JOURNAL OF VIROLOGY, June 1995, p. 3554–3559 Vol. 69, No. 6
0022-538X/95/$04.00+0

Identification and Characterization of a Serine-Like Proteinase of the Murine Coronavirus MHV-A59

YIQI LU, XIAOTAO LU, AND MARK R. DENISON*

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

Received 18 January 1995/Accepted 10 March 1995

Gene 1 of the murine coronavirus, MHV-A59, encodes approximately 800 kDa of protein products within two overlapping open reading frames (ORFs 1a and 1b). The gene is expressed as a polyprotein that is processed into individual proteins, presumably by virus-encoded proteinases. ORF 1a has been predicted to encode proteins with similarity to viral and cellular proteinases, such as papain, and to the 3C proteinases of the picornaviruses (A. E. Gorbalenya, A. P. Donchenko, V. M. Blinov, and E. V. Koonin, FEBS Lett. 243:103–114, 1989; A. E. Gorbalenya, E. V. Koonin, A. P. Donchenko, and V. M. Blinov, Nucleic Acids Res. 17:4847–4861, 1989). We have cloned into a T7 transcription vector a cDNA fragment containing the putative 3C-like proteinase domain of MHV-A59, along with portions of the flanking hydrophobic domains. The construct was used to express a polypeptide in a combined in vitro transcription-translation system. Major polypeptides with molecular masses of 38 and 33 kDa were detected at early times, whereas polypeptides with molecular masses of 32 and 27 kDa were predominant after 30 to 45 min and appeared to be products of specific proteolysis of larger precursors. Mutations at the putative catalytic histidine and cysteine residues abolished the processing of the 27-kDa protein. Translation products of the pGpro construct were able to cleave the 27-kDa protein in *trans* from polypeptides expressed from the noncleaving histidine or cysteine mutants. The amino-terminal cleavage of the 27-kDa protein occurred at a glutamine-serine dipeptide as previously predicted. This study provides experimental confirmation that the coronaviruses express an active proteinase within the 3C-like proteinase domain of gene 1 ORF 1a and that this proteinase utilizes at least one canonical QS dipeptide as a cleavage site in vitro.

The coronavirus mouse hepatitis virus strain A59 (MHV-A59) contains a single-stranded, positive-sense RNA genome 32 kb in length that is composed of seven distinct genes separated by translation termination codons and noncoding intergenic regions (20, 22). Replication of the MHV genome in the infected cell is likely initiated by translation of the 5′-most gene 1 into proteins necessary for replication of the coronavirus genome (5), as well as for transcription of six subgenomic mRNAs that are in turn translated into one or more structural or nonstructural proteins (3, 18, 19, 23, 24). At approximately 22 kb in length, gene 1 of MHV-A59 is by far the largest nonstructural gene among known RNA viruses. Sequence analysis of MHV and avian coronavirus infectious bronchitis virus (IBV) has been used to determine that gene 1 is translated into a single large polyprotein with a molecular mass of approximately 800 kDa via a ribosomal frameshift between two overlapping open reading frames, ORFs 1a and 1b (6–8, 20). Comparison of the gene 1 sequence of MHV with that of other positive-strand RNA viruses has revealed that this gene may encode several important functions, including two papain-like proteinases, a proteinase with similarities to the 3C proteinase of the picornaviruses, helicase, polymerase, and nucleoside triphosphate binding activities (15, 20). The first papain-like proteinase is responsible for the cleavage of the amino-terminal p28 protein from the gene 1 polyprotein during in vitro translation of genome RNA or synthetic gene 1 transcripts (2).

The putative coronavirus 3C-like proteinase (3CLpro) has not been identified or characterized, either in vitro or in virus-infected cells. The existence of 3CLpro is predicated upon sequence comparison of the MHV gene 1 sequence with that of known proteinases of other positive-strand RNA virus families such as the picornaviruses, potyviruses, and comoviruses, as well as with that of cellular serine proteinases (11, 20). These comparisons have revealed that the MHV 3C-like proteinase domain is contained within a 300-residue region demarcated on either side by glutamine-serine (QS) dipeptides with similarity to the known cleavage sites utilized by picornavirus 3C proteinases (Fig. 1). In addition, the predicted 3CLpro coding region is flanked by large (200- to 300-residue) predominantly hydrophobic regions predicted to be membrane-spanning domains (MP1 and MP2), which in turn are bounded by basic dipeptides similar to those bordering the predicted 3CLpro itself (20). There is conservation of histidine and cysteine residues in the MHV 3CLpro at positions which correspond to catalytic residues of other confirmed cellular serine and viral serine-like proteinases. In contrast, the overall amino acid sequence of the 3CLpro has no significant similarity to that of known proteinases of other RNA viruses (11).

In this study, we have cloned the predicted MHV-A59 3C-like proteinase domain, along with portions of the neighboring hydrophobic domains, into a transcription vector under the control of a T7 promoter. Translation of the resulting construct (pGpro) in rabbit reticulocyte lysates resulted in polypeptides which possess proteolytic activity. Site-directed mutagenesis of the putative catalytic histidine and cysteine residues eliminated proteolytic activity, but the noncleaved mutant polypeptides could still be processed in *trans* by the protein expressed from the parental pGpro construct. The proteinase encoded in this

* Corresponding author. Mailing address: Department of Pediatrics, Vanderbilt Medical School, D7235 MCN, Nashville, TN 37232-2581. Phone: (615) 322-2250. Fax: (615) 343-9723. Electronic mail address: denison@ctrvax.vanderbilt.edu.

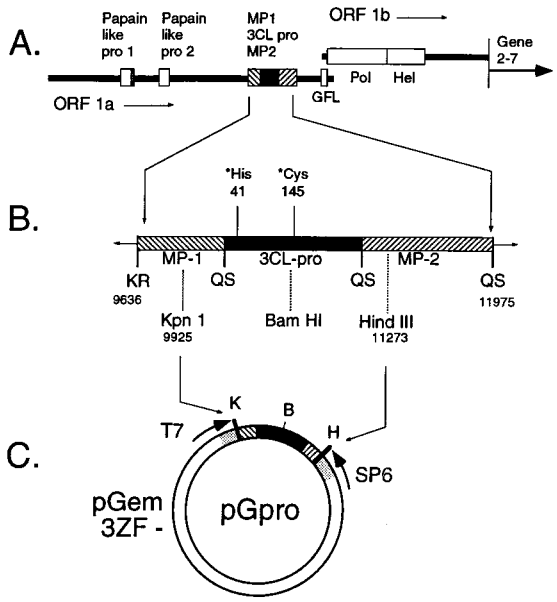

FIG. 1. Location, structure, and cloning of the 3C-like proteinase domain of MHV-A59. (A) Organization and predicted functional domains of gene 1. The gene is 21,739 nt in length, comprising two overlapping open reading frames (ORFs 1a and 1b). ORF 1a contains two papain-like proteinase domains and one 3C-like proteinase domain. GFL, murine epidermal growth factor-like domain; Pol, RNA-dependent RNA polymerase domain; Hel, helicase. (B) 3C-like proteinase with flanking hydrophobic membrane-spanning domains (MP1 and MP2). Dipeptides (KR and QS) indicate predicted cleavage sites for putative 3CLpro. The numbers below the dipeptides and restriction sites indicate nucleotides on gene 1 from the 5′ end. The residues above the line indicate the positions of amino acids numbered from the first QS site. *, amino acids predicted to be involved in catalytic activity. (C) Cloning of the 3CLpro domain. *Kpn*I-*Hin*dIII fragment of MP1–3CLpro–MP2 was cloned into pGEM-3Zf(−) to generate the pGpro construct.

region utilizes at least one QS dipeptide as a cleavage site, which is consistent with previous predictions.

MATERIALS AND METHODS

Construction of pGpro and opt-pGpro plasmids. The ORF 1a fragment from *Kpn*I (nucleotide [nt] 9925) to *Bam*HI (nt 10638) was derived from plasmid 917, which had been provided by Susan Weiss (Fig. 1) (22). The *Bam*HI-to-*Hin*dIII fragment of ORF 1a (nt 10638 to 11273) was excised from a cDNA obtained by reverse transcription-PCR amplification of the region between nt 9926 and 11479. Reverse transcription was performed with random hexamer oligonucleotides and MHV-A59 virion RNA which was purified as previously described (10). PCR was performed with specific primers based on the MHV-A59 sequence (4). The two fragments were ligated into pGEM-3Zf(−) (Promega) behind the T7 promoter to obtain the intact *Kpn*I-to-*Hin*dIII fragment which constituted pGpro. This construct encoded an estimated 9.7-kDa fragment of MP1, the entire 33.6-kDa 3CLpro, and a 5.8-kDa fragment of MP2 (see also Fig. 6B). The pGpro plasmid was used to transform *Escherichia coli* JM109 in the presence of 50 μg of carbenicillin per ml. Purified plasmid was bidirectionally sequenced across the *Kpn*I-to-*Hin*dIII region, and the sequence was compared with that of ORF 1a of MHV-A59 (4). An optimal ATG-containing sequence also was introduced between *Eco*RI and *Kpn*I on the vector by subcloning double-stranded oligonucleotide TGCCGCCATG, yielding the opt-pGpro construct.

PCR site-directed mutagenesis of pGpro. Synthetic oligonucleotides containing mismatches were used to perform PCR site-directed mutagenesis of the His-41, His-127, Cys-142, and Cys-145 residues, according to the method of Higuchi et al. (16). The numbering of residues was based on the predicted amino-terminal QS cleavage site at residue 1 (amino acid 3335 from the amino terminus of the ORF 1a polyprotein) (4). In brief, oligonucleotide His-41→Gln right (5′-GATAACTTGTCTTGGGCAAT-3′), containing a mismatch at nucleotide position 10334, and a T7 promoter primer were used to prime DNA synthesis from the pGpro construct. Similarly, oligonucleotide His-41 → Gln left (5′-ATTGCCCAAGACAAGTTATC-3′), containing a mismatch at nucleotide position 10334, and an SP6 promoter primer were used to prime DNA synthesis from the pGpro construct. The products of the two reactions were run on a 0.8% low-melting-point agarose gel to purify the PCR products. The resulting bands were excised and isolated from the agarose, and the purified PCR products were combined, denatured, reannealed, and allowed to amplify one cycle. The SP6 and T7 promoter primers were then added and PCR amplification was performed, resulting in a 1.4-kb band containing the entire *Kpn*I-to-*Hin*dIII fragment with a T-to-A change at nt 10334, which in turn encoded an amino acid substitution from His-41 to Gln. The amplified DNA was digested with *Kpn*I and *Hin*dIII and ligated into the *Kpn*I and *Hin*dIII sites of pGEM-3Zf(−). Mutagenesis of other residues at His-41 and His-127 was performed in a similar manner as described above. Since there is a *Bam*HI site between Cys-142 and Cys-145, substitution of Arg for Cys-145 was performed by PCR site-directed mutagenesis with oligonucleotide Cys-145 left (5′-GTGGATCCCGCGGTTCTGTA-3′) and an SP6 promoter primer. The resultant fragment was digested with *Bam*HI and *Hin*dIII and ligated into the pGpro plasmid from which the *Bam*HI-to-*Hin*dIII piece had been excised. Mutagenesis of Ile at Cys-142 was performed by priming with oligonucleotide Cys-142 right (5′-GCAGGATCCAATTAGAAAGG-3′) and a T7 promoter primer. The resultant product was digested with *Kpn*I and *Bam*HI and cloned into the pGpro plasmid from which the *Kpn*I-to-*Bam*HI piece had been excised. Mutagenesis of other residues at Cys-142 and Cys-145 was similarly performed. All specific mutations were confirmed by bidirectional sequencing (Sequenase II; U.S. Biochemicals) according to the manufacturer's instructions).

In vitro transcription and translation. Recombinant plasmids were transcribed by using T7 RNA polymerase in a coupled in vitro transcription-translation reticulocyte lysate system (TnT; Promega), according to the manufacturer's instructions. Approximately 0.5 μg of plasmid DNA was incubated at 30°C with 12.5 μl of TnT lysate, 1 μl of TnT reaction buffer, 0.5 μl of T7 RNA polymerase, 20 U of RNasin, 0.5 μl of 1 mM methionine-free amino acid mixture, and 20 μCi of [^{35}S]methionine in a final volume of 25 μl. Samples were taken at various time points and mixed with an equal volume of 2× sodium dodecyl sulfate (SDS) loading buffer (17) and electrophoresed on an SDS-polyacrylamide (5 to 18% gradient) gel. During pulse-chase experiments, transcription and translation were terminated by the addition of cycloheximide (5 μg/ml), RNase (10 μg/ml), and excess cold methionine at the times indicated for the individual experiments.

In *trans* proteinase cleavage assay. Site-directed mutants of pGpro were expressed in the presence of [^{35}S]Met. The parental pGpro construct was similarly expressed, but in the presence of unlabeled L-Met only, for 40 min; this was followed by termination of transcription and translation by adding RNase (10 μg/ml) and cycloheximide (5 μg/ml) for 5 min. Following termination of transcription and translation, labeled mutant and unlabeled pGpro translation reaction lysates were mixed 1:1 and incubated for an additional 135 min. Reaction products were electrophoresed on SDS-polyacrylamide (5 to 18% gradient) gels, and the gels were treated with dimethyl sulfoxide-PPO (2,5-diphenyloxazole) and visualized by fluorography. The reaction mixtures were checked for expression and processing from the pGpro construct by the addition of [^{35}S]Met to an aliquot of the unlabeled reaction mixture after treatment with RNase and cycloheximide and incubation for an additional 135 min.

Peptide radiosequencing. In vitro transcription and translation were performed in a total volume of 200 μl with 8.0 μg of pGpro DNA in the presence of 160 μCi of [^{35}S]methionine (Dupont, NEN) or 400 μCi of [^{3}H]valine (Amersham) for 120 min at 30°C. The products were mixed with Laemmli sample buffer and separated on an SDS-polyacrylamide (5 to 18% gradient) gel for 16 h at 60 V. After electrophoresis, the products were transferred to a polyvinylidene difluoride membrane at 50 V at 4°C for 6 h in transfer buffer containing 25 mM Tris-base, 192 mM glycine, and 10% (vol/vol) methanol. After transfer, the polyvinylidene difluoride membrane was air dried and exposed to X-ray film. Radiolabeled proteins were identified by autoradiography, and the corresponding bands were excised from the polyvinylidene difluoride membrane and subjected to amino-terminal sequencing on an Advanced Biotechnologies microsequencer. The amino acid fraction from each cycle was quantitated in a Beckman scintillation counter.

RESULTS

Protein expression and processing during in vitro translation of pGpro. The purified pGpro plasmid was in vitro transcribed and translated in the TnT reticulocyte lysate system (Promega) in the presence of [^{35}S]methionine, and the samples were analyzed by gradient SDS-polyacrylamide gel electrophoresis (PAGE) at time points between 15 and 300 min (Fig. 2A). Two proteins with molecular masses of 38 and 33 kDa were detected within 15 min after the addition of DNA. By 45 min, the 38- and 33-kDa bands became less prominent and new proteins with estimated molecular masses of 32 and 27 kDa were detectable. Within 180 min, the 38- and 33-kDa proteins were barely detectable, while the 32- and 27-kDa species continued to accumulate for up to 150 min and were

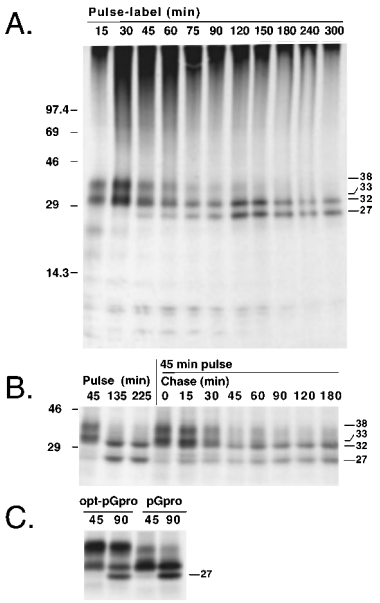

FIG. 2. Pulse-label and pulse-chase analyses of protein processing by pGpro and opt-pGpro constructs. In vitro transcription-translation of pGpro and opt-pGpro were carried out in the presence of [^{35}S]methionine. All samples were electrophoresed on an SDS-polyacrylamide (5 to 18% gradient) gel. Molecular mass markers are indicated on the left, and the estimated masses (in kilodaltons) of translation products are shown on the right. (A) Pulse-label translation of pGpro. Samples were taken from the reaction mixture at the time points indicated above the lanes. (B) Pulse-label translation and pulse-chase translation of pGpro. Samples were obtained at the times indicated. During the pulse-chase, transcription and translation were terminated at 45 min by the addition of cycloheximide, RNase, and cold methionine as described in Materials and Methods. Samples were taken from the reaction mixture at the times indicated above the lanes. (C) Pulse-label translation of pGpro and opt-pGpro. Samples were taken at 45 and 90 min.

stably detectable for up to 5 h. This pattern of expression was reproduced with several independently cloned *Kpn*I-*Hin*dIII fragments in pGEM-3Zf(−), using the TnT system. Three polypeptides with masses of less than 14.3 kDa also were detected during this prolonged pulse-label experiment, beginning at 15 min. Cleavage at the first QS dipeptide in the pGpro construct predicts a 9.7-kDa amino-terminal product, which may be consistent with the size of one of these polypeptides. However, the times of appearance and diminution of these products did not directly correlate with those of p27, and thus they may instead have been premature termination products which were subsequently degraded.

The result of the pulse-label experiment indicated that the 32- and 27-kDa proteins might be proteolytically cleaved from the 38- or 33-kDa precursors. In order to further examine the precursor-product relationship, pulse-chase assays were performed (Fig. 2B). Since we were interested in the fate of the 38- and 33-kDa proteins, a 45-min labeling period was chosen in order to maximize incorporation into larger proteins, which then could be monitored during the prolonged chase of up to 180 min. The pattern of expression and processing in the pulse-chase assay was nearly identical to that in the pulse-label experiment; specifically, a decrease in 38- and 33-kDa species was noted with stable accumulation of 32- and 27-kDa proteins, supporting the conclusion that the 32- and 27-kDa proteins were processed polypeptides. A repeat pulse-label translation performed in parallel with the pulse-chase reaction (Fig. 2B) gave a clear pulse-chase pattern of processing, indicating that de novo translation of the 38- and 33-kDa proteins probably ceased within 120 min. The results of this experiment were completely consistent with those of the initial pulse-label translation but more clearly demonstrated the disappearance of the 38- and 33-kDa proteins and stable accumulation of the 32- and 27-kDa proteins.

It was interesting that no protein of the predicted full length of the clone (48 kDa) was detected, suggesting either that rapid processing of proteins was occurring or that premature termination of transcription or translation was the rule with this plasmid. In order to eliminate the possibility that a stop codon might be introduced by reverse transcription-PCR, the sequence of pGpro was examined, revealing a 3-nt insertion at nucleotide positions 11073 to 11074 compared with the published sequence (4), resulting in an additional in-frame leucine residue (data not shown). No change was found which would cause termination of translation of pGpro.

Although the first AUG of pGpro was in a relatively good translation initiation context (GAGGAAATGG), we wanted to investigate the possibility of alternate translation initiation sites. We therefore inserted an optimal AUG immediately before the *Kpn*I site (TGCCGCCATG-G), in order to determine if this change would direct translation of a full-length polypeptide with a molecular mass of 48 kDa. In vitro expression and processing of this opt-pGpro was performed, with samples collected at 45- and 90-min pulse points (Fig. 2C). The 33-, 32-, and 27-kDa proteins were detected as with pGpro; however, a 39-kDa protein band was detected instead of the 38-kDa protein seen with pGpro. This size difference was attributable to the distance between the inserted optimal AUG and the natural first AUG of pGpro, strongly suggesting that translation of the 38-kDa protein from pGpro was initiating at the first methionine in the construct and further suggesting that the 38-kDa protein either terminated prior to the second QS dipeptide or was cleaved upstream from the second QS site. Because there was no difference in the putative proteolytic cleavage products between opt-pGpro and pGpro and since pGpro appeared to initiate translation at the 5′ end of the construct, we chose to use pGpro as the basis for subsequent experiments. In addition, since the 27-kDa protein was the putative cleavage product which was the most reproducibly detectable and most easily distinguishable from other proteins for all the constructs tested, we chose to use it as a marker of proteolytic activity in all subsequent experiments.

Mutations at putative catalytic amino acid residues abolish protein processing. Site-directed mutants of pGpro were constructed and utilized to determine if protein processing was dependent on proteinase activity encoded in pGpro and, if so, to identify residues essential for catalytic activity of the proteinase. Residues were selected for PCR site-directed mutagenesis on the basis of the predictions of Lee et al. (20) and Gorbalenya et al. (14, 15). His-41 and His-127 were chosen for mutagenesis since His-41 was a predicted catalytic residue and His-127 was the closest His residue to His-41. Cys-142 and Cys-145 were chosen because Cys-145 was a predicted catalytic residue and Cys-142 was close to Cys-145 and was located in an identical tripeptide motif (C-142GSC-145GS) in both the A59 and JHM sequences. All residue numbers are based on labeling Ser-3335 from the amino terminus of the gene 1 polyprotein as residue 1.

In vitro transcription-translation of pGpro resulted in p27 cleavage by 45 min (Fig. 3). The pGpro mutants pGproQ_{41} (H-41→Q), pGproG_{41} (H-41→G), pGproC_{41} (H-41→C), and pGproR_{145} (C-145→R) failed to cleave p27, demonstrating that mutations at putative catalytic sites His-41 and Cys-145 abolished the proteinase activity. In contrast, the pGproM_{127} (H-127→M), pGproR_{142} (C-142→R), and pGproI_{142} (C-142→I)

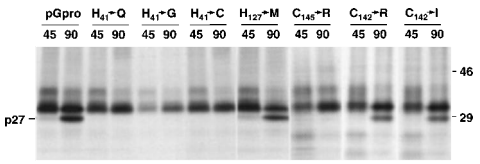

FIG. 3. Pulse-label translation of pGpro mutants. In vitro transcription-translation of the pGpro construct and pGpro mutants were carried out in the presence of [^{35}S]methionine as described in Materials and Methods. Samples were taken at the times (in minutes) indicated above the lanes and analyzed by SDS-PAGE (5 to 18%) gradient and fluorography. H-41→Q indicates that the histidine residue at position 41 was mutated to glutamine. Other mutants are similarly labeled. Molecular mass markers (in kilodaltons) are indicated on the right. The location of p27 is indicated on the left.

constructs resulted in a pattern of cleavage identical to that of the pGpro construct. The failure of protein processing by the mutant constructs indicated that the proteolytic activity was encoded by pGpro rather than derived from the proteinase within the reticulocyte lysate. It also demonstrated that His-41 and Cys-145 were essential residues for proteinase activity. Finally, these results confirmed that the pGpro construct also encoded at least one substrate cleavage site in the expressed polypeptide. We, therefore, next sought to determine whether the proteinase expressed from pGpro was active in *trans*.

Cleavage activity in *trans* by the pGpro-encoded protein. In order to examine the *trans* cleavage activity, the mutants pGproQ_{41} and pGproR_{145} were translated in the presence of [^{35}S]Met, and the parental pGpro construct was concurrently translated in the presence of unlabeled methionine. Following termination of pGpro translation as described in Materials and Methods, the labeled mutant polypeptides were mixed with the unlabeled translation products of pGpro and incubated for an additional 135 min (Fig. 4). As anticipated, p27 was efficiently cleaved during translation of labeled pGpro. In contrast, when the mutant constructs pGproQ_{41} and pGproR_{145} alone were translated in the presence of [^{35}S]Met, after which followed termination of the reaction and additional incubation, no cleavage products were seen. The addition of unlabeled pGpro translation products to reaction mixtures containing either of these mutant polypeptides resulted in cleavage of p27. During this experiment, *trans* cleavage activity was relatively inefficient on the His-41→Gln mutant compared with that on pGpro, but on the Cys-145→Arg mutant it was very efficient, with amounts of p27 comparable to those seen during translation of pGpro

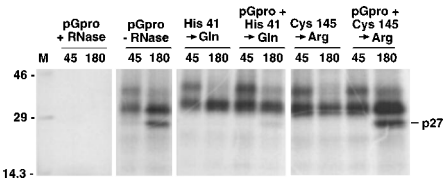

FIG. 4. In *trans* proteolytic activity of the pGpro-encoded polypeptide. In vitro transcription-translation of pGpro and mutants was carried out as described in Materials and Methods. All mutant constructs were translated in the presence of [^{35}S]methionine for 40 min; this was followed by termination of reactions with cycloheximide and RNase. The radiolabeled mutant polypeptides were then mixed with unlabeled pGpro polypeptide for an additional 135 min, when final samples were obtained (180 min). pGpro+RNase, unlabeled translation for 40 min, termination with RNase and cycloheximide for 5 min, and addition of [^{35}S]Met for 135 min (samples were taken at the times [in minutes] indicated). pGpro-RNase, translation of pGpro in the presence of [^{35}S]Met for 180 min (samples were taken at the times [in minutes] indicated). Samples were analyzed by SDS-PAGE (5 to 18%) gradient and fluorography. The position of p27 is indicated on the right, and molecular mass markers (M; in kilodaltons) are indicated on the left.

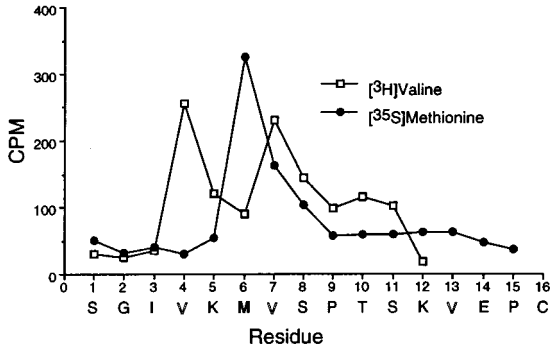

FIG. 5. Radiosequencing of p27 amino terminus. Peptide radiosequencing was performed as described in Materials and Methods, either with [^{3}H]valine- or [^{35}S]methionine-labeled protein. The amino acid fraction from each cycle was quantitated in a Beckman scintillation counter. Total sample counts per minute of either radioisotope are along the *y* axis, and the residue number and determined sequence are along the *x* axis.

alone. When the pGpro construct was translated with unlabeled methionine for 40 min, treated with RNase and cycloheximide for 5 min, and then incubated in the presence of [^{35}S]Met for another 135 min, no [^{35}S]Met incorporation occurred and no protein bands were detected. Thus, the possibility that the p27 detected during the *trans* cleavage assay was the product of de novo translation and processing of pGpro after mixing was eliminated. This experiment demonstrated that the pGpro-encoded proteinase could act in *trans* to cleave the mutant polypeptide and that the *trans* activity was efficient in the cleavage of p27. Although changes in the amino acid sequence introduced during mutagenesis may have affected the conformation and thus the *trans* cleavage efficiency, it was clear that the pGpro-encoded proteinase could cleave in *trans* and it was confirmed that the expressed polypeptide is a substrate for the proteinase.

Identification of proteinase cleavage site within the pGpro-expressed polypeptide. Because p27 was a discreet proteolytic product of the proteinase encoded by pGpro, we sought to determine its amino terminus. The protein was expressed and transferred, and the amino-terminal 15 amino acid residues were determined by radiosequencing with [^{3}H]valine and [^{35}S]methionine as described in Materials and Methods (Fig. 5). The pattern of peaks indicated that cleavage occurred between the Gln-3334 and Ser-3335 residues as numbered from the amino terminus of the ORF 1a polyprotein. There are at least three other locations within the pGpro coding region which could have an identical pattern of methionine residues, but none with a similar pattern of valine residues; the combination of the two is seen nowhere else within the pGpro-encoded protein. Based on the nucleotide sequence of gene 1, the −5-to-+5 sequence surrounding this cleavage site is TSF LQSGIVK, confirming the prediction of Lee et al. (20).

DISCUSSION

Our study demonstrates that the region of MHV-A59 gene 1 between nt 9925 and 11273 from the 5′ end encodes a proteinase which is active in vitro. In addition, the polypeptide encoded by this region contains at least one cleavage site for the proteinase, and the proteinase is able to cleave at this glutamine-serine dipeptide in *trans*. We also have confirmed that previously predicted histidine and cysteine residues are essential for proteolytic activity of the proteinase.

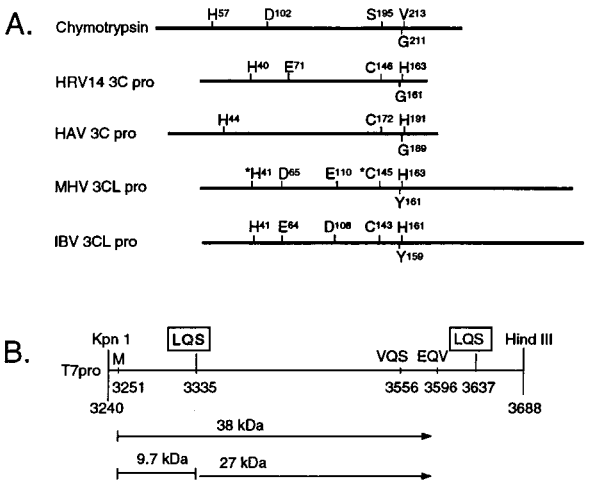

FIG. 6. Comparison of coronavirus 3C-like proteinases with chymotrypsin and serine-like proteinases of plus-strand RNA viruses. (A) Schematic drawing of known or predicted catalytic and substrate binding domains of several serine or serine-like proteinases. The names of the proteinases are shown on the left. The catalytic and substrate binding residues in chymotrypsin, human rhinovirus 14 (HRV14), and hepatitis A virus (HAV) have been experimentally confirmed by crystallography. In MHV-A59 3CLpro, only the *His-41 and *Cys-145 have been experimentally confirmed as essential residues. D-65(E-64) and E-110(D-108) are relatively conserved Asp or Glu residues of MHV-A59 and IBV which could serve as third members of the catalytic subunit of the 3CLpro, but they have not been investigated. The putative binding domain YXH has not been confirmed for any coronavirus. (B) Schematic of pGpro showing coding region, initiating methionine (M), location of confirmed LQS cleavage site (LQS-3335), possible alternative cleavage sites (VQS and EQV), and predicted carboxy terminus of 3CLpro (LQS-3637). Residue numbers numbered from the amino terminus of the gene 1 polyprotein are shown below the line. Also shown are the probable relationships between p38 and p27.

Translation and processing of the pGpro polypeptide. We can draw several conclusions from the data in this study concerning the translation, processing, and activity of the MHV-A59 3CLpro. It appears that translation of pGpro initiated at the first natural AUG, yielding a protein with a molecular mass of 38 kDa. On the basis of its size, this polypeptide would be predicted to terminate approximately 40 amino acids upstream from the predicted carboxy-terminal QS dipeptide cleavage site at residues 3637 to 3638 (Fig. 6B). We have identified the amino-terminal sequence of p27, which is cleaved at the previously predicted glutamine-serine dipeptide at amino acids 3334 to 3335. On the basis of its size, this protein also would be predicted to terminate 30 to 50 residues upstream from the QS dipeptide at residues 3637 to 3638. Finally, the amino-terminal serine of p27 is approximately 10 kDa downstream from the initial methionine in the pGpro construct. Together these results suggest that p38 is the precursor from which p27 is cleaved; this conclusion also would be consistent with the observation during pulse-chase experiments of transient products with molecular masses of less than 14.3 kDa.

While it is not yet possible to definitively conclude that p27 is 3CLpro, it is clear that the active proteinase is translated and is present within the maximum-size product, whose molecular mass is 38 kDa. Several other proteins, notably the 33-kDa early and 32-kDa late products, which are more consistent with the predicted size of the mature proteinase, are seen. However, we cannot account for these proteins on the basis of cleavage from p38, nor have we been able to sequence the amino termini, because of the heterogeneous nature of these proteins during prolonged translation reactions. It is interesting that p27 would contain the conserved essential His-41 and Cys-145 residues, as well as the predicted YXH substrate binding site (Fig. 6). In addition, secondary-structure algorithms predict a series of 12 to 13 beta sheets with amino- and carboxy-terminal helices within the p27 coding region, similar to the secondary structure of other viral serine-like proteinases. Finally, there are two potential alternate cleavage sites in the predicted region of p27 termination, specifically, QV at residues 3595 to 3596 and QS at residues 3555 to 3556, predicting proteins with molecular masses of 28 and 25 kDa, respectively. We are currently determining the carboxy terminus of p27 and working to identify the exact coding sequence of 3CLpro.

The role of the flanking hydrophobic domains, MP1 and MP2, is not yet clear. Because they are of equivalent size (approximately 300 amino acids), are demarcated by predicted 3CLpro cleavage sites, and directly flank the predicted proteinase, they have been predicted to play a role in membrane anchoring and presentation of the proteinase (20). The pGpro clone eliminated most of the amino and carboxy hydrophobic domains, and p38 would not be expected to contain any of the downstream domain. Thus, the full-length MP1 and MP2 domains are clearly not required for activity in vitro. Since we are investigating only one cleavage site in this study and the activity is not occurring in the context of the entire polyprotein, it is still possible that the hydrophobic domains may be involved in maturation of the polyprotein or in other proteinase-substrate interactions in the virus-infected cell.

We have shown that both His-41 and Cys-145 of MHV-A59 are essential for catalytic activity of the proteinase domain. His-41 is the only His in the N-terminal 127 residues of the proteinase counting from the predicted amino-terminal serine and is conserved between the IBV and MHV sequences. Cys-145 also is conserved in IBV (Cys-143) and MHV, although MHV has a duplication of the Cys-containing tripeptide (C-142GSC-145GS) whereas this is present only once in IBV (AGAC-143GS). The substitution mutations, His-41→Gln, His-127→Met, Cys-145→Arg, and Cys-142→Ile were carefully chosen to preserve the consensus secondary structure surrounding the mutagenized residues, so that the effect examined would be mainly based on amino acid changes rather than potential structural alteration (9, 12). Other residues also were substituted for His-41, His-127, Cys-145, and Cys-142 to directly compare identical substitutions at different potential catalytic sites. Mutagenesis of His-127 did not alter cleavage activity, supporting the role of His-41 as a catalytic residue. The specificity of Cys-145 as an essential residue in catalytic activity was confirmed by the fact that the Cys-142 mutation did not affect cleavage activity despite its close proximity to the catalytic residue.

Comparison of MHV-A59 3CLpro with cellular and viral proteinases. The coronavirus 3C-like proteinases have previously been predicted to belong to the serine-like superfamily of proteinases (14, 15). They are similar to cellular chymotrypsin structurally but utilize a cysteine instead of a serine as the active nucleophile residue. Like chymotrypsin, viral serine-like proteinases, including the MHV-A59 3CL proteinase, have been predicted to consist of a bilobular polypeptide, with each domain consisting of a unique beta-barrel structure. The catalytic triad residues His, Cys, and Asp(Glu) in most cases are located in the interface region between the two lobes. The Cys is substituted for the catalytic serine of chymotrypsin, but the protein is predicted to have a structure and mode of action similar to those of the cellular serine proteinase (Fig. 6). The 3CLpro of both IBV and MHV-JHM have been predicted to have conserved His and Cys residues within the catalytic domain, and we have confirmed the essential nature of these residues; however, there is no strong prediction of a well-

conserved Asp(Glu), though it has been predicted that Asp-65 of MHV might act as a third member of the triad (13).

The lack of a predicted conserved Asp(Glu) in a position analogous to that of Asp-102 of chymotrypsin is interesting for several reasons. The crystal structures of hepatitis A virus 3C proteinase (HAV 3Cpro) and human rhinovirus 3C proteinase (HRV 3Cpro) have recently been determined. The crystal structure of the HAV 3Cpro, as defined by Allaire et al. (1), revealed the predicted bilobular structure and interdomain cleft. Their results indicate that the amino and carboxy termini of the HAV 3Cpro are distant from the active site, making it improbable that the 3Cpro is released from the polyprotein by *cis* autoproteolysis. The overall structure suggests that an intermolecular (*trans*) cleavage releases the proteinase. This is intriguing in light of our data demonstrating that p27 is cleaved efficiently in *trans* by the MHV 3CLpro and suggests that a similar mechanism may be in operation.

The structure of the HAV 3Cpro also strongly indicates that Asp-84 does not interact with His-44 as predicted. The structure analysis suggests that the catalytic apparatus of HAV 3Cpro involves only a dyad, with Cys-172 as the nucleophile and His-44 as the general base. In contrast, the crystal structure of the HRV 3Cpro, as determined by Matthews et al., indicates that residues His-40, Glu-71, and Cys-146 form a linked cluster of three amino acids which have a geometry similar to that of the Ser-His-Asp catalytic triad found in cellular serine proteinases (21). Thus, the formation of a catalytic triad may not be a universal feature of viral serine-like proteinases, and the lack of a well-conserved and predictive Asp(Glu) in the coronaviruses may indicate that they do not require this residue for catalytic activity. It is possible, however, that they utilize a residue other than Asp(Glu) or that they utilize an Asp(Glu) in a different orientation relative to the His and Cys residues. Comparison of the MHV-A59 and IBV sequences reveals a conserved Asp-53 in both MHV and IBV, but it is in rather close proximity to His-41. More interesting are the Asp-65 and Glu-110 of MHV, which are inverted in IBV to Glu-64 and Asp-108, respectively (Fig. 6). The IBV Glu-64 has previously been aligned with the catalytic Asp-85 of poliovirus (15), and the MHV Asp-65 has recently been predicted to be a member of a catalytic triad (13). We are currently investigating whether any of these residues play a role in the catalytic activity of 3CLpro.

This study confirms the existence of a serine-like proteinase in MHV-A59 ORF 1a and demonstrates that predicted His-41 and Cys-145 residues are essential for its activity. During in vitro translation, neither of the complete flanking hydrophobic regions (MP1 or MP2) is required for proteolytic activity. In addition, this proteinase is able to cleave efficiently in *trans* at a glutamine-serine dipeptide. All of these observations, in addition to the fact that this proteinase is located in the midst of the largest known viral polyprotein, suggest that the study of the MHV 3C-like proteinase will add significantly to an understanding of both coronavirus replication and the functions of viral proteinases in general.

ACKNOWLEDGMENTS

We thank Susan Weiss for ORF 1a clone 917 and T. Dermody for helpful discussions.

This work was supported by Public Health Service grant AI-26603 from the National Institutes of Health and by the Elizabeth B. Lamb Center for Pediatric Research.

REFERENCES

1. **Allaire, M., M. M. Chernala, B. A. Malcolm, and N. G. James.** 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature (London) **369:**72–76.
2. **Baker, S. C., K. Yokomori, S. Dong, R. Carlisle, A. E. Gorbalenya, E. V. Koonin, and M. M. C. Lai.** 1993. Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. J. Virol. **67:**6056–6063.
3. **Baric, R. S., S. A. Stohlman, M. K. Razavi, and M. M. C. Lai.** 1985. Characterization of leader-related small RNAs in coronavirus infected cells: further evidence for leader-primed mechanism of transcription. Virus Res. **3:**19–33.
4. **Bonilla, P., A. Gorbalenya, and S. Weiss.** 1994. Mouse hepatitis virus strain A59 RNA polymerase gene ORF 1a: heterogeneity among MHV strains. Virology **198:**736–740.
5. **Brayton, P. R., M. M. C. Lai, C. C. Patton, and S. A. Stohlman.** 1982. Characterization of two RNA polymerase activities induced by mouse hepatitis virus. J. Virol. **42:**847–853.
6. **Bredenbeek, P. J., C. J. Pachuk, A. F. H. Noten, J. Charite, W. Luytjes, S. R. Weiss, and W. J. M. Spaan.** 1990. The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. Nucleic Acids Res. **18:**1825–1832.
7. **Brierley, I., M. E. G. Boursnell, M. M. Binns, B. Billimoria, V. C. Blok, T. D. K. Brown, and S. C. Inglis.** 1987. An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J. **6:**3779–3785.
8. **Brierley, I., P. Digard, and S. C. Inglis.** 1989. Characterization of an efficient coronavirus ribosomal frameshift signal: requirement for an RNA pseudoknot. Cell **57:**537–547.
9. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. Relat. Areas Mol. Biol. **47:**45–148.
10. **Denison, M. R., and S. Perlman.** 1986. Translation and processing of mouse hepatitis virus virion RNA in a cell-free system. J. Virol. **60:**12–18.
11. **Dougherty, W. G., and B. L. Semler.** 1993. Expression of virus-encoded proteinase: functional and structural similarities with cellular enzymes. Microbiol. Rev. **57:**781–822.
12. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120:**97–120.
13. **Gorbalenya, A., and E. Koonin.** 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. Sect. D. Physiochem. Biol. Rev. **11:**1–81.
14. **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. FEBS Lett. **243:**103–114.
15. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Coronavirus genome: prediction of putative functional domains in the nonstructural polyprotein by comparative amino acid sequence analysis. Nucleic Acids Res. **17:**4847–4861.
16. **Higuchi, R., B. Krummel, and R. K. Saiki.** 1988. A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res. **16:**7351–7367.
17. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
18. **Lai, M. M. C.** 1990. Coronavirus: organization, replication, and expression of genome. Annu. Rev. Microbiol. **44:**303–333.
19. **Lai, M. M. C., C. D. Patton, R. S. Baric, and S. A. Stohlman.** 1983. Presence of leader sequence in the mRNA of mouse hepatitis virus. J. Virol. **46:**1027–1033.
20. **Lee, H.-J., C.-K. Shieh, A. E. Gorbalenya, E. V. Koonin, N. LaMonica, J. Tuler, A. Bagdzhadhzyan, and M. M. C. Lai.** 1991. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology **180:**567–582.
21. **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.
22. **Pachuk, C. J., P. J. Bredenbeek, P. W. Zoltick, W. J. M. Spaan, and S. R. Weiss.** 1989. Molecular cloning of the gene encoding the putative polymerase of mouse hepatitis coronavirus, strain A59. Virology **171:**141–148.
23. **Shieh, C., L. Soe, S. Makino, M. Chang, S. Stohlman, and M.-C. Lai.** 1987. The 5′-end sequence of the murine coronavirus genome: implications for multiple fusion sites in leader-primed transcription. Virology **156:**321–330.
24. **Spaan, W. J. M., D. Cavanagh, and M. C. Horzinek.** 1988. Coronaviruses: structure and genome expression. J. Gen. Virol. **69:**2939–2952.