Coronavirus polyprotein processing

S. R. Weiss[1], S. A. Hughes[1], P. J. Bonilla[1], J. D. Turner[1], J. L. Leibowitz[2],
and M. R. Denison[3]

[1] Department of Microbiology, University of Pennsylvania School of Medicine,
Philadelphia, Philadelphia, [2] Department of Pathology and Laboratory Medicine,
University of Texas Health Sciences Center, Houston, Texas, [3] Department of
Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee, U.S.A.

Summary. MHV gene 1 contains two ORFs in different reading frames. Translation proceeds through ORF 1a into ORF 1b via a translational frame-shift. ORF 1a potentially encodes three protease activities, two papain-like activities and one poliovirus 3C-like activity. Of the three predicted activities, only the more amino terminal papain-like domain has been demonstrated to have protease activity. ORF 1a polypeptides have been detected in infected cells by the use of antibodies. The order of polypeptides encoded from the 5′ end of the ORF is p28, p65, p290. p290 is processed into p240 and p50. Processing of ORF1a polypeptides differs during cell free translation of genome RNA and in infected cells, suggesting that different proteases may be active under different conditions. Two RNA negative mutants of MHV-A59 express greatly reduced amounts of p28 and p65 at the non-permissive temperature. These mutants may have defects in one or more viral protease activities. ORF 1b, highly conserved between MHV and IBV, potentially contains polymerase, helicase and zinc finger domains. None of these activities have yet been demonstrated. ORF 1b polypeptides have yet been detected in infected cells.

Introduction

The coronavirus mouse hepatitis virus (MHV) contains the largest known viral RNA genome, a 31 kilobase (kb) single stranded positive sense RNA [1, 2]. The replication strategy includes the replication of full-length genome RNA via a full-length negative strand RNA, as well as synthesis of a nested set of six positive-stranded subgenomic mRNAs [3]. These mRNAs are thought to be synthesized via a leader priming mechanism. Recent data have identified subgenomic negative strand

RNAs [4], but the mechanism of synthesis of these RNAs and their possible role in expression of mRNAs is not yet understood.

Gene 1 of the coronaviruses is presumed to encode the viral polymerase. The complete nucleotide sequences for gene 1 are now available for the coronavirus, avian infectious bronchitis virus (IBV) [5] and the JHM [2] and A59 ([6]; unpubl. res.) strains of MHV. The general structures of the genes for the avian and murine viruses are similar; most of the specific information described in this manuscript will be for MHV-A59. Gene 1 contains two large ORFs of approximately 14 (ORF 1a) and 8 (ORF 1b) kb. These ORFs overlap by about 75 nucleotides. ORF 1b is translated in the −1 frame with respect to ORF 1a. Translation is thought to begin at nucleotide 210 of ORF 1a and proceed through the end of ORF 1a where a translational frame shift allows translation of ORF 1b [2, 5–7]. Translation of ORF 1a of MHV-A59 predicts a polypeptide of 4 469 amino acids while a fusion polypeptide translated from ORFs 1a and 1b would predict a polypeptide of 7 202 amino acids [6]. While these very large polypeptides have not yet been detected in coronavirus-infected cells, we describe below the detection of smaller polypeptides presumably processed from these very large precursors by proteases encoded in gene 1. We will discuss the structure of gene 1 and the functional domains predicted from analysis of the sequence and then describe our recent work on the detection of the polypeptides encoded in gene 1 of MHV-A59.

Structure of gene 1, the putative polymerase gene

Figure 1 is a schematic diagram of the functional domains predicted to be encoded in gene 1 of MHV. There are two predicted papain-like protease domains and one polio 3C-like protease domain within ORF 1a ([2]; Bonilla et al., unpubl. data). The two papain-like domains contain the expected cysteine and histidine catalytic residues at similar spacing to the cellular papain-like enzymes; however, sequence comparison suggests that the viral and cellular enzymes are only distantly related [8]. An "X" domain located downstream of the first papain domain, is also found in IBV gene 1 and conserved among other groups of viruses, such as the alpha and rubi viruses [8]. The function of this domain has not yet been demonstrated. The coronavirus 3C-like protease domain sequence predicts a protease similar to the poliovirus 3C protease, but different from the chymotrypsin-like serine proteases in that the catalytic serine residue is replaced by a cysteine. The coronavirus 3C-like protease domain also differs from the poliovirus enzyme and from chymotrypsin like serine proteases in the following ways: 1) there is a tyrosine substituted for glycine in the putative substrate binding region of the coro-
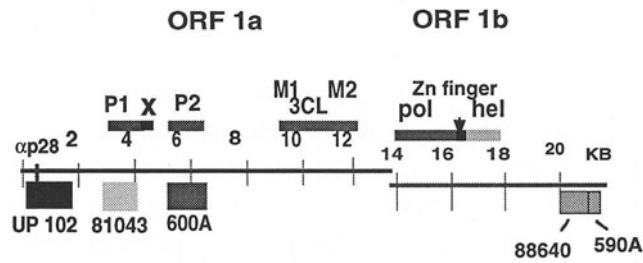
Fig. 1. Predicted functional domains encoded within gene 1 of MHV. ORF 1a and ORF 1b are shown with the predicted functional domains ([2] Bonilla et al., unpubl.). *P1* and *P2* designate the putative papain-like protease domains. *X* Conserved domain discussed in the text; *M1* and *M2* predicted membrane spanning domains; *3CL* polio 3C-like protease domain; *Pol* predicted polymerase domain; *hel* predicted dNTP binding (helicase) domain. Below the lines are shown the regions of genome against which gene 1 specific antisera (as discussed in the text) are directed

navirus sequences [2]; and 2) the predicted coronavirus enzymes lack the third catalytic residue of aspartic/glutamic acid [2]. The MHV 3C-like protease domain is bounded on either side by a potential membrane spanning domain [2]. Of these three predicted protease domains in MHV, only the more amino terminal papain-like enzyme has been demonstrated to have activity ([9]; see below).

ORF 1b is predicted to contain polymerase, dNTP binding (or helicase) and zinc finger domains [2, 5]. An unusual feature of the predicted polymerase domain in both MHV and IBV [2, 6] is that the usual GDD (glycine, aspartic acid, aspartic acid) core polymerase motif is replaced by SDD (serine, aspartic acid, aspartic acid). The NTP binding domain has several clusters of conserved amino acids. The most conserved "A" site motif (GKS) and "B" motifs (aspartic acid residue preceeded by three out five hydrophobic residues) thought to be the Mg^{+2} binding domain are present in the coronavirus sequences [10]. This predicted NTP binding domain is also thought to contain helicase activity because NTP binding proteins have homology with some bacterial helicases [10]. Between the polymerase and NTP binding domains is a cysteine rich putative zinc finger domain. This is thought to interact with nucleic acids during viral replication [2, 11].

Predictions of actual polypeptides to be generated from gene 1 were made by examining putative cleavage sites for the poliovirus 3C protease [2, 11]. Analysis of glutamine/glycine, glutamine/serine and glutamine/ alanine dipeptides (cleavage sites for the poliovirus 3C protease) and surrounding sequences both in IBV and MHV suggests that cleavages may occur to generate separate polypeptides containing the 3C-like protease, each of the potential membrane spanning domains, the poly-

merase and helicase/zinc finger domains. However, none of this has yet been confirmed experimentally. This analysis also predicts several other similar cleavage sites in other portions of ORF 1a and ORF 1b, for which functional domains have not yet been predicted.

ORF 1b is thought to be translated via a frame shift. Indeed, several labs have described a conserved "slippery sequence" (UUUUUAAAC in MHV) at the end of ORF 1a followed by a pseudoknot structure, both thought to be elements in frame shifting. Both the MHV [2, 6] and the IBV [7] sequences have been demonstrated to have frame shifting activity both during in vitro translation and in eukaryotic cells. We have also demonstrated that prokaryotic ribosomes will also frame shift when translating this sequence. The coronavirus frame shift is about 40% efficient [2, 6, 7]. It is interesting that ORF 1a, containing the proteases necessary to process the polymerase polypeptides, is probably expressed at a higher level than ORF 1b, which is believed to encode the actual polymerase activity.

Comparison of gene 1 of MHV-JHM, MHV-A59 and IBV

The above predicted functional domains are highly conserved between MHV-JHM and MHV-A59 and also between MHV and IBV. In fact, ORF 1b is 52.8% conserved at the amino acid level between JHM and IBV and 94.4% between JHM and A59 [2]. ORF 1a is less conserved than ORF 1b when comparing either A59 and JHM or MHV and IBV. ORF 1a of IBV and MHV have considerable differences, particularly near the 5′ end of the ORF. In fact IBV contains only one of the papain domains and is almost 2 kb shorter than MHV ORF 1a [2, 5]. There are some interesting differences between the ORF 1a sequences of JHM and A59. The A59 sequence has several regions containing small deletions and insertions resulting in frame shifts of up to 40 amino acids. There is also an 18 amino acid in frame deletion in the 5′ portion of the gene (Bonilla et al., in prep.). These frame shifts and deletions are all outside of the predicted functional domains, with the exception of the presence of one frame shift in the predicted membrane spanning domain upstream of the 3C-like protease domain. These differences between A59 and JHM in ORF 1a suggest that these regions are perhaps dispensable and that not all of the 21 kb of gene 1 is essential for function.

Antisera directed against proteins encoded in gene 1

Very little is known about the proteins encoded in gene 1 of the coronaviruses. This is probably due to the lack of antisera specific for gene 1 encoded polypeptides. We have made a considerable effort toward raising antisera against polypeptides encoded in this gene and have

begun to analyze these proteins. Figure 1 illustrates the regions of the genome encoding proteins against which these antisera were raised. Anti p28 serum is an anti-peptide serum directed against 14 amino acids encoded by the JHM genome (nucleotides 287–329). UP102 is directed against a viral/bacterial fusion protein expressed in the pET 3b vector [12], containing an approximately 2 kb fragment of the A59 genome as shown in Fig. 1. The antisera 81043, 600A, 88640 and 590A are directed against bacterial/viral tripartite proteins containing viral sequences encoded by the A59 genome (see Fig. 1) inserted into the plasmid pGE374 between truncated recA and lacZ sequences [13]. The fusion protein antisera were raised in rabbits, which had been inoculated with both native and denatured fusion proteins.

Synthesis of ORF 1 proteins by in vitro translation

Denison and Perlman [14] have shown that translation of MHV genome RNA results in the synthesis of p220 and p28. In the presence of leupeptin, an inhibitor of serine and cysteine proteases there is less p28 detected and a new polypeptide, p250 is observed. This suggests that p250 is a precursor to p220 and p28 [14]. As expected, p28 is immunoprecipitated with anti-p28 and UP102 antisera. The p220 is immunoprecipitated by UP102 and 81043, but not by anti-p28 antiserum because the p28 sequences are not present in this polypeptide. The precursor, p250, is immunoprecipitated by anti-p28, as well as UP102 and 81043 antisera, consistent with this polypeptide containing the p28 sequences. Baker et al. [9] have demonstrated that the cleavage of p250 into p220 and p28 is carried out by the first papain-like protease, encoded in ORF 1a. They showed that in vitro translation of a synthetic RNA representing the first 5.3 kb of the JHM genome resulted in the cleavage of p28. They also showed that translation products of a shorter, 3.9 kb RNA did not include p28, suggesting that the protease activity is encoded between 3.9 and 5.3 kb, and that the first papain-like protease was responsible for cleavage. They also concluded that this was a *cis*-acting protease only. We have carried out similar experiments with A59 in which we in vitro translated RNAs representing various portions of the 5' end of ORF 1a. As shown in Table 1, we observed that RNAs representing from 4.2 kb to 5.2 kb of ORF 1a were capable of synthesis of p28. However 3.6 kb and 1.9 kb transcripts were not. This confirms the results of Baker [9] that the first papain-like protease is likely to be responsible for p28 synthesis. Furthermore, the conserved "X" domain, which is encoded downstream of the first papain-like protease, would not be present in the translation products of the 4.2 kb RNA; this demonstrates that the "X" domain is not necessary for cleavage of p28.

Table 1. Cleavage of P28 during in vitro translation of RNAs translated from MHV-A59 ORF 1a

ORF 1a (nucleotide)	Location (amino acid)	P28 Cleavage
1985	591	−
3690	1160	−
4242	1344	+
4664	1484	+
4934	1574	+
5220	1670	+

We have used antisera to detect polypeptides encoded in ORF 1b of the genome in the products of in vitro translation [15]. We detected products of ORF 1b only when translations were allowed to proceed for long time periods, that is, longer than 60 min: this suggests that the ORF 1b products are synthesized by initiation of translation in ORF 1a followed by frame shifting into ORF 1b. In summary the two antisera directed against proteins encoded in the 3' end of ORF 1b (88640 and 590A, see Fig. 1) detected several polypeptides of 90 kd, 74 kd, 53 kd, 44 kd, 32 kd. We did not detect any larger proteins, even when leupeptin was included during the translation. We do not known if these polypeptides are similar to those in infected cells because we have not yet been able to detect the ORF 1b polypeptides during infections. In preliminary experiments, translation of a full-length RNA transcribed from an ORF 1b cDNA resulted in the synthesis of many polypeptides; this was unexpected since a protease domain in ORF 1b has not yet been predicted from its sequence. The observation of many polypeptides in the cell free translation products of ORF 1b could due to sub optimal conditions for the in vitro translation of very long mRNAs; alternatively there may actually be a protease activity encoded in ORF 1b.

Synthesis of ORF1a proteins in MHV-A59 infected cells

We have used the antisera described above to detect the polypeptide products of ORF 1a in infected cells (Fig. 2). Anti p28 antiserum immunoprecipitates p28; we do not observe larger polypeptides immunoprecipitated with this antiserum, suggesting that p28 is rapidly cleaved from its precursor. UP102 detects p28 and also p65, a polypeptide not observed during in vitro translation of genome RNA; the observation that p65 is detected with UP102 serum, but not with anti p28 serum, suggests that p65 is encoded downstream of p28. Pulse chase experiments support this conclusion. During pulse-chase experiments, the
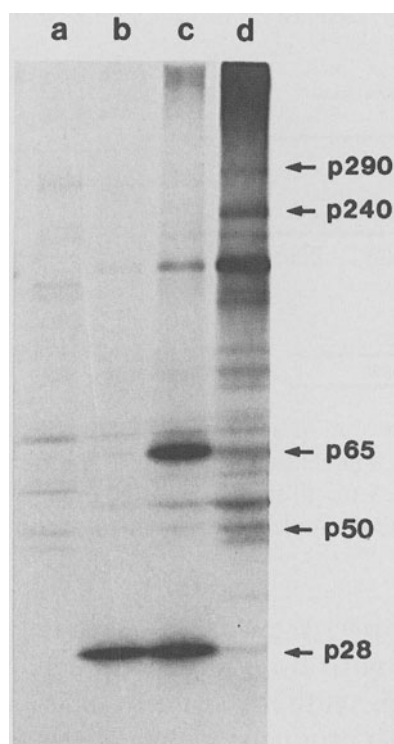
Fig. 2. Immunoprecipitation of MHV-A59 ORF1a products. Immunoprecipitates of [^{35}S] methionine labeled, (a) mock infected and (b,c,d) MHV-A59 infected murine fibroblast DBT cells were analyzed by 5–18% gradient polyacrylamide gel electrophoresis. a,c Antiserum UP102; b anti-p28 antiserum; d 81043 antiserum. The molecular weights of the specific precipitation products are indicated to the right of the gel with arrows

UP102 antiserum detects a high molecular weight polypeptide before p65 is apparent, suggesting that cleavage of p65 is not as rapid as p28 cleavage. The 81043 and 600 antisera detect p290 and p240, while only 81043 detects p50. These data, along with previously published kinetic data [16] suggest that p290 is a precursor of p240 and p50 and that p50 represents the amino terminus of p290 [16].

The products of ORF 1a observed in infected cells differ from those found in the products of cell free translation of genome RNA (see Fig. 3). Most strikingly, p65 is not observed during in vitro translation. Furthermore the proteins encoded downstream of differ. While in vitro translation products include p250 and p220, the intracellular polypeptides include p290, p240 and p50. As shown in Fig. 3, p290 would represent nearly all of ORF 1a. We do not know why these differences are observed between in vitro translation products and intracellular polypeptides. However, it is possible that there is temporal regulation of
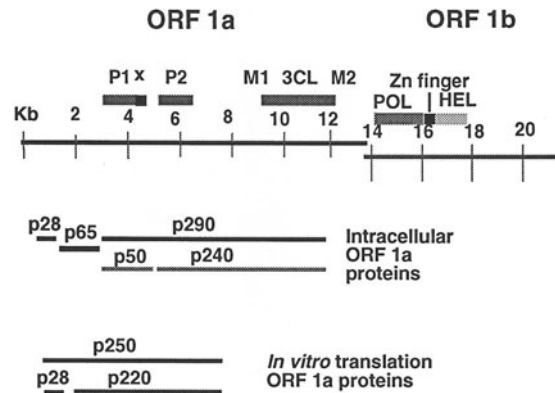
Fig. 3. Model of ORF1a translation and processing in MHV-A59 infected cells and in *in vitro* translation of MHV-A59 genome RNA. The predicted sizes of the polypeptides and the possible cleavage sites are all shown to scale. General alignment of polypeptides is based on apparent molecular weights and antisera specificity

processing during infection. We can only observe ORF 1a polypeptides late in infection and in vitro translation products might be more representative of polypeptides synthesized early in infection. Also, while we have shown that ORF 1b products are synthesized in vitro [15], this may not be very efficient. It is possible that there are not high enough levels of the downstream proteases to process the ORF 1a proteins properly. Thus, p65 may be cleaved either by the second papain protease or by the 3C-like protease and these are simply not present in high enough concentrations during in vitro translation. Another explanation for the differences in vivo and in vitro is the possibility of cellular enzymes participating in the processing of viral polypeptides. We feel this is unlikely as expression of the 5.2 kb ORF 1a construct via a vaccinia virus expression system in mouse cells results in the synthesis of p28 but not p65 (data not shown). Even in eukaryotic cells, p65 is not synthesized in the absence of infection.

Temperature sensitive mutants with possible processing defects

We have identified two RNA negative mutants of MHV-A59 than may have processing defects in the synthesis of ORF 1a polypeptides. Both NC11 (group B mutant) and LA16 (group A/B mutant) [17] synthesize greatly reduced levels of p28 and barely detectable amounts of p65 at the non-permissive temperature. Furthermore, p28 and p65 are stable at the non-permissive temperature. Since, in preliminary experiments these mutants do appear to synthesize the p290 and p240 ORF 1a polypeptides, we suggest that these viruses may be defective in processing of p28 and p65. It is interesting that both the synthesis of p28 and p65 appear to be

defective in these mutants because, as we have discussed above, these proteins are apparently processed by different protease activities. These results, in combination with the kinetics of accumulation of p28 and p65, suggest the possibility that p28 cleavage is necessary for p65 cleavage to occur.

Acknowledgements

References

1. Pachuk CJ, Bredenbeek PJ, Zoltick PW, Spaan WJM, Weiss SR (1989) Molecular cloning of the gene encoding the putative polymerase of mouse hepatitis virus strain A59. Virology 171: 141–148
2. Lee HJ, Shieh CK, Gorbalenya AE, Koonin EV, LaMonica N, Tuler J, Bagdzhadzhyan A, Lai MMC (1991) The complete sequence of the murine coronavirus gene 1 encoding the putative protease and RNA polymerase. Virology 180: 567–582
3. Spaan WJM, Cavanagh D, Horzinek MC (1988) Coronaviruses. Structure and genome expression. J Gen Virol 69: 2939–2952
4. Sethna PB, Hung S-L, Brian DA (1989) Coronavirus subgenomic minus-strand RNAs and the potential for RNA replicons. Proc Natl Acad Sci USA 86: 5626–5630
5. Boursnell MEG, Brown TDK, Foulds IJ, Green PH, Tomley FM, Binns MM (1987) Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J Gen Virol 68: 57–77
6. Bredenbeek PJ, Pachuk CJ, Noten AFH, Charite J, Luytjes W, Weiss SR, Spaan WJM (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. Brierly I, Boursnell MEG, Binns MM, Bilimoria B, Blok VC, Brown TDK, Inglis SC (1987) An efficient ribosomal frame-shifting signal in the polymerase encoding region of the coronavirus IBV. EMBO J 6: 3779–3785
8. Gorbalenya AE, Koonin EV, Lai MMC (1991) Putative papain-related thiol proteases of positive strand RNA viruses. FEBS Lett 288: 201–205
9. Baker SC, Shieh CK, Chang MF, Vannier DM, Lai MMC (1989) Identification of a domain required for autoproteolytic cleavage of murine coronavirus gene A polyprotein. J Virol 63: 3693–3699
10. Gorbalenya AE, Blinov VM, Donchenko AP, Koonin EV (1989) An NTP-binding domain is the most conserved seq uence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication. J Mol Evol 28: 256–268

11. Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM (1992) Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. Nucleic Acids Res 17: 4847–4861
12. Studier WF, Rosenberg AH, Dunn JJ, Dubendorf JW (1990) Use of T7 polymerase to direct the expression of cloned genes. Methods Enzymol 185: 60–89
13. Zoltick PW, Leibowitz JL, DeVries JR, Weinstock GM, Weiss SR (1989) A general method for the induction and screening of antisera for cDNA-encoded polypeptides: antibodies specific for a coronavirus putative polymerase encoding gene. Gene 85: 413–420
14. Denison MR, Perlman S (1986) Translation and processing of mouse hepatitis virus virion RNA in a cell-free system. J Virol 60: 12–18
15. Denison MR, Zoltick PW, Leibowitz JL, Pachuk CJ, Weiss SR (1991) Identification of polypeptides encoded in open reading frame 1b of the putative polymerase gene of the murine coronavirus mouse hepatitis virus A59. J Virol 65: 3076–3082
16. Denison MR, Zoltick PW, Hughes SA, Giangreco B, Olson AL, Perlman S, Leibowitz JL, Weiss SR (1992) Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. Virology 189: 274–284
17. Shaad MC, Stohlman SA, Egbert J, Lum K, Fu K, Wei T, Baric RS (1990) Genetics of mouse hepatitis virus transcription: Identification of cistrons which may function in positive and negative strand RNA synthesis. Virology 177: 634–645

Authors' address: Dr. S. R. Weiss, Department of Microbiology, 319 Johnson Pavilion, University of Pennsylvania, Philadelphia, PA 19104-6076, U.S.A.