# 16

# PROCESSING OF THE MHV-A59 GENE 1 POLYPROTEIN BY THE 3C-LIKE PROTEINASE

M. R. Denison,[1,2,3] A. C. Sims,[2,3] C. A. Gibson,[2,3] and X. T. Lu[1,3]

[1]Department of Pediatrics
[2]Department of Microbiology and Immunology
[3]Elizabeth B. Lamb Center for Pediatric Research
Vanderbilt University Medical Center
Nashville, Tennessee

## 1. ABSTRACT

The 3C-like proteinase of mouse hepatitis virus (MHV-3CLpro) is predicted to cleave at least 10 sites in the gene 1 polyprotein, resulting in processing of proteinase, polymerase and helicase proteins from the polyprotein. We have used *E. coli* expressed recombinant 3CLpro (r3CLpro) to define cleavage sites in carboxy-terminal region of the ORF 1a polyprotein. Polypeptides containing one or more putative 3CLpro cleavage site were translated *in vitro* from subcloned regions of gene 1, and the polypeptides were incubated with r3CLpro. Analysis of the cleavage products confirmed several putative cleavage sites, as well as identifying cleavage sites not previously predicted by analysis of the MHV sequence. Antibodies directed against a portion of the ORF 1a polyprotein were used to probe virus infected cells, and detected proteins that correspond to the cleavage sites used by 3CLpro *in vitro*. These results suggest that MHV 3CLpro cleaves at least 7 sites in the ORF 1a polyprotein, and that the specificity of 3CLpro for cleavage site dipeptides may be broader than previously predicted.

## 2. INTRODUCTION

Gene 1 of MHV-A59 encodes a fusion polyprotein with a predicted mass of 803 kDa (Gorbalenya *et al.*, 1989; Lee *et al.*, 1991; Bonilla *et al.*, 1994). Expression of the entire gene 1 polyprotein requires translation of two open reading frames, ORFs 1a and 1b. Since these ORFs are in different reading frames, ORF 1b can only be expressed if a ribo-

*Coronaviruses and Arteriviruses*, edited by Enjuanes *et al.*
Plenum Press, New York, 1998

somal frameshift occurs at the end of ORF 1a (Brierley, 1989). The ORF 1a portion of the gene 1 polyprotein encodes two experimentally confirmed proteinases, PLP-1 and 3CLpro, as well as an additional proteinase, PLP-2, for which no activity has yet been identified. The MHV 3CLpro has been shown to autoproteolytically liberate itself from the nascent polyprotein, both *in vitro* and in virus-infected cells (Lu *et al.*, 1995; Lu *et al.*, 1996). Multiple other cleavage sites have been predicted for 3CLpro, all of which are carboxy terminal to 3CLpro in the polyprotein (Lee *et al.*, 1991). At least five sites appear to be highly conserved among four sequenced coronaviruses.

We were interested in the region between the end of the 3CLpro encoding region and the end of ORF 1a (nt 11115–13589). This region is predicted to encode the predominantly hydrophobic MP-2 domain, as well as the remainder of the ORF 1a polyprotein with 4 predicted 3CLpro cleavage sites. We therefore undertook a study to define the cleavage sites in this portion of the ORF 1a polyprotein. In addition we sought to determine if any of the putative protein products could be identified in virus-infected cells.

## 3. METHODS

### 3.1. Cloning and Expression of Recombinant 3CLpro (r3CLpro)

Active MHV 3CLpro was cloned and expressed in *E. coli* as described in the accompanying paper. Briefly, the 3CLpro domain was expressed as a fusion protein with maltose binding protein (MBP-r3CLpro), partially purified on an amylose affinity column, and separated from MBP using factor Xa. The partially purified r3CLpro preparation was used for all experiments.

### 3.2. Cloning and Expression of Polypeptides Containing Putative 3CLpro Cleavage Sites

Subclones of the region of ORF 1a between nt 11991 and 13118 were constructed using PCR from a cDNA of the 3' 2.5 kb of the ORF 1a region of gene 1 (Figure 2). Left primers included an Nco 1 site and right primers an Xho 1 site, allowing for ligation into pET-23d (Novagen). This vector contains an optimized AUG at the insertion site, allowing for full-length expression of all constructs. The recombinant plasmids were used to program expression of polypeptides in an combined *in vitro* transcription/translation reticulocyte lysate system (TnT-Promega) in the presence of [$^{35}$S]methionine and [$^{35}$S]cysteine. Reactions were terminated by addition of SDS-PAGE sample buffer, or in the case of *trans* cleavage assays, by the addition of cycloheximide.

### 3.3. *Trans* Cleavage Assays

Radiolabeled polypeptides were incubated with elutate from the amylose affinity column containing r3CLpro, maltose binding protein and factor Xa at 30°C for 4 hours. Typically, 1–2 ul or *in vitro* translation reaction was incubated with 5 ul of r3CLpro containing solution. Reactions were terminated by addition of SDS-PAGE sample buffer and boiling for 5 minutes, prior to electophoresis on SDS- 5–18% gradient polyacrylamide gels.

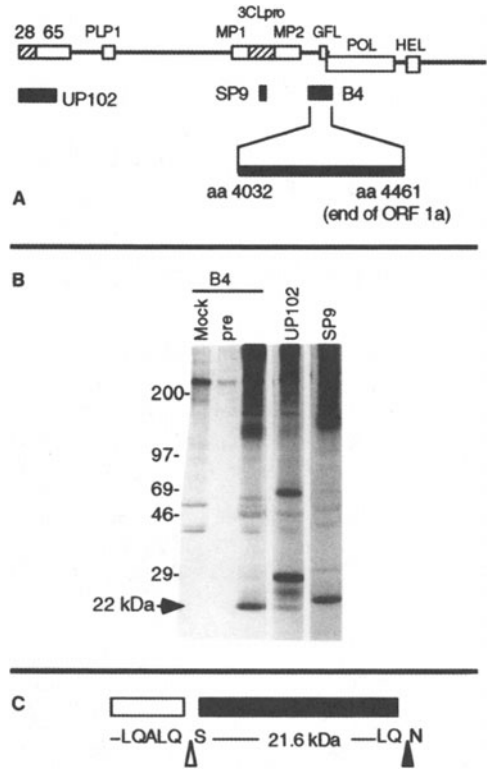


**Figure 1.** Identification of gene 1 proteins in MHV infected DBT cells. A. Scheme of gene 1 showing location of fusion proteins or peptides used to induce antisera UP102, SP9 and B4. The expression of the B4 fusion protein and antibody induction is described in the methods section. B. Immunoprecipitation of radiolabeled gene 1 proteins from virus infected cells. Mock infected (Mock) or infected cells were radiolabeled with [$^{35}$S]methionine. Lysates were immunoprecipitated with preimmune serum (pre) or immune sera (B4, UP102, or SP9). Molecular mass markers are to the left of the gel and the location of the B4-precipitated 22 kDa protein is noted. C. The location of the confirmed amino terminal cleavage site of the 22 kDa protein is shown (open arrowhead) and the putative carboxy terminal site is indicated (black arrowhead).

## 3.4. Antibody Production and Immunoprecipitation

The B4 protein was expressed in *E. coli* in the pQE-30 vector (Quiagen). the B4 subclone extended from the Xba 1 site at nt 12302 to the Kpn 1 site at 13901 (aa 4032 - 4461) (Figure 1). This clone contains the ORF 1a/ ORF 1b frameshift and potentially encodes 114 aa of ORF 1b; however the size of the *E. coli* expressed protein indicated that only the ORF 1a portion of this construct was translated. The 46 kDa protein was excised from SDS-polyacrylamide gels and used to induce polyclonal antibodies in rabbits. This antiserum was called B4. Antisera UP102 and SP9 have been previously described (Denison *et al.*, 1995; Lu *et al.*, 1996). Virus infections and immunoprecipitation experiments were performed as previously described (Denison *et al.*, 1992).

## 3.5. Amino Terminal Radiosequencing

The 22 kDa protein immunoprecipitated from MHV-infected DBT cells by the B4 antiserum was excised from the polyacrylamide gel following localization by autoradiography, and was transferred to a polyvinylidene difluoride (PVDF) membrane. Transfer of the protein and radiosequencing was performed on a ABI 470 sequencer as previously described (Lu *et al.*, 1995).

# 4. RESULTS

## 4.1. Detection of Gene 1 Proteins in MHV-Infected DBT Cells

The B4 antiserum was used to detect proteins at the carboxy terminus of ORF 1b in MHV-A59 infected DBT cells, along with the previously characterized UP102 and SP9 antisera (Figure 1). UP102 detected the amino terminal p28 and p65 proteins; SP9 precipitated 3CLpro (p27) from virus-infected cells. B4 detected a product of 22 kDa that was not detected by preimmune serum or in mock infected cells. Because of the extent of the B4 protein used to induce the B4 antiserum, it was not possible to define the amino and carboxy termini based on the antibody specificity alone.

Thus the [$^{35}$S]met labeled 22 kDa protein was immunoprecipitated, blotted to PVDF membrane, and used for radioterminal sequencing. The result indicated that the amino terminus of the B4 precipitated 22 kDa protein was cleaved at its amino terminus between Glutamine$_{4013}$ and Serine$_{4014}$ residues in the ORF 1a polyprotein. This Q_S dipeptide is within the sequence NTVLQALQ_SEFVN, which also contains the only directly adjacent possible 3CLpro cleavage sites in the gene 1 polyprotein, LQ_ALQ_S. The LQ_S is conserved in the gene 1 sequence of infectious bronchitis virus, transmissible gastroenteritis virus, and the human coronavirus 229E, whereas the LQA is not present in any of these sequences (Boursnell *ét al.*, 1987; Herold *et al.*, 1993; Eleouet *et al.*, 1995).

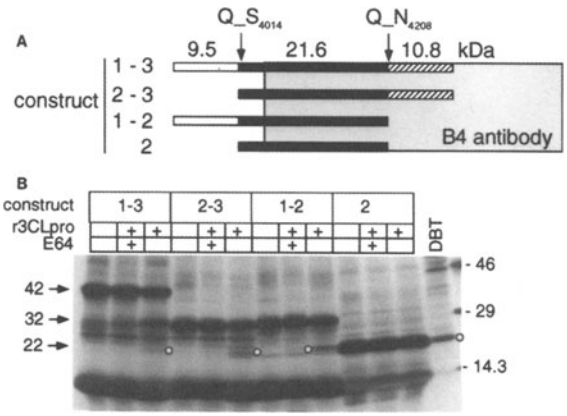


**Figure 2.** ORF 1a subclones, *in vitro* translation and 3CLpro *trans* cleavage. A. The ORF 1a subclones were constructed as described in Methods. All constructs had an optimal AUG at the 5' end in the construct. The constructs are labeled by the portions of ORF 1a potentially expressed: 1- 9.5 kDa fragment amino terminal to the LQ_S$_{4014}$; 2- 22 kDa protein precipitated by B4 antibody (location shown by grey box); 3 - 10.8 kDa fragment carboxy terminal to putative LQ_N$_{4208}$. B. *In vitro* translation and *trans* cleavage assay. The translation of the ORF 1a subclones is shown. Incubation of IVT products with recombinant 3CLpro (r3CLpro) and the cysteine proteinase inhibitor E64d is indicated by (+). Molecular mass markers are to the right of the gel and the location of specific translation products and cleavage products is indicated by numbers with arrows to the left of the gel. Location of 22 kDa proteins translated (construct 2) or cleaved are noted by clear circles. DBT indicates immunoprecipitation of proteins from MHV infected DBT cells by B4 antiserum.

## 4.2. Identification of a New 3CLpro Cleavage Site and Determination of the Carboxy Terminus of the 22 kDa Protein

Based on the apparent mass of the B4 precipitated protein and the determined amino terminus, we predicted that the 22 kDa protein must have its carboxy terminal cleavage at the sequence $VVLQ_{4207}\_N_{4208}NEL$. The LQNNE sequence is completely conserved in MHV, IBV, HCV 229E and TGEV, and the LQ_N has recently been shown to be a cleavage site for the IBV 3CLpro (Liu *et al.*, 1997). We therefore constructed a series of subclones of ORF 1b containing 1 or 2 cleavage sites in this region, in order to determine if cleavage in fact could occur *in trans* at the LQ_N peptide by recombinant 3Clpro (Figure 2).

The construct 2 contained the precise region between the amino-terminus of the B4-precipitated 22 kDa protein ($Q\_S_{4014}$) and the predicted $LQ\_N_{4208}$ cleavage site. *In vitro* translation of this construct resulted in a 22 kDa protein that had the exact migration of the protein isolated from virus-infected cells. Incubation of this polypeptide with r3CLpro did not result in any cleavage of the protein. The construct 1–2 contained the 22 kDa protein, the amino-terminal LQ_S site and an amino terminal flanking fragment with a predicted mass of 9.5 kDa. Expression of this protein resulted in a dominant protein of 32 kDa, consistent with the total mass predicted for this polypeptide. Incubation with r3CLpro resulted in cleavage of a 22 kDa protein. Similarly, the 2–3 construct contained the 22 kDa protein, the putative carboxy terminal LQ_N cleavage site and a 10.8 kDa carboxy terminal fragment. *In vitro* translation of this construct also resulted in an approximately 32 kDa protein, and a 22 kDa cleavage product was seen after incubation with r3CLpro. Finally, the 1–3 construct containing both cleavage sites and amino and carboxy terminal fragments was translated to a 42 kDa precursor, and could be cleaved, albeit less efficiently, into 32 and 22 kDa products. All of the cleavages observed in this experiment could be blocked by the addition of the proteinase inhibitor E64. Thus the cleavages were specific, were mediated by 3CLpro, and occurred at the predicted sites.

## 5. DISCUSSION

We have previously shown that 3CLpro mediates autoproteolytic cleavage of the proteinase from the polyprotein (Lu *et al.*, 1995; Lu *et al.*, 1996; Lu and Denison, 1997), and similar results have been obtained for 3CLpro of HCV-229E and IBV (Ziebuhr *et al.*, 1995; Liu and Brown, 1995). In this study we have confirmed two additional cleavage sites, including one, LQ_N4208, that had not previously been predicted for MHV. We have also shown the product of cleavage seen *in vitro* is also a stable mature product of polyprotein processing in MHV infected cells. This suggests that we will be able to determine with accuracy the boundaries of mature proteins processed from the remainder of gene 1 polyprotein. It also suggests that additional sites may be used by the proteinase that do not conform to the predicted Q_ (S,A,G) consensus cleavage sites.

Based on the sites we have confirmed, along with the remaining predicted cleavage sites for r3CLpro, it is likely that up to 12 proteins are processed by 3CLpro during gene 1 translation. (Figure 3).

It is particularly intriguing that the region of the ORF 1a polyprotein between the MP-2 domain and the end of ORF 1a is likely to incorporate four mature proteins of 10, 22, 12, and 15 kDa. Only the putative 15 kDa protein has been predicted to have any similarity to known proteins, in this case a similarity to murine epidermal growth factor recep-

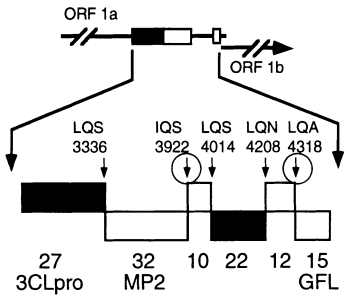


**Figure 3.** Model of 3CLpro ORF 1a polyprotein cleavage sites and protein products. A schematic of gene 1 with predicted or confirmed functional domains is shown at the top of the figure. The enlarged section shows the location of putative (black arrows) or confirmed (circled arrowheads) 3CLpro cleavage sites. Proteins that have been detected both *in vitro* and in virus-infected cells are shown as black boxes.

tor binding ligands (GFL). Thus it is likely that these proteins, including the confirmed 22 kDa protein, will mediate unique functions during MHV replication.

## ACKNOWLEDGMENTS

This work was supported by Public Health Service grant AI-26603 from the National Institute of Allergy and Infectious Diseases. Protein sequencing was performed in the shared resource of the Vanderbilt University Cancer Center (IP30CA68485).

## REFERENCES

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**: 736–40.

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.

Brierley, I., Digard, P., and Inglis, S. C., 1989, Characterization of an efficient coronavirus ribosomal framshift signal: requirement for an RNA pseudoknot, *Cell* **57**: 537–547.

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**: 316–20.

Denison, M. R., Zoltick, P. W., Hughes, S. A., Giangreco, B., Olson, A. L., Perlman, S., Leibowitz, J. L., and Weiss, S. R., 1992, Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events, *Virology* **189**: 274–84.

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**: 817–22.

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.

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.

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., Xu, H. and Brown, T., 1997, Proteolytic processing of the coronavirus infectious bronchitis virus 1a polyprotein: identification of a 10-kilodalton polypeptide and determination of its cleavage sites, *J Virol* **71**: 1814–1820.

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**: 3554–9.

Lu, Y. Q. and Denison, M. R., 1997, Determinants of mouse hepatitis virus 3C-like proteinase activity, *Virology* **230**: 335–342.

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.