Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

VIROLOGY 230, 335–342 (1997)
ARTICLE NO. VY978479

Determinants of Mouse Hepatitis Virus 3C-like Proteinase Activity

YIQI LU* and MARK R. DENISON*,†,[1]

†*Department of Pediatrics,* **Department of Microbiology and Immunology, and the Elizabeth B. Lamb Center for Pediatric Research, Vanderbilt University Medical Center, Nashville, Tennessee 37232-2581*

Received December 3, 1996; returned to author for revision January 14, 1997; accepted January 29, 1997

The coronavirus, mouse hepatitis virus strain A59 (MHV), expresses a chymotrypsin-like cysteine proteinase (3CLpro) within the gene 1 polyprotein. The MHV 3CLpro is similar to the picornavirus 3C proteinases in the relative location of confirmed catalytic histidine and cysteine residues and in the predicted use of Q/(S, A, G) dipeptide cleavage sites. However, less is known concerning the participation of aspartic acid or glutamic acid residues in catalysis by the coronavirus 3C-like proteinases or of the precise coding sequence of 3CLpro within the gene 1 polyprotein. In this study, aspartic acid residues in MHV 3CLpro were mutated and the mutant proteinases were tested for activity in an *in vitro trans* cleavage assay. MHV 3CLpro was not inactivated by substitutions at Asp_{3386} (D53) or Asp_{3398} (D65), demonstrating that they were not catalytic residues. MHV 3CLpro was able to cleave at a glutamine–glycine (QG_{3607-8}) dipeptide within the 3CLpro domain upstream from the predicted carboxy-terminal QS_{3635-6} cleavage site of 3CLpro. The predicted full-length 3CLpro (S_{3334} to Q_{3635}) had an apparent mass of 27 kDa, identical to the p27 3CLpro in cells, whereas the truncated proteinase (S_{3334} to Q_{3607}) had an apparent mass of 24 kDa. This 28-amino-acid carboxy-terminal truncation of 3CLpro rendered it inactive in a *trans* cleavage assay. Thus, MHV 3CLpro was able to cleave at a site within the putative full-length proteinase, but the entire predicted 3CLpro domain was required for activity. These studies suggest that the coronavirus 3CL-proteinases may have a substantially different structure and catalytic mechanism than other 3C-like proteinases.

INTRODUCTION

The coronavirus, mouse hepatitis virus, strain A59 (MHV-A59), contains a chymotrypsin-like proteinase within the 750-kDa gene 1 polyprotein (Fig. 1) (Lu *et al.,* 1995). The 3C-like proteinases of the coronaviruses MHV-A59, infectious bronchitis virus (IBV), and the human coronavirus 229E (HCV-229E) are encoded in a conserved region of ORF 1a (Boursnell *et al.,* 1987; Gorbalenya *et al.,* 1989; Lee *et al.,* 1991; Herold *et al.,* 1993). MHV, HCV-229E, and IBV encode 3CLpro molecules with apparent masses of 27, 34, and 35 kDa, respectively, as determined by SDS–PAGE analysis (Lu *et al.,* 1995; Ziebuhr *et al.,* 1995; Tibbles *et al.,* 1996). The classification of the coronavirus proteinases as "3C-like" is supported by mutagenesis studies of predicted catalytic cysteine or histidine residues. We have demonstrated that mutations at His_{41} or Cys_{145} of MHV 3CLpro abolish proteolytic activity. Similar results have been obtained for 3CLpro of IBV and HCV-229E, confirming the essential nature of these residues and demonstrating that His and Cys residues are in positions similar to those of the picornavirus 3C proteinases (Lu *et al.,* 1995; Liu and Brown, 1995; Ziebuhr *et al.,* 1995; Tibbles *et al.,* 1996).

[1] To whom correspondence and reprint requests should be addressed at Department of Pediatrics, Vanderbilt University Medical Center, D7235 MCN, Nashville, TN 37232-2581. E-mail: mark.denison@mcmail.vanderbilt.edu.

The role of aspartic acid or glutamic acid residues in 3CLpro activity is less well understood. Mutagenesis of Asp/Glu residues of 3C and 3C-like proteinases suggests that they might not participate as catalytic residues in all cases. Aspartic/glutamic acid residues have been shown to be essential for proteinase activity of tobacco etch virus (TEV), poliovirus, and human rhinovirus 14 (HRV-14) (Gorbalenya and Koonin, 1993). The positioning of His, Cys, and Glu residues in the HRV-14 3Cpro has been shown to be very similar to that of cellular trypsin by analysis of the HRV crystal structure (Matthews *et al.,* 1994). In contrast, analysis of the crystal structure of hepatitis A virus 3Cpro indicates that Asp84 most likely does not participate directly in catalysis (Allaire *et al.,* 1994). More directly relevant to MHV, it has been shown that mutagenesis at Glu residues in IBV 3CLpro does not abolish activity of the expressed proteinase (Liu and Brown, 1995).

Several 3CLpro cleavage sites within the gene 1 polyprotein recently have been defined for MHV, IBV, and 229E. The experimentally confirmed coronavirus 3CLpro cleavage sites have a leucine, isoleucine, or valine at position P2, glutamine at position P1, and serine or alanine at position P1′ (Liu and Brown, 1995; Lu *et al.,* 1995, 1996; Ziebuhr *et al.,* 1995; Tibbles *et al.,* 1996). In MHV, LQ/S_{3332-4} has been shown to be the amino terminus of 3CLpro (Lu *et al.,* 1995). Other predicted MHV 3CLpro cleavage sites possess phenyl-

0042-6822/97 $25.00

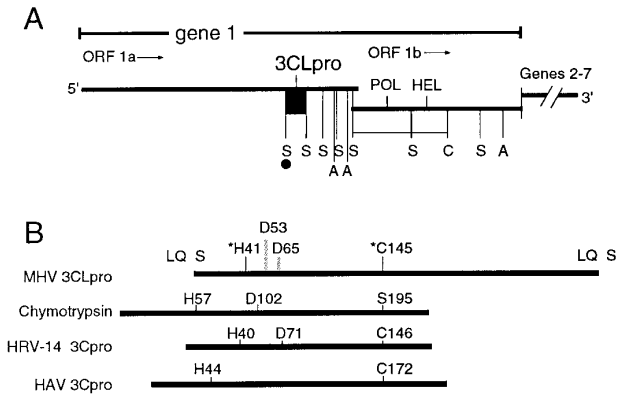

FIG. 1. MHV 3CLpro location, cleavage sites, and comparisons. (A) The linear schematic of the MHV genome shows the organization of the overlapping open reading frames, ORF 1a and ORF 1b, connected by a ribosomal frameshift. The location of 3CLpro in ORF 1a is shown by the black box, and the putative polymerase (POL) and Helicase (HEL) domains in ORF1b are shown by white boxes. Locations of the confirmed (·) and predicted cleavage sites of 3CLpro are indicated by the P1′ residue; S, A, and C indicate Ser, Ala and Cys, respectively. All P1′ residues are preceded by Gln at P1. (B) Comparison of 3CLpro with bovine chymotrypsin (Gorbalenya and Koonin, 1993), 3Cpro of human rhinovirus 14 (Matthews *et al.,* 1994), and hepatitis A virus 3Cpro (Allaire *et al.,* 1994). Black bars reflect the number of amino acid residues in the proteinase. The proteins are aligned at the confirmed Cys or Ser catalytic residues to show the carboxy-terminal extent of the protein. Catalytic Asp residues of HRV-14 and chymotrypsin have been confirmed by crystallography. Asp53 and Asp65 of MHV are predicted catalytic residues. LQS indicates cleavage sites at amino terminus (confirmed) and carboxy terminus (putative) of 3CLpro.

alanine at P2 and glycine at P1′ (Lee *et al.,* 1991). The precise determinants of 3CLpro cleavage site selection remain to be determined. Finally, comparison of the coronavirus 3CLpro sequences with those of other proteinases suggests that they may differ from other viral and cellular proteinases in their size and structure. The coronavirus 3CLpro domains contain significantly more amino acids downstream from the putative substrate binding site than other 3C or 3C-like proteinases (Fig. 1) (Gorbalenya and Koonin, 1993). The role of this additional region of polypeptide in structure or activity of the coronavirus proteinases is not known.

In this study, we demonstrate that aspartic acid residues in MHV 3CLpro are not required for catalytic activity. In addition, we show that the MHV 3CLpro is able to cleave at a Gln-Gly cleavage site upstream from the predicted carboxy terminal Gln-Ser cleavage site and within the proteinase itself. The 3CLpro extending from the confirmed amino-terminal serine to the internal glutamine/glycine cleavage site is inactive *in vitro.* In contrast, the "full-length" protein extending to the predicted glutamine/serine cleavage site is identical in size to 3CLpro expressed *in vitro* and in virus-infected cells and is an active proteinase. Thus, it appears that the entire predicted coding region is required for 3CLpro activity.

MATERIALS AND METHODS

Amino acid alignment and comparison

The amino acid sequences of the coronaviruses were obtained from GenBank; MHV (X73559) (Bonilla *et al.,* 1994), IBV (M94356) (Boursnell *et al.,* 1987), HCV-229E (X69721) (Herold *et al.,* 1993), and TGEV (Z34093) (Eleouet *et al.,* 1995). Deduced amino acid sequences of coronavirus 3CLpro domains were compared using a pam 250 scoring matrix and a word size of 2 (MacVector 4.5.3, IBI-Kodak) (Fig. 2). Numbering of MHV amino acid sequences was from the beginning of ORF 1a. Numbering of amino acid residues within MHV 3CLpro is based on labeling the confirmed amino terminal serine$_{3334}$ as Ser 1 (Lu *et al.,* 1995). Analogous numbering was used for the other coronaviruses, as used in the original reports. The positions of conserved residues T_{135} and H_{163} had been previously predicted (Lee *et al.,* 1991; Gorbalenya and Koonin, 1993).

Site-directed mutagenesis of pGpro

Asp53 and Asp65 were mutagenized by the Chameleon double-stranded, site-directed mutagenesis kit per the manufacturer's instruction (Stratagene). Two primers were simultaneously annealed to the template. One selection primer changed one nonessential unique restriction site *Alw*NI on the pGpro vector to a new restriction site. The other primer encoded for a specific mutation. After annealing and extension, all new plasmid DNA was incubated with the restriction enzyme *Alw*NI. The digested mixtures were then transformed into repair-deficient XL*mutS* cells, and the resultant colonies were isolated. The mutant plasmids were then purified and digested with *Alw*NI again, and the resultant DNA digestion was transformed into XL1-Blue cells. All specific mutations were confirmed by bidirectional sequencing (Sequenase II, U.S. Biochemicals, per the manufacturer's instructions).

Constructs expressing full-length and truncated versions of 3CLpro

The Ser3334 to Gln 3607(pG-S/FQ) and Ser3334 to Gln3635(pG-S/LQ) fragments were obtained by PCR amplification of the region between nt 10212 and 11034 and that between nt 10212 and 11117, respectively (Fig. 5). The pG-S/FQ left primer with an added on *Xba*I restriction site (5′-ATTCTAGATGTCTGGTATAGTGAAGATGG TGTCG-3′) and pG-S/FQ right primer with an added on *Hin*dIII restriction site (5′-TAAAATAAGCTT TCACTGGAA-TCCAGAATGCAGCCT-3′) were used to prime DNA synthesis from the pGpro construct. The PCR products were digested by *Xba*I and *Hin*dIII for 2 hr and then run on an 0.8% low-melting-point agarose gel. The product band was excised and ligated into the *Eco*RI and *Hin*dIII sites

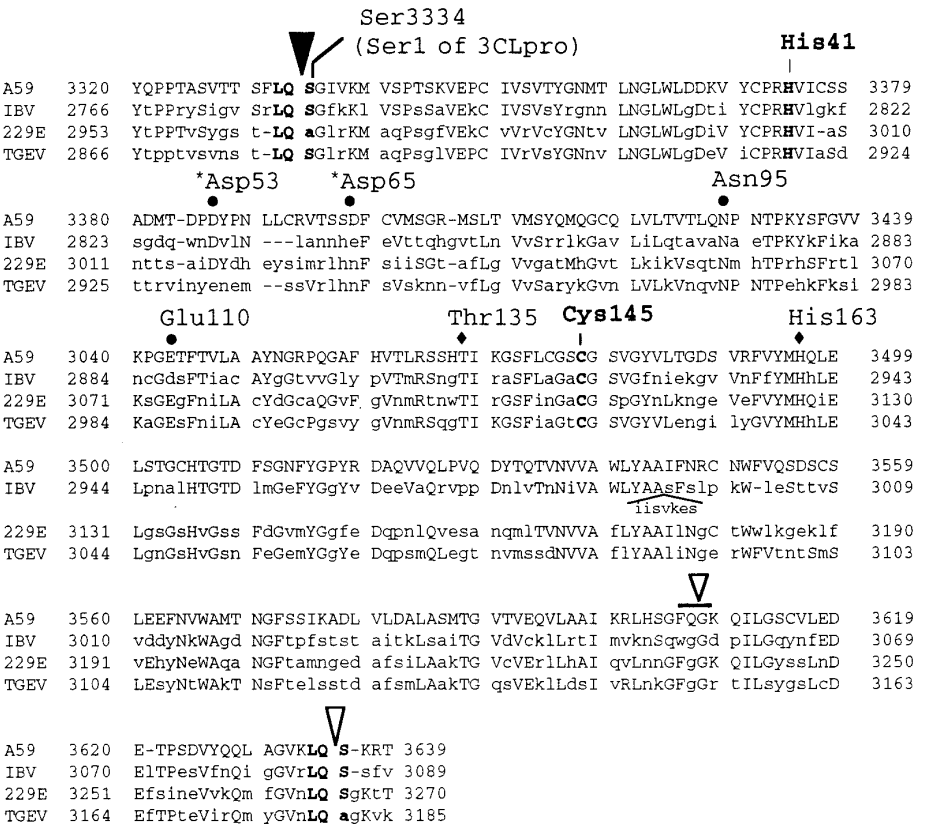

FIG. 2. Comparison of 3C-like proteinase domains of the coronaviruses. The derived amino acid sequences of the 3C-like proteinases from mouse hepatitis virus A59 (MHV-A59), the avian infectious bronchitis virus (IBV), the human coronavirus HCV-229E, and the porcine transmissible gastroenteritis virus (TGEV) were aligned by Mac Vector 4.5.3. with a pam 250 scoring matrix. The MHV-A59 His41 and Cys145 residues, and the corresponding residues of IBV, TGEV, and 229E are shown in boldface letters. The locations of aspartic acid residues (Asp53 and Asp65) of MHV-A59 are shown by asterisks. Other conserved asparagine (N95) and aspartic/glutamic acid residues (D110) are indicated by a dot. Residues predicted to be involved in substrate binding (Thr135 and His163) are indicated by a diamond. The solid arrowhead indicates the experimentally confirmed amino-terminal cleavage site of the MHV and 229E 3CLpro (Lu *et al.,* 1995; Ziebuhr *et al.,* 1995). The open arrowhead indicates the predicted carboxyl terminal LQ_S/A cleavage sites of the proteinases (Gorbalenya *et al.,* 1989; Lee *et al.,* 1991; Gorbalenya and Koonin, 1993). The location of the FQ_G sequence in MHV is indicated as an underlined arrowhead. Numbering of MHV His41, Cys145, Asp53, Asp65, Asn95, and Glu110 residues is based on identifying Ser_{3334} of the ORF 1a polyprotein as Ser1 of 3CLpro. MHV-A59 amino acid numbers were derived from the submitted nucleotide sequence of Bonilla, *et al.* (1994). "iisvkes" is a seven-amino-acid region present only in the IBV sequence.

of pGEM-3Zf(−) (Promega) behind the T7 promoter which constituted pG-S/FQ. pGopt-S/FQ was similarly constructed using a left primer with an optimal ATG (5′-GGGCGAATTCGCCACCATGAGTGGTATAGTGAAGAT GGTGTCG-3′). pC-S/FQ was constructed by using a left primer with an added *Nco*I restriction site (5′-TCATCC-ATGGCCTCTGGTATAGTGAAGAT G-3′) and a right primer with an added *Eco*RI restriction site (5′-AATTT-GAATTCACTGGAATCCAGAATGCAGCCT-3′). The fragment was then subcloned into pCITE (Novagen). pC-S/LQ was similarly constructed by using a left primer with an added *Nco*I restriction site (5′-TCATCCATGGCCTCT-GGTATAGTGAAGATG-3′) and a right primer with an added *Eco*RI restriction site (5′-TGTGCG AATTCACTG-TAGCTTGACACCAGCTA-3′). The fragment was then subcloned into the *Nco*I and *Eco*RI sites of pCITE (Novagen) behind the T7 promoter.

In vitro transcription and translation

Recombinant plasmids were transcribed and translated using a coupled *in vitro* transcription/translation rabbit reticulocyte lysate system (TnT, Promega), as previously described (Lu *et al.,* 1995, 1996). Approximately 0.5 μg of plasmid DNA was incubated at 30° with 12.5 μl TnT lysate, 1 μl TnT reaction buffer, 0.5 μl T7 RNA polymerase, 20 units RNasin, 0.5 μl 1 m*M* methionine-free amino acid mixture, and 20 μCi [^{35}S]methionine in a final volume of 25 μl. Samples were taken at various time points and electrophoresed on an SDS 5–18% gradient polyacrylamide gel (SDS–PAGE).

Trans cleavage assay

Inactive site-directed mutants of pGpro (pGproH41G or pGproH41Q) were translated in the presence of [^{35}S]-

methionine. The parental pGpro construct, plasmids encoding the predicted full-length 3CLpro (pC-S/LQ), and plasmids encoding the truncated forms of 3CLpro (pG-S/FQ, pGopt-S/FQ, and pC-S/FQ) were transcribed and translated in the presence of nonradiolabeled L-methionine. After 40 min, transcription and translation were terminated by the addition of RNase (10 μg/ml) and cycloheximide (5 μg/ml) for 5 min. Following termination of transcription and translation, labeled mutant and unlabeled 3CLpro reaction lysates were mixed 1:1 and incubated for an additional 135 min. The reaction mixtures were checked for residual expression and processing from the pGpro construct by the addition of [^{35}S]-methionine to an aliquot of the unlabeled reaction mixture after treatment with RNase and cycloheximide and incubation for an additional 135 min. All products were analyzed by electrophoresis by SDS gradient PAGE, followed by fluorography.

Peptide radiosequencing

In vitro transcription and translation were performed in a total volume of 200 μl with 8.0 μg pGpro DNA in the presence of 160 μCi [^{35}S]methionine (DuPont NEN), 400 μCi [^{3}H]valine (Amersham), or 160 μCi [^{3}H]leucine (Amersham) for 120 min at 30°. The products were separated on 5–18% gradient polyacrylamide gels, transferred to a polyvinylidene difluoride (PVDF) membrane at 50 V at 4° for 6 hr in transfer buffer containing 25 m*M* Tris-base, 192 m*M* glycine, and 10% (v/v) methanol. After transfer, the PVDF membrane was air dried and exposed to X-ray film. Radiolabeled proteins were identified by autoradiography, and the corresponding bands were excised from the PVDF membrane and subjected to amino-terminal sequencing on an ABI 470 sequencer. The amino acid fraction from each cycle was quantitated in a Beckman scintillation counter.

RESULTS

Alignment and comparison of coronavirus 3C-like proteinase domains

Predictions of catalytic residues of the coronavirus 3C-like proteinases have not strongly predicted aspartic or glutamic acid residues. Comparison of the deduced amino acid sequences of 3CLpro from the coronaviruses MHV-A59, IBV, HCV-229E, and TGEV revealed no completely conserved Asp or Glu residues at positions analogous to catalytic Asp or Glu residues of other 3C or 3C-like proteinases (Fig. 2). There was relative conservation of Asp, Glu, or Asn among the coronaviruses at the residue analogous to Asp_{3398} (D65) of MHV. It has been shown that the analogous residues within the IBV 3CLpro at Asp_{2841} or Asp_{2843} (D62 or D64) are not required for proteolytic activity (Liu and Brown, 1995). The MHV Asp_{3386} (D53) was conserved as either Asp or Glu among the four viruses. Glu_{3443} (E110) of MHV was conserved as Glu or Asp, and Asn_{3428} (N95) of MHV was identical in all four coronavirus sequences; however, the location of these residues relative to the essential His and Cys residues makes them less appealing as potential catalytic residues.

Comparison of the coronavirus 3CLpro amino acid sequences with chymotrypsin confirmed the additional amino acids between the putative substrate binding residue His_{3496} (H163) and the probable carboxy-terminal QS_{3635-6} cleavage site of 3CLpro. The comparison of the four coronavirus 3CLpro sequences revealed two potential cleavage sites present only in MHV, QS_{3554-5}, and QG_{3607-8}. Overall, the comparison of the coronavirus sequences indicated that there was variation among the proteinases in the location of potential catalytic residues and cleavage sites.

Mutagenesis of aspartic acid residues

Based on the analysis of the protein alignments, we chose Asp53 and Asp65 residues for mutagenesis studies. Asp65 has been considered the most likely candidate for a third residue to be involved in catalysis. Asp53 was in a less favorable position relative to the His, but was conserved among the coronaviruses and provided a good control. In addition, studies of other viruses have demonstrated that deviation from predictions of active residues is not uncommon. The construct used for these studies (pGpro) encoded amino acids 3239–3687 of MHV gene 1, including 3CLpro (3334–3635) and portions of the flanking domains (Fig. 3A). We have previously shown that translation of pGpro *in vitro* results in a precursor polypeptide from which active 3CLpro is autoproteolytically cleaved and that 3CLpro has an apparent mass of 27–29 kDa (p27) following SDS–PAGE (Lu *et al.,* 1995). Liberation of p27 3CLpro was therefore used as a marker of proteolytic activity of proteins expressed from different constructs *in vitro.*

The wild-type proteinase construct (pGpro) and mutant proteinase constructs were transcribed and translated in a rabbit reticulocyte lysate (Fig. 3B). The proteinase expressed from pGpro was able to process p27 3CLpro (Fig. 3B, lane 1), whereas the proteinase with the His41 to Gln mutation (H41Q) did not cleave p27 (Fig. 3B, lane 2). Mutation of Asp65 to Pro or Ala (D65P and D65A) resulted in a proteinase with activity comparable to that expressed from wild-type pGpro (Fig. 3B, lanes 3 and 4). Substitution of Asp53 by Glu (D53E) did not affect 3CLpro activity (Fig. 3B, lane 5), whereas the substitution of Asp53 by Pro (D53P) impaired processing of p27 approximately 70% relative to pGpro (Fig. 3B, lane 6). The D53P change might be expected to cause a change in the proteinase structure with a concomitant alteration of ac-

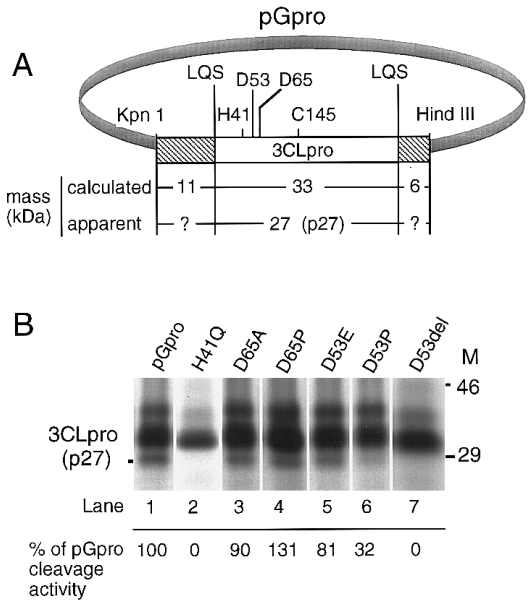

FIG. 3. Mutagenesis of aspartic acid residues and *trans* cleavage activity *in vitro.* (A) The organization of the pGpro construct is shown. *Kpn*I and *Hin*dIII are the cloning sites of the MHV gene 1 subclone in pGEM 3ZF-. LQS cleavage sites flank the 3CLpro region. The calculated and apparent masses of 3CLpro and the flanking polypeptide fragments are shown. (B) The proteinase constructs with substitutions at Asp53 and Asp65 were expressed in a combined transcription and translation lysate as previously described (Lu *et al.,* 1995). Samples were taken at 120 min for analysis by 5–18% SDS gradient PAGE. The wild-type pGpro construct and the His41 to Gln mutant (H41Q) were used as controls. D65A refers to an Ala substitution at Asp53; other constructs are similarly labeled. Mass markers are to the right of the gel and the location of p27 is shown to the left of the gel. Processing of p27 by 3CLpro expressed from pGpro was considered as 100%, and the percentage of proteinase activity of each expressed protein is shown beneath the lane markers.

tivity; however, the D53P substitution diminished but did not abolish activity, indicating that D53 was not an indispensable residue. It was interesting that the D65P change did not diminish proteolytic activity, suggesting that even a major change at this location was inconsequential for liberation of p27 3CLpro in the *in vitro* system. Deletion of Asp53 (D53del) resulted in complete loss of proteinase activity (lane 7). This was not surprising since such a change might be expected to significantly alter the structure of the proteinase. These experiments demonstrated that neither Asp53 nor Asp65 was directly involved in catalysis with His41 or Cys145.

Identification of a 3CLpro cleavage site

During *in vitro* translation of pGpro several proteins with apparent masses of less than 14.3 kDa were seen along with p27 (Fig. 4). The pulse–label expression (Fig. 4A) showed that the smallest of these polypeptides appeared concurrently with p27 3CLpro but then decreased over a 4-hr period (Fig. 4, lanes 1–3). Translation of pGpro in the presence of leupeptin blocked cleavage of p27 and also completely blocked processing of the small polypeptide fragment (Fig. 4, lanes 4–5). The small cleavage fragment indicated by the arrow was consistently seen when pGpro or proteolytically active mutants were translated, but not when proteolytically inactive mutants were expressed (Fig. 4B). The cleavage fragment was detected following translation of pGpro, H127Q, H127M, and C142R (Fig. 4B, lanes 1, 2, 3, and 4, respectively), all of which also processed p27. In contrast, no small fragment was seen after translation of C145G or H41Q (Fig. 4B, lanes 5 and 6), both of which are inactive in p27 processing. Together these results indicated that this cleavage fragment was processed by products expressed from the proteinase constructs *in vitro,* rather than by proteinases in the reticulocyte lysate.

The smallest proteolytic fragment was used for amino terminus radiosequencing since it was the most discrete and abundant. The pGpro construct was transcribed and translated in the presence of [^{3}H]leucine, [^{3}H]valine, or [^{35}S]cysteine. The proteins were transferred to PVDF membranes and cleaved by Edman degradation, and radioactivity of individual amino acids was quantitated (Fig. 4C). Within the first 14 residues, the peaks of radioactivity were consistent with leucine at residues 5 and 10, cysteine at residue 8, and valine at residue 9. The only cleavage site within the pGpro expression product that could result in a product with this pattern was Gln-Gly$_{3607\text{-}8}$.

The QG$_{3607\text{-}8}$ was not conserved in any of the other coronaviruses and previously had not been predicted as a cleavage site for 3CLpro. Radiosequencing with three different amino acids confirmed specific cleavage between glutamine$_{3607}$ and glycine$_{3608}$ by the *in vitro* translated proteinase. We could not define the presumed carboxy-terminal fragment containing the predicted QS$_{3635\text{-}6}$ cleavage site, possibly due to the compression of proteins in this region (5.8 kDa) of the gel by the nonlabeled globin protein from the lysate. Additionally, the protein from the construct may have been targeted for rapid degradation. We also did not detect any prominent alternative form of p27 3CLpro. Since we do know the order of cleavages or pattern of precursors expressed from pGpro it has not been possible to determine when the FQ/G site is cleaved. Direct comparison of these cleavage sites will require constructs expressing single cleavage sites to determine specificity.

Truncation of 3CLpro and *trans* cleavage activity *in vitro*

Since 3CLpro was able to cleave upstream of the predicted QS$_{3635\text{-}6}$ cleavage site, we determined whether Ser$_{3334}$ to Gln$_{3607}$ was the entire coding region for the active p27 3CLpro protein detected in virus-infected cells

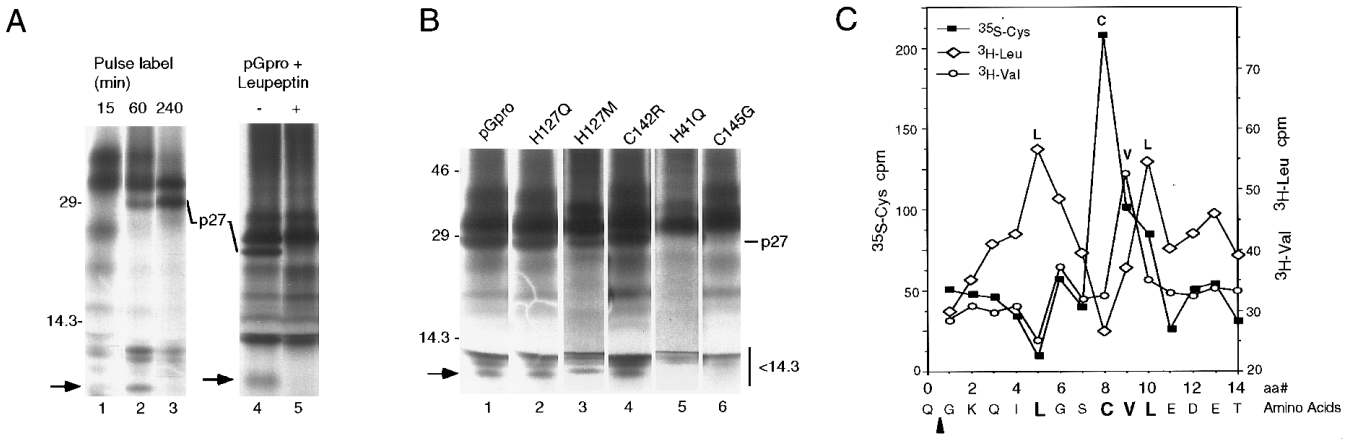

FIG. 4. Identification and sequencing of a 3CLpro cleavage site. (A) The pGpro construct was labeled with [^{35}S]met for 4 hr, with samples taken at the times (in minutes) shown (lanes 1–3). The arrow shows the location of the discrete fragment of <14.3 kDa. pGpro was translated in the absence or in the presence of 2 m*M* leupeptin for 1 hr, and the location of the proteolytic fragment is similarly indicated. Markers are to the left of the gel. (B) pGpro and pGpro mutants were translated for 1 hr in the presence of [^{35}S]met and analyzed by SDS–PAGE. The location of p27, proteins <14.3 kDa, and the specific proteolytic fragment are shown. Lanes 1, 2, 4, and 6 are from a separate experiment from lanes 3 and 5. (C) Radioactivity was quantitated following cleavage of radiolabeled amino acids in the small cleavage fragment shown by the arrows in A and B. The cpm of fractions containing [^{35}S]Cys are to the left of the figure and the cpm of fractions containing [^{3}H]Val or [^{3}H]Leu are to the right of the figure. The region within the pGpro expressed protein corresponding to the pattern is shown below the figure, with an arrowhead indicating the probable cleavage site.

and during *in vitro* translation of pGpro (Fig. 5). The amino acid sequence extending from Ser_{3334} to Gln_{3635} would predict a protein of with a calculated mass of 33 kDa, whereas cleavage at Gln_{3607} would predict a protein of 30 kDa in mass, somewhat closer in size to the apparent mass of p27 (Fig. 5A). We constructed a panel of plasmids containing cDNAs encoding amino acids from S_{3334}–Gln_{3607} or S_{3334}–Gln_{3635} (Fig. 5A). The cDNAs were expressed in a variety of plasmids, using either the first natural AUG (pG-S/FQ) or an optimized AUG before Ser_{3334} (pGopt-S/FQ). We also used vectors containing EMCV IRES elements to ensure that translation initiated before the Ser_{3334} (pC-S/FQ and pC-S/LQ).

The constructs were used to direct translation *in vitro* and the proteins either were radiolabled or were translated in nonlabeled medium and used in a *trans* cleavage assay of the inactive mutant pGproH41G (Fig. 5B). Translation of pC-S/LQ, encoding Ser_{3334} to Gln_{3635}, resulted in a single 27-kDa protein, the same migration pattern as p27 3CLpro detected after expression of pGpro (Fig. 5B, lanes 1 and 2). In contrast, translation of three different constructs encoding the truncated 3CLpro domain from Ser_{3334} to Gln_{3607} resulted in a single 24-kDa protein (Fig. 5B, lanes 3, 4, and 5). These data indicated that proteins expressed from the 3CLpro domain differed in their calculated and apparent masses by 6 to 10 kDa. The results also supported the conclusion that the active 3CLpro in MHV-infected cells and from *in vitro* translation products incorporated Ser_{3334} to Gln_{3635}.

We assessed the *in vitro* cleavage activity of the 24-kDa Ser_{3334}–Gln_{3607} and the 27-kDa Ser_{3334}–Gln_{3635} proteins by incubating the nonradiolabeled translation products of these constructs with radiolabeled substrate expressed from the inactive proteinase mutant pGproH41G (His to Gly) (Fig. 5B, lane 6). The "full-length" 27-kDa 3CLpro expressed from the Ser_{3334}–Gln_{3635} construct was able to cleave the pGproH41G expressed protein in *trans* (Fig. 5B, lane 8), whereas the 24-kDa truncated 3CLpro expressed from the Ser_{3334}–Gln_{3607} constructs did not process the mutant protein (Fig. 5B, lanes 9, 10, and 11). This result demonstrated that a 28-amino-acid carboxy terminal truncation of 3CLpro abolished proteolytic activity.

DISCUSSION

MHV 3CLpro is postulated to mediate the majority of cleavages in the gene 1 polyprotein during virus replication. We have shown that aspartic acid residues of MHV 3CLpro in locations analogous to essential Asp/Glu residues of other 3C and 3C-like proteinases are not necessary for processing of substrate by 3CLpro *in vitro.* Results similar to ours have been reported for infectious bronchitis virus (IBV) (Liu and Brown, 1995). Our study demonstrates that conservation of Asp/Glu in this region of the coronavirus 3C-like proteinases is not due to an indispensable catalytic role. The mechanism of the MHV 3CLpro may be more similar to that of hepatitis A virus, in which the Asp is on an external motif and not directly involved in the catalytic unit (Allaire *et al.,* 1994). It is possible that this variation in the use of a third residue may have coevolved with the specificity for cleavage

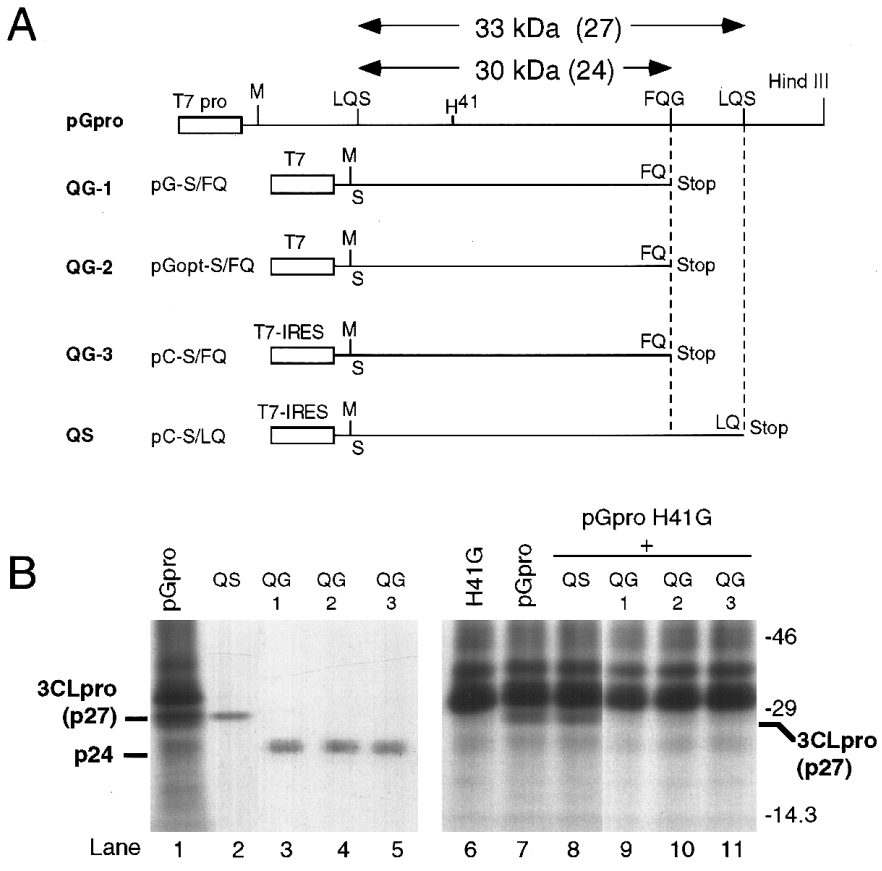

FIG. 5. Truncation of 3CLpro and trans cleavage activity. (A) The pGpro construct is shown, with location of T7 promoter, first AUG (M), location of possible cleavage sites, location of catalytic H41 residue, and calculated mass of full-length (33 kDa) and truncated (30 kDa) versions of 3CLpro. The apparent masses (27 and 24 kDa) are shown in parentheses. Plasmids encoding the predicted full-length (pC-S/LQ) or truncated (pG-S/FQ, pGopt-S/FQ and pC-S/FQ) proteins were constructed as described under Materials and Methods. IRES indicates the inclusion of an EMCV internal ribosome entry site upstream from the initiating methionine. The abbreviated names for the constructs are indicated at the far left of the figure. (B) Expression of full-length or truncated 3CLpro and *trans* cleavage activity. The results of translation of pGpro are in lane 1. Lanes 1–5 show the proteins expressed from the constructs from A during *in vitro* transcription and translation, as indicated by the abbreviated name. Lane 6 is translation of the inactive mutant pGproH41G. Lane 7 is incubation of pGpro alone; lanes 8–11 show the results of incubation of pGproH41G translation products with nonradiolabeled proteins as shown in lanes 2–5.

sites. There is a precedent in other virus systems for a contribution of Asp residues to proteinase specificity even though they may not be involved in catalysis. For example, poliovirus contains an FRD (D85) sequence that was initially thought to be involved in catalytic activity but subsequently was found to be in a flanking turn domain and to be involved in autocatalytic cleavage of 3CD (Hammerle *et al.,* 1992). Despite the lack of use of an Asp residue, the MHV 3CLpro should still be classified as a chymotrypsin-like enzyme because of the localization of histidine and cysteine residues as well as flanking residues considered to be important in protein structure and substrate binding (Gorbalenya and Koonin, 1993).

Analysis of the full-length 3CLpro domain (Ser_{3334} to Gln_{3635}) reveals several possible differences between the MHV 3CLpro and other viral proteinases in the group of cysteine-containing enzymes (Fig. 1). First, most of these enzymes terminate within 30 amino acids following the consensus substrate binding residues, whereas the MHV sequence extends an additional 137 residues to its carboxy terminus (Lee *et al.,* 1991). Our study demonstrates that a small deletion of this part of the proteinase abolishes its ability to cleave new molecules of p27 3CLpro *in trans,* demonstrating that the entire carboxy-terminal region is essential for 3CLpro activity.

Analysis of confirmed and predicted cleavage sites in the gene 1 polyproteins of MHV-A59, HCV-229E, IBV, and TGEV has revealed a preference for a Gln at P1 and Leu, Ile, Val, or less often Phe or Met at P2 (Boursnell *et al.,* 1987; Breedenbeek *et al.,* 1990; Lee *et al.,* 1991; Herold *et al.,* 1993; Bonilla *et al.,* 1994; Eleouet *et al.,* 1995). Although the Phe-Gln-Gly (FQ/G) 3CLpro cleavage site we identified within 3CLpro has similarities to other predicted 3CLpro sites from P4 to P1′, it is not present in the other sequenced coronaviruses. We have not determined if the FQ/G_{3608} site can be cleaved in virus-infected

cells. It is possible that the fragment of gene 1 used in these studies allows presentation of this site in a manner that would not occur during virus replication. If the FQ/G cleavage site is used by the proteinase in cells, it might represent a pathway for regulation of the proteinase concentrations, since the truncated 24-kDa protein is inactive. Overexpression of 3C proteinases can have detrimental effects on host cells and also on virus replication, such as with the picornavirus FMDV 3C (Martinez-Salas and Domingo, 1995). Alternatively, different pathways of cleavage may regulate availability of different forms of the proteinase, such as has been reported for the poliovirus 3CD to 3C′ and 3D′ cleavage (Lawson and Semler, 1992).

In conclusion, we have identified several unique features of MHV-A59 3CLpro that provide insights into the relationship of structure and function in this expanding family of enzymes. These results will also allow us to further characterize the role of 3CLpro in MHV polymerase gene polyprotein processing and virus replication.

ACKNOWLEDGMENTS

This project was supported by Public Health Service Grant RO1 AI26603 (M.R.D.) from the NIH. We thank Xiaotao Lu for excellent technical support. We appreciate the critical reading and insights of Anne Gibson and Amy Sims. Peptide sequencing was performed by Eric Howard and Dr. Masaki Tamura in the protein sequencing shared resource of the Vanderbilt University Cancer Center (IP30CA68485).

REFERENCES

Allaire, M., Chernala, M. M., Malcolm, B. A., and James, N. G. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* 369, 72–76.

Bonilla, P. J., Gorbalenya, A. E., and Weiss, S. R. (1994). Mouse hepatitis virus strain A59 RNA polymerase gene ORF 1a: Heterogeneity among MHV strains. *Virology* 198, 736–740.

Boursnell, M. F. G., Brown, T. D. K., Foulds, I. J., Green, P. F., Tomley, F. M., and Binns, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J. Gen. Virol.* 68, 57–77.

Breedenbeek, P. J., Pachuk, C. J., Noten, A. F. H., Charite, J., Luytjes, W., Weiss, S. R., and Spaan, W. J. M. (1990). The primary structure and expression of the second open reading frame of the polymerse gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res.* 18, 1825–1832.

Denison, M. R., Hughes, S. A., and Weiss, S. R. (1995). Identification and characterization of a 65-kDa protein processed from the gene 1 polyprotein of the murine coronavirus MHV-A59. *Virology* 207, 316–320.

Eleouet, J. F., Rasschaert, D., Lambert, P., Levy, L., Vende, P., and Laude, H. (1995). Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. *Virology* 206, 817–822.

Gorbalenya, A., and Koonin, E. (1993). Comparative analysis of amino-acid sequences of key enzymes of replication and expression of positive-strand RNA viruses: Validity of approach and functional and evolutionary implications. *Sov. Sci. Rev. D Physiochem. Biol.* 11, 1–81.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., and Blinov, V. M. (1989). Coronavirus genome: Prediction of putative functional domains in the nonstructural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.* 17, 4847–4861.

Hammerle, T., Molla, A., and Wimmer, E. (1992). Mutational analysis of the proposed FG look of poliovirus proteinase 3C identifies amino acids that are necessary for 3CD cleavage and might be determinants of a function distinct from proteolytic activity. *J. Virol.* 66, 6028–6034.

Herold, J., Raabe, T., Schelle, P. B., and Siddell, S. G. (1993). Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology* 195, 680–91.

Herold, J., Raabe, T., and Siddell, S. G. (1993). Characterization of the human coronavirus 229E (HCV 229E) gene 1. *Adv. Exp. Med. Biol.* 342, 75–9.

Lawson, M. R., and Semler, B. L. (1992). Alternate poliovirus nonstructural protein processing cascades generated by primary sites of 3C proteinase cleavage. *Virology* 191, 309–320.

Lee, H.-J., Shieh, C.-K., Gorbalenya, A. E., Koonin, E. V., LaMonica, N., Tuler, J., Bagdzhadhzyan, A., and Lai, M. M. C. (1991). The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* 180, 567–582.

Liu, D. X., and Brown, T. D. K. (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.

Lu, X., Lu, Y., and Denison, M. R. (1996). Intracellular and *in vitro* translated 27-kDa proteins contain the 3C-like proteinase activity of the coronavirus MHV-A59. *Virology* 222, 375–382.

Lu, Y., Lu, X., and Denison, M. R. (1995). Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *J. Virol.* 69, 3554–3559.

Martinez-Salas, E., and Domingo, E. (1995). Effect of expression of the aphthovirus protease 3C on viral infection and gene expression. *Virology* 212, 111–120.

Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., Gribskov, C. L., and Worland, S. (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.

Tibbles, K. W., Brierley, I., Cavanaugh, D., and Brown, T. D. K. (1996). Characterization in vitro of an autocatalytic processing activity associated with the predicted 3C-like proteinase domain of the coronavirus avian infectious bronchitis virus. *J. Virol.* 70, 1923–1930.

Ziebuhr, J., Herold, J., and Siddell, S. G. (1995). Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *J. Virol.* 69, 4331–8.