JOURNAL OF VIROLOGY, Jan. 2003, p. 1415–1426 Vol. 77, No. 2
0022-538X/03/$08.00+0 DOI: 10.1128/JVI.77.2.1415–1426.2003
Copyright © 2003, American Society for Microbiology. All Rights Reserved.

The 3C-Like Proteinase of an Invertebrate Nidovirus Links Coronavirus and Potyvirus Homologs

John Ziebuhr,[1]* Sonja Bayer,[1] Jeff A. Cowley,[2] and Alexander E. Gorbalenya[3]

Institute of Virology and Immunology, University of Würzburg, Würzburg, Germany[1]; Cooperative Research Center for Aquaculture, CSIRO Livestock Industries, Long Pocket Laboratories, Indooroopilly, Australia[2]; and Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands[3]

Received 27 June 2002/Accepted 15 October 2002

Gill-associated virus (GAV), a positive-stranded RNA virus of prawns, is the prototype of newly recognized taxa (genus *Okavirus*, family *Roniviridae*) within the order *Nidovirales*. In this study, a putative GAV cysteine proteinase (3C-like proteinase [3CL^{pro}]), which is predicted to be the key enzyme involved in processing of the GAV replicase polyprotein precursors, pp1a and pp1ab, was characterized. Comparative sequence analysis indicated that, like its coronavirus homologs, 3CL^{pro} has a three-domain organization and is flanked by hydrophobic domains. The putative 3CL^{pro} domain including flanking regions (pp1a residues 2793 to 3143) was fused to the *Escherichia coli* maltose-binding protein (MBP) and, when expressed in *E. coli*, was found to possess N-terminal autoprocessing activity that was not dependent on the presence of the 3CL^{pro} C-terminal domain. N-terminal sequence analysis of the processed protein revealed that cleavage occurred at the location ^{2827}LVTHE↓VRTGN^{2836}. The *trans*-processing activity of the purified recombinant 3CL^{pro} (pp1a residues 2832 to 3126) was used to identify another cleavage site, ^{6441}KVNHE↓LYHVA^{6450}, in the C-terminal pp1ab region. Taken together, the data tentatively identify VxHE↓(L,V) as the substrate consensus sequence for the GAV 3CL^{pro}. The study revealed that the GAV and potyvirus 3CL^{pro}s possess similar substrate specificities which correlate with structural similarities in their respective substrate-binding sites, identified in sequence comparisons. Analysis of the proteolytic activities of MBP-3CL^{pro} fusion proteins carrying replacements of putative active-site residues provided evidence that, in contrast to most other 3C/3CL^{pro}s but in common with coronavirus 3CL^{pro}s, the GAV 3CL^{pro} employs a Cys^{2968}-His^{2879} catalytic dyad. The properties of the GAV 3CL^{pro} define a novel RNA virus proteinase variant that bridges the gap between the distantly related chymotrypsin-like cysteine proteinases of coronaviruses and potyviruses.

Gill-associated virus (GAV) is an enveloped, rod-shaped, positive-stranded RNA virus that infects *Penaeus monodon* (black tiger) prawns in Australia (8, 41). While subclinical GAV infections, originally reported as lymphoid organ virus, are highly prevalent in both wild and farmed *P. monodon* (41), acute infections causing mortality have also been reported (40). GAV is closely related morphologically and genetically to yellow head virus (7, 10, 41), which is associated with yellow head disease and has caused considerable production losses in *P. monodon* farmed throughout southeast Asia (5).

GAV and yellow head virus have recently been placed in a new genus, *Okavirus*, within a new family, *Roniviridae* (8, 11), that, together with the *Coronaviridae* and *Arteriviridae*, forms the order *Nidovirales* (6, 12). The phylogenetic relationship between GAV and nidoviruses became evident from comparative sequence analyses of the 20-kb 5′-terminal region of the GAV genome (8), which revealed striking similarities in the organization and expression of the viral replicase genes. In common with nidoviruses, the 5′-terminal replicase gene of GAV encodes two large open reading frames, ORF1a and ORF1b, comprising 12,248 and 7,941 nucleotides, respectively. In vitro data also demonstrated that the downstream ORF1b, which overlaps ORF1a by 99 nucleotides, is expressed by ribosomal frameshifting, as in all nidoviruses. Most probably, slippage into the −1 frame occurs at the sequence ^{12215}AAAUUUU^{12221} and involves an RNA pseudoknot located immediately downstream of this slippery sequence (8). Accordingly, ORFs 1a and 1b are translated as two polyproteins, pp1a (460 kDa) and its C-terminally extended form, pp1ab (758 kDa), which are expected to mediate the functions required for genome replication and transcription of a 3′-coterminal nested set of subgenomic mRNAs encoding the viral structural proteins (9).

Comparative sequence analysis revealed several putative functional domains in the GAV polyproteins, including helicase and polymerase motifs, ordered similarly to the cognate domains in the viral polyproteins of other nidoviruses (8). This observation, combined with the fact that the GAV polymerase domain contains the SDD motif unique to nidovirus polymerases, strongly suggested that GAV (infecting invertebrates) and nidoviruses (infecting vertebrates) have a common ancestor (14). However, the presence of a number of regions with low sequence similarity in ORF1b and, in particular, the extremely poor pp1a conservation suggested that GAV has diverged significantly from the vertebrate nidoviruses (corona- and arteriviruses). Indeed, the only region in pp1a with significant sequence similarity proved to be a putative chymotrypsin-like (3C-like) proteinase domain (3CL^{pro}), flanked by hydrophobic (probably membrane-spanning) domains.

* Corresponding author. Mailing address: Institute of Virology and Immunology, University of Würzburg, Versbacher Str. 7, 97078 Würzburg, Germany. Phone: 49 931 20149966. Fax: 49 931 20149553. E-mail: ziebuhr@vim.uni-wuerzburg.de.

In vertebrate nidoviruses, the 3CL^{pro} cleaves the viral polyproteins at multiple conserved sites and is responsible for posttranslational release of the key replicative proteins. It has therefore also been referred to as the main proteinase (M^{pro}) to distinguish it from accessory nidovirus proteinases, which cleave at only a few sites in the N-terminal pp1a/pp1ab regions (51). Although no 3CL^{pro} cleavage sites could be readily predicted in the pp1a/pp1ab polyproteins of this invertebrate nidovirus, it seems likely that this GAV proteinase may have a similar critical role in viral replication, as has been demonstrated conclusively for its vertebrate nidovirus homologs (8, 51). Based on sequence comparisons, it has been proposed that the GAV 3CL^{pro} is distantly related to the main proteinases of arteri- and coronaviruses as well as the NIa proteinases of plant potyviruses, which all have an (E,Q) ↓ (G,S,A) substrate specificity (8). 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. If various residues are found at a given position, these are listed in parentheses.

In this report, we provide direct evidence for the predicted proteolytic function of GAV 3CL^{pro}. Predictions of putative active-site residues identified by sequence comparisons were substantiated by site-directed mutagenesis, and information on the GAV 3CL^{pro} substrate specificity was obtained. The theoretical and experimental data presented in this study define a new member of the constantly growing group of viral 3C-like proteinases, which may combine the Cys-His catalytic dyad of the main proteinase of coronaviruses with a potyvirus-like substrate-binding pocket.

MATERIALS AND METHODS

Expression of GAV pp1a/pp1ab sequences. The cDNA clones pGCLP7.6 and pGAV1b-3′-9 (J. A. Cowley, unpublished data) were used as templates for PCR amplification of GAV sequences. The sequences of pGCLP7.6 and pGAV1b-3′-9 deviated at several positions from the sequence reported previously (GenBank accession number AF227196), which was derived from multiple random reverse transcription-PCR products generated from total RNA isolated from pooled lymphoid organs of GAV-infected *P. monodon* (8). The pp1a/pp1ab sequence used in this study contained nucleotide changes that led to four amino acid substitutions: Cys^{3073}Arg, Ala^{3110}Thr, Ser^{3127}Leu, and His^{6631}Tyr.

DNA sequences encoding different GAV pp1a/pp1ab regions were amplified by PCR with the primers listed in Table 1. The PCR products were treated with T4 DNA polymerase, phosphorylated with T4 polynucleotide kinase, digested with *Eco*RI, and inserted into the *Xmn*I and *Eco*RI sites of pMal-c2 (New England Biolabs, Frankfurt, Germany). The resulting plasmids, which are shown in Table 1, allowed the expression of GAV pp1a/pp1ab sequences fused to the maltose-binding protein (MBP) of *Escherichia coli* (Fig. 1). Site-directed mutagenesis was done by a recombination-PCR method (19, 47). *E. coli* TB1 cells transformed with the appropriate pMal-c2 derivatives (Table 1) were grown at 37°C in Luria-Bertani (LB) medium containing 100 μg of ampicillin per ml until they reached a culture density (A_{595}) of 0.6. Expression of the recombinant proteins was induced by addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 h at 24°C. For analysis of recombinant protein expression, aliquots of the cell cultures were suspended in 2× Laemmli sample buffer and heated at 94°C for 3 min, and the lysates were analyzed by electrophoresis in sodium dodecyl sulfate (SDS)-polyacrylamide gels and Western immunoblotting with standard protocols.

Proteins 2832-3126, 2832-3126_C^{2968}A, MBP-2948-3143, and MBP-6338-6673 (Table 1) were purified by amylose affinity chromatography as described previously (17, 50). Two of the proteins, 2832-3126 and 2832-3126_C^{2968}A, were purified further. To this end, the affinity-purified fusion proteins were subjected to cleavage by factor Xa (Amersham Biosciences, Freiburg, Germany) and loaded onto phenyl-Sepharose HP columns (Amersham Biosciences) that had been preequilibrated with buffer containing 20 mM Tris-HCl (pH 7.5), 600 mM NaCl, 1 mM dithiothreitol, and 0.1 mM EDTA. The GAV-specific proteins were eluted with 20 mM Tris-HCl (pH 7.5)–1 mM dithiothreitol–0.1 mM EDTA, concentrated (Centricon-3; Millipore), and loaded onto a Superdex 75 column (Amersham Biosciences), which was run under isocratic conditions with 20 mM Tris-HCl (pH 7.5)–150 mM NaCl–1 mM dithiothreitol–0.1 mM EDTA. The purified proteins were concentrated to 5 mg/ml (Centricon-3) and stored at −80°C.

N-terminal protein sequence analysis. Following SDS-polyacrylamide gel electrophoresis (PAGE), the proteins were transferred to polyvinylidene difluoride membranes (162-0180; Bio-Rad Laboratories, Munich, Germany) and subsequently stained with Coomassie brilliant blue. The membrane regions containing the proteins of interest were isolated as described previously (49), and the proteins were subjected to six cycles of Edman degradation by use of a pulsed-liquid protein sequencer (ABI 467A; Applied Biosystems, Weiterstadt, Germany).

Preparation of antiserum α-MBP-2948-3143. The MBP-2948-3143 fusion protein was purified by amylose affinity chromatography from TB1[pMal-GAV-2948-3143] cells as described above. The protein was cleaved with factor Xa (Amersham Biosciences) and used to immunize rabbits as described previously (49). The antiserum was designated α-MBP-2948-3143.

***trans*-cleavage assay.** Typical 20-μl reaction mixes contained recombinant GAV 3CL^{pro} (2832-3126 or 2832-3126_C^{2968}A) and the substrate protein, MBP-6638-6673 (each at 1.6 μM), in a buffer containing 20 mM Tris-HCl (pH 7.5), 200 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol. Following incubation at 22°C for 16 h, the reaction products were separated on SDS–15% polyacrylamide gels that were stained with Coomassie brilliant blue R-250.

Computer-aided comparative sequence analyses. Amino acid sequences were derived from the Genpeptides database. 3CL^{pro} sequence alignments were produced with the Clustal X program (42) and the Blossum series of scoring interresidue tables (18). The virus interfamily alignments were generated in the profile mode. The alignments obtained were used in the PhD program (34, 35) to predict secondary structures and also to build profiles with the Profileweight program (43). These profiles were compared in pairs with the Proplot program (43). Two profiles, where one profile may be a sequence, were compared by sliding a window of the selected length along each possible register for a given dot plot. Several window lengths were tested. Matches between two profiles that were within the top 0.05% or between the top 0.1% and 0.05% were marked by two different types of dots.

RESULTS

Comparative sequence analysis of GAV 3CL^{pro} with chymotrypsin-like cysteine proteinases of positive-stranded RNA viruses. We first sought to refine the previously published sequence comparison of the putative GAV proteinase (8) to provide a theoretical basis of sufficient reliability for subsequent experimental studies. Specifically, we tried to gain initial insight into the substrate specificity and possible active-site residues.

Comparison of the entire replicase gene revealed that, among all viruses sequenced to date, the *Coronaviridae* represent the most closely related family to GAV (unpublished data). In the case of the 3CL^{pro}, however, the most significant matches were found in homologs from the *Potyviridae* family (8) (data not shown). Comparison of the GAV 3CL^{pro} with both corona- and potyvirus 3CL^{pro}s revealed conservation of two regions: (i) the segment containing the catalytic His residue, which is most similar between the GAV and coronavirus 3CL^{pro}s, and (ii) the segment containing the catalytic Cys residue, which is most similar between the GAV and potyvirus 3CL^{pro}s (Fig. 2). No conservation was evident in the segment between the catalytic His and Cys residues, which contains the catalytic Asp residue of potyvirus (and many other) 3C-like proteinases.

To dissect this region further, the GAV 3CL^{pro} was compared with a combined and structurally corrected (2) alignment of corona- and potyvirus 3CL^{pro}s (17) with the global

TABLE 1. Oligonucleotides used for the amplification or mutagenesis of GAV sequences

Oligonucleotides used for cloning or mutagenesis (5′ → 3′)[a]	Plasmid[b]	pp1a/pp1ab amino acids[c]	Amino acid substitution
AACGCATATGCCCAGGCAATCGATTC AAAGAATTCTTAGCAACGGAATCTGGTGAGAGGA	pMal-GAV-2793–3143	2793–3143	
AACGCATATGCCCAGGCAATCGATTC AAAGAATTCTTACTGATAGTTGGTGGGGAGCTTTGGTGTTG	pMal-GAV-2793–3059	2793–3059	
AACGCATATGCCCAGGCAATCGATTC AAAGAATTCTTAGACGGGCCAGACCTTTGGTGGATCGAC	pMal-GAV-2793–3028	2793–3028	
ATCAGGCTCGGCTCAATGTCCACT AAAGAATTCTTAGCAACGGAATCTGGTGAGAGGA	pMal-GAV-2948–3143	2948–3143	
GTTCGTACAGGTAACGCCACCACGGTC AAAGAATTCTTAGTTGCTGAGTGGAGAAAGGTCAGCAATA	pMal-GAV-2832–3126	2832–3126	
AGGATGGTGATGCTGGTTCCATCATCTTCGACCACC ATGATGGAACCAGCATCACCATCCTTGGTGGAGATG	pMal-GAV-2832–3126_C^{2968}A	2832–3126	Cys^{2968} → Ala
GTTCGTACAGGTAACGCCACCACGGTC AAAGAATTCTTACTGATAGTTGGTGGGGAGCTTTGGTGTTG	pMal-GAV-2832–3059	2832–3059	
GTTCGTACAGGTAACGCCACCACGGTC AAAGAATTCTTAGACGGGCCAGACCTTTGGTGGATCGAC	pMal-GAV-2832–3028	2832–3028	
AACACTAACAATTGGGAACAAATAC AAAGAATTCTTAAAATTTGATGAATCTGGGAGAT	pMal-GAV-6338–6673	6338–6673[d]	
CACTTCCCTCGACGCATCTTCGACACCTGCACTGACA TGTCGAAGATGCGTCGAGGGAAGTGGAGGGATTTGC	pMal-GAV-2793–3143_H^{2879}R	2793–3143	His^{2879} → Arg
CACTTCCCTCGACTCATCTTCGACACCTGCACTGACA TGTCGAAGATGAGTCGAGGGAAGTGGAGGGATTTGC	pMal-GAV-2793–3143_H^{2879}L	2793–3143	His^{2879} → Leu
AGGATGGTGATGCTGGTTCCATCATCTTCGACCACC ATGATGGAACCAGCATCACCATCCTTGGTGGAGATG	pMal-GAV-2793–3143_C^{2968}A	2793–3143	Cys^{2968} → Ala
AGGATGGTGATTCTGGTTCCATCATCTTCGACCACC ATGATGGAACCAGAATCACCATCCTTGGTGGAGATG	pMal-GAV-2793–3143_C^{2968}S	2793–3143	Cys^{2968} → Ser
TGAGTGAAGAATATGCTGCTACACCATTCATCAAAGTTG GAATGGTGTAGCAGCATATTCTTCACTCAAAAGCTCGATG	pMal-GAV-2793–3143_D^{2912}A	2793–3143	Asp^{2912} → Ala
TGAGTGAAGAATATGAGGCTACACCATTCATCAAAGTTG GAATGGTGTAGCCTCATATTCTTCACTCAAAAGCTCGATG	pMal-GAV-2793–3143_D^{2912}E	2793–3143	Asp^{2912} → Glu
TGAGTGAAGAATATCAAGCTACACCATTCATCAAAGTTG GAATGGTGTAGCTTGATATTCTTCACTCAAAAGCTCGATG	pMal-GAV-2793–3143_D^{2912}Q	2793–3143	Asp^{2912} → Gln
CGTCGGTGCCGCTATCGTCGGTATCTCCTGCATCCCT TACCGACGATAGCGGCACCGACGACATTACCGAGGTG	pMal-GAV-2793–3143_H^{2983}A	2793–3143	His^{2983} → Ala
CGTCGGTGCCTTTATCGTCGGTATCTCCTGCATCCCT TACCGACGATAAAGGCACCGACGACATTACCGAGGTG	pMal-GAV-2793–3143_H^{2983}F	2793–3143	His^{2983} → Phe
TCGTCGGTATCGCCTGCATCCCTCCAGTCAACGGTG GGAGGGATGCAGGCGATACCGACGATATGGGCACCGA	pMal-GAV-2793–3143_S^{2988}A	2793–3143	Ser^{2988} → Ala
TCGTCGGTATCCACTGCATCCCTCCAGTCAACGGTG GGAGGGATGCAGTGGATACCGACGATATGGGCACCGA	pMal-GAV-2793–3143_S^{2988}H	2793–3143	Ser^{2988} → His

[a] Underlined residues in the oligonucleotide sequence indicate mutant codons.
[b] GAV sequences were inserted into the unique *Xmn*I and *Eco*RI restriction sites of pMal-c2 plasmid DNA (New England Biolabs).
[c] The GAV pp1a/pp1ab residues given were expressed as fusions with *E. coli* MBP. The amino acid residues are numbered according to the sequence published by Cowley et al. (8) (GenBank accession no. AF227196).
[d] Amino acid numbering of the ORF1b-encoded portion of pp1ab is based on the prediction that −1 ribosomal frameshifting occurs at the sequence ^{12215}AAAUUUU^{12221} (8).

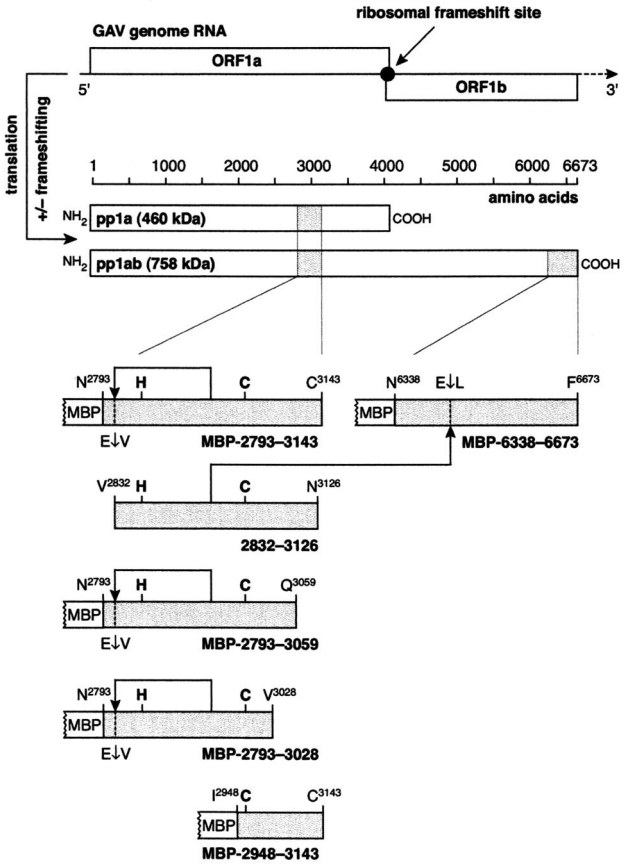


FIG. 1. Expression of GAV replicase gene. The ≈20,000-nucleotides gene comprises ORFs 1a and 1b, which occupy the 5′-terminal region of the GAV genome and encode two replicase polyproteins, pp1a and pp1ab. Expression of pp1ab requires a −1 frameshift during translation, which is predicted to be mediated by a slippery heptanucleotide sequence and an RNA pseudoknot structure (8). The primary GAV pp1a/pp1ab-derived protein constructs used in this study are shown schematically. The N- and C-terminal residues of the GAV-specific amino acid sequences are given in the one-letter code. The numbering of pp1a/pp1ab amino acids is based on predictions on the GAV frameshift site, AAAUUUU (nucleotides 12215 to 12221 of the GAV genome) (8) (GenBank accession number AF227196). Fusions of GAV pp1a/pp1ab amino acids with *E. coli* MBP are indicated. Also, the positions of putative active-site Cys and His residues and the GAV 3CL^{pro} cleavage sites characterized in this study are given (C, H, and E ↓ V, E ↓ L, respectively).

alignment tool Clustal X (42). In this study, Ala^{2913} was identified as a plausible candidate to occupy the main chain position equivalent to that of the catalytic Asp residue of potyvirus 3CL^{pro}s, suggesting that GAV 3CL^{pro}, like coronavirus 3CL^{pro}s (2, 17), may lack a catalytic acidic residue in this region (Fig. 2).

The computer-aided analysis of putative substrate-binding residues of 3CL^{pro} produced a low-resolution model. GAV His^{2983}, the previously proposed counterpart to the key S1 subsite His residues of other 3C/3CL^{pro}s (8), was either at the edge or even outside of a stretch of matching residues in the GAV-versus-potyvirus and GAV-versus-coronavirus dot plots, respectively (Fig. 2). The low similarity in this region is due to the unusually short size of this segment in GAV 3CL^{pro} and unique amino acid replacements in the immediate vicinity of GAV His^{2983} and the corresponding His residues in coronavirus 3CL^{pro}s (15, 17) (Fig. 3). Accordingly, when the GAV 3CL^{pro} was compared separately with each of the two proteinase groups with Clustal X, another closely located residue of GAV, Ser^{2988}, was aligned with the substrate-binding His (not shown). Five residues upstream of the catalytic Cys, a Thr/Ser residue which, in many 3C/3CL^{pro}s, together with His, makes contact with the substrate's P1 Gln/Glu side chain (3, 16, 28–30), was found to be conserved in the GAV sequence (GAV Thr^{2963}), suggesting that His (rather than Ser) is the most probable candidate to assume the key position in the S1 subsite.

Apart from Thr^{2963}, three other residues (His^{2959}, Ile^{2961}, and Gly^{2981}) located nearby were revealed to be conserved among GAV and potyvirus but not coronavirus 3CL^{pro}s (Fig. 3). Based on the available 3C/CL^{pro} structure information (2, 4, 29, 30), these residues are likely to be part of the extended substrate-binding pocket. The observed sequence conservation suggested that the well-defined substrate specificity of potyvirus 3CL^{pro}s (21) may, at least in part, be shared by the GAV enzyme.

Nidovirus 3CL^{pro}s comprise two catalytic β-barrels and an extra C-terminal domain. In the viral polyprotein, they are flanked by well-conserved cleavage sites that are used to release the proteinase from adjacent transmembrane domains (15, 51). A similar domain organization was unraveled in GAV, although the sequence conservation was rather low, especially outside the catalytic domains (Fig. 3). In striking contrast to other nidoviruses, we were unable to identify conservation in the immediate flanking regions of 3CL^{pro} or, at least, dipeptides conforming to canonical 3CL^{pro} cleavage sites [(Glu,Gln) ↓ (Ser,Ala,Gly)], indicating that the GAV 3CL^{pro} may have a deviant specificity and release itself from the precursor in a unique fashion.

Proteolytic activity of GAV 3CL^{pro} domain. To address the predicted proteolytic activity of the GAV 3CL^{pro}, pp1a/pp1ab residues 2793 to 3143 (containing the presumed 3CL^{pro} and a short N-terminal flanking region) were expressed as part of an MBP fusion protein (MBP-2793-3143) in *E. coli*. Based on studies on the related human coronavirus 3CL^{pro} (49), the N-terminal region was expected to contain a 3CL^{pro} site that could be autoprocessed in *E. coli*. As Fig. 4A (lanes 2 and 3) shows, induction of expression resulted in the synthesis of two proteins of ≈47 and ≈38 kDa that were not detectable in the noninduced control, suggesting proteolytic cleavage of the primary translation product, for which a molecular mass of 82 kDa was calculated. The fact that the control protein, MBP-2793-3143_H^{2879}R, in which Arg replaced the putative active-site His^{2879} residue, gave rise to the full-length protein (Fig. 4A, lanes 4 and 5) provided conclusive evidence that, as predicted, GAV pp1a/pp1ab residues 2793 to 3143 contain a functional proteinase domain.

To identify the N- and C-terminal portions of the cleaved protein, the lysate obtained from IPTG-induced *E. coli* TB1[pMal-GAV-2793-3143] cells was analyzed by Western blotting with specific antiserum. The data presented in Fig. 4B revealed that the 47-kDa protein was the N-terminal (that is, MBP-containing) cleavage product and that the 38-kDa pro-

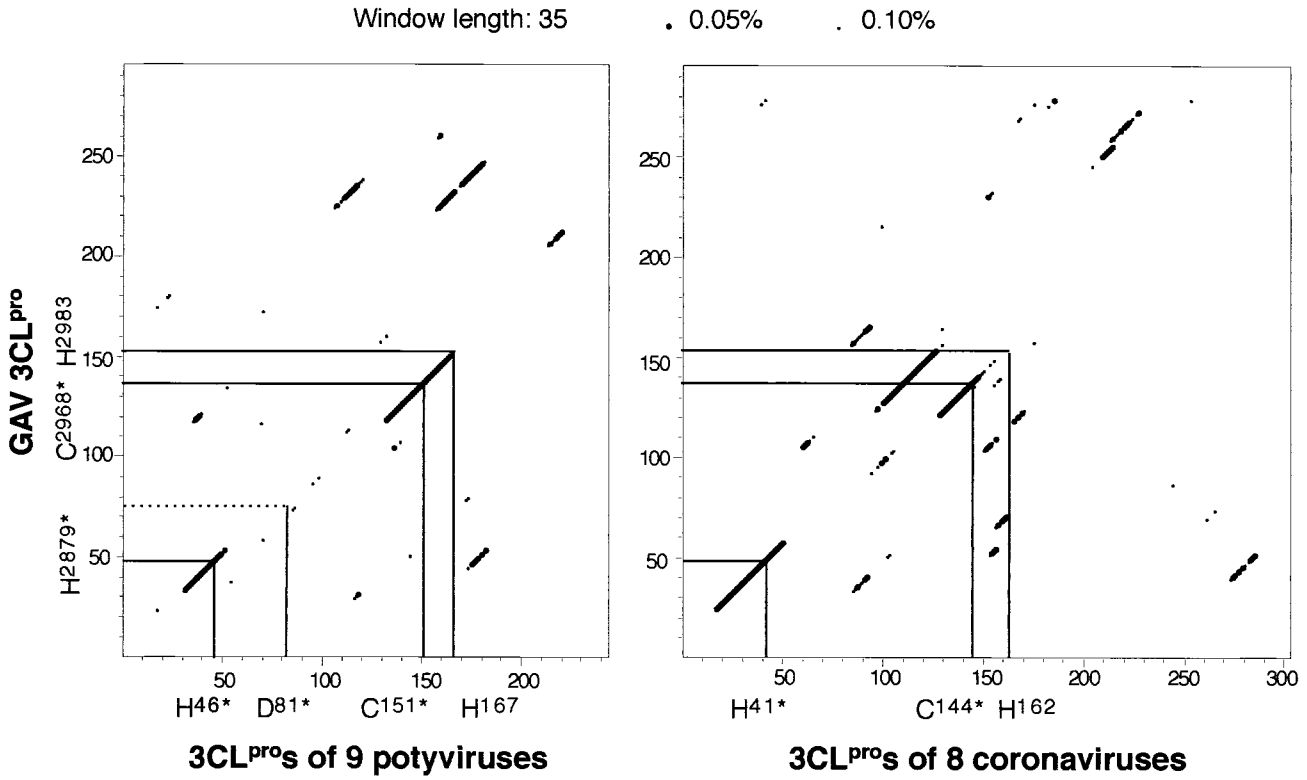


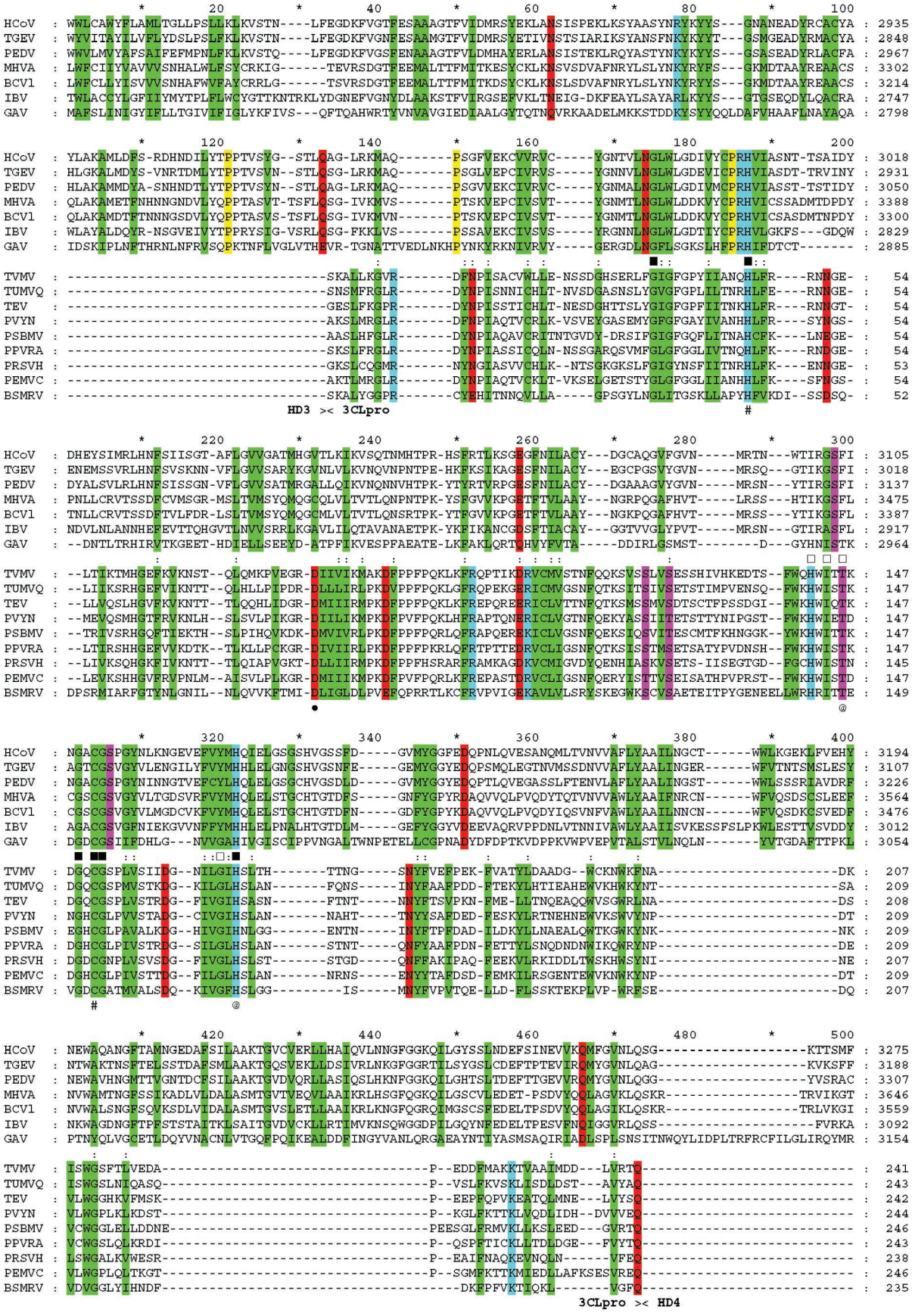
FIG. 2. Profile-versus-profile dot plot cross-comparisons of GAV 3CL^{pro} with coronavirus and potyvirus 3CL^{pro}s. Alignments of coronavirus and potyvirus 3C-like proteinases were converted into profiles and compared in a dot plot fashion, as described in Materials and Methods. Shown are the dot plots generated with a window of 35 amino acid residues. The projected positions of the catalytic residues (H^{46}/H^{41} versus H^{2879}, D^{81}, C^{151}/C^{144} versus C^{2968}), as well as the substrate-binding H^{167}/H^{162} residues versus H^{2983}, are shown at each axis. Putative catalytic residues are designated by asterisks. Those dots, which lay at any of the four possible crosses of projections of two functionally equivalent residues (e.g., H^{46} and H^{2879}) or close to a nonvisible diagonal passing these crosses, belong or may belong to the true matches between two profiles. The rest of the dots are background hits (false-positives).

tein was the C-terminal cleavage product containing the GAV pp1a/pp1ab sequence 2948 to 3143.

***trans*-cleavage activity of recombinant GAV 3CL^{pro}.** From the data presented above, it could not be concluded whether the N-terminal 3CL^{pro} cleavage had occurred in *cis* or was mediated by *trans*-acting precursors. Although the high cleavage efficiency indicated by the virtual absence of detectable precursors strongly suggested a cotranslational monomolecular reaction, we expected that the recombinant 3CL^{pro} might also have *trans*-cleavage activity required by the native proteinase to process the full spectrum of cleavage sites assumed to exist in the 460-kDa and 758-kDa GAV replicase polyproteins. The demonstration of such *trans*-cleavage activity would also formally exclude the involvement of *E. coli* proteinases in the processing described in Fig. 4.

trans-cleavage activity was examined with purified, recombinant 3CL^{pro} (for details, see Materials and Methods). Because of the uncertainty regarding the C-terminal border of 3CL^{pro} (see below), we initially tested bacterially expressed proteins with C termini of different lengths (2832 to 3143 and 2832 to 3126). Both proteins had proteolytic activity. We decided to use 2832-3126 in subsequent *trans*-cleavage experiments because of its superior stability. As a control, a protein with the same sequence but containing a substitution of the putative nucleophilic active-site Cys^{2968} residue (2832-3126_C^{2968}A) was produced (Fig. 5). The purified proteins were incubated with bacterially expressed MBP-6338-6673 containing the C-terminal GAV pp1ab sequence corresponding to the coronavirus pp1ab region with the most C-terminal 3CL^{pro} cleavage site (20, 25, 51). The data (Fig. 5) revealed that the wild-type proteinase but not the active-site mutant was active in *trans*, proving that GAV 3CL^{pro} is indeed a proteinase.

Substrate specificity of GAV 3CL^{pro}. To obtain information on 3CL^{pro}'s substrate specificity, the structure of two cleavage sites was determined with mono- and bimolecular cleavage reactions. First, we determined the N-terminal sequence of the 38-kDa C-terminal processing product of the MBP-2793-3143 fusion protein precursor (Fig. 6). Proteins in the *E. coli* lysate analyzed in Fig. 4A (lane 3) were separated by SDS-PAGE, transferred electrophoretically to a polyvinylidene difluoride membrane, and stained with Coomassie brilliant blue, and the 38-kDa protein was isolated and subjected to six cycles of Edman degradation. The data shown in Fig. 6 clearly indicated that cleavage occurred at the sequence ^{2827}LVTHE↓VRTGN^{2836}, which identifies Val^{2832} as the N terminus of 3CL^{pro}. The observed molecular mass of the 3CL^{pro}-containing cleavage product (38 kDa) slightly surpassed that calcu-


HCoV
TGEV
PEDV
MHVA
BCV1
IBV
GAV
TVMV
TUMVQ
TEV
PVYN
PSBMV
PPVRA
PRSVH
PEMVC
BSMRV
HD3 >< 3CLpro
3CLpro >< HD4

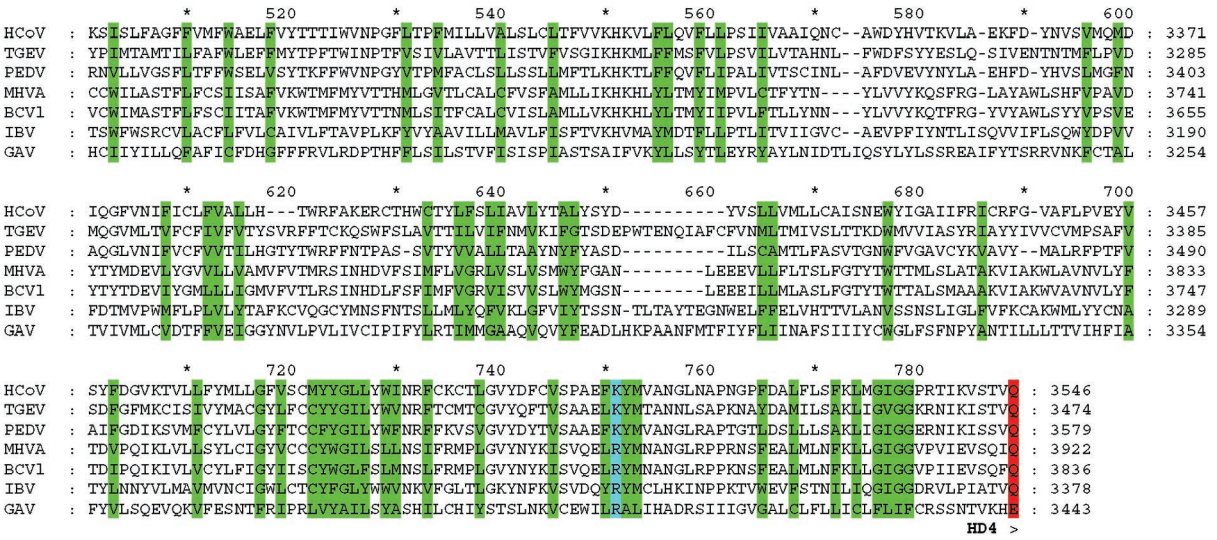


FIG. 3. Multiple sequence alignment of GAV, coronavirus, and potyvirus 3CL^{pro} domains. The Clustal X-based alignment of corona- and potyvirus 3CL^{pro}s produced previously (17) was modified slightly to accommodate the results of the tertiary-structure analysis of a porcine coronavirus 3CL^{pro} (2) and used to align the GAV 3CL^{pro} sequence. For GAV and coronaviruses, this alignment was further expanded by including upstream and downstream sequences with Clustal X. Shown are the regions enriched in hydrophobic amino acid residues and flanking the 3CL^{pro} from both the N terminus (C-terminal part of hydrophobic domain [HD3]) and the C terminus (entire HD4). These hydrophobic domains are conserved in all nidoviruses (14). For GAV and coronaviruses, the pp1a/1ab amino acid positions are given on the right; for potyviruses, the numbers refer to the amino acid positions in the 3CL^{pro}. The column conservation in the two groups of coronavirus/GAV versus potyvirus sequences was highlighted separately with different colors for the following groups of amino acids: green for G, A, L, I, V, M, F, Y, and W; blue for H, K, and R; red for N, Q, E, and D; yellow for P; and violet for S and T. Columns with conserved or identical residues in all sequences are indicated by colons and solid squares, respectively, in the line separating the coronavirus/GAV and potyvirus groups. Empty squares highlight columns with identical residues in the GAV and potyvirus sequences. #, conserved catalytic Cys and His residues; @, P1-binding His residue conserved in all sequences and Thr residue conserved among GAV and potyviruses; solid circle, catalytic Asp residue of potyviruses. ><, positions of cleavage sites separating 3CL^{pro} from flanking domains in corona- and potyviruses. Abbreviations of virus names and DDBJ/EMBL/GenBank accession numbers for the sequences are as follows: HCoV, human coronavirus (strain 229E) (X69721); TGEV, transmissible gastroenteritis virus (strain Purdue 115) (Z34093); PEDV, porcine epidemic diarrhea virus (strain CV777) (NC_003436); MHVA, murine hepatitis virus (strain A59) (NC_001846); BCoVl, bovine coronavirus (isolate LUN) (AF391542); IBV, avian infectious bronchitis virus (strain Beaudette) (M95169); TVMV, tobacco vein mottling virus (P09814); TUMVQ, turnip mosaic virus (strain Quebec) (Q02597); TEV, tobacco etch virus (P04517); PVY, potato virus Y (strain N) (P18247); PSBMV, pea seed-borne mosaic virus (strain DPD1) (P29152); PPVRA, plum pox virus (strain Rankovic) (P17767); PRSVH, papaya ringspot virus (strain P/mutant HA) (Q01901); PEMVC, pepper mottle virus (California isolate) (Q01500); BSMRV, Brome streak mosaic rymovirus (strain 11-Cal) (Q65730).

lated for this peptide sequence (34.8 kDa), making a second, C-terminal cleavage of MBP-2793-3143 unlikely.

Second, we conducted a similar N-terminal sequence analysis (data not shown) of the ≈27-kDa C-terminal cleavage product from the *trans*-cleavage reaction documented in Fig. 5. This analysis unambiguously identified the scissile bond as ^{6441}KVNHE↓LYHVA^{6450}. As no other processing product was detected, it is reasonable to assume that the C-terminal processing product of GAV pp1ab is a 27-kDa protein encompassing amino acids 6446 to 6673. The data provided additional information on the GAV 3CL^{pro} substrate specificity, which allows us to preliminarily propose VxHE↓(L,V) as the consensus sequence of GAV 3CL^{pro} cleavage sites. Although the picture is still incomplete, our data indicate that the substrate specificity of the GAV 3CL^{pro} is well defined, as in vertebrate nidovirus main proteinases and many of their viral relatives, but differs from that of typical 3C/3C-like enzymes.

Dispensability of C-terminal sequences for 3CL^{pro} autoprocessing activity. The observed preference for substrates containing HEL or HEV tripeptides lends additional support to our hypothesis that there is no cleavage site between the 3CL^{pro} domain and the downstream putative membrane-spanning domain. It is thus tempting to speculate that, in contrast to the main proteinases of vertebrate nidoviruses, the GAV 3CL^{pro} is the N-terminal component of a larger protein. To determine whether the sequences downstream of the predicted two-β-barrel domain are essential for 3CL^{pro} cleavage activity, we compared the proteolytic activities of two C-terminal MBP-2793-3143 deletion mutants with that of the parental protein. As Fig. 7 shows, the two C-terminally truncated proteins had reduced but clearly detectable proteolytic activities, suggesting that the N-terminal region from 1 to 197 contains all the structural elements and residues required for substrate binding and catalysis. Furthermore, comigration of the processed N-terminal product (Fig. 7) suggests that, in all three proteins with proteolytic activity, cleavage occurred at the same peptide bond.

Active center of GAV 3CL^{pro}. In a final set of experiments, the predictions of possible active-site residues (8) (Fig. 3) were tested by site-directed mutagenesis. The MBP-2793-3143 protein encoded by the parental plasmid construct pMal-GAV-2793-3143 (Table 1, Fig. 1) and characterized in the experiments shown in Fig. 4 was used as a positive control. Single-amino-acid substitutions were introduced into this con-

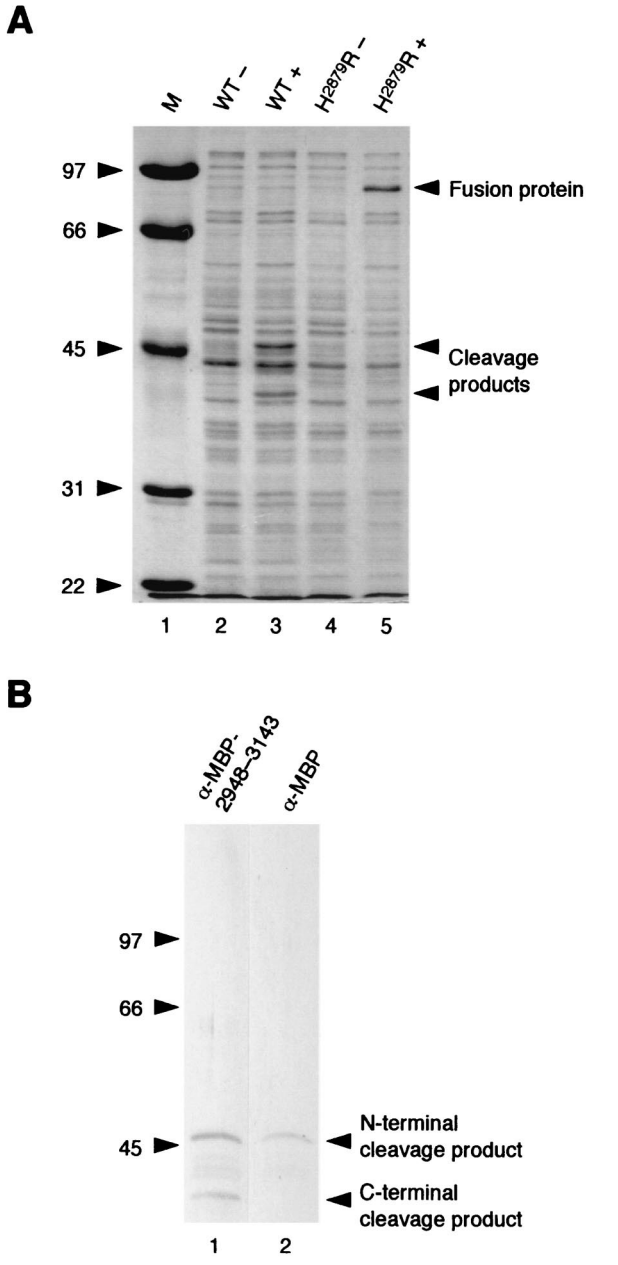


FIG. 4. Proteolytic activity of GAV pp1a/pp1ab amino acids 2793 to 3143. (A) Total cell lysates from *E. coli* TB1 cells transformed with pMal-GAV-2793-3143 (lanes 2 and 3, WT) and pMal-GAV-2793-3143_H^{2879}R (lanes 4 and 5, H^{2879}R) were separated by SDS-PAGE in a 12.5% polyacrylamide gel and stained with Coomassie brilliant blue R-250. The bacteria were mock induced (lanes 2 and 4) or induced with 1 mM IPTG for 3 h (lanes 3 and 5). The positions of the fusion proteins and cleavage products are indicated, and the molecular masses of marker proteins (lane 1) are given (in kilodaltons). (B) The protein lysate shown in panel A (lane 3) was separated by SDS-PAGE in a 10% polyacrylamide gel, transferred to a nitrocellulose membrane, and immunostained with MBP-2948-3143-specific rabbit antiserum (lane 1) or MBP-specific antiserum (New England Biolabs) (lane 2). The positions of the N-terminal (i.e., MBP-containing) and C-terminal cleavage products are indicated, and the positions of marker proteins are given (with masses in kilodaltons).

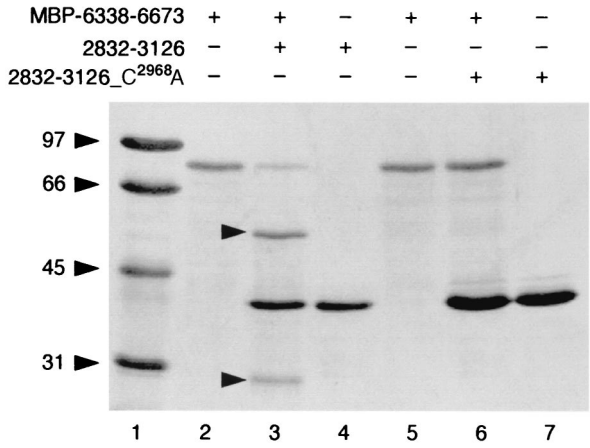


FIG. 5. *trans*-cleavage activity of GAV $3CL^{pro}$. Recombinant GAV $3CL^{pro}$ encompassing 295 amino acids (2832 to 3126) and an active-site mutant (2832-3126_C^{2968}A) were bacterially expressed, purified, and incubated with an MBP fusion protein substrate, MBP-6338-6673, containing the C-terminal GAV pp1ab sequence (see Materials and Methods for details). Lanes: 1, marker proteins, with molecular masses indicated in kilodaltons; 2, MBP-6338-6673 incubated with buffer; 3, MBP-6338-6673 incubated with 2832-3126; 4, 2832-3126 incubated with buffer; 5, MBP-6338-6673 incubated with buffer; 6, MBP-6338-6673 incubated with 2832-3126_C^{2968}A; 7, 2832-3126_C^{2968}A incubated with buffer. Cleavage products of MBP-6338-6673 are indicated by arrowheads.

struct, and their effects were studied by analyzing the autoprocessing activities of the MBP-2793-3143 mutants. The data shown in Fig. 8 revealed that replacements of the predicted catalytic His^{2879} (by Arg and Leu) and Cys^{2968} (by Ala and Ser) residues completely abolished proteolytic activity, supporting the proposed catalytic function of these residues. In contrast, all the Asp^{2912} mutants (D^{2912}A, D^{2912}E, and D^{2912}Q) retained their activities in the assay used. This result is consistent with our sequence comparison data, which also contradicted a catalytic function of this residue (see Fig. 3).

Mutagenesis of His^{2983} resulted in proteolytically inactive proteins, whereas the Ser^{2988} mutants retained wild-type activity. These data make His^{2983} the most probable candidate for the key position in the S1 subsite of the $3CL^{pro}$ substrate-binding pocket. We speculate that His^{2983} may cooperate with a threonine residue (Thr^{2963}) that, as in many other 3C/3C-like proteinases (3, 16, 29, 30), is located 5 residues upstream of the presumed GAV $3CL^{pro}$ principal nucleophile (Cys^{2968}) and, together with the imidazole side chain of histidine, may contact the P1 side chain of the substrate. The results thus fully support our predictions on GAV $3CL^{pro}$ putative active-site residues (see above and Fig. 3).

DISCUSSION

GAV is the first invertebrate nidovirus to be characterized at the molecular level. It infects black tiger prawns and represents the prototype of newly established taxa, genus *Okavirus*, family *Roniviridae*, within the order *Nidovirales* (8, 11). In this study, the viral main proteinase, a 3C-like cysteine proteinase, was characterized. Despite the wealth of information available for diverse $3CL^{pro}$s, predictions of the key features of the GAV

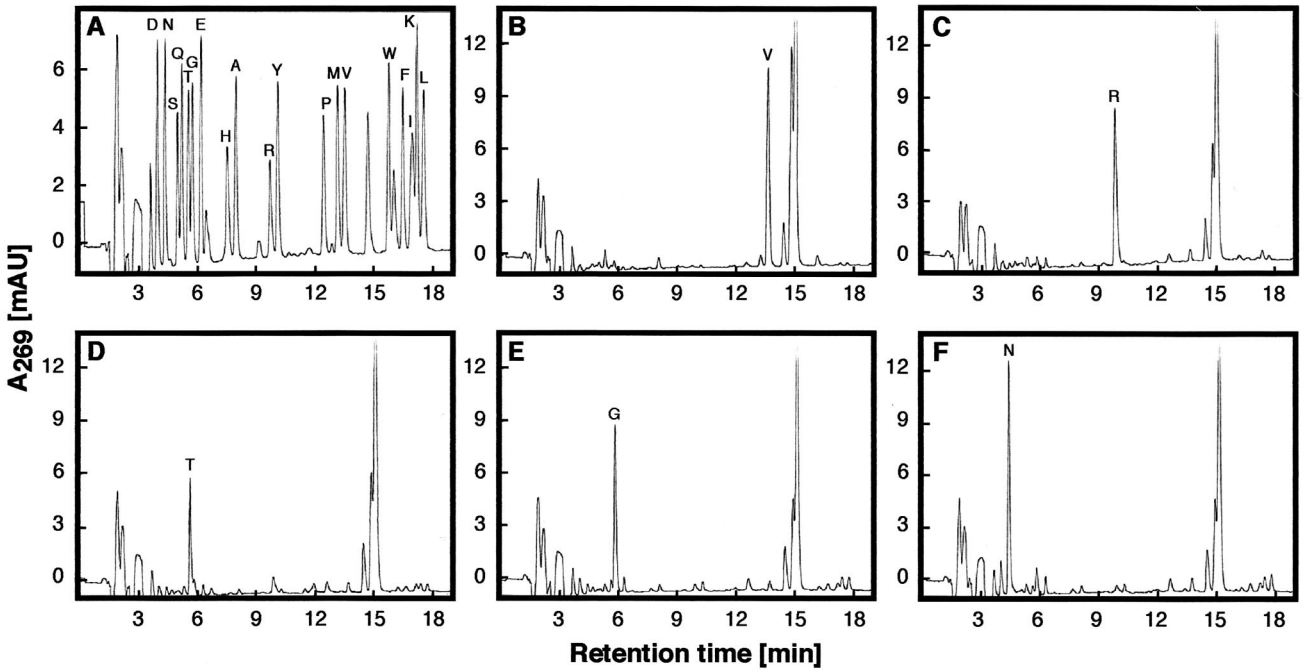


FIG. 6. Characterization of N-terminal GAV 3CL^{pro} autoprocessing site by protein sequencing. The C-terminal MBP-2793-3143 cleavage product (Fig. 4A, lane 3) was subjected to Edman degradation, and phenylthiohydantoin (PTH)-amino acids generated during each reaction cycle were detected by their absorbance at 269 nm (expressed as milliabsorption units) and identified by their characteristic retention times 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 of PTH-amino acids are indicated by the single-letter code.

3CL^{pro} proved to be challenging because of the unique phylogenetic position of this invertebrate nidovirus. Nevertheless, we were able to produce a coherent picture with a combination of bioinformatics and biochemical and genetic methods.

Previous studies of coronavirus 3CL^{pro}s suggested that ancestors of these enzymes accepted unprecedented substitutions in most of the conserved positions of the catalytic system and the substrate pocket, making this group of enzymes an outlier among the huge family of viral and cellular chymotrypsin-like homologs (2, 15, 17). We now provide evidence that GAV 3CL^{pro} provides an evolutionary link between the 3CL^{pro}s of coronaviruses and (all the) other positive-stranded RNA viruses. Specifically, our data indicate that the unique replacements in coronavirus 3CL^{pro}s of otherwise strictly conserved residues must have been acquired gradually in the nidovirus lineage. In this context, the GAV 3CL^{pro} seems to emerge as an important model to study (separately) the functional effects of the (abridged) Cys-His catalytic system. This is possible because, in contrast to coronavirus 3CL^{pro}s, which feature both a Cys-His catalytic center and a noncanonical substrate pocket, the GAV 3CL^{pro} Cys-His catalytic center seems to be combined with a canonical (potyvirus-like) substrate pocket (see below and Fig. 9).

Catalytic system of 3CL^{pro}. Sequence comparisons revealed that the GAV 3CL^{pro} has very little similarity to other RNA viral 3C-like proteinases (Fig. 2) (8). Even with the closest known relatives, potyvirus NIa and coronavirus main proteinases, similarities to the GAV 3CL^{pro} are restricted essentially to the regions containing the putative Cys and His active-site residues (Fig. 2), which made sequence alignments in other regions less robust. Our experimental evidence strongly suggests that 3CL^{pro} employs a catalytic dyad composed of Cys^{2968} and His^{2879}. The mutagenesis data did not corroborate earlier predictions of a third catalytic residue (Asp^{2912}) (8). Instead, the acidic residue appears to be replaced in GAV by the neutral Ala^{2911} residue (Fig. 3).

It should be noted that an equivalent of the Asp residue of the chymotrypsin catalytic triad is also missing in coronavirus 3CL^{pro}s (2, 17, 24, 50). Also, in the crystal structure of the hepatitis A virus 3C proteinase, the side chain of the conserved Asp residue adopts an unexpected orientation (1, 4). Even though the hepatitis A virus Asp^{84} residue occupies the expected position in the main chain, it forms a salt bridge with the ϵ amino group of a Lys side chain from strand fII (4) rather than interacting with the catalytic His^{44}, and thus, a catalytic function is unlikely. Apparently, in an appropriate environment, the relatively low pK_a of the Cys nucleophile (compared to that of Ser) may fully or partially relieve some 3C/3C-like cysteine proteinases from dependence on an Asp (Glu) carboxylate group, which is usually required to stabilize the developing positive charge on the catalytic histidine residue during serine proteinase catalysis (13, 23, 27).

Substrate specificity. In this study, initial information on the substrate specificity of the GAV 3CL^{pro} was obtained by determining the N-terminal 3CL^{pro} autoprocessing site and a second 3CL^{pro} cleavage site in the C-terminal region of pp1ab. The sequences flanking the scissile bonds, ^{2827}LVTHE↓VRTGN^{2836} and ^{6441}KVNHE↓LYHVA^{6450}, share the VxHE↓(L,V) motif. Inspection of coronavirus/GAV replicase alignments (A. E. Gorbalenya and J. Ziebuhr, unpublished

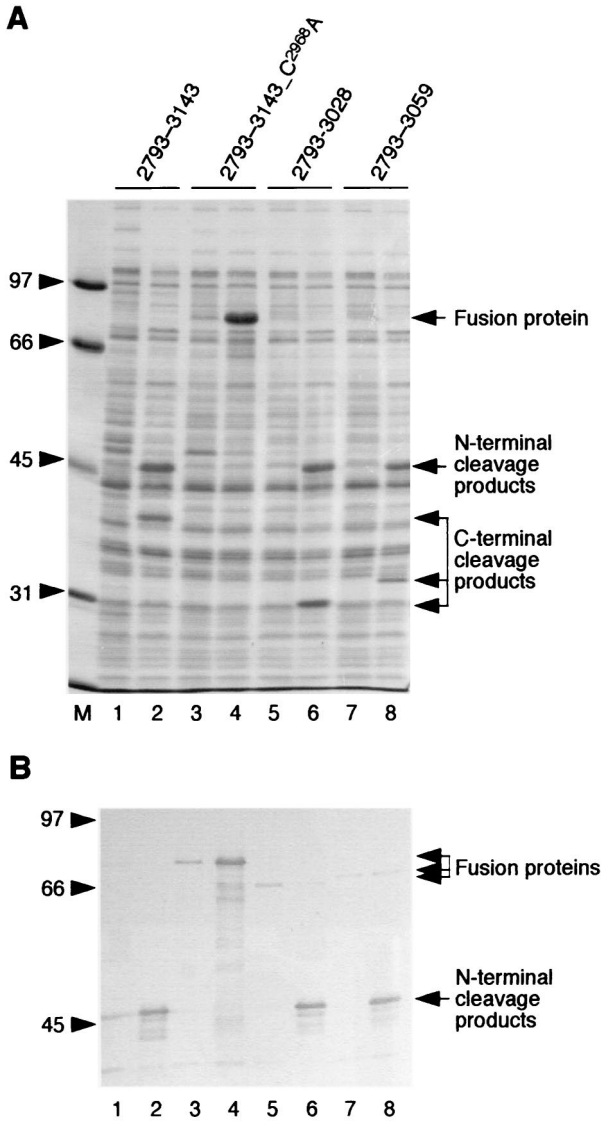


FIG. 7. Effect of C-terminal deletions on the self-processing activity of MBP-2793-3143. (A) Total cell lysates from *E. coli* TB1 cells transformed with pMal-GAV-2793-3143 (lanes 1 and 2; 2793-3143), pMal-GAV-2793-3143_C^{2968}A (lanes 3 and 4; 2793-3143_C^{2968}A), pMal-GAV-2793-3028 (lanes 5 and 6; 2793-3028), and pMal-GAV-2793-3059 (lanes 7 and 8; 2793-3059) were separated by SDS-PAGE in a 12.5% polyacrylamide gel and stained with Coomassie brilliant blue R-250. The bacteria were mock induced (lanes 1, 3, 5, and 7) or induced with 1 mM IPTG for 3 h (lanes 2, 4, 6, and 8). The positions of the fusion proteins and cleavage products are indicated, and the molecular masses of marker proteins (lane M) are given (in kilodaltons). (B) The cell lysates shown in panel A were separated by SDS-PAGE, transferred to a nitrocellulose membrane, and immunostained with anti-MBP antiserum (New England Biolabs). The positions of the uncleaved fusion proteins and the N-terminal (i.e., MBP-containing) cleavage products are indicated, and the positions of marker proteins are given (with masses in kilodaltons).

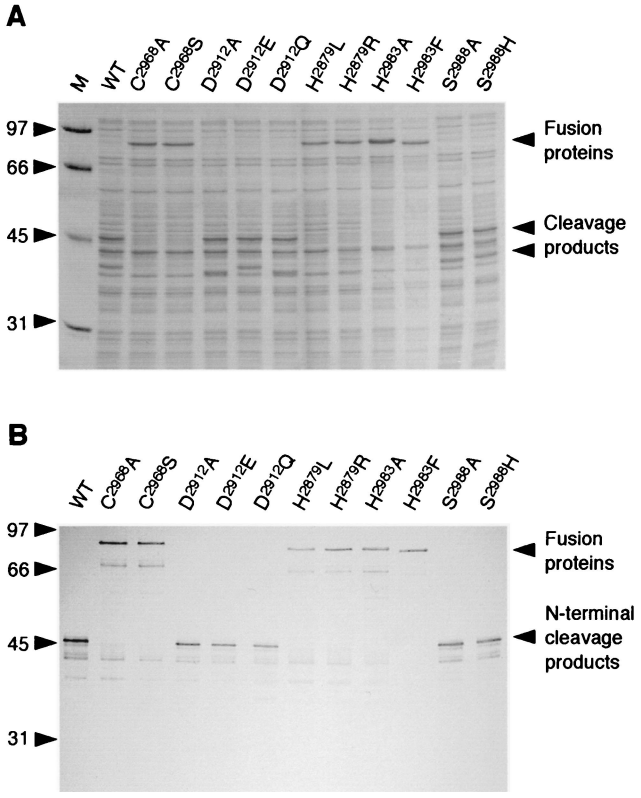


FIG. 8. Mutational analysis of active center of GAV $3CL^{pro}$. (A) The proteolytic activities of bacterially expressed MBP-2793-3143 proteins carrying substitutions of putative active-site residues were examined by SDS-PAGE of cell lysates obtained after IPTG-induced (3 h, 24°C) protein expression. The introduced amino acid substitutions and the positions of both uncleaved fusion proteins and cleavage products are indicated. The proteolytic activity of the wild-type MBP-2793-3143 (WT) (see also Fig. 4) served as a positive control. (B) The cell lysates shown in panel A were separated by SDS-PAGE, transferred to a nitrocellulose membrane, and immunostained with anti-MBP antiserum (New England Biolabs). The positions of the uncleaved fusion proteins and the N-terminal (that is, MBP-containing) cleavage products are indicated. Also shown are the positions of molecular mass markers (with masses given in kilodaltons).

data) leads us to believe that Val/Thr/Ser and Leu/Val/Ile/Gly/Ser/Ala at the substrate P4 and P1′ positions, respectively, may be compatible with proteolysis by GAV $3CL^{pro}$. This conservation pattern suggests that the P4, P2, P1, and P1′ positions are the major $3CL^{pro}$ specificity determinants. The same positions are critical in corona- and potyvirus $3CL^{pro}$ cleavage sites, which provides further support to combine the GAV, corona-, and potyvirus $3CL^{pro}$s in a separate group.

Whereas the presence of Glu (or Gln) at the P1 position is a typical feature of RNA virus 3C/$3CL^{pro}$ substrates (16, 36), the GAV $3CL^{pro}$ preferences at the other conserved positions are less common and, taken together, give this proteinase a unique substrate specificity formula. Interestingly, some plant potyvirus NIa 3C-like proteinases (21, 33, 44, 48) share the P2 His substrate specificity with the GAV $3CL^{pro}$. It is also noteworthy that, unlike most other 3C/3C-like proteinases, GAV $3CL^{pro}$ seems to possess a relatively large (hydrophobic) S1′ subsite, which would accommodate the branched side chains of valine and leucine.

A striking parallel between GAV $3CL^{pro}$ and various well-characterized positive-stranded RNA virus homologs (3, 16, 28–30) is the conservation of the pair of His/Thr residues in the S1 subsite. Our hypothesis that the corresponding GAV

	CATALYTIC RESIDUES					
	*	*		*		
Enterovirus (PV)	H	E	T	C	G	H
Hepatovirus (HAV)	H	(D)	G	C	G	H
Nepovirus (TBRV)	H	E	S	C	G	L
Potyvirus (PEMV)	H	D	T	C	G	H
Ronivirus (GAV)	H		T	C	G	H
Coronavirus (HCoV)	H			C	Y	H
Arterivirus (EAV)	H	D	T	S	G	H
			#		#	#
	SUBSTRATE POCKET RESIDUES					

FIG. 9. Variations in catalytic and substrate-binding residues of RNA viral chymotrypsin-like proteinases. PV, poliovirus; HAV, hepatitis A virus; TBRV, tomato black ring virus; PEMV, pepper mottle virus; HCoV, human coronavirus; EAV, equine arteritis virus. The key catalytic (*) and substrate-binding pocket (#) residues are indicated. The catalytic Asp residue of hepatitis A virus is shown in brackets because its side chain orientation in the hepatitis A virus $3C^{pro}$ crystal structure (1, 4) argues against the proposed catalytic function (see text for details).

$3CL^{pro}$ residues (Thr^{2963} and His^{2983}) may play an equivalent role is further supported by the local conservation of the corresponding region among GAV and potyvirus $3CL^{pro}$s (Fig. 2 and 3) and our mutagenesis data (see above). Despite these similarities, it is likely that additional (poorly recognized) determinants may tune the P1 specificity in a virus-specific manner. Thus, for example, it is conceivable that the $3CL^{pro}$s of GAV and arteriviruses, which both recognize a P1 Glu (rather than Gln) side chain (38; this paper), have similarly organized S1 subsites.

Cleavage at C terminus of $3CL^{pro}$. RNA virus (including vertebrate nidovirus) 3C/$3CL^{pro}$s are commonly released from the replicase polyproteins by autocatalytic processing. In some cases, the N- and C-terminal sites are cleaved with different kinetics. Thus, for example, C-terminal 3C/$3CL^{pro}$ cleavage occurs more slowly (picornaviruses) (36), is tightly regulated (arteriviruses) (45), or is totally lacking (some caliciviruses) (39, 46). In our experiments, no evidence was obtained for cleavage in the region immediately downstream of the GAV $3CL^{pro}$ which, according to comparative sequence analysis (Fig. 3), also does not contain potential [that is, VxHE ↓ (L,V)] cleavage sites.

It is possible that a site immediately downstream of the proteinase domain might be cleaved by a cellular proteinase. However, this would be unprecedented based on data for other viral $3CL^{pro}$s. Alternatively, domains from other regions of the viral polyprotein, which are missing in our constructs, might assist in autoprocessing at a C-terminal $3CL^{pro}$ site with a deviant structure. For instance, studies of the arterivirus equine arteritis virus have revealed that the C-terminal release of the nsp4 proteinase from the nsp4-8 precursor requires nsp2 as a cofactor (45). Further studies with larger GAV $3CL^{pro}$-containing precursor proteins and alternative expression systems, including insect cells and primary crustacean cells (31), may help to address this question more rigorously.

If GAV $3CL^{pro}$ and the downstream hydrophobic domain are not separated by proteolytic cleavage, as our results suggest, then the proteinase would remain anchored to intracellular membranes throughout the replication cycle. To some extent, this association would resemble the situation in the arterivirus equine arteritis virus and the coronavirus mouse hepatitis virus, in which significant amounts of nsp4 and $3CL^{pro}$, respectively, are known to remain part of long-lived (or even stable) precursors which possess flanking hydrophobic domains on either one or both sides (22, 37, 45).

Domain structure of $3CL^{pro}$. In contrast to other 3C/$3CL^{pro}$s, which consist of two catalytic β-barrel domains (1, 4, 28–30), nidovirus and potyvirus $3CL^{pro}$s possess an extra C-terminal domain of variable size (51). This additional domain is also present in the GAV $3CL^{pro}$, although its precise size remains to be determined. In coronavirus $3CL^{pro}$s, the C-terminal domain is involved in *trans*-cleavage activity (2, 26, 32, 50). Recent crystal structure analysis of the transmissible gastroenteritis virus $3CL^{pro}$ showed that the domain adopts a unique α-helical structure that interacts with the enzyme's N terminus. This interaction fixes the orientation of a loop region involved in substrate binding (2).

The fact that the C-terminally truncated, 197-residue GAV $3CL^{pro}$ (Fig. 1 and 7) retained significant autoprocessing activity when expressed as an MBP fusion protein argues against an equally important role for the C-terminal domain of GAV $3CL^{pro}$, at least in *cis* reactions. The effects of C-terminal deletions on the activity in *trans* remain to be determined. This experiment is of special interest because coronavirus $3CL^{pro}$s have been shown to be differentially affected by C-terminal deletions in *cis*- versus *trans*-cleavage reactions (2, 26, 32, 50).

Taken together, the differences and similarities revealed in this study between the main proteinase of a crustacean nidovirus and its viral homologs indicate a novel pattern of functional and structural conservation that has not been observed in any of the previously characterized proteinases from mammalian and plant pathogens. We are confident that, from an evolutionary perspective, the characterization of proteins of positive-stranded RNA viruses isolated from less-characterized habitats will allow valuable insights into the evolution of viruses and help identify both missing phylogenetic links and evolutionary forces operating in specific biological systems.

ACKNOWLEDGMENTS

The work was supported by grants from the Deutsche Forschungsgemeinschaft awarded to J.Z. (Zi 618/1 and Zi 618/2).

We thank Viviane Hoppe for protein sequence data.

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. **Barrette-Ng, I. H., K. K. Ng, B. L. Mark, D. van Aken, 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.
4. **Bergmann, E. M., S. C. Mosimann, M. M. Chernaia, B. A. Malcolm, and M. N. 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.

5. **Boonyaratpalin, S., K. Supamataya, J. Kasornchandra, S. Direkbusaracom, U. Aekpanithanpong, and C. Chantanachookin.** 1993. Non-occluded baculo-like virus, the causative agent of yellow-head disease in the black tiger shrimp (*Penaeus monodon*). Fish Pathol. **28:**103–109.
6. **Cavanagh, D.** 1997. Nidovirales: a new order comprising Coronaviridae and Arteriviridae. Arch. Virol. **142:**629–633.
7. **Chantanachookin, C., S. Boonyaratpalin, J. Kasornchandra, D. Sataporn, U. Ekpanithanpong, K. Supamataya, S. Riurairatana, and T. W. Flegel.** 1993. Histology and ultrastructure reveal a new granulosis-like virus in *Penaeus monodon* affected by yellow-head disease. Dis. Aquat. Org. **17:**145–157.
8. **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.
9. **Cowley, J. A., C. M. Dimmock, and P. J. Walker.** 2002. Gill-associated nidovirus of *Penaeus monodon* prawns transcribes 3′-coterminal subgenomic mRNAs that do not possess 5′-leader sequences. J. Gen. Virol. **83:**927–935.
10. **Cowley, J. A., C. M. Dimmock, C. Wongteerasupaya, V. Boonsaeng, S. Panyim, and P. J. Walker.** 1999. Yellow head virus from Thailand and gill-associated virus from Australia are closely related but distinct prawn viruses. Dis. Aquat. Org. **36:**153–157.
11. **Cowley, J. A., and P. J. Walker.** 2002. The complete genome sequence of gill-associated virus of *Penaeus monodon* prawns indicates a gene organisation unique among nidoviruses. Arch. Virol. **147:**1977–1987.
12. **de Vries, A. A., M. C. Horzinek, P. J. 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.
13. **Dodson, G., and A. Wlodawer.** 1998. Catalytic triads and their relatives. Trends Biochem. Sci. **23:**347–352.
14. **Gorbalenya, A. E.** 2001. Big nidovirus genome. When count and order of domains matter. Adv. Exp. Med. Biol. **494:**1–17.
15. **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.
16. **Gorbalenya, A. E., and E. J. Snijder.** 1996. Viral cysteine proteinases. Perspect. Drug Discov. Des. **6:**64–86.
17. **Hegyi, A., A. Friebe, A. E. Gorbalenya, and J. Ziebuhr.** 2002. Mutational analysis of the active centre of coronavirus 3C-like proteases. J. Gen. Virol. **83:**581–593.
18. **Henikoff, S., and J. G. Henikoff.** 1994. Position-based sequence weights. J. Mol. Biol. **243:**574–578.
19. **Herold, J., S. Siddell, and J. Ziebuhr.** 1996. Characterization of coronavirus RNA polymerase gene products. Methods Enzymol. **275:**68–89.
20. **Heusipp, G., C. Grötzinger, J. Herold, S. G. Siddell, and J. Ziebuhr.** 1997. Identification and subcellular localization of a 41 kDa, polyprotein 1ab processing product in human coronavirus 229E-infected cells. J. Gen. Virol. **78:**2789–2794.
21. **Kang, H., Y. J. Lee, J. H. Goo, and W. J. Park.** 2001. Determination of the substrate specificity of turnip mosaic virus NIa protease with a genetic method. J. Gen. Virol. **82:**3115–3117.
22. **Kanjanahaluethai, A., and S. C. Baker.** 2000. Identification of mouse hepatitis virus papain-like proteinase 2 activity. J. Virol. **74:**7911–7921.
23. **Kraut, J.** 1977. Serine proteases: structure and mechanism of catalysis. Annu. Rev. Biochem. **46:**331–358.
24. **Liu, D. X., and T. D. Brown.** 1995. Characterisation and mutational analysis of an ORF 1a-encoding proteinase domain responsible for proteolytic processing of the infectious bronchitis virus 1a/1b polyprotein. Virology **209:** 420–427.
25. **Liu, D. X., S. Shen, H. Y. Xu, and S. F. Wang.** 1998. Proteolytic mapping of the coronavirus infectious bronchitis virus 1b polyprotein: evidence for the presence of four cleavage sites of the 3C-like proteinase and identification of two novel cleavage products. Virology **246:**288–297.
26. **Lu, Y., and M. R. Denison.** 1997. Determinants of mouse hepatitis virus 3C-like proteinase activity. Virology **230:**335–342.
27. **Matthews, B. W., P. B. Sigler, R. Henderson, and D. M. Blow.** 1967. Three-dimensional structure of tosyl-α-chymotrypsin. Nature **214:**652–656.
28. **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.
29. **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.
30. **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.
31. **Mulford, A. L., F. Lyng, C. Mothersill, and B. Austin.** 2000. Development and characterization of primary cell cultures from the hematopoietic tissues of the Dublin Bay prawn, *Nephrops norvegicus*. Methods Cell Sci. **22:**265–275.
32. **Ng, L. F., and D. X. Liu.** 2000. Further characterization of the coronavirus infectious bronchitis virus 3C-like proteinase and determination of a new cleavage site. Virology **272:**27–39.
33. **Nicolas, O., and J. F. Laliberte.** 1992. The complete nucleotide sequence of turnip mosaic potyvirus RNA. J. Gen. Virol. **73:**2785–2793.
34. **Rost, B.** 1996. PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol. **266:**525–539.
35. **Rost, B., R. Casadio, P. Fariselli, and C. Sander.** 1995. Transmembrane helices predicted at 95% accuracy. Protein Sci. **4:**521–533.
36. **Ryan, M. D., and M. Flint.** 1997. Virus-encoded proteinases of the picornavirus super-group. J. Gen. Virol. **78:**699–723.
37. **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.
38. **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.
39. **Sosnovtseva, S. A., S. V. Sosnovtsev, and K. Y. Green.** 1999. Mapping of the feline calicivirus proteinase responsible for autocatalytic processing of the nonstructural polyprotein and identification of a stable proteinase-polymerase precursor protein. J. Virol. **73:**6626–6633.
40. **Spann, K. M., J. A. Cowley, P. J. Walker, and R. J. Lester.** 1997. Gill-associated virus (GAV), a yellow head-like virus from *Penaeus monodon* cultured in Australia. Dis. Aquat. Org. **31:**169–179.
41. **Spann, K. M., J. E. Vickers, and R. J. Lester.** 1995. Lymphoid organ virus of *Penaeus monodon* from Australia. Dis. Aquat. Org. **23:**127–134.
42. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
43. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Improved sensitivity of profile searches through the use of sequence weights and gap excision. Comput. Appl. Biosci. **10:**19–29.
44. **Vance, V. B., D. Moore, T. H. Turpen, A. Bracker, and V. C. Hollowell.** 1992. The complete nucleotide sequence of pepper mottle virus genomic RNA: comparison of the encoded polyprotein with those of other sequenced potyviruses. Virology **191:**19–30.
45. **Wassenaar, A. L., W. J. 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.
46. **Wei, L., J. S. Huhn, A. Mory, H. B. Pathak, S. V. Sosnovtsev, K. Y. Green, and C. E. Cameron.** 2001. Proteinase-polymerase precursor as the active form of feline calicivirus RNA-dependent RNA polymerase. J. Virol. **75:** 1211–1219.
47. **Yao, Z., D. H. Jones, and C. Grose.** 1992. Site-directed mutagenesis of herpesvirus glycoprotein phosphorylation sites by recombination polymerase chain reaction. PCR Methods Appl. **1:**205–207.
48. **Yeh, S. D., F. J. Jan, C. H. Chiang, T. J. Doong, M. C. Chen, P. H. Chung, and H. J. Bau.** 1992. Complete nucleotide sequence and genetic organization of papaya ringspot virus RNA. J. Gen. Virol. **73:**2531–2541.
49. **Ziebuhr, J., J. Herold, and S. G. Siddell.** 1995. Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. J. Virol. **69:**4331–4338.
50. **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.
51. **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.