic residues. Although our findings do not prove that Glu^{3347} is included in a catalytic triad with Ser^{3416} and His^{3304}, they would at least be consistent with this notion. This view might seem at odds with observations made for various other viral and cellular chymotrypsin-like proteinases, where site-directed mutagenesis of the third residue resulted in a far more dramatic loss of proteolytic activity (11, 16, 33, 41, 64, 65). Still, in the case of severe acute respiratory syndrome (SARS) coronavirus M^{pro}, replacement of the active-site Cys by Ser yields a functional enzyme. Although this mutant proteinase relies exclusively on a Ser-His catalytic dyad, its activity is merely 40-fold reduced as compared to that of wild-type SARS coronavirus M^{pro} (36). A reduction in activity of this magnitude might have only a modest effect on processing even in our *trans*-cleavage assay under conditions of vTF7-3-driven EToV M^{pro} overexpression. As the mapping of active-site residues by the combination of comparative sequence analysis and site-directed mutagenesis is inherently fraught with a measure of uncertainty, definitive answers with respect to catalytic center composition and the presence/identity of a third catalytic residue will require the elucidation of the torovirus M^{pro} crystal structure.

Comparison of the sequences flanking the scissile bonds of the internal and the N- and C-terminal cleavage sites tentatively identifies FxxQ ↓ (S, A) as the substrate consensus sequence of EToV M^{pro}. Its cleavage site specificity thus resembles that of other nidoviral M^{pro}s, which, as is typical for 3CLPs, all require Gln or Glu at the P1 position (74, 76). The specificity for P1-(Gln, Glu) of the picorna-, arteri-, and coronavirus enzymes is determined by a strictly conserved His, which via its imidazole side chain hydrogen bonds to the carbonyl/carboxylate oxygens of the P1 side chain (1, 2, 4, 7, 44, 45, 47, 73, 74, 76). A corresponding His residue is also present in the torovirus M^{pro} (EToV His^{3430}) as part of a Gly-X-His sequence (Fig. 1B), a motif well conserved in most 3CLPs (including arteri- and ronivirus M^{pro}), but saliently absent in the M^{pro}s of coronaviruses; the latter have a Tyr-Met-His motif instead (2, 4, 34, 64, 74, 75). In many 3CLPs, conserved Ser/Thr residues, located five residues upstream of the catalytic Ser/Cys, are also part of the substrate-binding subsite S1 and assist in binding by forming additional hydrogen bonds with the carbonyl/carboxylate oxygen atoms of the P1 side chain (4, 44, 47, 48, 74). We propose that, in the torovirus M^{pro}, Thr^{3411} plays a similar role. This would again be a difference from the coronavirus M^{pro}, which lacks a corresponding Ser/Thr residue (2, 74).

The arteri- and ronivirus M^{pro}s prefer glutamic acid at the P1 position of their substrates, whereas coronavirus M^{pro} requires glutamine (64, 74, 76). In arterivirus M^{pro}s, this substrate preference is presumably conferred by a Ser residue (located three residues downstream of the S1 subsite His; S^{137} in EAV M^{pro}; see Fig. 1B), which would hydrogen bond with the carboxylate oxygen atom (that is, the one not interacting with the subsite S1 His/Thr pair) of the P1-Glu side chain (4). This Ser is not conserved in the corona- and torovirus M^{pro}s; instead, at the corresponding positions, Glu and Leu residues, respectively, are found (Fig. 1b). The S1 specificity site of coronavirus M^{pro} consists of the main chain atoms of Met^{165}, Glu^{166}, and His^{172} and the side chains of Phe^{140} and His^{163}, with the imidazole of the latter residue interacting with the P1-Gln side chain (2, 3, 73) (residue numbering according to SARS-CoV M^{pro}). Also with respect to S1 subsite composition, the torovirus main proteinase differs from that of coronaviruses. Although EToV M^{pro}, like its coronavirus homologues, apparently prefers P1-Gln, none of the coronavirus S1 subsite residues are conserved, except for His^{163}. As yet, it is not clear from the EToV M^{pro} sequence how the P1-Gln is accommodated.

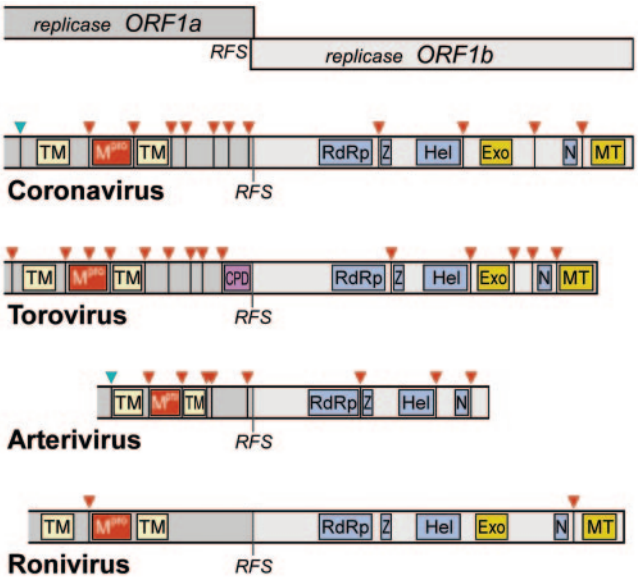

FIG. 7. Comparison of the pp1a/pp1ab domain organization in different nidoviruses and a tentative M^{pro} processing scheme of the torovirus replicase polyproteins. The C-terminal half of pp1a (dark gray) and the ORF1b-encoded part of pp1ab (light gray) are shown schematically, to scale, for corona-, toro-, arteri-, and roniviruses. The border between amino acids encoded in ORF1a and ORF1b is indicated as RFS (ribosomal frameshift). The locations of domains that have been identified as structurally or functionally related are highlighted and indicated as follows: TM, putative transmembrane domains; M^{pro}, main proteinase; RdRp, RNA-dependent RNA polymerase motif; Z, zinc finger motif; Hel, helicase; Exo, (putative) 3′-to-5′ exonuclease (ExoN); N, uridylate-specific endoribonuclease (NendoU); MT, (putative) 2′-*O*-methyl transferase (26, 30; see also reference 57 and references therein). Also indicated is the (predicted) cyclic phosphodiesterase domain (CPD) that resides near the C terminus of the torovirus pp1a. Arrowheads represent known cleavage sites for accessory papain-like proteinase (cyan) or M^{pro} (red), as established for corona-, arteri-, and roniviruses and as established casu quo predicted here for toroviruses (see also Table 2). The figure was adapted from reference 56.

Tentative M^{pro} processing scheme of torovirus pp1a/pp1ab. Although each of the three established EToV M^{pro} cleavage sites conforms to the consensus FxxQ ↓ (S, A), we believe, on the basis of alignments of pp1a/pp1ab of coronaviruses and the toroviruses EToV and BToV, that Tyr/Met/Leu and Gly/Lys may also be tolerated at the P4 and P1′ positions, respectively. Although the picture is still incomplete, the substrate specificity of the torovirus M^{pro} is reminiscent of and yet distinct from those of the corona-, arteri-, and roniviruses, which mainly conform to LQ ↓ (S, A), E ↓ (G, S, A), and VxHE ↓ (L, V), respectively (74, 76). In particular, a preference for a bulky hydrophobic residue at P4 and no apparent predilection for certain residues at P2 would give the EToV M^{pro} a unique cleavage site formula. A tentative M^{pro} processing scheme of EToV pp1a/1b is presented in Fig. 7, with the sequences and locations of potential cleavage sites listed in Table 2. Several points are of note. (i) All cleavage sites predicted for EToV are conserved in BToV (22). (ii) A putative cleavage site, ^{2870}FKKQ ↓ SV^{2875} (Table 2, site 1), identical in sequence to the one C terminal of the M^{pro} domain (Table 2, site 4), is found halfway in EToV pp1a at a position topologically similar to that of an accessory papain-like proteinase 2 site within coronavirus pp1a. (iii) Of the predicted sites in the ORF1b-encoded part of the replicase (Table 2, sites 10 to 14), only one, located between the RdRp and zinc finger/helicase domains, conforms to the proposed consensus FxxQ ↓ (S, A) sequence (site 10; Table 2). Future studies should determine whether the predicted sites within pp1a/pp1ab are actually cleaved by M^{pro}. (iv) At the very C terminus of torovirus pp1a, there is a domain that is absent from all other nidovirus pp1a proteins and that shares identity with a predicted cyclic phosphodiesterase (CPD) motif in the ns2A protein of group 2 coronaviruses and in the rotavirus VP3 guanylyl transferase (Fig. 7) (57, 60). Puzzlingly, a cleavage site C terminal of this domain is not immediately apparent, raising the question whether M^{pro} utilizes a noncanonical site, whether another viral or host proteinase is involved, or whether processing of pp1ab yields an RdRp with the CPD homologue attached; note that in the latter case, processing of pp1a would still produce an additional, "free" CPD.

Evolutionary implications. EToV M^{pro} combines several traits of its closest relatives. (i) As is typical for nidovirus M^{pro}s, the EToV proteinase has a predicted three-domain structure, with two catalytic β-barrel domains and an additional C-terminal domain of unknown function. (ii) With respect to substrate specificity, the EToV M^{pro} resembles its coronavirus homologue in its preference for P1-Gln. (iii) Then again, it has what appears to be a canonical substrate-binding subsite, S1, resembling that of arteri- and ronivirus M^{pro}s, but clearly distinct from that of coronavirus M^{pro}. (iv) Finally, in contrast to the M^{pro}s of corona- and roniviruses, but like that of arterivirus, the torovirus M^{pro} uses serine instead of cysteine as its principal nucleophile. Rather than clarifying the evolutionary history of the nidoviral M^{pro}s, the analysis of a torovirus M^{pro} seems to make the picture even more complex. If anything, our findings further emphasize the unique properties of the coronavirus M^{pro}, which on the basis of its catalytic system and substrate-binding pocket, was already considered an outlier among viral and cellular chymotrypsin-like homologues (2, 74). It is still quite possible, however, that striking similarities between corona- and toroviral M^{pro}s will become

TABLE 2. Tentative cleavage sites in pp1a/pp1ab for the torovirus main proteinase

Site no.	Cleavage site[a]	Position[b]
1	GNFKKQ ↓ SVAY	2868–2877
2	SNFSFQ ↓ SVFS	3246–3255
3	SEFATQ ↓ AWQT	3392–3401
4	QPFKKQ ↓ SVSN	3538–3547
5	EVFSPQ ↓ AGSY	3797–3806
6	PTLMWQ ↓ AEDV	3974–3983
7	GIFKQQ ↓ ANIG	4156–4165
8	LQYRQQ ↓ GVRL	4241–4250
9	PVFKPQ ↓ SIVY	4394–4403
10	CEFKQQ ↓ ANFD	5405–5414
11	FGMEKQ ↓ SDFN	5963–5972
12	SDYVEQ ↓ ALSP	6295–6304
13	LYYLQQ ↓ GKGE	6437–6446
14	QTFYPQ ↓ KDFV	6587–6596

[a] Amino acid sequence of the putative torovirus M^{pro} cleavage sites.
[b] Amino acid position as counted from Met[1] of torovirus pp1a/pp1ab.

overt once the crystal structure of toroviral M^{pro} has been solved.

Corona- and toroviruses share many traits (e.g., with regard to virion composition, genome organization, and the extent of colinearity in the ORF1b-encoded region of pp1ab), which would justify their grouping together, separate from the arteri- and roniviruses (15, 19, 25, 26). However, in a recent study of Gonzalez et al. (25), entailing a systematic quantitative analysis of sequence conservation among nidoviral proteins, the expected sequence affinity between corona- and toroviruses was not evident. Thus, with the topology of the nidovirus tree undecided, it was suggested that "the nidovirus phylogeny must be verified later when the number and diversity of torovirus and ronivirus sequences will match those of arteriviruses and coronaviruses" (25). If it were to be firmly established that of the nidoviruses, corona- and toroviruses are, after all, the most closely related, and if this evolutionary relationship also holds for the M^{pro} (i.e., if no RNA recombination-based exchange of proteinase domains—or parts thereof—has occurred during nidoviral divergence), our observations would have one more major implication: namely that the transition from serine- to cysteine-based proteolytic catalysis (or vice versa) has happened more than once in the course of nidovirus evolution. In this respect, it is of interest that a mutant EToV M^{pro} with a Ser^{3416}→Cys substitution retained enzymatic activity and that, conversely, the catalytic site Cys→Ser substitution in SARS-CoV M^{pro} also yields an active enzyme (36).

ACKNOWLEDGMENTS

We acknowledge the contribution of Alexander E. Gorbalenya to the bioinformatical analysis of EToV ORF1a-derived cDNA clones 133 and 135 and EToV pp1a, as well as to the design and analysis of some of the experiments. We are grateful to Peter J. M. Rottier and Jolanda D. F. de Groot-Mijnes for critical reading of the manuscript.

REFERENCES

1. **Allaire, M., M. M. Chernaia, B. A. Malcolm, and M. N. James.** 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature **369:**72–76.
2. **Anand, K., G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld.** 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J. **21:**3213–3224.
3. **Anand, K., J. Ziebuhr, P. Wadhwani, J. R. Mesters, and R. Hilgenfeld.** 2003. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science **13:**13.
4. **Barrette-Ng, I. H., K. K. Ng, B. L. Mark, D. Van Aken, M. M. Cherney, C. Garen, Y. Kolodenko, A. E. Gorbalenya, E. J. Snijder, and M. N. James.** 2002. Structure of arterivirus nsp4. The smallest chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. J. Biol. Chem. **277:**39960–39966.
5. **Bazan, J. F., and R. J. Fletterick.** 1989. Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. Virology **171:**637–639.
6. **Bazan, J. F., and R. J. Fletterick.** 1988. Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. Proc. Natl. Acad. Sci. USA **85:**7872–7876.
7. **Bergmann, E. M., S. C. Mosimann, M. M. Chernaia, B. A. Malcolm, and M. N. G. James.** 1997. The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. J. Virol. **71:**2436–2448.
8. **Bhardwaj, K., L. Guarino, and C. C. Kao.** 2004. The severe acute respiratory syndrome coronavirus Nsp15 protein is an endoribonuclease that prefers manganese as a cofactor. J. Virol. **78:**12218–12224.
9. **Brierley, I., P. Digard, and S. C. Inglis.** 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell **57:**537–547.
10. **Cavanagh, D.** 1997. Nidovirales: a new order comprising Coronaviridae and Arteriviridae. Arch. Virol. **142:**629–633.
11. **Chambers, T. J., C. S. Hahn, R. Galler, and C. M. Rice.** 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. **44:**649–688.
12. **Chirnside, E. D., P. M. Francis, and J. A. Mumford.** 1995. Expression cloning and antigenic analysis of the nucleocapsid protein of equine arteritis virus. Virus Res. **39:**277–288.
13. **Choi, H. K., L. Tong, W. Minor, P. Dumas, U. Boege, M. G. Rossmann, and G. Wengler.** 1991. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. Nature **354:**37–43.
14. **Cornelissen, L. A. H. M.** 1999. Molecular characterization of ungulate toroviruses. Ph.D. thesis. Utrecht, The Netherlands.
15. **Cowley, J. A., C. M. Dimmock, K. M. Spann, and P. J. Walker.** 2000. Gill-associated virus of Penaeus monodon prawns: an invertebrate virus with ORF1a and ORF1b genes related to arteri- and coronaviruses. J. Gen. Virol. **81:**1473–1484.
16. **Craik, C. S., S. Roczniak, C. Largman, and W. J. Rutter.** 1987. The catalytic role of the active site aspartic acid in serine proteases. Science **237:**909–913.
17. **Demler, S. A., and G. A. de Zoeten.** 1991. The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus. J. Gen. Virol. **72:**1819–1834.
18. **den Boon, J. A., E. J. Snijder, E. D. Chirnside, A. A. F. de Vries, M. C. Horzinek, and W. J. M. Spaan.** 1991. Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. J. Virol. **65:**2910–2920.
19. **de Vries, A. A. F., M. C. Horzinek, P. J. M. Rottier, and R. J. de Groot.** 1997. The genome organization of the Nidovirales: similarities and differences between Arteri-, Toro-, and Coronaviruses. Semin. Virol. **8:**33–47.
20. **Dodson, G., and A. Wlodawer.** 1998. Catalytic triads and their relatives. Trends Biochem. Sci. **23:**347–352.
21. **Dougherty, W. G., and B. L. Semler.** 1993. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. Microbiol. Rev. **57:**781–822.
22. **Draker, R., R. L. Roper, M. Petric, and R. Tellier.** 2005. The complete sequence of the bovine torovirus genome. Virus Res. **115:**56–68. (First published, 30 August 2005; doi:10.1016/jvirusres.2005.07.005.)
23. **Elroy-Stein, O., and B. Moss.** 1990. Cytoplasmic expression system based on constitutive synthesis of bacteriophage T7 RNA polymerase in mammalian cells. Proc. Natl. Acad. Sci. USA **87:**6743–6747.
24. **Fuerst, T. R., E. G. Niles, F. W. Studier, and B. Moss.** 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. Proc. Natl. Acad. Sci. USA **83:**8122–8126.
25. **Gonzalez, J. M., P. Gomez-Puertas, D. Cavanagh, A. E. Gorbalenya, and L. Enjuanes.** 2003. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. Arch. Virol. **148:**2207–2235.
26. **Gorbalenya, A. E.** 2001. Big nidovirus genome. When count and order of domains matter. Adv. Exp. Med. Biol. **494:**1–17.
27. **Gorbalenya, A. E., V. M. Blinov, and A. P. Donchenko.** 1986. Poliovirus-encoded proteinase 3C: a possible evolutionary link between cellular serine and cysteine proteinase families. FEBS Lett. **194:**253–257.
28. **Gorbalenya, A. E., A. P. Donchenko, V. M. Blinov, and E. V. Koonin.** 1989. Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. FEBS Lett. **243:**103–114.
29. **Gorbalenya, A. E., E. V. Koonin, V. M. Blinov, and A. P. Donchenko.** 1988. Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. FEBS Lett. **236:**287–290.
30. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. Nucleic Acids Res. **17:**4847–4861.
31. **Gorbalenya, A. E., E. V. Koonin, and M. M. Lai.** 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.
32. **Gosert, R., A. Kanjanahaluethai, D. Egger, K. Bienz, and S. C. Baker.** 2002. RNA replication of mouse hepatitis virus takes place at double-membrane vesicles. J. Virol. **76:**3697–3708.
33. **Hedstrom, L.** 2002. Serine protease mechanism and specificity. Chem. Rev. **102:**4501–4524.
34. **Hegyi, A., and J. Ziebuhr.** 2002. Conservation of substrate specificities among coronavirus main proteases. J. Gen. Virol. **83:**595–599.
35. **Horton, R. M., and L. R. Pease.** 1991. Recombination and mutagenesis of DNA sequences using PCR, p. 217–247. *In* M. J. McPherson (ed.), Directed mutagenesis, a practical approach. IRL Press, Oxford, United Kingdom.
36. **Huang, C., P. Wei, K. Fan, Y. Liu, and L. Lai.** 2004. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. Biochemistry **43:**4568–4574.
37. **Ivanov, K. A., T. Hertzig, M. Rozanov, S. Bayer, V. Thiel, A. E. Gorbalenya, and J. Ziebuhr.** 2004. Major genetic marker of nidoviruses encodes a replicative endoribonuclease. Proc. Natl. Acad. Sci. USA **101:**12694–12699.
38. **Jiang, B., S. S. Monroe, E. V. Koonin, S. E. Stine, and R. I. Glass.** 1993. RNA

sequence of astrovirus: distinctive genomic organization and a putative retrovirus-like ribosomal frameshifting signal that directs the viral replicase synthesis. Proc. Natl. Acad. Sci. USA **90:**10539–10543.
39. **Kanjanahaluethai, A., and S. C. Baker.** 2000. Identification of mouse hepatitis virus papain-like proteinase 2 activity. J. Virol. **74:**7911–7921.
40. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44:**283–292.
41. **Lawson, M. A., and B. L. Semler.** 1991. Poliovirus thiol proteinase 3C can utilize a serine nucleophile within the putative catalytic triad. Proc. Natl. Acad. Sci. USA **88:**9919–9923.
42. **Li, X., M. D. Ryan, and J. W. Lamb.** 2000. Potato leafroll virus protein P1 contains a serine proteinase domain. J. Gen. Virol. **81:**1857–1864.
43. **Lu, Y., X. Lu, and M. R. Denison.** 1995. Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. J. Virol. **69:**3554–3559.
44. **Matthews, D. A., P. S. Dragovich, S. E. Webber, S. A. Fuhrman, A. K. Patick, L. S. Zalman, T. F. Hendrickson, R. A. Love, T. J. Prins, J. T. Marakovits, R. Zhou, J. Tikhe, C. E. Ford, J. W. Meador, R. A. Ferre, E. L. Brown, S. L. Binford, M. A. Brothers, D. M. DeLisle, and S. T. Worland.** 1999. Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. Proc. Natl. Acad. Sci. USA **96:**11000–11007.
45. **Matthews, D. A., W. W. Smith, R. A. Ferre, B. Condon, G. Budahazi, W. Sisson, J. E. Villafranca, C. A. Janson, H. E. McElroy, C. L. Gribskov, et al.** 1994. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. Cell **77:**761–771.
46. **Mijnes, J. D. F., L. M. van der Horst, E. van Anken, M. C. Horzinek, P. J. M. Rottier, and R. J. de Groot.** 1996. Biosynthesis of glycoproteins E and I of feline herpesvirus: gE-gI interaction is required for intracellular transport. J. Virol. **70:**5466–5475.
47. **Mosimann, S. C., M. M. Cherney, S. Sia, S. Plotch, and M. N. James.** 1997. Refined X-ray crystallographic structure of the poliovirus 3C gene product. J. Mol. Biol. **273:**1032–1047.
48. **Nienaber, V. L., K. Breddam, and J. J. Birktoft.** 1993. A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. Biochemistry **32:**11469–11475.
49. **Paetzel, M., and R. E. Dalbey.** 1997. Catalytic hydroxyl/amine dyads within serine proteases. Trends Biochem. Sci. **22:**28–31.
50. **Reed, S. B., C. A. Wesson, L. E. Liou, W. R. Trumble, P. M. Schlievert, G. A. Bohach, and K. W. Bayles.** 2001. Molecular characterization of a novel *Staphylococcus aureus* serine protease operon. Infect. Immun. **69:**1521–1527.
51. **Rice, C. M., and J. H. Strauss.** 1981. Nucleotide sequence of the 26S mRNA of sindbis virus and deduced sequence of the encoded virus structural proteins. Proc. Natl. Acad. Sci. USA **78:**2062–2066.
52. **Ryan, M. D., and M. Flint.** 1997. Virus-encoded proteinases of the picornavirus super-group. J. Gen. Virol. **78:**699–723.
53. **Ryan, M. D., S. Monaghan, and M. Flint.** 1998. Virus-encoded proteinases of the Flaviviridae. J. Gen. Virol. **79:**947–959.
54. **Schiller, J. J., A. Kanjanahaluethai, and S. C. Baker.** 1998. Processing of the coronavirus MHV-JHM polymerase polyprotein: identification of precursors and proteolytic products spanning 400 kilodaltons of ORF1a. Virology **242:** 288–302.
55. **Shanks, M., and G. P. Lomonossoff.** 1990. The primary structure of the 24K protease from red clover mottle virus: implications for the mode of action of comovirus proteases. J. Gen. Virol. **71:**735–738.
56. **Siddell, S. G., J. Ziebuhr, and E. J. Snijder.** 2005. Coronaviruses, toroviruses and arteriviruses, p. 823–856. *In* B. W. Mahy and V. ter Meulen (ed.), Topley and Wilson's microbiology and microbial infections. Virology. Hodder Arnold, London, United Kingdom.
57. **Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. Spaan, and A. E. Gorbalenya.** 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. **331:**991–1004.
58. **Snijder, E. J., J. A. den Boon, P. J. Bredenbeek, M. C. Horzinek, R. Rijnbrand, and W. J. Spaan.** 1990. The carboxyl-terminal part of the putative Berne virus polymerase is expressed by ribosomal frameshifting and contains sequence motifs which indicate that toro- and coronaviruses are evolutionarily related. Nucleic Acids Res. **18:**4535–4542.
59. **Snijder, E. J., J. A. den Boon, M. C. Horzinek, and W. J. Spaan.** 1991. Characterization of defective interfering RNAs of Berne virus. J. Gen. Virol. **72:**1635–1643.
60. **Snijder, E. J., and M. C. Horzinek.** 1993. Toroviruses: replication, evolution and comparison with other members of the coronavirus-like superfamily. J. Gen. Virol. **74:**2305–2316.
61. **Snijder, E. J., M. C. Horzinek, and W. J. M. Spaan.** 1990. A 3′-coterminal nested set of independently transcribed mRNAs is generated during Berne virus replication. J. Virol. **64:**331–338.
62. **Snijder, E. J., S. G. Siddell, and A. E. Gorbalenya.** 2005. The order Nidovirales, p. 390–404. *In* B. W. Mahy and V. ter Meulen (ed.), Topley and Wilson's microbiology and microbial infections. Virology. Hodder Arnold, London, United Kingdom.
63. **Snijder, E. J., A. L. M. Wassenaar, and W. J. M. Spaan.** 1994. Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. J. Virol. **68:**5755–5764.
64. **Snijder, E. J., A. L. Wassenaar, L. C. van Dinten, W. J. Spaan, and A. E. Gorbalenya.** 1996. The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. J. Biol. Chem. **271:**4864–4871.
65. **Sprang, S., T. Standing, R. J. Fletterick, R. M. Stroud, J. Finer-Moore, N. H. Xuong, R. Hamlin, W. J. Rutter, and C. S. Craik.** 1987. The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis. Science **237:**905–909.
66. **ten Dam, E., M. Flint, and M. D. Ryan.** 1999. Virus-encoded proteinases of the Togaviridae. J. Gen. Virol. **80:**1879–1888.
67. **van der Meer, Y., H. van Tol, J. K. Locker, and E. J. Snijder.** 1998. ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex. J. Virol. **72:**6689–6698.
68. **Van Vliet, A. L., S. L. Smits, P. J. Rottier, and R. J. De Groot.** 2002. Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. EMBO J. **21:**6571–6580.
69. **Vennema, H., R. Rijnbrand, L. Heijnen, M. C. Horzinek, and W. J. Spaan.** 1991. Enhancement of the vaccinia virus/phage T7 RNA polymerase expression system using encephalomyocarditis virus 5′-untranslated region sequences. Gene **108:**201–209.
70. **Wang, R. F., and S. R. Kushner.** 1991. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in Escherichia coli. Gene **100:**195–199.
71. **Wassenaar, A. L. M., W. J. M. Spaan, A. E. Gorbalenya, and E. J. Snijder.** 1997. Alternative proteolytic processing of the arterivirus replicase ORF1a polyprotein: evidence that NSP2 acts as a cofactor for the NSP4 serine protease. J. Virol. **71:**9313–9322.
72. **Wellink, J., and A. van Kammen.** 1988. Proteases involved in the processing of viral polyproteins. Brief review. Arch. Virol. **98:**1–26.
73. **Yang, H., M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, and Z. Rao.** 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc. Natl. Acad. Sci. USA **100:** 13190–13195.
74. **Ziebuhr, J., S. Bayer, J. A. Cowley, and A. E. Gorbalenya.** 2003. The 3C-like proteinase of an invertebrate nidovirus links coronavirus and potyvirus homologs. J. Virol. **77:**1415–1426.
75. **Ziebuhr, J., G. Heusipp, and S. G. Siddell.** 1997. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. J. Virol. **71:**3992–3997.
76. **Ziebuhr, J., E. J. Snijder, and A. E. Gorbalenya.** 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J. Gen. Virol. **81:**853–879.