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 180, 567–582 (1991)

The Complete Sequence (22 Kilobases) of Murine Coronavirus Gene 1 Encoding the Putative Proteases and RNA Polymerase

HAN-JUNG LEE,* CHIEN-KOU SHIEH,* ALEXANDER E. GORBALENYA,† EUGENE V. KOONIN,‡ NICOLA LA MONICA,* JEREMY TULER,* ANUSH BAGDZHADZHYAN,* AND MICHAEL M. C. LAI*,[1]

*Howard Hughes Medical Institute, Department of Microbiology, University of Southern California School of Medicine, Los Angeles, California 90033; †Institute of Poliomyelitis and Viral Encephalitides, The USSR Academy of Medical Sciences, 142782 Moscow Region, USSR; and ‡Institute of Microbiology, The USSR Academy of Sciences, 117811 Moscow, USSR

Received September 4, 1990; accepted October 5, 1990

The 5′-most gene, gene 1, of the genome of murine coronavirus, mouse hepatitis virus (MHV), is presumed to encode the viral RNA-dependent RNA polymerase. We have determined the complete sequence of this gene of the JHM strain by cDNA cloning and sequencing. The total length of this gene is 21,798 nucleotides long, which includes two overlapping, large open reading frames. The first open reading frame, ORF 1a, is 4488 amino acids long. The second open reading frame, ORF 1b, overlaps ORF 1a for 75 nucleotides, and is 2731 amino acids long. The overlapping region may fold into a pseudoknot RNA structure, similar to the corresponding region of the RNA of avian coronavirus, infectious bronchitis virus (IBV). The *in vitro* transcription and translation studies of this region indicated that these two ORFs were most likely translated into one polyprotein by a ribosomal frameshifting mechanism. Thus, the predicted molecular weight of the gene 1 product is more than 800,000 Da. The sequence of ORF 1b is very similar to the corresponding ORF of IBV. In contrast, the ORF 1a of these two viruses differ in size and have a high degree of divergence. The amino acid sequence analysis suggested that ORF 1a contains several functional domains, including two hydrophobic, membrane-anchoring domains, and three cysteine-rich domains. It also contains a picornaviral 3C-like protease domain and two papain-like protease domains. The presence of these protease domains suggests that the polyprotein is most likely processed into multiple protein products. In contrast, the ORF 1b contains polymerase, helicase, and zinc-finger motifs. These sequence studies suggested that the MHV gene 1 product is involved in RNA synthesis, and that this product is processed autoproteolytically after translation. This study completes the sequence of the MHV genome, which is 31 kb long, and constitutes the largest viral RNA known.

INTRODUCTION

Mouse hepatitis virus (MHV), a murine coronavirus, contains a single-stranded, positive-sense RNA genome (Lai and Stohlman, 1978; Wege *et al.*, 1978). The genomic organization is well understood (Spaan *et al.*, 1988; Lai, 1990). It contains 8 genes, each of which is expressed from the 5′-end of a polycistronic mRNA species. These mRNAs have a 3′-coterminal, nested-set structure (Lai *et al.*, 1981). Starting from the 5′-end of the genome, the genes are named 1, 2a, 2b, 3, and so on until gene 7 (Cavanagh *et al.*, 1990). Genes 2b, 3, 6, and 7 encode the four known viral structural proteins, i.e., HE (hemagglutinin-esterase), S (spike), M (membrane), and N (nucleocapsid) proteins, respectively. The remaining genes presumably encode nonstructural proteins, most of which are yet to be identified in the virus-infected cells. The nucleotide sequences of genes 2 to 7 have been determined for two strains, A59 and JHM, of MHV (Armstrong *et al.*, 1983, 1984; Skinner *et al.*, 1985; Skinner and Siddell, 1983, 1985; Schmidt *et al.*, 1987; Luytjes *et al.*, 1987, 1988; Shieh *et al.*, 1989). Altogether these seven genes account for roughly 9.5 kb. The remaining gene, gene 1, which is the 5′-most gene, has been estimated to be longer than the size of all of the other genes combined (Pachuk *et al.*, 1989; Baker *et al.*, 1990). Only the 5′-terminal 5.3 kb in JHM strain and the 3′-terminal 8.4 kb of this gene in A59 strain have so far been sequenced (Soe *et al.*, 1987; Baker *et al.*, 1989; Pachuk *et al.*, 1989; Bredenbeek *et al.*, 1990). The corresponding gene of an avian coronavirus, infectious bronchitis virus (IBV), has been completely sequenced and shown to be 20 kb long (Boursnell *et al.*, 1987). This IBV gene consists of two open reading frames (ORFs), which can be translated into a polyprotein via a ribosomal frameshifting mechanism (Brierley *et al.*, 1987, 1989). Again, the gene products have yet to be detected in the virus-infected cells. The size of MHV gene 1 has not been determined. From the approximate sizes of the cDNA clones, it has been estimated to be roughly 22–23 kb (Pachuk *et al.*, 1989; Baker *et al.*, 1990). Comparison of the published partial sequences of gene 1 showed that IBV and MHV share sequence similarity in the 3′-terminus of the gene (Bredenbeek *et*

Sequence data from this article have been deposited with the EMBL/GenBank under Accession No. M55148.

[1] To whom correspondence should be addressed.

0042-6822/91 $3.00

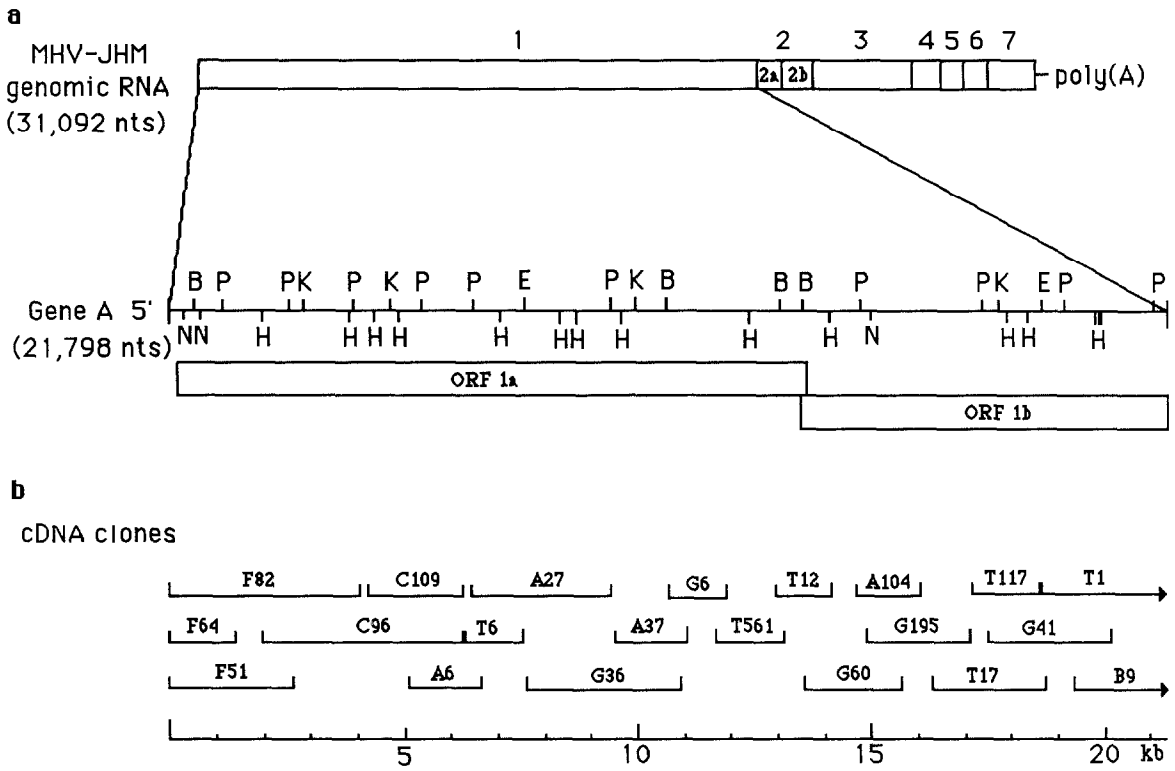

FIG. 1. Molecular clones and restriction map of the gene 1 of the genomic RNA of MHV-JHM. (a) Schematic diagram of the MHV-JHM genome and restriction map of the cDNA clones. (b) The cDNA clones used for sequencing. Abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; N, *Nco*I; P, *Pst*I. Lengths are expressed in kilobase pairs.

al., 1990), and yet their 5'-ends are diverged (Soe *et al.*, 1987; Baker *et al.*, 1989). Thus, the evolutionary relationship of these two viruses in gene 1 is not clear.

Several pieces of evidence suggest that gene 1 may encode proteins which are directly involved in viral RNA synthesis: First, since MHV does not contain RNA polymerase (Brayton *et al.*, 1982), this enzyme has to be synthesized from the incoming virion genomic RNA. This translation is only possible if the gene is located at the 5'-end of the genome. Second, RNA recombination studies using temperature-sensitive (ts) mutants indicated that the ts lesions affecting RNA synthesis are localized within the gene 1 region (Keck *et al.*, 1987). This conclusion has been confirmed by RNA recombination mapping studies (Baric *et al.*, 1990). Third, the 3'-half of the gene 1 sequences of IBV and MHV-A59 contains the sequence motifs for RNA polymerase and helicase, which are the activities expected to be involved in RNA synthesis (Boursnell *et al.*, 1987; Gorbalenya *et al.*, 1989b; Bredenbeek *et al.*, 1990). However, these postulated functions have not been directly demonstrated. At least one enzymatic activity, i.e., an autoprotease (Baker *et al.*, 1989), has been associated with the gene product. The presence of the protease activity suggests that the gene 1 product is likely to be processed into multiple proteins.

The properties of the RNA polymerase of coronavirus are of considerable interest since the coronavirus RNA synthesis utilizes an unusual mechanism of discontinuous transcription, probably involving a free leader RNA species (Lai, 1988). The understanding of the RNA polymerase should shed further light on the mechanism of RNA synthesis. To this end, we have obtained the complete sequence of gene 1 of the JHM strain of MHV. This gene is nearly 22,000 nucleotides long and contains two overlapping ORFs, similar to the corresponding IBV gene. Sequence analysis shows that the MHV gene may have undergone extensive divergence from the IBV gene, particularly at its 5'-half. Several functional domains were identified, which may be important for the processing and the enzymatic activities of its gene product.

MATERIALS AND METHODS

Virus and cells. The plaque-cloned JHM strain of MHV (Makino *et al.*, 1984) was used throughout this study. The virus was propagated on DBT cells (Hirano *et al.*, 1974) at m.o.i. of 1. Virus was harvested and purified from the medium, and viral RNA was prepared as previously described (Makino *et al.*, 1984).

cDNA cloning. The cDNA clones encompassing

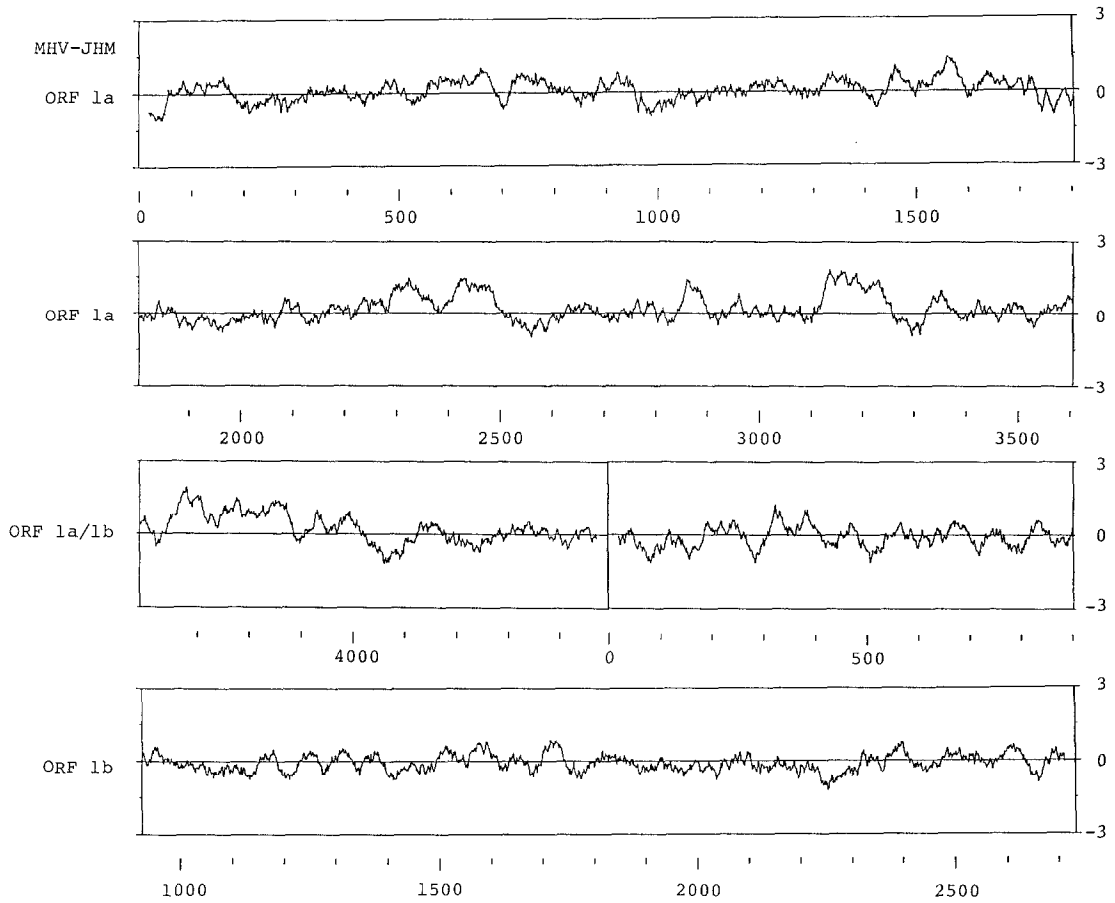

FIG. 2. Hydropathy profiles of the predicted amino acid sequences of ORF 1a and ORF 1b. Values above the line are hydrophobic and values below the line are hydrophilic. The hydropathicity was calculated using a moving window of 40 amino acids, with a value plotted every 16 residues (Kyte and Doolittle, 1982).

gene 1 were obtained by using specific synthetic oligonucleotides as primers and purified virion genomic RNA as template. Initially, the sequences of these oligonucleotides were derived from RNA sequence analysis of the RNase T1-resistant oligonucleotides which had been mapped to either gene 1 or 2 (Shieh *et al.*,

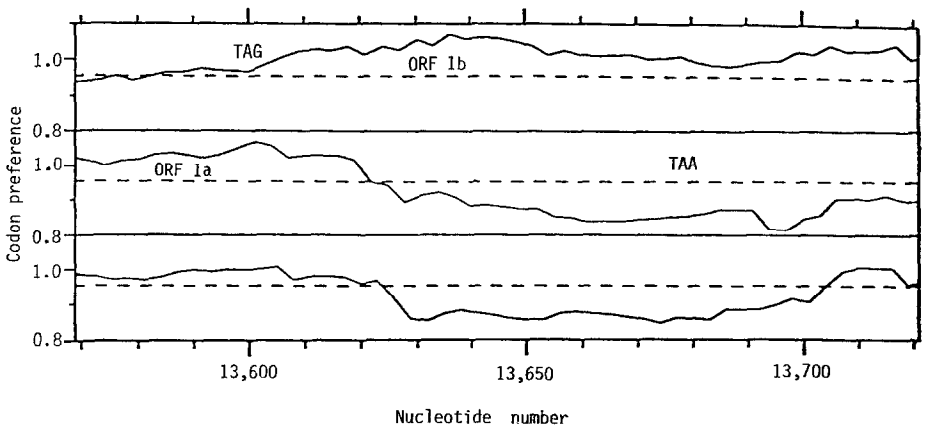

FIG. 3. Diagram of the codon preference in the region between ORF 1a and ORF 1b. The codon usage patterns for the three reading frames of the predicted amino acid sequences at the junction between the ORF 1a and ORF 1b are shown. The two stop codons at 13600 (TAG) and 13679 (TAA) are marked. The codon usage table was generated for genes 3, 6, and 7, which encode the viral structural proteins, of MHV-JHM (Schmidt *et al.*, 1987; Skinner and Siddell, 1983), and used for comparison with ORFs 1a and 1b. The parameters used are a window length of 25 and a maximum scale of 1.1 (Gribskov *et al.*, 1984).

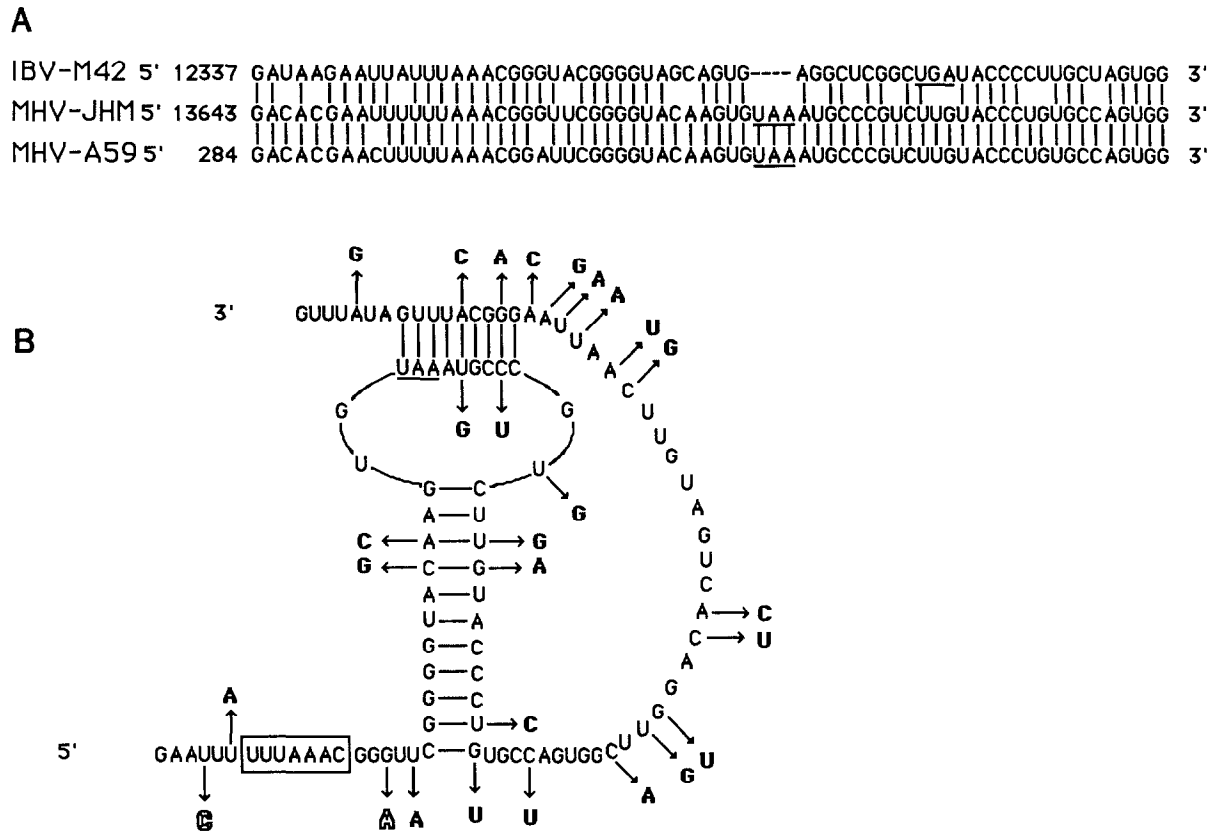

FIG. 4. Comparison of the RNA sequences and the proposed secondary structure of the MHV-JHM, MHV-A59 and IBV RNAs at the junction between ORF 1a and ORF 1b. (A) Alignment of nucleotide sequences. The first nucleotides are numbered according to Boursnell *et al.* (1987), for IBV, and Bredenbeek *et al.* (1990), for MHV-A59, and termination codons are underlined. (B) Tertiary RNA structure at the region of ribosomal frameshifting. The potential signal for ribosomal frameshifting is boxed, and the stop codon is underlined. Arrows indicate the differences in the RNA sequence of MHV-JHM in comparison with that of IBV (boldfaced) and MHV-A59 (outlined).

1987, 1989; Soe *et al.*, 1987). cDNA synthesis was performed by the general method of Gubler and Hoffman (1983). The double-stranded cDNA molecules were trimmed with T4 DNA polymerase and ligated to pTZ18U (United States Biochemical Corp.) either by blunt-end ligation or *Eco*RI linker ligation. The recombi-

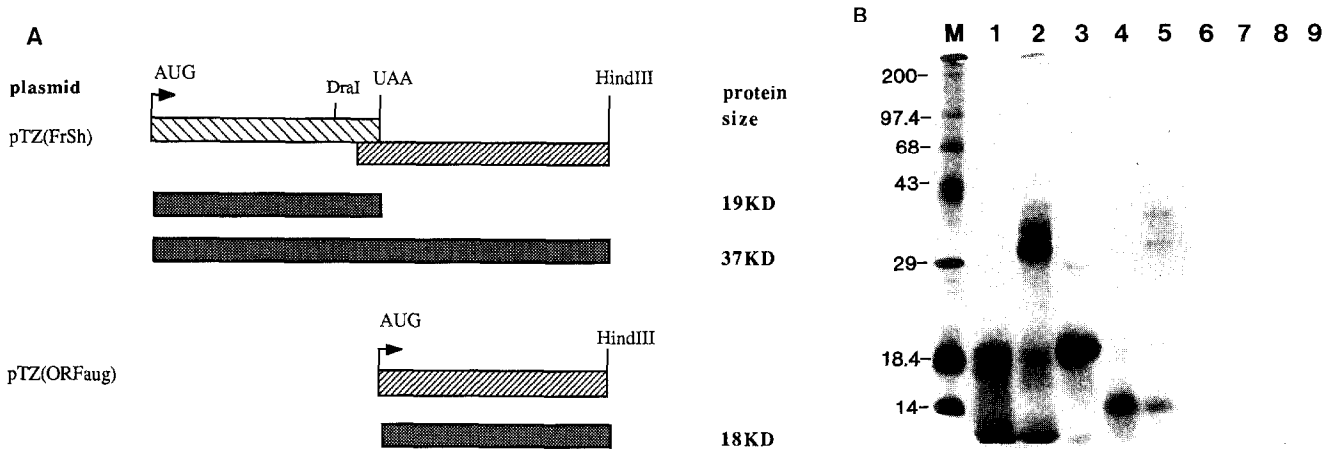

FIG. 5. SDS–PAGE analysis of *in vitro* translated products. (A) Diagram of the plasmids used and the predicted sizes of the translation products from the transcribed RNAs. (B) Plasmid pTZ(FrSh) was linearized with either *Hin*dIII (lanes 2, 5, and 8), generating a full-length transcript by T7 RNA polymerase, or with *Dra*I (lanes 1, 4, and 7), generating a 0.5-kb RNA. Translation was performed in a rabbit reticulocyte lysate system using [^{35}S]methionine. Translation products were analyzed directly (lanes 1–3) or after immunoprecipitation using the ORF 1a-specific antiserum (lanes 4–6) or rabbit preimmune serum (lanes 7–9). M indicates molecular weight markers in kilodaltons; lanes 3, 6, and 9, translation of pTZ(ORF^{aug}).

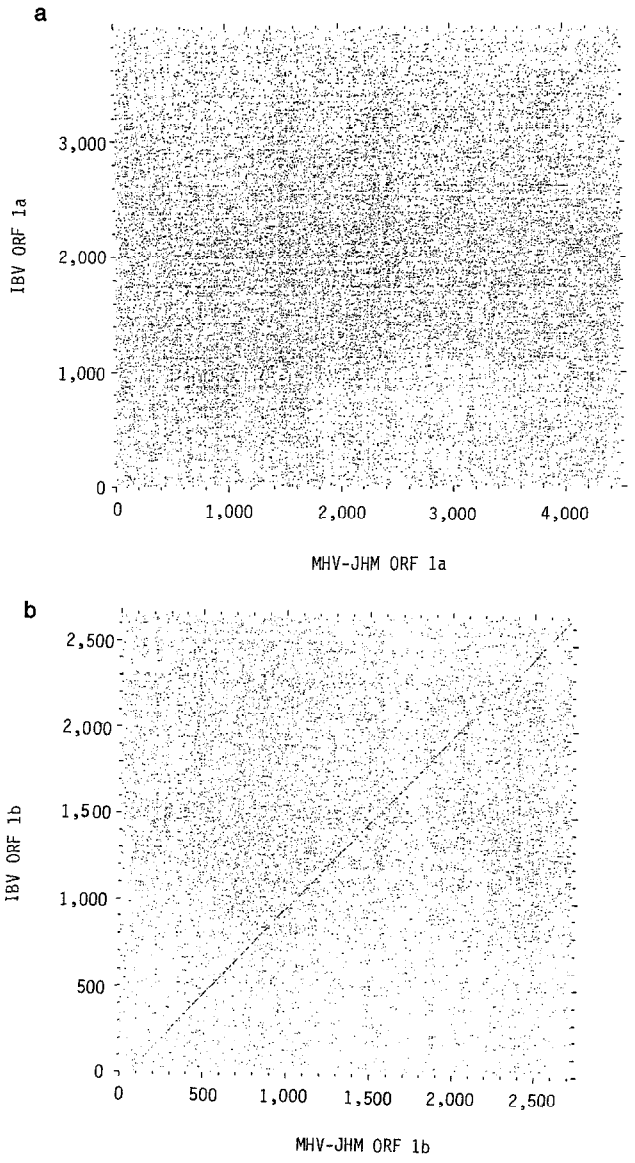

FIG. 6. Dot matrix comparison of the predicted amino acid sequences of ORF 1a and ORF 1b of MHV-JHM and IBV. The profiles were generated by the compare/word option from the Genetics Computer Group program (Devereux *et al.*, 1984) with a word-size of 2 and alphabet of 20 for ORF 1a (a) and 21 for ORF 1b (b).

nant DNAs were transformed into *Escherichia coli* strain MV1190 competent cells (Dagert and Ehrlich, 1979). Homopolymer dC tailing to the 3'-end of the cDNAs using terminal transferase were also used to anneal to *Pst*I-linearized pBR322 with oligo(dG) tails and transformed into *E. coli* strain MC1061. Specific cDNA clones were identified using 5'-end-labeled oligonucleotides as probes and confirmed by subsequent hybridization to viral mRNA (Shieh *et al.*, 1987). Once the sequences of the cDNA clones were obtained, oligonucleotides complementary to the 5'-ends of these clones were synthesized to serve as primers for additional cDNA cloning to obtain overlapping cDNA clones.

DNA sequencing. Sequencing was performed as previously described (Shieh *et al.*, 1987, 1989). Both chemical modification (Maxam and Gilbert, 1980) and dideoxynucleotide chain termination (Sanger *et al.*, 1977) methods were used directly on plasmid DNA (Chen and Seeburg, 1985).

Construction of recombinant plasmids for the frameshifting analysis. Subcloning and mutagenesis of cDNA clone T-12 was accomplished using synthetic oligonucleotides and polymerase chain reaction (PCR). Briefly, oligomer #166 (5'-GATCGAATTCCTTTACATGGTGAAGGGGTG-3'), which extends from nucleotide 13,147 to 13,167 of gene 1 and contains mismatches at both nucleotides 13,154 and 13,156, and oligomer #199 (5'-CATATGACACAGGATCCTTTATGCC-3'), which is complementary to nucleotides 13,529 to 13,553 and includes the *Bam*HI site at nucleotide 13,537, were used for DNA amplification by PCR according to the standard procedures (Saiki *et al.*, 1988). The resulting PCR DNA product encompasses sequences from nucleotide 13,147 to 13,537 with a specific mutation (T to A) at nucleotide 13,154 and another (T to G) at nucleotide 13,156, resulting in the introduction of an ATG codon. The DNA was then digested with

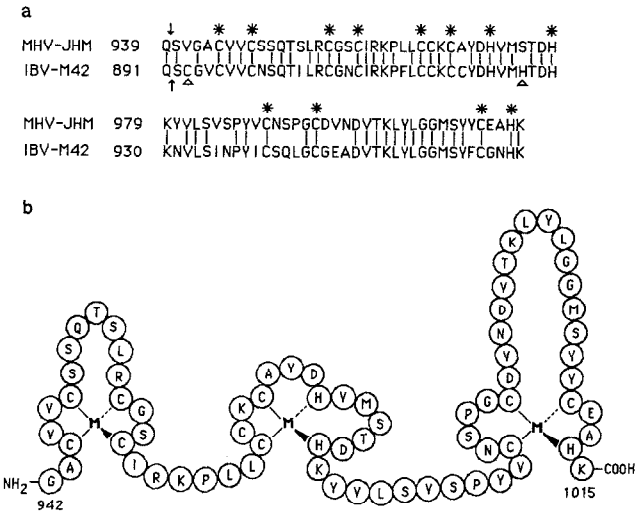

FIG. 7. Comparison of the sequence and structure of the putative metal-binding domain of ORF 1b from MHV-JHM and IBV. (a) Alignment of amino acid sequences. The amino acid residues are numbered with respect to ORF 1b. Asterisks indicate the conserved Cys and His residues. Arrows show the putative cleavage sites for the 3C-like proteases. The open triangles indicate the residues putatively liganded with the metal ion in the case of IBV (Gorbalenya *et al.*, 1989b). These amino acids are substituted in MHV, but neighboring residues preserve the metal-binding domain. (b) Predicted structure of the metal-binding domain of MHV-JHM ORF 1b. M, metal cation (Zn^{2+}). Only one of the several possible foldings of this domain is shown.

FIG. 8. Alignment of the ORF 1a of MHV-JHM and IBV. The overall alignment was generated by combining segments aligned by programs OPTAL (Gorbalenya *et al.*, 1989a) and MULTALIN (Corpet, 1988). It consists of four distinct pieces separated by regions that could not be aligned with certainty. For the latter regions, only the total numbers of amino acid residues are indicated. The amino acid numbers of the first and the last residues of each aligned segment are indicated in parentheses. Two dots, identical residues; single dots, similar residues. Conserved Cys residues are highlighted by boldface. Asterisks, putative catalytic residues of proteases; arrows, putative cleavage sites for 3C-like proteases. Box, the putative cleavage site for $3CL^{pro}$ in IBV substituted by a KR dipeptide in MHV-JHM. The IBV sequence was from Boursnell *et al.* (1987). MHVA: ORF 1a of MHV. IBVF1: ORF 1a of IBV.

FIG. 8—*Continued*

*Eco*RI and *Bam*HI and subcloned into pTZ18U, yielding pTZ(FS^{aug}). The specific mutations were confirmed by DNA sequencing.

Plasmid pTZ(FS^{aug}) was digested with *Bam*HI and *Hin*dIII (*Hin*dIII site in the polylinker of pTZ18U) and ligated to a 626-bp *Bam*H–*Hin*dIII DNA fragment derived from the clone T-12. The resulting plasmid pTZ(FrSh) consists of the sequence from nucleotides 13,147 to 14,164 of gene 1.

Plasmid pTZ(ORF^{aug}) consists of the sequences from nucleotide 13,671 to 14,164 of gene 1. An ATG codon was introduced at nucleotide 13,678–13,680 by PCR-

FIG. 8—*Continued*

mediated mutagenesis in a similar method as for pTZ(FS^{aug}).

In vitro transcription and translation. Recombinant plasmids pTZ(ORF^{aug}) and pTZ(FrSh) were linearized by digestion with restriction enzymes *Hin*dIII or *Dra*I and transcribed *in vitro* with T7 RNA polymerase as previously described (Soe *et al.*, 1987). The resulting RNA was translated in the mRNA-dependent rabbit reticulocyte lysate (Promega Biotech) in the presence of [^{35}S]methionine. Reactions were carried out in a final

FIG. 8—*Continued*

volume of 25 μl under conditions recommended by the manufacturer. The translation products were immunoprecipitated by the method of Shin and Morrison (1989) and analyzed by electrophoresis on 7.5 to 15% polyacrylamide gel.

Computer analysis of nucleotide and amino acid sequences. Sequence data were analyzed on a VAX 1852 using the GCG sequence analysis software package developed by Genetics Computer Group of University of Wisconsin. Detailed comparative analyses of coronavirus protein sequences were done by programs MULTALIN (Corpet, 1988), OPTAL (Gorbalenya *et al.,* 1989a), DOTHELIX (Leontovich *et al.,* 1990), and SITE (Koonin *et al.,* 1990). The programs DOTHELIX and SITE are parts of the GENBEE program package for biopolymer sequence analysis.

RESULTS

Molecular cloning of the gene 1 of the genomic RNA of MHV-JHM. To clone the gene 1 region, which represents more than two thirds of the MHV genome, a synthetic oligonucleotide (oligo 30; 5′-CTGAATTTGGGG-GTTGGG-3′) was initially used as a primer for cDNA synthesis (Shieh *et al.,* 1987). The sequence of this oligonucleotide was based on the sequence analysis of the RNase T1-resistant oligonucleotide No. 30, which had previously been mapped to gene 2 (Makino *et al.,* 1984). The resulting cDNA clones contained inserts ranging from 0.5 to 3 kb in size. These cDNA clones detected only the genomic RNA on Northern blots of intracellular RNA from MHV-infected cells (data not shown). Based on the nested-set structure of MHV

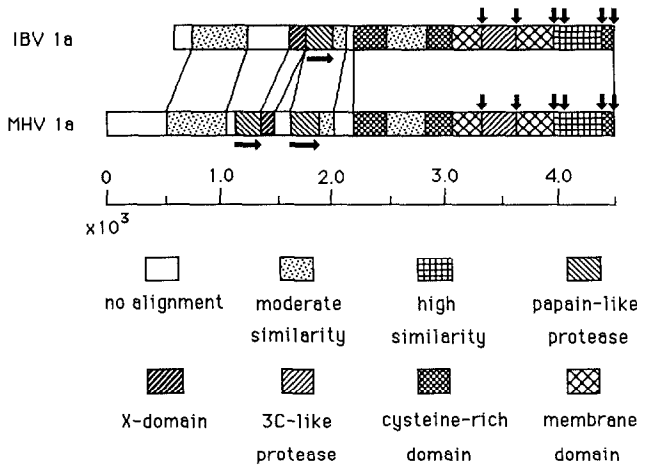

FIG. 9. A schematic presentation of the relationship between the ORF 1a of MHV-JHM and IBV. The two ORF 1a are shown to scale. The designation of regions, for which specific functional predictions could be made, and of regions of similarity between the two viruses are shown in the bottom of the figure. High similarity, statistical significance over 10 SD (standard deviation), when aligned by the program OPTAL (Gorbalenya *et al.*, 1989a,b); moderate similarity, significance of 3 to 10 SD. The alignments in the regions, with predicted functions, were significant at the level of at least 5 SD. Regions of similarity between the two viruses are joined. Vertical arrows, putative cleavage sites for $3CL^{pro}$. Horizontal arrows, putative papain-like proteases (two copies in MHV-JHM, and one copy in IBV).

mRNAs (Lai *et al.*, 1981), this result indicated that these cDNA clones represent part of gene 1. The 5′-ends of these DNAs were sequenced, and synthetic oligonucleotides complementary to these sequences were generated to prime further cDNA synthesis for walking toward the 5′-end of gene 1. In this way, overlapping DNA clones which encompass about 11 kb at the 3′-end of gene 1 were obtained (Fig. 1). cDNA clones representing the 5′-terminal 6.2 kb of gene 1 were derived as described (Shieh *et al.*, 1987; Baker *et al.*, 1989). The cDNA clones spanning the gap between the two cDNA groups were obtained by using specific primers representing both the sequences downstream and upstream of the gap as primers for first-strand and second-strand cDNA synthesis, respectively. The overlap of these cDNA clones was determined by Southern blotting and confirmed by DNA sequencing. The complete cloning of JHM gene 1 indicated that the size of gene 1 is approximately 22 kb in length (Fig. 1), longer than that of IBV (Boursnell *et al.*, 1987), and agrees with the previous estimate for the gene 1 of the A59 strain of MHV (Pachuk *et al.*, 1989).

Analysis of the nucleotide sequence and the predicted amino acid sequence. The complete MHV-JHM gene 1 sequence was obtained from the cDNA clones as indicated in Fig. 1. This sequence has been deposited with GenBank (Accession No. M55148), and will not be duplicated in this publication. The complete sequence of gene 1 contains 21,798 nucleotides preceding the UCUAUAC, which is the transcriptional initiation site for gene 2 (Shieh *et al.*, 1989). Analysis of the sequence revealed two large, overlapping open reading frames (ORFs), ORF 1a and ORF 1b (Fig. 1a). ORF 1a is 4488 amino acids long and has a predicted molecular weight of 499,319, which includes the coding region for p28 protein at its N-terminus (Soe *et al.*, 1987). The hydropathy plot (Kyte and Doolittle, 1982) shows that ORF 1a has several long stretches of hydrophobic regions at the carboxy-terminal region, which indicate potential membrane-spanning domains (Fig. 2). ORF 1b, which overlaps ORF 1a for 75 nucleotides but is located at a different reading frame, is 2731 amino acids long with a predicted molecular weight of 308,483. The ORF 1b sequence is very similar to that of MHV-A59 in both nucleotide and predicted amino acid sequences (Bredenbeek *et al.*, 1990). Only minor substitutions were noted between the two strains (data not shown). The ORF 1b starts with CUG instead of AUG. The first potential initiator codon AUG is located 399 nucleotides downstream of the first amino acid

```
                          *           *
PV1    30 HDNVAILPTHA 102 AGQCGG-VITCT-G---KVIGMHVGG  19
HRV2   30 YDRFVVVPTHA 102 SGYCGG-VLYKI-G---QVLGIHVGG  19
EMCV   38 RGRTLVVNRHM 108 KGWCGSALLADL-GGSKKILGIHSAG 25
FMDV   38 FGTAYLVPRHL 112 AGYCGGAVLAKD-GADTFIVGTHSAG 29
HAV    38 KDDWLLVPSHA 123 PGMCGGALVSSNQSIQNAILGIHVAG 23
CPMV   30 PGRRFLACKH- 116 PEDCGSLVIAHIGG-KHKIVGVHVAG 21
TBRV   28 KNKSVRMTRHQ 120 NDDCGMIILCQIKG-KMRVVGMLVAG 19
BWYV   29 ENA-LMTATHV 101 GGHSGSPYF-NGKT----ILGVHSCA  ?
SBMV   40 MDV-LMVPHHV  97 KGWSGTPLY-TRDG----IVGMHTGY  ?
TEV   224 FGPFIITNKHL  99 DGQCGSPLVSTRDG---FIVGIHSAS 72
                                               ↕
IBV    31 LGDTIYCPRHV 105 AGACGSVGFNIEKGV-VNFFYMHHLE 143
MHV    31 LDDKVYCPRHV 109 CGSCGSVGYVLTGDS-VRFVYMHQLE 137
```

FIG. 10. Alignment of the segments surrounding the putative catalytic His and Cys residues of the coronavirus 3C-like protease with the respective segments of other viral $3CL^{pro}$. The figure is an excerpt of the complete alignment generated by program OPTAL. The complete amino acid sequences of each viral $3CL^{pro}$ are indicated, but only the sequences around the catalytic residues are shown. The numbers of amino acid residues to the known or postulated termini of the respective viral $3CL^{pro}$ and between the aligned segments are indicated. For MHV $3CL^{pro}$, the postulated N-terminus is at amino acid residue 3350 (Fig. 8 and Table 1). Residues identical or similar to those in the coronavirus sequences are highlighted by boldface. The arrow shows the Gly to Tyr substitution in the putative substrate-binding sites of the coronavirus proteases. Asterisks, (putative) catalytic residues. Abbreviations: PV1, poliovirus type 1, Mahoney strain; HRV2, human rhinovirus type 2; EMCV, encephalomyocarditis virus; FMDV, foot-and-mouth disease virus type A10; HAV, hepatitis A virus; CPMV, cowpea mosaic virus; TBRV, tomato black ring virus; BWYV, beet western yellows virus; SBMV, southern bean mosaic virus; TEV, tobacco etch virus. For sources of the sequences, see Gorbalenya *et al.* (1989b), except BWYV (Veidt *et al.*, 1988) and SBMV (Wu *et al.*, 1987).

TABLE 1

SUMMARY OF THE PREDICTED CLEAVAGE SITES FOR THE 3C-LIKE PROTEASES IN CORONAVIRUS POLYPROTEINS

IBV			Putative protein next to the C-terminus of the cleavage site	MHV		
aa sequence	aa position	Size (# of aa)		aa position	Size (# of aa)	aa sequence
			ORF 1a			
V I K F Q ↓ G V F K A	2583	196	MP1	3160	190	L I K L K ↓ R A F G D ??
V S R L Q S G F K K	2779	307	$3CL^{pro}$	3350	303	T S F L Q S G I V K
G V R L Q S S F V R	3086	293	MP2	3652	288	G V K L Q S K R T R
I A T V Q A K L S D	3379	83	?	3941	89	V S Q I Q S R L T D
S T V L Q S V T Q E	3462	322	?	4030	307	N T V L Q A L Q S E
N V V V Q S K G H E	3784	144	GFL	4337	137	T V R L Q A G T A T
K S S V Q S V A G A	3928	931	POL	4474	918	G S Q F Q S K D T N
			ORF 1b			
P T T L Q S C G V C	891	601	HEL	939	600	S A V M Q S V G A C
E T S L Q G T G L F	1492	520	?	1539	519	N P R L Q C T T N L ?
F S A L Q S I D N I	2012	338	?	2058	374	F T R L Q S L E N V
Y P Q L Q S A W T C	2350	302	?	2432	299	Y P R L Q A A A D W

Note. In the aa position columns, the amino acid positions of the respective Q residues are indicated. The arrows show the predicted cleavage sites, Abbreviations: MP1, MP2, putative membrane proteins flanking the $3CL^{pro}$ at the N- and C-sides, respectively. POL: polymerase motif. HEL: helicase motif. GFL: growth factor-like domain. The data on IBV was obtained from Gorbalenya *et al.* (1989b). The sequence analysis was performed using the computer program as described under Materials and Methods.

codon in ORF 1b. Nevertheless, the codon preference plot suggests that the 399 nucleotides upstream of the first AUG are most likely translated together with the downstream sequences using the same reading frame (Fig. 3). In light of the corresponding sequences of IBV and MHV-A59 (Boursnell *et al.,* 1987; Bredenbeek *et al.,* 1990), this result suggests that this region could be translated via a ribosomal frameshifting mechanism (Brierley *et al.,* 1989).

Comparison of tertiary structure of RNA in the frameshift regions. It has been proposed that the nucleotide sequences in the overlapping regions between ORF 1a and ORF 1b in IBV and MHV-A59 RNAs are able to fold into a pseudoknot tertiary structure, which is essential for efficient frameshifting and, thus, expression of the downstream ORF 1b (Brierley *et al.,* 1989; Bredenbeek *et al.,* 1990). Comparison of the primary sequence revealed that the corresponding region of MHV-JHM contains a "slippery" sequence, UUUAAAC, similar to that of IBV (Fig. 4A). The possible folding of RNA in this region into a pseudoknot tertiary structure is similar among IBV, MHV-A59, and MHV-JHM (Fig. 4B). It is interesting to note that the nucleotide changes between MHV-JHM and IBV in either the stem or loop regions are compensated by mutations at the complementary positions (Fig. 4B). This suggests the significance of the putative tertiary structure in ribosomal frameshifting. Only two nucleotides differ between MHV-JHM and MHV-A59 in this region; they are located at the regions immediately upstream and downstream of the UUUAAAC sequence.

Ribosomal frameshifting in vitro. To confirm that the ORF 1a and 1b of MHV-JHM could be translated into one polypeptide by ribosomal frameshifting, we cloned the region spanning from nucleotide 13,147 to 14,164 of gene 1 into an expression vector under the control of the T7 promoter for *in vitro* translation studies. Because of the lack of a translational initiation codon, an ATG codon was introduced by PCR-mediated mutagenesis at nucleotide 13,154–13,156. If the translation of this transcript terminates at the UAA stop codon in ORF 1a, a 19-kDa protein will be produced. However, if the −1 translational frameshift occurs, a 37-kDa protein will be synthesized. As shown in Fig. 5, the *in vitro* translation of this RNA yielded both proteins (lane 2). The 37-kDa protein was heterogeneous; the smaller proteins may represent aberrant translational initiation or specific processing of the translation products. The addition of protease inhibitors in the rabbit reticulocyte lysates did not alter this translation pattern (data not shown). The antiserum prepared against the amino acid sequence just upstream of the frameshift (unpublished) precipitated both proteins (lane 5). Surprisingly, the major products precipitated by this anti-

```
                                10        20        30        40        50
IBV     (1236-1497)     GLDA---QKY VIYLQTLAQ- KWNVQYRDNF LILEWRDGNC WISSAIVLLQ
MHV     (1696-1953)     GFDE---PQL LQYYSMLGMC KWPVVVCGNY FAFKQSNNNC YINVACLMLQ
MHV     (1100-1349)     AFDAIYSETL SAFYAVPSD- ETHFKVCG-F YSPAIERTNC WLRSTLIVMQ
                                                          *

        60        70        80        90       100       110       120
AAKIRFKGF- LTEAWAKLLG GDPTDFVAWC YASCTAKVGD FSDANWLLAN LAEHFDADYT NAFLKKRVSC
HLSLKFPKWQ WRRPGNEFRS GKPLRFVSLV LAKGSFKFNE PSDST-DFIR VELR-EADLR SATCDLEFIC
SLPLEFKDLG MQKLWLSYKA GYDQCFVDK- LVKSAPKSII LPQGG-YVAD FAYFFLSQ-C SFKVHANWRC

       130       140       150       160       170       180       190
-NCGIKSYEL RGLEACIQPV RATNLLHFKT QYSNCPTCGA NNTDEVIEAS LPYLLLFATD GPATVDCDED
-KCGVKQEQR KGVDA-VMHF GTLDKSGLVK GYNIACTCG- DKLVHCTQFN VPFLI--CSN TPEGKKLPDD
LKCGM-ELKL QGLDA-VFFY GDV-VSHM-- -----CKCG- NSMT-LLSAD IPYTFDFGVR DDKFCAFYTP

       200       210       220       230       240       250       260
AVGTVVFVG- STNSGHCYTQ AAGQA-FD-- NLAKDRKFGK -KSPYITAMY TRFAFKNE-T S-LPVAKQSK
VVAANIFTG- G-SVGH-YTH VKCKPKYQLY DACNVSKVSE AKGNFTDCLY LK-NLKQTFS SVLTTYYLDD
RKVFRAACAV DVNDCHSMAV VDGK---QI- DGKVVTKFNG DKFDFMVGHG MTFSMSPFEI AQLYGSCITP
                  *

       270
GKSKSVKEDV SN
VKCVAYKPDL SQ
NVCF-VKGDV IK
```

FIG. 11. Alignment of the putative coronavirus papain-like proteases. The numbers of the first and last residues of the aligned segments are indicated in parentheses. Both of the two papain-like proteases of MHV are shown. Residues conserved in all the three sequences (identical or similar) are highlighted by boldface. Asterisks, putative catalytic residues.

serum migrated faster than the respective primary translation products, suggesting that protein processing had occurred. None of the proteins was immunoprecipitated by the preimmune serum. As controls, the transcripts containing either the 5′- or the 3′-halves

```
                      *                 *
MCP          14 PVKNQGQCGSCWAFSA 128 NLDHGVLLVGYG 49
catH         15 PVKNQGACGSCWTFST 130 KVNHAVLAVGYG 45
aleurain     14 PVKNQAHCGSCWTFST 128 DVNHAVLAVGYG 45
actinidin    14 DIKSQGECGGCWAFSA 128 AVDHAIVIVGYG 52
papain       14 PVKNQGSCGSCWAFSA 127 KVDHAVAAVGYN 46
DCP          14 PVKNQGQCGSCWSFST 138 SLDHGILIVGYS 50
catB         18 QIRDQGSCGSCWAFGA 164 MGGHAIRILGWG 55
catL         14 PVKDQGACGSCWAFNT 128 DLDHGVLVVGYG 44
CDP          98 DIC-QGALGDCWLLAA 146 VKGHAYSVTAPK 431

MHVpro1       ? FYSPAIERTNCWLRST 142 NDCHSMAVVDKG ?
MHVpro2       ? YFAFKQSNNNCYINVA 148 SVGH-YTHVKCK ?
IBVpro        ? FLILEWRDGNCWISSA 154 NSGHCYTQAAGQ ?
```

FIG. 12. Alignment of the segments around the putative catalytic residues of coronavirus papain-like proteases with the respective segments of papain-like proteases of cellular origin. The designations are as in Fig. 10. Abbreviations: MCP, mouse cysteine protease; catH, rat cathepsin H; DCP, *Dyctiostelium* cysteine protease; catB, rat cathepsin B; catL, rat cathepsin L; CDP, chicken calcium-dependent protease. The sources of the sequences: Portnoy *et al.* (1986) (MCP, aleurain, actinidin, papain, DVP, catB); Dufour *et al.* (1988) (catL); Ohno *et al.* (1984) (CDP).

[pTZ(ORF^{aug})] of the ORF did not yield the 37-kD protein. As predicted, only the products of the 5′-half were precipitated by this antibody (Fig. 5B, lane 4). These results are in agreement with the results obtained with IBV (Brierly *et al.*, 1987) and MHV-A59 (Bredenbeek *et al.*, 1990).

Analysis of sequence homology among MHV-JHM, MHV-A59, and IBV. The comparison of nucleotide and predicted amino acid sequences between MHV-JHM and IBV revealed considerable similarity between the two. The dot matrix comparison of the amino acid sequences shows that ORF 1b is very similar between MHV and IBV (Fig. 6). Overall, there are 47.7% similarity at nucleotide level and 52.8% at amino acid level. Similar to the ORF 1b of IBV, the MHV ORF contains the polymerase and helicase motifs at the corresponding positions (Gorbalenya *et al.*, 1989b) (data not shown). The putative zinc-binding domain is also largely conserved between the two viruses. On the other hand, two of the residues implicated in metal binding for IBV (Gorbalenya *et al.*, 1989b) are replaced in MHV, suggesting that the specific structures of the putative "fingers" may differ (Fig. 7). The ORF 1b of MHV-JHM and MHV-A59 are also very similar (95.9% at nucleotide level, and 94.9% at amino acid level) (data not shown).

In contrast, the ORF 1a is more diverged (Fig. 8). The MHV ORF 1a is longer than the corresponding IBV ORF by 537 amino acids. The C-terminal half of the ORF 1a is relatively conserved between MHV-JHM and IBV, while the N-terminal half is very diverged (Fig. 6). The alignment of amino acids in ORF 1a of MHV-JHM and IBV showed that there are four possible stretches of moderate homology which are separated by highly diverged sequences (Fig. 8).

Analysis of the functional domains of ORF 1a. Although ORF 1a is highly diverged between MHV-JHM and IBV, common functional domains could be identified in this ORF of both viruses by detailed amino acid sequence analysis (see Materials and Methods) (Fig. 9). Two hydrophobic, potentially membrane-anchoring regions are present in the C-terminal half. There are three cysteine-rich domains, one of which contains a segment distantly resembling growth factors and their receptors (Gorbalenya *et al.,* 1989b). In both coronaviruses, homologous domains of about 300 residues each have been identified to be related to the putative 3C-like proteases ($3CL^{pro}$) of picorna-, como-, nepo-, poty-, sobemo- and luteoviruses (Gorbalenya *et al.,* 1989b). The sequences of the putative coronavirus 3C-like proteases possess certain unusual features distinct from that of other viral 3C-like proteases (Fig. 10, and see Discussion). The search for sequences resembling the cleavage sites for the 3C-like proteases revealed six conserved putative target sites for the MHV and IBV 3C-like proteases (Table 1) (see Discussion). These potential cleavage sites are localized in the ORF 1b and the C-terminal half of the ORF 1a. Interestingly, the N-terminal one of these cleavage sites marks the N-end of the putative 3C-like protease itself. Finally, there is a region of moderate conservation between MHV and IBV, which contains short segments resembling those around the catalytic Cys and His residues of papain-like proteases (Fig. 11). This region is duplicated in the MHV genome, but not in IBV, at an upstream site in the ORF 1a. This upstream papain-like cysteine protease has been identified as the one responsible for the cleavage of p28 from the N-terminus of the gene 1 protein (Baker *et al.,* 1989). A domain of considerable conservation between MHV and IBV (X domain in Fig. 9) has been found next to the putative coronavirus papain-like proteases. Interestingly, a homologous conservative domain also flanks the putative thiol proteases of alpha- and rubiviruses (A. E. Gorbalenya, unpublished observations).

DISCUSSION

The complete sequence of gene 1 of MHV presented in this paper shows that this gene is probably the largest known viral gene among RNA viruses. Evidence was presented suggesting that the two ORFs in this gene may be translated into a large polyprotein. This interpretation is consistent with the lack of the transcriptional initiation signal (UCUAAAC) in the entire gene 1 sequence except at the extreme 5′-end. Although the putative "slippery" sequence (UUUAAAC) between the ORF 1a and 1b (Brierly *et al.,* 1989) is similar to the transcriptional initiation signal, no major subgenomic mRNAs have been detected within this gene. Thus, this gene most likely encodes a single polyprotein of at least 800 kDa. The total size of the RNA genome of MHV is approximately 31 kb, which is considerably larger than any of the other known viral RNA. The evolution of the coronavirus RNA genome into such a large RNA may have reflected the unusual mechanism of coronavirus RNA synthesis. The complexity of the discontinuous mode of coronavirus RNA synthesis (Lai, 1988) suggests that the coronavirus RNA polymerase needs a variety of different enzymatic activities.

The amino acid sequence of gene 1 of MHV shows considerable similarity to that of IBV. The ORF 1b is particularly conserved. Its degree of conservation between MHV and IBV is higher than that for any of the other genes in the coronavirus genomes. The ORF 1b contains the polymerase, helicase, and metal-binding motifs (Gorbalenya *et al.,* 1989b), suggesting that this region may be directly involved in RNA synthesis. These structural features are conserved between these viruses. The proposed pseudoknot structure which is important for the ribosomal frameshifting for cotranslation of ORF 1a and ORF 1b (Brierley *et al.,* 1989) is also highly conserved. This fact has previously been recognized in the partial sequence of gene 1 of MHV-A59 (Bredenbeek *et al.,* 1990). The sequence differences between MHV-A59 and MHV-JHM within this junction region are located at the nucleotides which do not affect the putative pseudoknot structure. In contrast, ORF 1a is much more diverged. It is nearly 2 kb longer than the ORF 1a of IBV, and contains several stretches of sequence which are not present in the IBV genome. These nonhomologous stretches of sequence are interspersed between the conserved regions. Furthermore, a papain-like protease domain, which is present once in the IBV genome, is duplicated in the 5′-half of the ORF 1a of MHV. The N-terminal sequence including p28, which is cleaved by the papain-like protease of MHV (Baker *et al.,* 1989), is also highly diverged between MHV and IBV. Thus, it appears that the 5′-end of ORF 1a has undergone considerable sequence rearrangement and possibly recombination, while the remaining sequences in gene 1 are almost colinear between MHV and IBV.

In contrast to the ORF 1b which contains sequence motifs related to the synthesis of RNA, the ORF 1a contains several domains suggestive of other functions. First of all, there are two long stretches of hydrophobic domains, which are conserved between IBV and MHV. The presence of these domains suggests that the gene 1 products may be anchored to the membrane. This possibility is consistent with the finding that MHV RNA synthesis occurs on the membrane fractions in the infected cells (Brayton *et al.,* 1982). Second, there are three cysteine-rich regions, which are also homologous between MHV and IBV. The function of the Cys-rich domains is still not clear. However, it has been noted with IBV that the C-terminal Cys-rich domain is related to that of the growth factors and their receptors (Gorbalenya *et al.,* 1989b). Third, there is a 3C-like protease domain ($3CL^{pro}$) in the 3′-half of ORF 1a, which is also conserved in IBV. The putative catalytic His and Cys residues previously predicted in IBV have also been observed in MHV (Fig. 10). However, the putative coronavirus proteases remain unique in that they do not contain a conserved Asp(Glu) residue that could serve as the third catalytic residue as suggested for the other 3C-like proteases (Gorbalenya *et al.,* 1989b). Furthermore, the unusual substitution of Tyr for Gly in the putative substrate-binding region, described previously in IBV, is also observed in the putative MHV $3CL^{pro}$ (Fig. 10). The potential cleavage sites for this 3C-like protease have been identified to be mainly in ORF 1b and the C-terminus of ORF 1a (Gorbalenya *et al.,* 1989b). These sites (QS) are either conserved or converted to QA in MHV (Table 1). The potential cleavage at Q/S and Q/A sites by picornavirus $3CL^{pro}$ has been demonstrated previously (Parks and Palmenberg, 1987). Two QG dipeptides proposed to be cleaved in IBV were substituted in MHV by QC in one case, and by KR dipeptide in another (Table 1). Substitution of a C (unlike several other residues) for G in a cleavage site for encephalomyocarditis virus protease did not abolish processing in an *in vitro* system (Parks *et al.,* 1989). Dibasic dipeptides are cleaved in the polyproteins of flaviviruses (Strauss and Strauss, 1988). Thus, these postulated cleavage sites are potentially cleavable by MHV $3CL^{pro}$ despite the divergence. These cleavages could separate different functional domains of the gene 1 polyprotein into distinct protein products. Whether these sites are indeed cleaved in MHV-infected cells remains to be studied. Fourthly, the N-terminal portion, which is the most diverged region, contains a papain-like protease domain as pointed out previously for IBV (Gorbalenya *et al.,* 1989b). The papain protease domain is duplicated in the MHV ORF 1a (Fig. 11) and is homologous with the known proteases (Fig. 12). This protease is probably involved in the cleavage of the N-terminus of the gene 1 polyprotein (Baker *et al.,* 1989), which has been demonstrated in MHV-infected cells (Denison and Perlman, 1987). Site-specific mutagenesis studies demonstrated that this protease has Cys and His at its active site (unpublished observation).

The possible presence of the protease domains suggests that the gene 1 polyprotein is processed into many proteins. It has been shown that there are at least five to six complementation groups involving MHV RNA synthesis, five of which have been mapped within gene 1 (Leibowitz *et al.,* 1982; Baric *et al.,* 1990). These proteins conceivably participate in various aspects of MHV RNA synthesis. None of the proteins have been detected so far.

ACKNOWLEDGMENTS

We thank Dr. Susan Baker for advice throughout the course of the study. We also thank Lisa Banner and Daphne Shimoda for editorial assistance. A.E.G. and E.V.K. are most grateful to Professor V. I. Agol for constant support, to Dr. L. I. Brodsky for supply of the GENBEE program package, and to Dr. F. Corpet for sending the program MULTALIN. This work was supported by Public Health Service Research Grants AI19244 and NS181146 from the National Institutes of Health. N.L.M. is a postdoctoral fellow of the National Multiple Sclerosis Society. M.M.C.L. is an Investigator of Howard Hughes Medical Institute.

REFERENCES

ARMSTRONG, J., SMEEKENS, S., and ROTTIER, P. (1983). Sequence of the nucleocapsid gene from murine coronavirus MHV-A59. *Nucleic Acids Res.* **11,** 883–891.

ARMSTRONG, J., NIEMANN, H., SMEEKENS, S., ROTTIER, P., and WARREN, G. (1984). Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature (London)* **308,** 751–752.

BAKER, S. C., LA MONICA, N., SHIEH, C.-K., and LAI, M. M. C. (1990). Murine coronavirus gene 1 polyprotein contains an autoproteolytic activity. *In* "Pathogenesis and Molecular Biology of Coronavirus" (D. Cavanagh, and T. D. K. Brown, Eds.). Plenum, New York (in press).

BAKER, S. C., SHIEH, C.-K., SOE, L. H., CHANG, M.-F., VANNIER, D. M., and LAI, M. M. C. (1989). Identification of a domain required for autoproteolytic cleavage of murine coronavirus gene A polyprotein. *J. Virol.* **63,** 3693–3699.

BARIC, R. S., FU, K., SCHAAD, M. C., and STOHLMAN, S. A. (1990). Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. *Virology* **177,** 646–656.

BOURSNELL, M. E. 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.

BRAYTON, P. R., LAI, M. M. C., PATTON, C. D., and STOHLMAN, S. A. (1982). Characterization of two RNA polymerase activities induced by mouse hepatitis virus. *J. Virol.* **42,** 847–853.

BREDENBEEK, 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 polymerase gene of the coronavirus MHV-A59; a highly conserved poly-

merase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res.* **18,** 1825–1832.

BRIERLEY, I., BOURSNELL, M. E. G., BINNS, M. M., BILIMORIA, B., BLOK, V. C., BROWN, T. D. K., and INGLIS, S. C. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J.* **6,** 3779–3785.

BRIERLEY, I., DIGARD, P., and INGLIS, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot. *Cell* **57,** 537–547.

CAVANAGH, D., BRIAN, D. A., ENJUANES, L., HOLMES, K. V., LAI, M. M. C., LAUDE, H., SIDDELL, S. G., SPAAN, W., TAGUCHI, F., and TALBOT, P. J. (1990). Recommendations of the coronavirus study group for the nomenclature of the structural proteins, mRNAs and genes of coronavirus. *Virology* **176,** 306–307.

CHEN, E. J., and SEEBURG, P. H. (1985). Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. *DNA* **4,** 165–170.

CORPET, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16,** 10,881–10,890.

DAGERT, M., and ERLICH, S. D. (1979). Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene* **6,** 23–29.

DENISON, M., and PERLMAN, S. (1987). Identification of putative polymerase gene product in cells infected with murine coronavirus A59. *Virology* **157,** 565–568.

DEVEREUX, J., HAEBERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12,** 387–395.

DUFOUR, E., OBLED, A., VALIN, C., and BECHET, D. (1987). Purification and amino acid sequence of chicken liver cathepsin L. *Biochemistry* **26,** 5689–5695.

GORBALENYA, A. E., BLINOV, V. M., DONCHENKO, A. P., and KOONIN, E. V. (1989a). An NTP-binding motif is the most conserved sequence in a highly diverged monophyletic group of proteins involved in positive strand RNA viral replication. *J. Mol. Evol.* **28,** 256–258.

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

GRIBSKOV, M., DEVEREUX, J., and BURGESS, R. R. (1984). The codon preference plot: Graphic analysis of protein coding sequences and prediction of gene expression. *Nucleic Acids Res.* **12,** 539–549.

GUBLER, U., and HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25,** 263–269.

HIRANO, N., FUJIWARA, K., HINO, S., and MATSUMOTO, M. (1974). Replication and plaque formation of mouse hepatitis virus (MHV-2) in mouse cell line DBT culture. *Arch. Gesamte Virusforsch.* **44,** 298–302.

KECK, J. G., STOHLMAN, S. A., SOE, L. H., MAKINO, S., and LAI, M. M. C. (1987). Multiple recombination sites at the 5′-end of murine coronavirus RNA. *Virology* **156,** 331–341.

KOONIN, E. V., CHUMAKOV, K. M., and GORBALENYA, A. E. (1990). A method for localization of motifs in amino acid sequences. *Biopolim. Kletka,* in press.

KYTE, J., and DOOLITTLE, R. F. (1982). A simple method for displaying the pathic character of a protein. *J. Mol. Biol.* **157,** 105–132.

LAI, M. M. C. (1988). Replication of coronavirus RNA. *In* "RNA Genetics" (E. Domingo, J. J. Holland, and P. Ahlquist, Eds.), Vol. I, pp. 115–136. CRC, Boca Raton, FL.

LAI, M. M. C. (1990). Coronavirus: Organization, replication and expression of genome. *Annu. Rev. Microb.,* **44,** 303–333.

LAI, M. M. C., BRAYTON, P. R., ARMEN, R. C., PATTON, C. D., PUGH, C., and STOHLMAN, S. A. (1981). Mouse hepatitis virus A59: Messenger RNA structure and genetic localization of the sequence divergence from the hepatotropic strain MHV 3. *J. Virol.* **39,** 823–834.

LAI, M. M. C., and STOHLMAN, S. A. (1978). The RNA of mouse hepatitis virus. *J. Virol.* **26,** 236–242.

LEIBOWITZ, J. L., DEVRIES, J. R., and HASPEL, M. V. (1982). Genetic analysis of murine hepatitis virus strain JHM. *J. Virol.* **42,** 1080–1087.

LEONTOVICH, A. M., BRODSKY, L. I., and GORBALENYA, A. E. (1990). A method for generation of complete local similarity maps between two amino acid sequences. DOTHELIX program of the GENBEE package. *Biopolim. Kletka,* in press.

LUYTJES, W., STURMAN, L. S., BREDENBEEK, P. J., CHARITE, J., VAN DER ZEIJST, B. A. M., HORZINEK, M. C., and SPAAN, W. J. (1987). Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and identification of the trypsin cleavage site. *Virology* **161,** 479–487.

LUYTJES, W., BREDENBEEK, P. J., NOTEN, A. F., HORZINEK, M. C., and SPAAN, W. J. (1988). Sequence of mouse hepatitis virus A59 mRNA 2: Indications for RNA recombination between coronavirus and influenza C virus. *Virology* **166,** 415–422.

MAKINO, S., TAGUCHI, F., HIRANO, N., and FUJIWARA, K. (1984). Analysis of genomic and intracellular viral RNAs of small plaque mutants of mouse hepatitis virus, JHM strain. *Virology* **139,** 138–151.

MAXAM, A. M., and GILBERT, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *In* "Methods in Enzymology" (L. Grossman and K. Moldave, Eds.), Vol. 65, pp. 499–560. Academic Press, San Diego, CA.

OHNO, S., EMORI, Y., IMAJOH, S., KAWASAKI, H., KISARAGI, M., and SUZUKI, K. (1984). Evolutionary origin of a calcium-dependent protease by fusion of genes for a thiol protease and a calcium-binding protein. *Nature (London)* **312,** 566–570.

PORTNOY, D. A., ERICKSON, A. H., KOCHAN, J., RAVETCH, J. V., and UNKELESS, J. C. (1986). Cloning and characterization of a mouse cysteine proteinase. *J. Biol. Chem.* **261,** 14,697–14,703.

PACHUK, C. J., BREDENBEEK, P. J., ZOLTICK, P. W., SPAAN, W. J. M., and WEISS, S. R. (1989). Molecular cloning of the gene encoding the putative polymerase of mouse hepatitis coronavirus strain A59. *Virology* **171,** 141–148.

PARKS, G. D., BAKER, J. C., and PALMENBERG, A. C. (1989). Proteolytic cleavage of encephalomyocarditis viral capsid region substrates by precursors to the 3C enzyme. *J. Virol.* **63,** 1054–1058.

PARKS, G. D., and PALMENBERG, A. C. (1987). Site-specific mutations at a picornavirus VP3/VP1 cleavage site disrupt in vitro processing and assembly of capsid precursors. *J. Virol.* **61,** 3680–3687.

SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARF, S. J., HIGUCHI, R., HORN, G. T., MULLIS, K. B., and ERLICH, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487–491.

SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

SCHMIDT, I., SKINNER, M., and SIDDELL, S. (1987). Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus MHV-JHM. *J. Gen. Virol.* **68,** 47–56.

SHIEH, C.-K., LEE, H.-J., YOKOMORI, K., LA MONICA, N., MAKINO, S., and LAI, M. M. C. (1989). Identification of a new transcriptional initiation site and the corresponding functional gene 2b in the murine coronavirus RNA genome. *J. Virol.* **63,** 3729–3736.

SHIEH, C.-K., SOE, L. H., MAKINO, S., CHANG, M.-F., STOHLMAN, S. A., and LAI, M. M. C. (1987). The 5′-end sequence of the murine coronavirus genome: Implications for multiple fusion sites in leader-primed transcription. *Virology* **56,** 321–330.

SHIN, S. U., and MORRISON, S. L. (1989). Production and properties of chimeric antibody molecules. *In* "Methods in Enzymology" (J. J.

Langone, Ed.), Vol. 178, pp. 459–476. Academic Press, San Diego, CA.

SKINNER, M. A., EBNER, D., and SIDDELL, S. G. (1985). Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. *J. Gen. Virol.* **66,** 581–592.

SKINNER, M. A., and SIDDELL, S. G. (1983). Coronavirus JHM: Nucleotide sequence of the mRNA that encodes nucleocapsid protein. *Nucleic Acids Res.* **11,** 5045–5054.

SKINNER, M. A., and SIDDELL, S. G. (1985). Coding sequence of coronavirus MHV-JHM mRNA 4. *J. Gen. Virol.* **66,** 593–596.

SOE, L. H., SHIEH, C.-K., BAKER, S. C., CHANG, M.-F., and LAI, M. M. C. (1987). Sequence and translation of the murine coronavirus 5′-end genomic RNA reveals the N-terminal structure of the putative RNA polymerase. *J. Virol.* **61,** 3968–3976.

SPAAN, W. M., CAVANAGH, D., and HORZINEK, M. C. (1988). Coronaviruses: Structure and genome expression. *J. Gen. Virol.* **69,** 2939–2952.

STRAUSS, J. H., and STRAUSS, E. G. (1988). Replication of the RNAs of alphaviruses and flaviviruses. *In* "RNA Genetics" (E. Domingo, J. J. Holland, and P. Ahlquist, Eds.), Vol. I, pp. 71–90. CRC, Boca Raton, FL.

VEIDT, I., LOT, H., LEISER, M., SCHEIDECKER, D., GUILLEY, H., RICHARDS, K., and JONARD, J. (1988). Nucleotide sequence of beet western yellows RNA. *Nucleic Acids Res.* **16,** 9917–9932.

WEGE, H., MULLER, A., and TER MEULEN, V. (1978). Genomic RNA of the murine coronavirus JHM. *J. Gen. Virol.* **41,** 217–227.

WU, S., RINEHART, C. A., and KAESBERG, P. (1987). Sequence and organization of southern bean mosaic virus RNA. *Virology* **161,** 73–80.