

Characterization of Bafinivirus Main Protease Autoprocessing Activities[∇]

Rachel Ulferts,¹ Thomas C. Mettenleiter,² and John Ziebuhr^{1,3*}

Centre for Infection and Immunity, The Queen's University of Belfast, Belfast, United Kingdom¹; Friedrich-Loeffler-Institut, Greifswald-Insel Riems, Germany²; and Institute of Medical Virology, Justus Liebig University, Giessen, Germany³

Received 13 August 2010/Accepted 1 November 2010

The production of functional nidovirus replication-transcription complexes involves extensive proteolytic processing by virus-encoded proteases. In this study, we characterized the viral main protease (M^{Pro}) of the type species, *White bream virus* (WBV), of the newly established genus *Bafinivirus* (order *Nidovirales*, family *Coronaviridae*, subfamily *Torovirinae*). Comparative sequence analysis and mutagenesis data confirmed that the WBV M^{Pro} is a picornavirus 3C-like serine protease that uses a Ser-His-Asp catalytic triad embedded in a predicted two-β-barrel fold, which is extended by a third domain at its C terminus. Bacterially expressed WBV M^{Pro} autocatalytically released itself from flanking sequences and was able to mediate proteolytic processing *in trans*. Using N-terminal sequencing of autoproteolytic processing products we tentatively identified Gln↓(Ala, Thr) as a substrate consensus sequence. Mutagenesis data provided evidence to suggest that two conserved His and Thr residues are part of the S1 subsite of the enzyme's substrate-binding pocket. Interestingly, we observed two N-proximal and two C-proximal autoprocessing sites in the bacterial expression system. The detection of two major forms of M^{Pro}, resulting from processing at two different N-proximal and one C-proximal site, in WBV-infected epithelioma papulosum cyprini cells confirmed the biological relevance of the biochemical data obtained in heterologous expression systems. To our knowledge, the use of alternative M^{Pro} autoprocessing sites has not been described previously for other nidovirus M^{Pro} domains. The data presented in this study lend further support to our previous conclusion that bafiniviruses represent a distinct group of viruses that significantly diverged from other phylogenetic clusters of the order *Nidovirales*.

The order *Nidovirales* comprises four profoundly separated clusters of plus-strand RNA viruses that have been assigned to three families called *Coronaviridae*, *Arteriviridae*, and *Roniviridae* (13, 20). The taxonomy of the family *Coronaviridae* has recently been revised and now includes two subfamilies called *Coronavirinae* (genera *Alpha-*, *Beta-*, and *Gamma-coronavirus*) and *Torovirinae* (genera *Torovirus* and *Bafinivirus*) (<http://ictvonline.org/virusTaxonomy.asp>). *White bream virus* (WBV) strain DF24/00 is the type species and, to date, sole member of the newly established genus *Bafinivirus* (38). With about 26.6 kb, the genome of WBV rivals the large genome sizes reported for many other nidoviruses. Further similarities between bafiniviruses and other nidoviruses include conserved (i) polycistronic genome organizations; (ii) genome expression strategies, including the synthesis of a nested set of subgenomic (sg) mRNAs; and (iii) colinear arrays of replicative enzymes, including RNA-processing activities that are not conserved in other RNA viruses. WBV (and yet-to-be-identified members of the genus *Bafinivirus*) and viruses from the genus *Torovirus* were classified as two separate genera in a common subfamily on the basis of phylogenetic analyses of the most conserved viral enzymes and conserved features of their structural proteins (18, 38). The available information suggests that members of the two genera originated from an immediate common ancestor but split early in evolution (38).

Also, the two groups of viruses infect very different hosts (mammals and fish, respectively) and use slightly different strategies to synthesize their subgenomic RNAs (38, 48). The available (although limited) data suggest that the bafinivirus cluster diverged profoundly from other nidoviruses (including toroviruses) and occupies a distant position in the evolutionary tree of nidoviruses. Molecular studies of the biology of bafiniviruses may therefore be expected to provide new insight into conserved and distinct biological features of the major clusters of nidoviruses and, possibly, help identify major driving forces involved in the evolution of this highly divergent group of RNA viruses. We therefore decided to embark on a systematic characterization of the molecular biology of bafiniviruses, and we report here the results of a first set of experiments aimed to investigate the expression and autoprocessing of the WBV main protease (M^{Pro}), an enzyme predicted to play a key role in the proteolytic processing of the WBV ORF1a/1b-encoded polyproteins, pp1a and pp1ab, and thus in the formation and functional maturation of the viral replication-transcription complex, as shown previously for other nidoviruses (reviewed in references 47 and 58).

Proteolytic processing of WBV pp1a and pp1ab is expected to yield more than a dozen nonstructural proteins (nsps). The majority of processing events are predicted to be mediated by the M^{Pro}, while other yet-to-be-identified proteases might cleave a limited number of sites in the N-terminal polyprotein regions (38). Because of similarities in both sequence and structure between nidovirus M^{Pro}s and picornavirus 3C proteases (3C^{Pro}), the nidovirus M^{Pro} is occasionally referred to as 3C-like protease (3CL^{Pro}) (19, 63), while yet another name, chymotrypsin-like proteases (CHL^{Pro}), is some-

* Corresponding author. Mailing address: Institute of Medical Virology, Justus Liebig University, Frankfurter Str. 107, 35392 Giessen, Germany. Phone: 49-641-9941200. Fax: 49-641-9941209. E-mail: john.ziebuhr@viro.med.uni-giessen.de.

[∇] Published ahead of print on 10 November 2010.

times used for the same group of proteases to refer to the chymotrypsin (two β -barrel) fold that is shared by all 3C and 3C-like proteases characterized to date (2–4, 7, 31, 32, 42, 46, 52, 54, 64).

We previously identified a putative M^{PRO} homolog in the WBV replicase polyproteins (38). Similar to its homologs in other nidoviruses (21, 59, 63), the WBV M^{PRO} was confirmed to be located within the C-terminal third of pp1a where it is flanked on both sides by hydrophobic domains (38). The present study provides initial insight into the autoprocessing characteristics and substrate specificity of the WBV M^{PRO} using information derived from heterologous expression systems and WBV-infected cells and revealed conserved residues predicted to be involved in catalysis and substrate specificity.

MATERIALS AND METHODS

Cell culture and propagation of virus. Epithelioma papulosum cyprini (EPC) cells (RIE173) were obtained from the cell culture collection of the Friedrich-Loeffler-Institut (Greifswald-Insel Riems, Germany). They were grown in minimum essential medium with Earle's salts, 25 mM HEPES, GlutaMAX I (Invitrogen, catalog no. 42360-024) supplemented with 1 mM sodium pyruvate (Invitrogen, catalog no. 11360-039) and 10% fetal bovine serum (PAA Laboratories) at 27°C in a 5% CO₂ atmosphere. WBV isolate DF24/00 (23) was propagated on EPC cells at 20°C.

Construction and mutagenesis of expression plasmids. Plasmids were generated by using standard cloning procedures. All PCR products were generated by using *Pfu* polymerase (Promega) according to the manufacturer's instructions. The correct sequences of the inserts were verified by sequencing using a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems). To generate pMal-WBV-3424-3725-GST and pMal-WBV-3424-3725_S3589A-GST, the coding sequence for the WBV pp1a/pp1ab amino acid residues Gln-3424 to Met-3725 was amplified by using the primers MalE24 (5'-GGTCGTCAGACTGTCGATGAA GCC-3') and M^{PRO}_BbsI (5'-AATAAGAAGACTGTGTGTTAACTGGTGATG TGAGTGT-3') and pMal-WBV-3CL_559-560 or pMal-WBV-3CL_S3589A (38) as a template, while the *Schistosoma japonicum* glutathione S-transferase (GST) coding sequence was amplified from pCITE-GST (44) by using the primers GSTup_BsaI (5'-AAAAGGTCTCTACAACATGAGCCCTATACTAGGTTA TTGG-3') and GSTdn_BamHI (5'-AAAGGATCCTTAACGCGGAACCAGA TCCGATTTTGG-3'). The resulting PCR products were digested with BbsI and BsaI, respectively, and subsequently ligated by using T4 DNA ligase (New England Biolabs [NEB]). The resulting pp1a-3424-3725-GST fusion sequence was amplified by using MalE24 and GSTdn_BamHI, digested with SacI and BamHI, and ligated with SacI/BamHI-digested pMal-c2X plasmid DNA (NEB).

To construct pMal-WBV-M^{PRO}-CHis, the coding region for pp1a/pp1ab amino acids Gln-3424 to Thr-3707 was amplified from pMal-pp1a-3424-3725-GST by using the primers JZ559 (5'-TCAGCATCAGAATGCATTCTGTAT-3') and RU32 (5'-TTTGATCCTTAGTGATGGTGATGGTGATGTTACTGGT GCTACTGGGAG-3'). The resulting PCR product was phosphorylated by using T4 polynucleotide kinase (PNK; NEB), digested with EcoRI and ligated with XmnI/EcoRI-digested pMal-c2X plasmid DNA (NEB).

The inserts for plasmids pMal-WBV-N1-C1, pMal-WBV-N1-C2, pMal-WBV-N2-C1, and pMal-WBV-N2-C2 were amplified from pMal-WBV-3424-3725_S3589A-GST with the primers RU36 (5'-GCCCTCAACGATCAGAC AAGC-3') in combination with RU177 (5'-ATAGAATTCTTATTGTCTGGT GATGGTGTAC-3'), RU36 in combination with RU191 (5'-AAAGGATCCT TATTGTCTACTGTGACAGATGT-3'), RU37 (5'-GCAGTCACAGTTGTT GTAACAC-3') in combination with RU177, and RU37 in combination with RU191, respectively. The 5' termini of the resulting PCR products were phosphorylated by using PNK. The inserts for pMal-WBV-N1-C1 and pMal-WBV-N2-C1 were digested with EcoRI and then ligated with XmnI/EcoRI-digested pMal-c2X plasmid DNA. The inserts for pMal-WBV-N1-C2 and pMal-WBV-N2-C2 were digested with BamHI and ligated with XmnI/BamHI-digested pMal-c2X plasmid DNA. Site-directed mutagenesis of plasmid constructs was carried out by PCR-based methods essentially as described previously (28, 55). Sequences of oligonucleotides used for mutagenesis are available on request.

Expression of fusion proteins. *Escherichia coli* TB1 (NEB) cells transformed with pMal-pp1a-3424-3725-GST or pMal-pp1a-3424-3725_S3589A-GST were grown in Luria-Bertani (LB) medium containing ampicillin at a concentration of 100 μ g per ml until the culture reached an optical density at 600 nm (OD₆₀₀) of

0.5. The culture was halved, and protein expression was induced through the addition of IPTG (isopropyl- β -D-thiogalactopyranoside) to a final concentration of 1 mM in one of the two cultures, while the other was mock induced. After induction of protein expression, the cultures were incubated for 3 h at 25°C. Aliquots of each culture were spun at 15,000 \times g for 1 min, the pelleted cells were resuspended in 2 \times Laemmli sample buffer and incubated at 94°C for 2 min. Protein expression was analyzed by SDS-PAGE and Coomassie staining.

GST-affinity purification and N-terminal sequencing of proteolytic processing products. The proteins were expressed in *E. coli* TB1 essentially as described above. Three hours after induction of expression, the cells were pelleted by centrifugation at 6,000 g for 10 min at 4°C, resuspended in GST purification buffer (20 mM Tris-Cl [pH 7.5], 200 mM NaCl, 1 mM dithiothreitol [DTT]), and stored at -20°C overnight. The cells were thawed, 1 \times complete protease inhibitor cocktail without EDTA (Roche) was added, and the cells were lysed by sonication. Insoluble material was removed by centrifugation at 18,000 \times g for 30 min at 4°C. GST-containing cleavage products were purified by using GST-Bind Resin (Novagen) and eluted with 10 mM reduced glutathione in GST purification buffer. For determination of the N terminus, the eluted protein was electrophoresed in a discontinuous SDS-10% polyacrylamide gel, transferred onto Sequi blot polyvinylidene difluoride membrane (Bio-Rad), and stained with Coomassie brilliant blue R-250. Determination of the N terminus by Edman degradation was carried out by the protein sequencing facility of the Functional Genomics Center, Zurich, Switzerland.

IMAC of M^{PRO}-CHis wild-type and mutant proteins. *E. coli* TB1 cells were transformed with the appropriate (wild-type or mutant) pMal-WBV-M^{PRO}-CHis plasmid DNA. The protein was expressed as described for GST-affinity purification with the exception that the immobilized metal-affinity chromatography (IMAC) purification buffer contained 20 mM Tris-Cl (pH 8.0), 200 mM NaCl, and 5 mM 2-mercaptoethanol. In addition, the lysis buffer contained 10 mM imidazole. The cleared lysate was incubated with Ni-NTA agarose (Invitrogen) for 1 h at 4°C, and nonspecifically bound proteins were subsequently removed by extensive washing with IMAC purification buffer containing 50 mM imidazole. Bound proteins were eluted using 250 mM imidazole in IMAC purification buffer.

Amylose affinity purification. Maltose-binding protein (MBP) fusion proteins were expressed in *E. coli* TB1 cells essentially as described above for GST fusion proteins with the exception that bacteria were suspended in MBP purification buffer (50 mM Tris-Cl [pH 7.5], 300 mM NaCl, 1 mM DTT). Cleared lysates were applied onto a column containing amylose resin (NEB). Nonspecifically bound proteins were removed by extensive washing with MBP purification buffer, and bound protein was eluted using MBP purification buffer supplemented with 10 mM maltose. The MBP-M^{PRO} marker fusion proteins were subsequently dialyzed against factor Xa cleavage buffer (20 mM Tris-Cl [pH 7.5], 200 mM NaCl, 5 mM CaCl₂) and cleaved with factor Xa (Novagen) according to the manufacturer's instructions.

Detection of proteolytic processing products by immunoblotting. To generate a WBV M^{PRO}-specific antiserum, M^{PRO}-CHis was expressed in *E. coli* TB1 and purified by IMAC as described above. The protein was then purified further by size exclusion chromatography using a HiLoad 16/60 Superdex 75 prep-grade column (GE Healthcare) and buffer containing 20 mM Tris-Cl (pH 8.0), 200 mM NaCl, 1 mM DTT. This purified protein was used to raise a WBV M^{PRO}-specific polyclonal rabbit antiserum (Eurogentec, Inc.). Anti-MBP monoclonal antibody (catalog no. E8032S) was purchased from NEB, and rabbit anti-GST antiserum (catalog no. ab9085) was purchased from Abcam. For immunodetection of proteolytic processing products of MBP-pp1a-3424-3725-GST wild-type or mutant proteins, *E. coli* TB1 cells were transformed with pMal-pp1a-3424-3725-GST plasmid DNA or the appropriate mutant derivative. As a control for MBP expression, pMal-c2X was used. The proteins were expressed essentially as described above with the exemption that cultures were chilled on ice for 10 min prior to addition of IPTG. As a control, untransformed *E. coli* TB1 were used which were treated likewise. Aliquots of the cultures corresponding to equal amounts of OD₆₀₀ were centrifuged at 15,000 \times g for 1 min. Bacterial pellets were resuspended in 1 \times Laemmli sample buffer supplemented with 100 mM MgCl₂ and agitated on a vortex for 5 min. After DNA was pelleted by centrifugation for 5 min at 15,000 \times g, the supernatant was heated for 2 min at 94°C. Proteins were electrophoresed in 18-by-16-cm discontinuous SDS-15% polyacrylamide gels and subsequently transferred onto nitrocellulose membranes by semidry blotting. Unspecific binding sites were blocked by incubation of the membrane in 5% (wt/vol) skimmed milk powder in phosphate-buffered saline (PBS). The membrane was incubated with rabbit anti-GST antiserum (1:10,000) in 5% (wt/vol) skimmed milk powder in PBS-T0.05 (PBS containing 0.05% [vol/vol] Tween 20) overnight at 4°C. After washing for 10 min in PBS-T0.05 three times, the membrane was incubated with IRDye 800CW (LI-COR Bio-

sciences) goat anti-rabbit IgG (1:20,000) in 5% (wt/vol) skimmed milk powder in PBS-T0.05 for 1 h at room temperature. The membrane was washed three times for 10 min in PBS-T0.05 and then rinsed with PBS. Images were obtained by using the Odyssey Imaging system and software (LI-COR Biosciences). To remove bound antibodies, the membrane was incubated in 0.1 N NaOH for 30 min at room temperature. The membrane was then washed three times for 5 min with PBS and incubated with blocking solution (5% [wt/vol] skimmed milk powder in PBS) for 30 min at room temperature. To confirm the efficient removal of antibodies, the membrane was incubated with IRDye 800CW goat anti-rabbit IgG (1:20,000 in 5% [wt/vol] skimmed milk powder in PBS-T0.05) for 1 h at room temperature, washed three times with PBS-T0.05, and analyzed by using the Odyssey system. The immunoblot was then repeated as described above, but this time anti-MBP monoclonal antibody (1:10,000) and anti-WBV-M^{PRO} (1:2,000) were used as primary antibodies and IRDye 800CW goat anti-rabbit IgG and IRDye 680 goat anti-mouse IgG (both at 1:20,000) were used as secondary antibodies.

Detection of M^{PRO} in WBV-infected EPC cells. For the detection of M^{PRO} in WBV-infected cells, EPC cells were infected at a multiplicity of infection (MOI) of 10 50% tissue culture infective dose(s) (TCID₅₀) per cell. At 24 h postinfection, the cells were lysed in 1× Laemmli sample buffer, and the proteins were denatured by heating for 3 min at 94°C. The samples were purified by using the SDS-PAGE sample prep kit (Pierce) according to the manufacturer's instructions. The protein was quantified with Coomassie Plus Protein Assay Reagent (Pierce) using bovine serum albumin (BSA) as a standard. A total of 100 µg of protein per sample was separated in an 18-by-16-cm discontinuous SDS-15% polyacrylamide gel and transferred onto nitrocellulose. To ensure even running across the lanes, the M^{PRO} markers had been mixed with lysates from mock-infected cells (100 µg of cellular protein per lane). WBV M^{PRO} expressed in virus-infected cells, and bacterially expressed M^{PRO} markers were detected by immunoblotting as described above with anti-WBV-M^{PRO} (1:1,000) as the primary antibody, followed by labeling with IRDye 800CW goat anti-rabbit IgG.

HPLC-based peptide cleavage assay. The peptide conforming to the N1 cleavage site (N1 peptide, H₂N-GNTITRQ↓ALQR-COOH, high-pressure liquid chromatography [HPLC] purified to ≥98%) was purchased from Eurogentec, Inc. The reactions contained 0.5 mM peptide in 50 mM Tris-Cl, 100 mM NaCl, 1 mM DTT (pH 8.0), and 0.3 µM protease or the equivalent volume of purification buffer (negative control) in a total reaction volume of 25 µl. After incubation at 30°C for the indicated length of time, the reaction was terminated by addition of one volume 2% trifluoroacetic acid (TFA). Reaction products were resolved in a 2 to 50% acetonitrile gradient in 0.1% TFA over 5 column volumes (CV) (3 h of incubation) or 9 CV (20 h of incubation) at a flow rate of 0.5 ml per min on a µRPC C2/C18 ST 4.6/100 column connected to an ÄKTA purifier 10 system (both GE Healthcare) and detected by measuring the absorption at 215 nm (A₂₁₅).

RESULTS

Amino- and carboxyl-terminal autoprocessing activities of the WBV M^{PRO} domain. For all members of the family *Coronaviridae* analyzed to date, the M^{PRO} domain has been shown (or is predicted) to autocatalytically release itself from the downstream transmembrane domain (39, 40, 50, 62, 63). In contrast, processing immediately C-terminal of the M^{PRO} domain could not be confirmed in the rotavirus Gill-associated virus (GAV) (60). We previously showed that expression of a fusion protein comprised of MBP and WBV pp1a/pp1ab residues Gln-3424 to Gln-3726 (that is, the putative M^{PRO} domain together with short flanking sequences) resulted in autoprocessing N-terminal of the protease domain (38). In these experiments, C-terminal M^{PRO} autoprocessing at the M^{PRO} carboxyl terminus could not be characterized because the expected small size of the potential C-terminal processing product likely precluded its detection. In the present study, we therefore extended the expression construct at the carboxyl terminus. To prevent the possible cytotoxicity that is often caused by the expression of hydrophobic sequences in *E. coli*, we did not extend the fusion protein with the transmembrane domain located downstream in WBV pp1a/pp1ab but inserted

a heterologous sequence. Plasmid pMal-WBV-3CL₅₅₉₋₅₆₀ (38) was modified to contain the GST coding sequence downstream of the MBP-pp1a-3424-3725 sequence (Fig. 1A), resulting in plasmid pMal-pp1a-3424-3725-GST. When protein expression of IPTG- versus mock-induced *E. coli* cells transformed with pMal-pp1a-3424-3725-GST was analyzed by SDS-PAGE of whole-cell lysates, overexpression of the ~46-kDa N-terminal processing product could again be observed. Although an additional product with an apparent molecular mass of ~27 kDa was also detectable, we could not reliably assign its identity at this point. No third processing product could be detected. Using GST-Bind Resin (Novagen), an ~27-kDa protein was purified from *E. coli* expressing the wild-type (wt) fusion protein (Fig. 1E, lane 3). A protein of similar apparent molecular mass was also detected by immunoblot analysis with GST-specific antiserum (data not shown and see Fig. 4D). At 27 kDa, the molecular mass of the additional observed product was higher than the calculated molecular mass of the GST domain alone (26 kDa), indicating that processing had occurred within the WBV pp1a/pp1ab part of the fusion protein (~10 amino acids from the C-terminal end of the WBV sequence). The data led us to conclude that the fusion protein containing the WBV pp1a/pp1ab wt sequence 3424 to 3725 was proteolytically cleaved at both the N and the C termini of the WBV M^{PRO} domain. The failure to detect all of the expected processing products in the initial expression analysis may be due to comigration of M^{PRO}- and GST-containing products (due to their similar sizes), instability of one of the products, or masking of one of the products through comigration with *E. coli* protein(s).

Determination of the M^{PRO} domain boundaries and substrate specificity of the protease. To determine the exact location of the processing sites and obtain information on the substrate specificity of the WBV M^{PRO}, we carried out a series of affinity purifications of different processing products and determined their N termini by Edman degradation. In the first instance, we sequenced the N terminus of the purified C-terminal (GST-containing) processing product of MBP-pp1a-3424-3725-GST (see above) by five cycles of Edman degradation (data not shown). The analysis revealed that cleavage had occurred at the WBV pp1a/pp1ab sequence ₃₇₁₃TVGQ↓TLT_{S3720}. 3CL^{PRO}s commonly show a preference for either Gln or Glu at the position immediately preceding the scissile bond (commonly referred to as the P1 residue according to the nomenclature introduced by Schechter and Berger [37], while the subsequent residue is referred to as P1'). The Gln/Glu residue is often followed by a small aliphatic residue at the P1' position (reviewed in references 36 and 63). The tentatively identified substrate specificity of WBV M^{PRO} for Gln followed by Thr thus conformed well to these previous observations, lending additional support to our identification of the C-terminal processing site.

As mentioned above, the P1 Gln/Glu residue is generally the major specificity determinant of 3C-like proteases (4, 7, 17, 26, 33, 45, 46, 52, 53). Substitution of the P1 residue of the identified cleavage site, Gln-3716, with Ala in the MBP-pp1a-3424-3725-GST fusion protein could therefore be expected to abolish processing C-terminal of the M^{PRO} domain. Expression of a protein containing this specific substitution (MBP-pp1a-3424-3725_Q3716A-GST) should then allow us to purify the M^{PRO}-

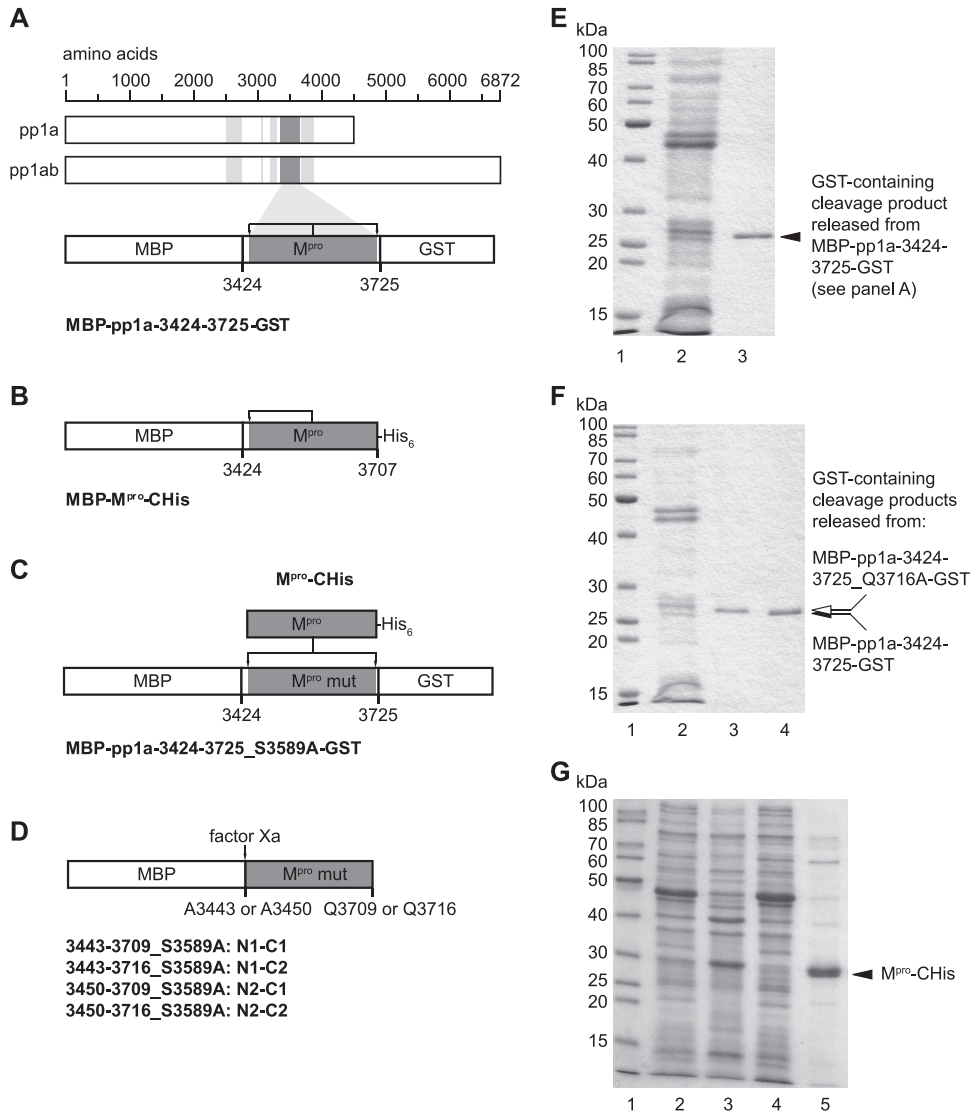


FIG. 1. (A to D) Proteins expressed in the present study. (A) Schematic showing the location of the putative M^{pro} domain (dark gray) within the pp1a/pp1ab replicase polyproteins. Also shown are the locations of the putative transmembrane domains in pp1a/pp1ab (light gray) as predicted by TMHMM (38). The putative WBV M^{pro} core domain was expressed with flanking regions (Gln-3424 to Met-3725) as a fusion with the maltose-binding protein (MBP) at its N-terminal and glutathione *S*-transferase (GST) at its C-terminal (MBP-pp1a-3424-3725-GST). This construct was presumed to contain both upstream and downstream flanking cleavage sites (indicated by arrows). (B) Expression of WBV pp1a/pp1ab residues Gln-3424 to Thr-3707 as an N-terminal fusion with MBP and containing a C-terminal hexahistidine (His₆) tag. This construct only contained the N-terminal autoprocessing site, thus allowing a single-step purification (by IMAC) of the M^{pro} domain (M^{pro}-CHis) following its intracellular autocatalytic release from the fusion protein precursor. (C) Intermolecular (*trans*) cleavage reaction of the proteolytically inactive MBP-pp1a-3424-3725_S3589A-GST fusion protein by M^{pro}-CHis. (D) To obtain molecular mass markers for the four theoretical forms of the M^{pro} domain resulting from cleavage at two alternative N- and C-terminal cleavage sites, respectively, the respective parts of the polyprotein (see also Fig. 3A) were expressed as N-terminal fusions with MBP. Cleavage of the purified fusion proteins with factor Xa yielded size markers with authentic N termini. To prevent autoprocessing at cleavage sites present in some of the proteins, the active-site Ser residue (Ser-3589) was substituted with Ala in all constructs. (E to G) Purification of processing products. (E) The MBP-pp1a-3424-3725-GST fusion protein was expressed in *E. coli* TB1 (for details, see Materials and Methods) and the C-terminal, GST-containing processing product was purified using GST-Bind resin. The cleared lysate (lane 2) and the purified protein (lane 3) were separated by SDS-PAGE and stained with Coomassie brilliant blue R-250. Lane 1: protein ladder. The filled arrowhead indicates the purified C-terminal processing product. (F) MBP-pp1a-3424-3725_Q3716A-GST was expressed in *E. coli* TB1 and the C-terminal processing product was affinity purified using GST-Bind resin. The cleared lysate (lane 2) and the purified protein (lane 3) were analyzed by SDS-PAGE and Coomassie blue staining. The C-terminal processing product of the MBP-pp1a-3424-3725-GST fusion protein purified in panel E was electrophoresed on the same gel for size comparison (lane 4). Lane 1, protein ladder. The open arrowhead indicates the purified C-terminal autoprocessing product of the MBP-pp1a-3424-3725_Q3716A-GST fusion protein; the filled arrowhead indicates the C-terminal autoprocessing product of the MBP-pp1a-3424-3725-GST fusion protein. (G) Purification of M^{pro}-CHis. The MBP-M^{pro}-CHis fusion protein was expressed in *E. coli* TB1 and the C-terminal, His-tagged processing product was purified by IMAC. Samples taken at different steps of the purification protocol were analyzed by SDS-PAGE and Coomassie blue staining. Lane 1, protein ladder; lane 2, total lysate; lane 3, insoluble protein fraction; lane 4, soluble supernatant fraction; lane 5, eluate fraction. M^{pro}-CHis is indicated by a filled arrowhead. Molecular masses (in kilodaltons) of the protein ladder proteins are given to the left for each of the gels shown.

GST fusion protein through glutathione affinity purification and determination of the protein's N terminus would reveal the N-terminal M^{Pro} autoprocessing site. Surprisingly, however, the apparent molecular mass of the purified product, as judged by SDS-PAGE, was only ~1 kDa larger than the C-terminal processing product of the MBP-pp1a-3424-3725-GST wt protein (Fig. 1F, compare lanes 3 and 4). N-terminal sequencing (data not shown) of this processing product revealed that processing had occurred at a second C-terminal processing site, ³⁷⁰⁶VTQQ ↓ TSVT₃₇₁₃, immediately N terminal of the previously identified ³⁷¹³TVGQ ↓ TLTS₃₇₂₀ processing site. Both sites possess identical residues at the P1 and P1' positions, Gln and Thr, respectively, but no further conserved features could be inferred for the two cleavage sites. In both cases, cleavage seemed to occur with high efficiency, as only the fully processed product was detectable in the eluate fraction obtained from the glutathione affinity purification. For brevity, the more N terminal of the two C-terminal M^{Pro} autoprocessing sites, ³⁷⁰⁶VTQQ ↓ TSVT₃₇₁₃, is referred to as the C1 site, and the more C terminal, ³⁷¹³TVGQ ↓ TLTS₃₇₂₀, is referred to as the C2 processing site.

To be able to purify M^{Pro} and determine its N terminus, amino acids Gln-3424 to Thr-3707 of WBV pp1a/1ab were expressed as a fusion protein with MBP. A hexahistidine tag was added at the C terminus (just upstream of the C1 processing site) to allow affinity purification of the M^{Pro} domain (Fig. 1B). The WBV pp1a/pp1ab sequence in this protein should contain the M^{Pro} core domain and the N-terminal autoprocessing site, allowing autocatalytic release and subsequent affinity purification of the hexahistidine-tagged M^{Pro} domain (M^{Pro}-CHis). The purified protein had an apparent molecular mass of ~28 kDa (Fig. 1G, lane 5). The theoretical molecular mass of the entire WBV pp1a/1ab-derived part (Gln-3424 to Thr-3707) of the fusion protein including the C-terminal hexahistidine tag is 31.3 kDa. The data thus led us to conclude that (I) WBV pp1a/1ab residues Gln-3424 to Thr-3707 are sufficient for autocatalytic processing at the N terminus and (II) processing occurred about 20 amino acids into the N-terminal region of the expressed WBV pp1a/1ab fragment.

Sequencing of the N terminus of this protein (data not shown) revealed that the band consisted in fact of two processing products. While cleavage had predominantly occurred at ³⁴³⁹ITRQ ↓ ALQR₃₄₄₆, a minor processing product, where cleavage had occurred seven amino acids further downstream, at ³⁴⁴⁶RIRQ ↓ AVTV₃₄₅₃, could also be detected. In keeping with the nomenclature introduced for the C-terminal WBV M^{Pro} processing sites (see above), the more N terminal of the two detected processing sites (³⁴³⁹ITRQ ↓ ALQR₃₄₄₆) are thus referred to as the N1 site and the more C-terminal processing site (³⁴⁴⁶RIRQ ↓ AVTV₃₄₅₃) as the N2 site.

The release of M^{Pro} from the polyprotein can occur either intramolecularly (*cis*) or intermolecularly (*trans*). To determine the substrate preferences of the WBV M^{Pro} in *trans*-cleavage reactions, we devised an assay that was suitable to determine whether M^{Pro} displays a preference for one of the two detected N- and C-terminal cleavage sites. The proteolytically inactive MBP-pp1a-3424-3725_S3589A-GST fusion protein was used as a substrate in this experiment. Heterologously expressed and affinity-purified substrate protein was incubated with purified M^{Pro}-CHis, and reaction products were analyzed

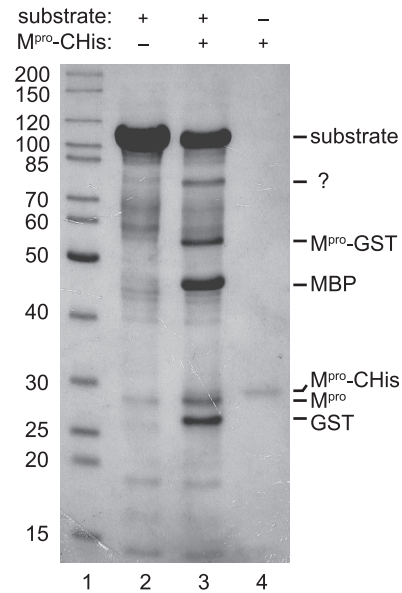


FIG. 2. *In trans* proteolytic processing of the MBP-pp1a-3424-3725_S3589A-GST fusion protein by M^{Pro}-CHis. MBP-pp1a-3424-3725_S3589A-GST was expressed in *E. coli* TB1 and affinity purified using amylose resin. The MBP-pp1a-3424-3725_S3589A-GST fusion protein was then incubated with (lane 3) or without (lane 2) M^{Pro}-CHis. Processing products were separated through SDS-PAGE and stained with Coomassie brilliant blue R-250. Lane 4, the amount of M^{Pro}-CHis equal to that present in lane 3; lane 1, protein ladder. The molecular masses (in kilodaltons) of the ladder proteins are indicated to the left of the gel.

by SDS-PAGE and Coomassie blue staining. Five major processing products could be identified (Fig. 2, lane 3). Based on the correlation of the apparent molecular masses of these products with the calculated molecular masses of theoretically possible processing products and the previously observed electrophoretic mobilities of some of these products, the four smaller products were postulated to be (I) the GST domain (and a few C-terminal residues of the WBV pp1a/1ab fragment) (~27 kDa), (II) the M^{Pro} domain (~29 kDa), (III) the MBP domain (and a few N-terminal residues of the WBV pp1a/1ab fragment) (~46 kDa) and the M^{Pro}-GST processing intermediate (~55 kDa). The largest processing product of ~80 kDa was originally thought to be the MBP-M^{Pro} processing intermediate which, as a consequence of cleavage at one of the C-terminal processing sites, lacks the C-terminal GST domain. However, the results of a subsequent immunoblot analysis led us to conclude that this product was the result of a processing event within the MBP domain (see below for details).

To determine the C-terminal M^{Pro} processing site that had been used under *trans*-cleavage reaction conditions, we subjected the 27-kDa processing product to five cycles of Edman degradation. Similarly, to determine the N-terminal M^{Pro} processing site cleaved under these conditions, we analyzed the N terminus of the 55-kDa M^{Pro}-GST processing intermediate. The M^{Pro}-GST processing intermediate was chosen instead of the M^{Pro} domain since the latter could not be sufficiently separated from the M^{Pro}-CHis included in the reaction (Fig. 2, compare lanes 3 and 4), which might have caused contamina-

tion problems in subsequent sequence analyses. The data obtained in this analysis allowed us to conclude that processing had occurred at the C1 (3706 VTOQ ↓ TSVTV₃₇₁₃) and N1 (3439 ITRQ ↓ ALQR₃₄₄₆) sites, respectively. However, upon extended incubation of the *trans*-cleavage reaction, processing at the C2 site also became evident (data not shown). In summary, the data show that WBV M^{Pro} (pp1a/pp1ab residues 3443 to 3707) is active in *trans* and preferentially cleaves the N1 and C1 sites under these conditions.

Identification of M^{Pro} in WBV-infected cells. The timely and spatially regulated release of replicase subunits is thought to be critical for plus-strand RNA virus replication, and defects in virus replication observed in coronaviruses carrying nonsynonymous mutations that affect polyprotein processing support this hypothesis (12, 14; N. Karl, T. Hertzog, and J. Ziebuhr, unpublished data). In the light of these observations, the multitude of processing sites identified in the bacterial expression system was all the more astonishing and raised the question of whether all of these processing sites are indeed processed during virus replication. To investigate M^{Pro} autoprocessing in virus-infected cells and to corroborate the autoprocessing data obtained in our bacterial expression systems, we performed another set of experiments.

As described above, M^{Pro} autoprocessing could occur at two N-terminal (N1 and N2) and two C-terminal (C1 and C2) sites, potentially resulting in four different processing products, N1-C1, N1-C2, N2-C1, and N2-C2 (Fig. 3A). To identify and distinguish these proteins in SDS-polyacrylamide gels, we devised size markers for the four theoretically possible forms of WBV M^{Pro}, N1-C1, N1-C2, N2-C1, and N2-C2. Fragments of WBV pp1a/pp1ab corresponding to the four putative forms of M^{Pro} were expressed as fusion proteins with MBP, affinity purified, and subsequently cleaved with factor Xa to release the desired M^{Pro} fragments without any additional, vector-derived sequences. The following WBV pp1a/1ab fragments were expressed: pp1a/1ab Ala-3443 to Gln-3709 for the N1-C1 marker, Ala-3443 to Gln-3716 for the N1-C2 marker, Ala-3450 to Gln-3709 for the N2-C1 marker, and Ala-3450 to Gln-3716 for the N2-C2 marker. To prevent autoprocessing at processing sites present in some of these proteins, all marker proteins contained the active-site Ser-3589-to-Ala substitution.

The electrophoretic mobility of M^{Pro} detected in WBV-infected EPC cells was compared to that of the size markers described above in an immunoblot with an M^{Pro}-specific antiserum. Two proteins with slightly different molecular masses were detected in infected but not mock-infected cells (Fig. 3B). The two proteins comigrated with the N1-C1 (open arrowhead) and N2-C1 (filled arrowhead) size markers, respectively. The data lead us to conclude that the M^{Pro} domain accumulates in two forms in virus-infected cells. Direct sequencing of the N termini and/or mass spectrometric analysis would be required to unequivocally confirm the identities of these two proteins. These types of analyses would require the isolation of the two forms from WBV-infected cells in comparatively high amounts and purity. Only limited options of flanking processing sites conforming to the detected substrate specificity of M^{Pro} and even more so products of fitting molecular mass resulting from processing at this site are conceivable. Based on the comigration pattern with marker proteins it seems reasonable to suggest that the two forms of M^{Pro} are produced by

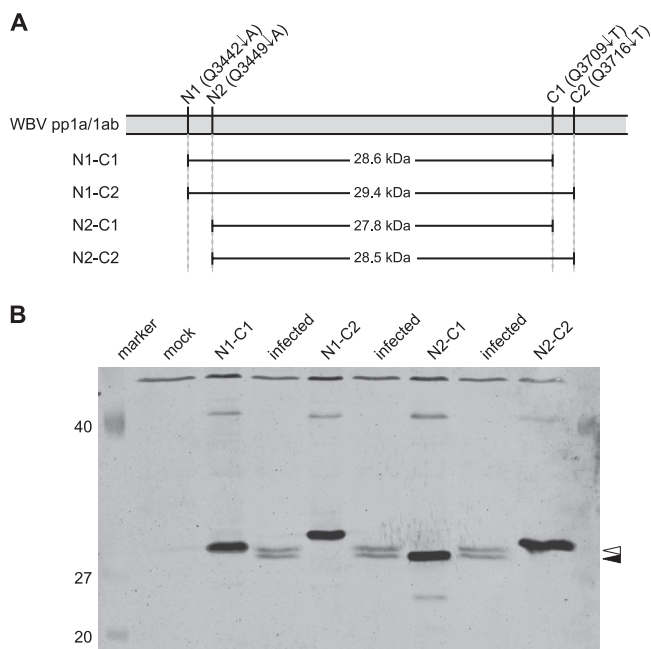


FIG. 3. Western blot analysis of M^{Pro} forms expressed in WBV-infected cells. (A) Schematic representation of the four putative forms of the M^{Pro} domain. The location of the two upstream (N1 and N2) and the two downstream (C1 and C2) cleavage sites identified after expression of the putative M^{Pro} and flanking sequences in *E. coli* are depicted. Numbers refer to the amino acid position in the WBV pp1a/pp1ab. Depending on the cleavage sites used, four forms of the M^{Pro} domain would be possible (N1-C1, N1-C2, N2-C1, and N2-C2). The theoretical molecular mass of each form is indicated. (B) EPC cells were infected with WBV at an MOI of 10 TCID₅₀ per cell or mock infected. At 24 h postinfection, lysates were prepared. These were separated in a 15% discontinuous SDS-polyacrylamide gel alongside heterologously expressed molecular mass markers that conform to the four theoretical forms of the M^{Pro} domain: N1-C1, N1-C2, N2-C1, and N2-C2. M^{Pro} present in infected cells and the M^{Pro} marker proteins were detected by immunoblotting with rabbit anti-WBV-M^{Pro} antiserum. The molecular masses of the protein ladder proteins (in kilodaltons) are indicated to the left. Two forms of M^{Pro} were detected in WBV-infected cells (indicated by arrowheads).

C-terminal processing at the C1 site and N-terminal processing at the N1 or N2 site. No product resulting from processing at the C2 site was detectable.

Mutagenesis analysis of residues predicted to be involved in catalysis and substrate specificity. Cellular and viral chymotrypsin-like serine proteases rely on a Ser-His-Asp catalytic triad (reviewed in reference 24). To analyze whether this characteristic is conserved in WBV M^{Pro} and to identify other conserved residues important for protease activity, the sequence of WBV M^{Pro} was compared to that of homologs from astroviruses and toroviruses, which we identified in database searches as the closest known relatives of the WBV M^{Pro}. We also included arterivirus main proteases in our analysis, because the active-site residues of these related nidovirus serine proteases have been characterized in detail in previous mutagenesis and structural studies. WBV pp1a/pp1ab residues Ser-3589, His-3492, and Asp-3518 (Fig. 4A) could be aligned with the active-site Ser, His, and Asp/Glu residues of other 3C-like proteases (19, 29, 41, 42, 46, 63), although sequence conservation was largely confined to the areas surrounding the

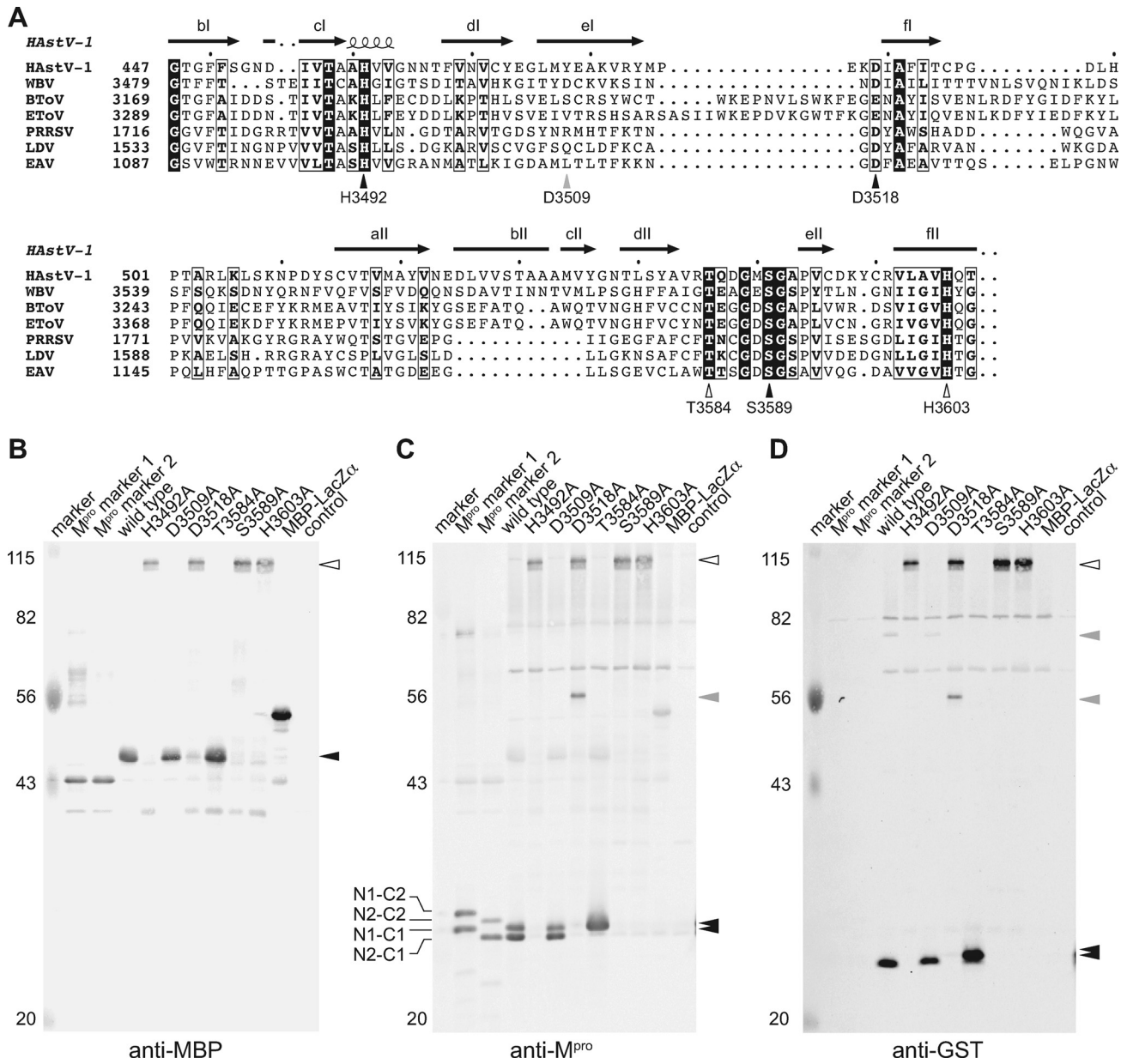


FIG. 4. Putative WBV M^{pro} active-site residues. (A) Multiple sequence alignment of astro-, arteri-, toro-, and bafinivirus serine proteases. Partial protease sequences were aligned by using ClustalX 2.0 (30) and rendered with ESPrnt 2.2 (<http://esprnt.ibcp.fr/ESPrnt/cgi-bin/ESPrnt.cgi>) (22). Abbreviations of virus names and accession numbers are as follows: HAStV-1, human astrovirus 1 (AAW51880); EToV, equine torovirus (X52374); BToV, bovine torovirus (AY427798); PRRSV, porcine reproductive and respiratory syndrome virus (Q04561); LDV, lactate dehydrogenase-elevating virus (NC_02534); EAV, equine arteritis virus (NP_127506). Secondary structure elements of the astrovirus M^{pro} are shown above the sequence (PDB code: 2W5E). These were extracted from the PDB file and added to the alignment using ESPrnt 2.2 and manually adjusted to conform to the numbering used by Speroni et al. (42). Conserved residues that were identified as possible catalytic or S1 subsite residues based on data available for the corresponding residues in astrovirus and nidovirus homologs (see the text for details) are marked with closed or open arrowheads, respectively. The gray arrowhead marks the (nonconserved) Asp residue that was used as a negative control in this mutagenesis study. Residues were numbered according to their position in the respective viral polyprotein sequence. (B to D) Autoprocessing activities of mutant forms of the WBV M^{pro} with substitutions of putative active-site residues. MBP-pp1a-3424-3725-GST and mutant proteins were expressed in *E. coli* TB1. After induction of expression for 3 h at 25°C, total cell lysates were prepared and separated in a 15% discontinuous SDS-polyacrylamide gel and transferred onto nitrocellulose. GST-containing unprocessed precursor proteins, processing intermediates, and processing end products were detected by Western blot analysis with rabbit anti-GST antiserum (D) as the primary antibody and IRDye800CW anti-rabbit IgG as the secondary antibody. After imaging, bound antibodies were removed. MBP- and M^{pro}-containing proteins were subsequently detected by using mouse anti-MBP monoclonal antibody (B) and rabbit anti-WBV M^{pro} antiserum (C), followed by labeling with IRDye680 anti-mouse IgG and IRDye 800CW anti-rabbit IgG, respectively. Images were acquired by using an LI-COR Odyssey system and software. Wild-type and mutant proteins analyzed in this experiment are indicated above the lanes. MBP-LacZ α , *E. coli* TB1 transformed with pMal-c2X and induced with 1 mM IPTG; control, untransformed and mock induced *E. coli* TB1; M^{pro} marker 1, M^{pro} markers N1-C1 and N1-C2; M^{pro} marker 2, M^{pro} markers N2-C1 and N2-C2. The molecular masses of the marker proteins (in kilodaltons) and the identities of the M^{pro} marker proteins (in panel C) are indicated to the left. Black arrowheads, processing end products; gray arrowheads, processing intermediates; open arrowhead, unprocessed precursor.

putative active-site triad described above. Two conserved residues, a His and a Thr/Ser, the latter located five residues upstream of the active site Ser/Cys, have been implicated in binding the substrate P1-Gln/Glu, thereby contributing to the partially conserved P1 specificity of picornavirus 3C and related viral proteases (7, 31–33, 41, 60, 64). These two residues, His-3603 and Thr-3584, were found to be conserved in the WBV M^{Pro} sequence.

The functional significance of putative catalytic and P1-binding residues for WBV M^{Pro} activity was assessed by characterizing the proteolytic activities of mutant MBP-pp1a-3424-3725-GST fusion proteins carrying specific Ala replacements of predicted active-site residues. Unprocessed precursor proteins and cleavage products were identified in total bacterial lysates by immunoblotting with antibodies specific for MBP (Fig. 4B), WBV M^{Pro} (Fig. 4C), and GST (Fig. 4D), respectively. Consistent with previously reported data and the predicted role of Ser-3589 as the M^{Pro} active-site nucleophile (38), unprocessed full-length fusion protein with an apparent molecular mass of 101 kDa could be detected in lysates obtained from *E. coli* TB1 cells expressing the S3589A mutant protein in all three immunoblots (Fig. 4B, C, and D, open arrowhead). In contrast, when the MBP-pp1a-3424-3725-GST fusion protein containing the wild-type WBV M^{Pro} sequence was expressed, proteolytic processing was observed. Thus, an MBP-containing N-terminal processing product of ~46 kDa (Fig. 4B, filled arrowhead) could be detected using anti-MBP antibody and two forms of the M^{Pro} domain equivalent to those identified in WBV-infected cells (that is, the N1-C1 and N2-C1 forms) could be detected using the M^{Pro}-specific antiserum (Fig. 4C, filled arrowheads). Furthermore, a C-terminal processing product of ~27 kDa could be detected using the GST-specific antiserum (Fig. 4D). In the latter case, an additional minor product of ~80 kDa could be specifically detected. A protein of the same size had also been detected in the M^{Pro}-CHis-mediated *trans*-cleavage reaction of the MBP-pp1a-3424-3725_S3589A-GST mutant protein reported above (Fig. 2, lane 3). This minor processing product was not detected by the anti-MBP antibody specific for an epitope within the N-terminal 75 amino acids of MBP (NEB technical support, unpublished data), confirming that the protein is a C-proximal rather than N-terminal processing product. Although the identity of this minor processing product was not determined unambiguously, its apparent molecular mass and immunological reactivity led us to suggest that the protein was produced by cleavage within the MBP sequence, which is irrelevant for pp1a/pp1ab processing in infected cells and, therefore, was not studied further.

When the putative active site His-3492 was substituted with Ala, only unprocessed precursor was detected (Fig. 4B, C, and D), supporting the proposed critical role of His-3492 in M^{Pro} activity, where the residue is predicted to act as a general base during catalysis. Similarly, unprocessed fusion protein was detected when the predicted acidic residue of the catalytic triad, Asp-3518, was substituted with Ala (Fig. 4B, D3518A). In this case, however, a small amount of the N-terminal processing product could also be detected. In addition, a protein with an apparent molecular mass of ~56 kDa was detected in these lysates using the anti-M^{Pro} or the anti-GST antiserum (Fig. 4C and D, gray arrowhead). The apparent molecular mass, as well as the antigenic properties, of the protein led us to conclude

that this protein is the M^{Pro}- and GST-containing processing intermediate resulting from processing of the full-length precursor at one (or both) of the N-terminal M^{Pro} autoprocessing sites. In contrast, substitution of Asp-3509 with Ala, which we used as a control in this experiment, did not impair proteolytic processing. We therefore concluded that Asp-3518 has a role in proteolytic activity and likely represents the acidic residue of the catalytic triad. Interestingly, substitution of this residue did not completely abolish activity and differentially affected the proteolytic processing at specific sites. Thus, processing at (at least one) N-terminal processing site did still occur (albeit relatively inefficiently), whereas virtually no C-terminal autoprocessing activity was detected. Together, the data lead us to suggest that the conserved Asp-3518 residue has an important but nonessential role in proteolytic activity, at least under these *in vitro* conditions. To further corroborate these conclusions, the potential effects of the Asp-3518-to-Ala substitution on protease activity were studied in a peptide-based *trans*-cleavage assay. The desired substitution was introduced into the pMal-WBV-M^{Pro}-CHis plasmid by site-directed mutagenesis and the mutant protein was expressed in *E. coli* TB1 and purified by IMAC. When analyzed by SDS-PAGE (Fig. 5A), the preparation of wt protein was found to contain two proteins with apparent molecular masses of ~29 and ~28 kDa, resulting from differential processing at either of the two N-terminal processing sites N1 and N2. In contrast, the preparation of the D3518A mutant protein contained only the larger (~29 kDa) form. In addition, small amounts of a protein with a molecular mass corresponding to that of the unprocessed fusion protein were present in the eluate fraction. Taken together, these observations confirmed that autoprocessing is impaired in the D3518A mutant. The expressed proteins, M^{Pro}-CHis wt and D3518A, were then used in a peptide-based cleavage assay (for details, see Materials and Methods) using a peptide whose sequence corresponded to the N1 processing site (N1 peptide, H₂N-GNTITRQ ↓ ALQR-COOH). No cleavage of the substrate peptide could be detected (even after overnight incubation) with the D3518A mutant protein (Fig. 5C), whereas wt protein digested the peptide to completion under equivalent conditions. The data confirmed the critical role for Asp-3518 in proteolytic activity. Substitution of this residue with Ala resulted in an overall reduced activity but differentially affected the processing at individual sites. In a bacterial expression system, processing at either of the C-terminal processing sites and the N2 site was more severely affected than processing at the N1 site. Peptide cleavage data confirmed and extended these conclusions by showing that, in contrast to the wt protein, the D3518A mutant was virtually inactive at the N1 processing site when tested in an *in vitro trans*-cleavage assay.

The effect of substitutions of the putative P1-Gln/Glu-binding residues, His-6303 and Thr-3584, with Ala was also analyzed in this experiment. No autoprocessing activity could be detected in the His-3603-to-Ala substitution mutant (Fig. 4B, C, and D), leading us to conclude that this residue, in accordance with our hypothesis, is crucial for proteolytic activity, likely by contributing to substrate binding. In contrast, no unprocessed precursor was detected in lysates of *E. coli* TB1 cells expressing the T3584A mutant protein, demonstrating that the role of this conserved residue in proteolysis is less critical.

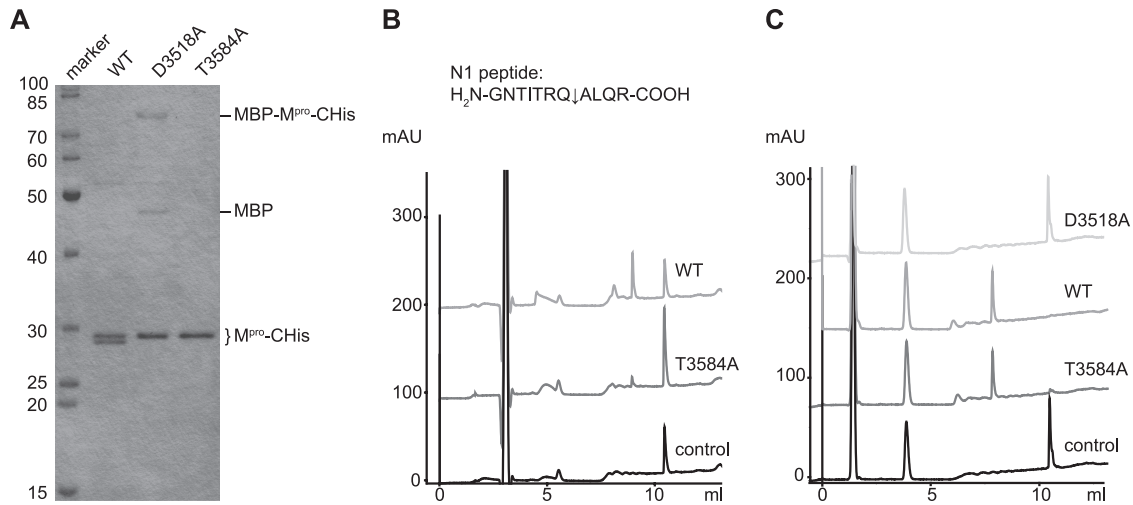


FIG. 5. Purification and proteolytic activities of M^{pro}-CHis wild-type and mutant proteins. (A) M^{pro}-CHis wild-type and mutant proteins D3518A and T3584A were expressed in *E. coli* and purified by IMAC. One microgram of each of the indicated proteins was electrophoresed in an SDS-15% polyacrylamide gel and stained with Coomassie brilliant blue R-250. The molecular masses (in kilodaltons) of marker proteins are given to the left. (B to C) A synthetic peptide (0.5 mM final concentration) whose sequence corresponded to the previously determined N1 cleavage site (N1 peptide) was incubated with 0.3 μ M M^{pro}-CHis wild-type or mutant protein or no protein (control) for 3 h (B) or 20 h (C). The reaction products were analyzed by reversed-phase chromatography. The y axis conforms to the control reaction, all other chromatograms are depicted as shifted up by 100 mAU (B) or 70 mAU (C) with respect to the one below.

However, we only detected the larger N1-C1 form of M^{pro} in lysates derived from *E. coli* cells expressing the T3584A mutant protein (Fig. 4C), while the smaller N2-C1 form, which was readily apparent in cells expressing the wt construct, was not detectable in cells expressing the mutant protein. Furthermore, the predominant C-terminal GST-containing processing product detected in lysates of *E. coli* expressing the T3584A mutant protein was of marginally lower electrophoretic mobility than that detected in the wt preparation (Fig. 4D). This could possibly be due to (predominant) processing at the C1 but not the C2 site. However, due to the minor differences in the molecular masses of the two proteins and the limited resolution of proteins of this size in SDS-PAGE, we cannot exclude that some processing at the C2 site did occur. The effect of the Thr-3584-to-Ala substitution on *trans*-processing activity was also tested. The M^{pro}-CHis T3584A mutant protein was expressed and purified as described above for the M^{pro}-CHis D3518A mutant protein. In contrast to the purified wt protein, only the larger form of M^{pro} resulting from processing at the N1 site was present in the eluate fraction of the T3584A mutant protein (Fig. 5A). This is in good agreement with the changes in N-terminal processing previously observed after expression of the MBP-pp1a-3424-3725_T3584A-GST mutant protein (Fig. 4C). When tested in the peptide-based *trans*-cleavage assay using N1 peptide as substrate, the activity of the T3584A mutant protein was markedly reduced compared to wt protein (Fig. 5B). However, some processing of the peptide could still be observed and, after overnight incubation, T3584A mutant protein digested the peptide to near completion (Fig. 5C). Thus, under the conditions used, the activity of the T3584A mutant protein was reduced compared to wt but less affected than that of the D3518A mutant discussed above.

DISCUSSION

The study revealed important new information on the auto-processing activities of an M^{pro} from bafiniviruses. We were able to show that WBV M^{pro} cleaves the polyprotein at N- and C-proximal sites to release (at least) two M^{pro}-containing processing products. Based on comparative sequence analysis data (Fig. 4A and data not shown) and crystal structure information available for 3C and 3C-like proteases (2, 3, 7, 31, 32, 42), it is reasonable to suggest that the N-proximal two-thirds of the WBV M^{pro} have a chymotrypsin-like two- β -barrel fold that is extended at the C terminus by an extra domain. Secondary structure predictions generated by using the Jpred 3, Predict-Protein, and I-TASSER servers (9, 34, 35, 56) consistently suggested that the C-terminal extra domain has a helical structure comprised of at least three α -helices (data not shown). With about 80 residues, the size of the C-terminal domain is larger than that of the arterivirus and smaller than that of the coronavirus homologs.

The data presented here suggest that pp1a/pp1ab residues Ser-3589, His-3492, and Asp-3518 residues constitute a prototypical Ser-His-Asp catalytic triad that is conserved in many cellular and viral CHL^{pro}s. Although the conserved Ser and His residues proved to be essential for activity, replacement of Asp-3518 with Ala resulted in reduced but clearly detectable activity, with cleavage at different sites being affected to a varying extent (Fig. 4B to D). The data suggest a supportive (rather than essential) role for this residue in catalysis, most likely by positioning the active-site His in the required orientation and neutralizing the developing positive charge of the His in the transition state (5, 6, 10, 15, 16, 43). Similar data have previously been reported for related chymotrypsin-like proteases, including EAV nsp4 (Asp-1129) and EToV M^{pro} (Glu-3347) (39, 41). For example, partial processing defects at

selected sites had also been seen in EAV nsp4 mutant proteins with equivalent Asp replacements. Although further studies, including the elucidation of the three-dimensional structure of the WBV M^{pro}, are needed to unequivocally confirm the role of the Asp residue in catalysis, the data provide strong evidence to propose that Asp-3518 is the third member of a Ser-His-Asp catalytic triad. With respect to the catalytic triad, the WBV M^{pro} is most similar to the arterivirus nsp4, which also uses a Ser-His-Asp triad, whereas the more closely related toroviruses appear to use an M^{pro} with a catalytic triad in which a Glu residue assumes the role of the third catalytic residue (7, 39, 41, 46). In marked contrast to the Ser-His-Asp/Glu catalytic triad used by arteri-, toro-, and bafinivirus M^{pro}s, the M^{pro}s of coronaviruses and roniviruses rely on a Cys-His catalytic dyad (3, 4, 52, 54, 60, 61).

The data suggest a critical involvement of His-3603 and Thr-3584 in mediating specificity for Gln at the P1 position of WBV M^{pro} cleavage sites as revealed in the present study (Fig. 4B, C, and D). Equivalent residues in other Gln/Glu-specific CHL^{pro}s were previously shown to be located in the S1 subsite, where their side chains form hydrogen bonds to the carboxamide/carboxylate moiety of the P1-Gln/Glu residue (31, 32, 41, 42). Our data strongly support the proposed critical role of His-3603 in substrate recognition and are in agreement with mutational analyses reported earlier for other 3C and 3C-like proteases (25, 27, 41). In the case of the T3584A mutant, proteolytic processing was not completely abolished. Although processing defects were readily detectable for the N2 and C2 sites (Fig. 4C and D and Fig. 5), cleavage at other sites was not evidently affected. Similar differential defects caused by substitutions of the two S1 subsite residues were previously seen in other Gln/Glu-specific CHL^{pro}s. Thus, whereas substitution of the equivalent His in EAV nsp4, pp1a/pp1ab His-1198, abolished autoproteolytic processing (see above), substitution of the corresponding Thr, Thr-1179, caused less profound and varying functional defects, depending on the nature of the substituting residue and the processing site (41). Similarly, poliovirus 3C^{pro} Thr-142 mutants displayed partial processing defects at specific sites, whereas His-161 mutants proved to be completely inactive (8, 27).

In the course of the present study, we identified a total of four (two N-terminal and two C-terminal) M^{pro} autoprocessing sites through N-terminal sequencing of relevant cleavage products. The use of three of these sites [two N-terminal and one C-terminal site(s)] during virus replication could be confirmed by studying M^{pro} expression in WBV-infected cells (Fig. 3), providing support for the biological relevance of the processing data obtained in bacterial expression systems. All cleavage sites identified in the present study conformed to the consensus Q↓(A,T), thus tentatively identifying the substrate specificity of the WBV M^{pro}. The vast majority of viral 3C and 3C-like proteases share a specificity for Gln or Glu at the P1 position and, apparently, the WBV M^{pro} is no exception to this general rule. Only a few exceptions have been described, most notably the M^{pro} of *Human coronavirus HKU-1* and the NIa protease of *Sweet potato mild mottle virus* (family *Potyviridae*) NIa, which are able to cleave after His (1, 57). To identify and determine the role of other residues possibly contributing to WBV M^{pro} substrate specificity, further studies are required, including, for example, the experimental identification of other WBV M^{pro}

cleavage sites in the viral polyproteins, site-directed mutagenesis studies, and protease assays using libraries of peptides with single substitutions of residues flanking the scissile bond on either side.

To our knowledge, the use of alternative N- and C-terminal M^{pro} autoprocessing sites has no precedent in other RNA virus systems, where proteolytic processing is generally thought to occur in a tightly controlled and spatially and temporally regulated manner. We therefore sought to further corroborate our biochemical data and studied M^{pro} autoprocessing during viral replication. The data we obtained suggest that proteolytic cleavage had occurred at (at least) three of the previously identified sites during viral replication in EPC cells, giving rise to two major forms of M^{pro} that differed in their N terminus (Fig. 3B). The data are consistent with information derived from our heterologous expression experiments and support the biological significance of the two WBV M^{pro} isoforms identified here.

Processing at both N-terminal M^{pro} processing sites was observed upon heterologous expression of the MBP-pp1a-3424-3725-GST fusion protein and in WBV-infected cells. In both experiments, processing could be the result of cleavage in *cis* or *trans* or a combination of both reaction modes. Heterologously expressed M^{pro}CHis is able to efficiently process the N1 site in *trans*-cleavage assays (Fig. 2 and Fig. 5B and C). No processing of the N2 site could be detected in *trans*-cleavage assays (Fig. 2 and data not shown), although we cannot fully exclude that processing occurred at a low level below the detection limit of our assay. Thus, it seems reasonable to suggest that processing at the N2 site primarily occurs in *cis*, although further experiments will be required to rigorously establish the processing mode at either site. At present, based on the limited M^{pro} processing data obtained in *E. coli* and WBV-infected cells, it is not possible to answer the question of whether cleavage at the two N-terminal sites occurs alternatively or consecutively. To date, we have been unable to separate N2-C2 from N1-C2 and determine their respective activities independently *in vitro*.

Both C-terminal M^{pro} processing sites were processed very efficiently in the MBP-pp1a-3424-3725-GST fusion protein expressed in *E. coli* (Fig. 1E and F and Fig. 4C and D). Notably, efficient processing at the C1 site in addition to efficient processing at the C2 site could be confirmed for the MBP-pp1a-3424-3725-GST wt protein in the immunoblot experiments (Fig. 4C), excluding the possibility that cleavage at the C1 site originally observed in the MBP-pp1a-3424-3725_Q3716A-GST mutant protein was due to (or even dependent on) the Gln-to-Ala substitution at the C2 site. However, in reactions that allow cleavage only in *trans*, the C2 site appears to be cleaved slightly less efficiently than the C1 site (Fig. 2). On the basis of the available data, different scenarios for C-terminal processing of the WBV M^{pro} domain are conceivable. Thus, cleavage might occur (I) first at the C1 site and subsequently at the C2 site by *trans* cleavage of the released C-terminal fragment, (II) first at the C2 site and subsequently within the M^{pro}-containing processing intermediate at the C1 site (in *cis* or *trans*) or (III) with no pronounced preference for either of the two sites. Although none of these possibilities can be formally excluded on the basis of the available data, we currently favor a model in which, with regard to C-terminal processing, the C1 site is

cleaved first. This is supported by the following observations. First, no M^{PRO}-containing processing intermediates carrying a C2 terminus could be detected for either the bacterially expressed MBP-pp1a-3424-3725-GST fusion protein or in WBV-infected cells. Second, M^{PRO}-CHis-mediated processing of the MBP-pp1a-3424-3725_S3589A-GST fusion protein in *trans* primarily led to processing at the C1 site, whereas processing at the C2 site was only observed after extended incubation. Third, in *E. coli* TB1 cells expressing the T3584A substitution mutant, processing occurred almost exclusively at the C1 site (Fig. 4), further supporting the idea that the C1 site is cleaved more efficiently than the C2 site. However, since the GST-containing, C1-cleaved product was produced by a mutant protease (T3584A) carrying a substitution of a presumed substrate-binding residue, we cannot formally exclude the possibility that inefficient cleavage at the C2 site was caused by subtle changes in substrate specificity rather than a generally reduced activity.

One of the questions that remain to be answered is whether or not processing at the C2 site occurs during virus replication. While the detection of two stable forms of M^{PRO}, N1-C1 and N2-C1, in WBV-infected cells is in perfect agreement with the data obtained for the bacterially expressed MBP-pp1a-3424-3725-GST fusion protein (where efficient processing had also occurred at the C2 site), the available data do not allow us to confirm or exclude that the C2 processing site in pp1a/pp1ab is indeed processed *in vivo*.

As outlined above, the data suggest that two forms of M^{PRO} are produced in WBV-infected cells. To our knowledge, this is the first report of alternative autoprocessing of a nidovirus protease domain to date. Previously, the use of two alternative M^{PRO} cleavage sites has been reported only for the nsp7/8 junction in *Mouse hepatitis virus A59* (MHV-A59). Both of these sites appear to be processed during MHV-A59 replication, but their individual significance remains elusive (11). Our *in vitro* assays were necessarily based on a mixture of both M^{PRO} forms and this mixture showed clear *trans*-processing activity toward both protein and peptide substrates. For M^{PRO}s of coronaviruses, it is known that the addition or deletion of amino acid residues at the N terminus severely decreases activity by disrupting critical interactions between the two protomers in the coronavirus M^{PRO} dimer, with the N terminus of one protomer contributing to the formation of a functional S1 site in the other protomer (49, 51, 52, 54). It is therefore conceivable that processing of the WBV M^{PRO} N2 site affects the enzyme's activity and/or specificity, a possibility that remains to be explored in future studies. In contrast to the coronavirus homologs, the M^{PRO}s of the arteriviruses EAV and *Porcine reproductive and respiratory syndrome virus* (PRRSV) were reported to be monomeric, both in solution and in crystals of these proteins. Also in the crystal structures reported for these enzymes, the N termini were found to be located well away from the active site (7, 46). The multimerization state of the M^{PRO} of WBV and the more closely related toroviruses and their structures remain to be determined but, clearly, the unique observation of two forms of M^{PRO} being produced during virus replication warrants further studies, including the use of reverse-genetic approaches, to establish possible roles for the two forms of M^{PRO} in the viral life cycle.

In conclusion, the study provides the first detailed characterization of a replicase gene-encoded protein from WBV, the

prototype virus of the newly established genus *Bafinivirus*. Although the data obtained here are generally consistent with WBV being a virus most closely related to toroviruses, they confirm our previous notion that WBV and yet-to-be-identified bafiniviruses diverged profoundly from other nidoviruses, including toroviruses and extending to the most conserved viral enzymes. Our study reveals a remarkable functional and structural variability of nidovirus main proteases, whose functional implications remain to be investigated in future studies.

ACKNOWLEDGMENTS

This study was supported by the Department of Employment and Learning (Northern Ireland) through its "Strengthening the All-Ireland Research Base," Queen's University Belfast, and the Deutsche Forschungsgemeinschaft (ZI 618/4).

We thank Barbara Schelle and Beatrix Loth for excellent technical assistance.

REFERENCES

- Adams, M. J., J. F. Antoniw, and F. Beaudoin. 2005. Overview and analysis of the polyprotein cleavage sites in the family *Polyviridae*. *Mol. Plant Pathol.* **6**:471–487.
- Allaire, M., M. M. Cherniaia, 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.
- 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.
- Anand, K., J. Ziebuhr, P. Wadhvani, J. R. Mesters, and R. Hilgenfeld. 2003. Coronavirus main proteinase (3CL^{PRO}) structure: basis for design of anti-SARS drugs. *Science* **300**:1763–1767.
- Ash, E. L., J. L. Sudmeier, E. C. De Fabo, and W. W. Bachovchin. 1997. A low-barrier hydrogen bond in the catalytic triad of serine proteases? Theory versus experiment. *Science* **278**:1128–1132.
- Bachovchin, W. W. 1986. 15N NMR spectroscopy of hydrogen-bonding interactions in the active site of serine proteases: evidence for a moving histidine mechanism. *Biochemistry* **25**:7751–7759.
- 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.
- Blair, W. S., J. H. Nguyen, T. B. Parsley, and B. L. Semler. 1996. Mutations in the poliovirus 3CD proteinase S1-specificity pocket affect substrate recognition and RNA binding. *Virology* **218**:1–13.
- Cole, C., J. D. Barber, and G. J. Barton. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**:W197–W201.
- Craik, C. S., S. Rocznik, C. Largman, and W. J. Rutter. 1987. The catalytic role of the active site aspartic acid in serine proteases. *Science* **237**:909–913.
- Deming, D. J., R. L. Graham, M. R. Denison, and R. S. Baric. 2007. Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication. *J. Virol.* **81**:10280–10291.
- Donaldson, E. F., R. L. Graham, A. C. Sims, M. R. Denison, and R. S. Baric. 2007. Analysis of murine hepatitis virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in polyprotein processing. *J. Virol.* **81**:7086–7098.
- Enjuanes, L., A. E. Gorbalenya, R. J. de Groot, J. A. Cowley, J. Ziebuhr, and E. J. Snijder. 2008. *Nidovirales*, p. 419–430. In B. W. J. Mahy and M. H. V. van Regenmortel (ed.), *Encyclopedia of virology*, 3rd ed., vol. 3. Elsevier, Oxford, England.
- Fang, S., H. Shen, J. Wang, F. P. Tay, and D. X. Liu. 2010. Functional and genetic studies of the substrate specificity of coronavirus infectious bronchitis virus 3C-like proteinase. *J. Virol.* **84**:7325–7336.
- Fodor, K., V. Harmat, R. Neutze, L. Szilagyi, L. Graf, and G. Katona. 2006. Enzyme:substrate hydrogen bond shortening during the acylation phase of serine protease catalysis. *Biochemistry* **45**:2114–2121.
- Frey, P. A., S. A. Whitt, and J. B. Tobin. 1994. A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* **264**:1927–1930.
- Gayathri, P., P. S. Satheshkumar, K. Prasad, S. Nair, H. S. Savithri, and M. R. Murthy. 2006. Crystal structure of the serine protease domain of Sesbania mosaic virus polyprotein and mutational analysis of residues forming the S1-binding pocket. *Virology* **346**:440–451.
- Gorbalenya, A. E. 2008. Genomics and evolution of the *Nidovirales*, p. 15–28. In S. Perlman, T. Gallagher, and E. J. Snijder (ed.), *Nidoviruses*. ASM Press, Washington, DC.

19. **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.
20. **Gorbalenya, A. E., L. Enjuanes, J. Ziebuhr, and E. J. Snijder.** 2006. *Nidovirales*: evolving the largest RNA virus genome. *Virus Res.* **117**:17–37.
21. **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.
22. **Gouet, P., E. Courcelle, D. I. Stuart, and F. Metz.** 1999. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**:305–308.
23. **Granzow, H., F. Weiland, D. Fichtner, H. Schütze, A. Karger, E. Mundt, B. Dresenkamp, P. Martin, and T. C. Mettenleiter.** 2001. Identification and ultrastructural characterization of a novel virus from fish. *J. Gen. Virol.* **82**:2849–2859.
24. **Hedstrom, L.** 2002. Serine protease mechanism and specificity. *Chem. Rev.* **102**:4501–4524.
25. **Hegyí, 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.
26. **Hegyí, A., and J. Ziebuhr.** 2002. Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **83**:595–599.
27. **Ivanoff, L. A., T. Towatari, J. Ray, B. D. Korant, and S. R. Petteway, Jr.** 1986. Expression and site-specific mutagenesis of the poliovirus 3C protease in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **83**:5392–5396.
28. **Jones, D. H., and B. H. Howard.** 1991. A rapid method for recombination and site-specific mutagenesis by placing homologous ends on DNA using polymerase chain reaction. *Biotechniques* **10**:62–66.
29. **Kiang, D., and S. M. Matsui.** 2002. Proteolytic processing of a human astrovirus nonstructural protein. *J. Gen. Virol.* **83**:25–34.
30. **Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins.** 2007. CLUSTAL W and CLUSTAL X version 2.0. *Bioinformatics* **23**:2947–2948.
31. **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, and S. Worland.** 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.
32. **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.
33. **Phan, J., A. Zdanov, A. G. Evdokimov, J. E. Tropea, H. K. Peters III, R. B. Kapust, M. Li, A. Wlodawer, and D. S. Waugh.** 2002. Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* **277**:50564–50572.
34. **Rost, B., G. Yachdav, and J. Liu.** 2004. The PredictProtein server. *Nucleic Acids Res.* **32**:W321–W326.
35. **Roy, A., A. Kucukural, and Y. Zhang.** 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**:725–738.
36. **Ryan, M. D., and M. Flint.** 1997. Virus-encoded proteinases of the picornavirus supergroup. *J. Gen. Virol.* **78**:699–723.
37. **Schechter, I., and A. Berger.** 1967. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**:157–162.
38. **Schütze, H., R. Ulferts, B. Schelle, S. Bayer, H. Granzow, B. Hoffmann, T. C. Mettenleiter, and J. Ziebuhr.** 2006. Characterization of *White bream virus* reveals a novel genetic cluster of nidoviruses. *J. Virol.* **80**:11598–11609.
39. **Smits, S. L., E. J. Snijder, and R. J. de Groot.** 2006. Characterization of a torovirus main proteinase. *J. Virol.* **80**:4157–4167.
40. **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.
41. **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.
42. **Speroni, S., J. Rohayem, S. Nenci, D. Bonivento, I. Robel, J. Barthel, V. B. Luzhkov, B. Coutard, B. Canard, and A. Mattevi.** 2009. Structural and biochemical analysis of human pathogenic astrovirus serine protease at 2.0 Å resolution. *J. Mol. Biol.* **387**:1137–1152.
43. **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.
44. **Tautz, N., A. Kaiser, and H. J. Thiel.** 2000. NS3 serine protease of bovine viral diarrhoea virus: characterization of active site residues, NS4A cofactor domain, and protease-cofactor interactions. *Virology* **273**:351–363.
45. **Thiel, V., K. A. Ivanov, A. Putics, T. Hertzog, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, and J. Ziebuhr.** 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**:2305–2315.
46. **Tian, X., G. Lu, F. Gao, H. Peng, Y. Feng, G. Ma, M. Bartlam, K. Tian, J. Yan, R. Hilgenfeld, and G. F. Gao.** 2009. Structure and cleavage specificity of the chymotrypsin-like serine protease (3CLSP/nsp4) of porcine reproductive and respiratory syndrome virus (PRRSV). *J. Mol. Biol.* **392**:977–993.
47. **Ulferts, R., I. Imbert, B. Canard, and J. Ziebuhr.** 2010. Expression and functions of SARS coronavirus replicative proteins, p. 75–98. *In* S. K. Lal (ed.), *Molecular biology of the SARS-coronavirus*. Springer, Berlin, Germany.
48. **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.
49. **Verschuere, K. H., K. Pumpor, S. Anemuller, S. Chen, J. R. Mesters, and R. Hilgenfeld.** 2008. A structural view of the inactivation of the SARS coronavirus main proteinase by benzotriazole esters. *Chem. Biol.* **15**:597–606.
50. **Woo, P. C., Y. Huang, S. K. Lau, H. W. Tsoi, and K. Y. Yuen.** 2005. In silico analysis of ORF1ab in coronavirus HKU1 genome reveals a unique putative cleavage site of coronavirus HKU1 3C-like protease. *Microbiol. Immunol.* **49**:899–908.
51. **Xue, X., H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, M. Bartlam, and Z. Rao.** 2007. Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.* **366**:965–975.
52. **Xue, X., H. Yu, H. Yang, F. Xue, Z. Wu, W. Shen, J. Li, Z. Zhou, Y. Ding, Q. Zhao, X. C. Zhang, M. Liao, M. Bartlam, and Z. Rao.** 2008. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* **82**:2515–2527.
53. **Yang, H., W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K. Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, and Z. Rao.** 2005. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **3**:e324.
54. **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. U. S. A.* **100**:13190–13195.
55. **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.
56. **Zhang, Y.** 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**:40.
57. **Zhao, Q., S. Li, F. Xue, Y. Zou, C. Chen, M. Bartlam, and Z. Rao.** 2008. Structure of the main protease from a global infectious human coronavirus, HCoV-HKU1. *J. Virol.* **82**:8647–8655.
58. **Ziebuhr, J.** 2008. Coronavirus replicative proteins, p. 65–81. *In* S. Perlman, T. Gallagher, and E. J. Snijder (ed.), *Nidoviruses*. ASM Press, Washington, DC.
59. **Ziebuhr, J.** 2005. The coronavirus replicase. *Curr. Top. Microbiol. Immunol.* **287**:57–94.
60. **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.
61. **Ziebuhr, J., G. Heussipp, and S. G. Siddell.** 1997. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *J. Virol.* **71**:3992–3997.
62. **Ziebuhr, J., and S. G. Siddell.** 1999. Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab. *J. Virol.* **73**:177–185.
63. **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.
64. **Zunzain, P. A., S. R. Knox, T. R. Sweeney, J. Yang, N. Roque-Rosell, G. J. Belsham, R. J. Leatherbarrow, and S. Curry.** 2010. Insights into cleavage specificity from the crystal structure of foot-and-mouth disease virus 3C protease complexed with a peptide substrate. *J. Mol. Biol.* **395**:375–389.