VIROLOGY 166, 415–422 (1988)

Sequence of Mouse Hepatitis Virus A59 mRNA 2: Indications for RNA Recombination between Coronaviruses and Influenza C Virus

WILLEM LUYTJES, PETER J. BREDENBEEK, ANS F. H. NOTEN, MARIAN C. HORZINEK, AND WILLY J. M. SPAAN[1]

Institute of Virology, Veterinary Faculty, State University of Utrecht, P.O. Box 80.165, Utrecht, The Netherlands

Received January 27, 1988; accepted May 30, 1988

The nucleotide sequence of the unique region of coronavirus MHV-A59 mRNA 2 has been determined. Two open reading frames (ORF) are predicted: ORF1 potentially encodes a protein of 261 amino acids; its amino acid sequence contains elements which indicate nucleotide binding properties. ORF2 predicts a 413 amino acids protein; it lacks a translation initiation codon and is therefore probably a pseudogene. The amino acid sequence of ORF2 shares 30% homology with the HA1 hemagglutinin sequence of influenza C virus. A short stretch of nucleotides immediately upstream of ORF2 shares 83% homology with the MHC class I nucleotide sequences. We discuss the possibility that both similarities are the result of recombinations and present a model for the acquisition and the subsequent inactivation of ORF2; the model applies also to MHV-A59-related coronaviruses in which we expect ORF2 to be still functional.
© 1988 Academic Press, Inc.

INTRODUCTION

Murine hepatitis virus (MHV) is the most widely studied member of the Coronaviridae. This family of enveloped, single-stranded RNA viruses causes considerable economic loss, since coronavirus infections can severely affect cattle, poultry, and pets. Human coronavirus OC43 causes the common cold in man. Murine coronaviruses are of particular interest because several strains can cause a (chronic) demyelinating disease in rats and mice. For this reason the pathogenesis of MHV infections is studied as an animal model for virus-induced demyelination (Wege *et al.*, 1982). MHV-A59 virions contain an infectious RNA genome, about 30 kb in length, associated with a nucleocapsid protein (N). Two membrane proteins have been identified: the transmembrane glycoprotein E1 and the large surface glycoprotein E2 (Siddell *et al.*, 1982). The MHV-A59 genome is composed of seven different regions (A to G), separated by short, very similar junction sequences (Bredenbeek *et al.*, 1987). The messenger RNAs that are synthesized during infection are 3′-coterminal, and each extends to a different junction sequence in the 5′-direction. This results in a nested set of mRNAs, including the genome, in which each has a different "unique" region at its 5′-end (Leibowitz *et al.*, 1981; Lai *et al.*, 1981; Spaan *et al.*, 1982). All mRNAs share a leader sequence of about 72 nucleotides (Spaan *et al.*, 1983; Lai *et al.*, 1984). *In vitro* translated MHV mRNAs encoding the structural proteins N, E1, and E2 and the 14.5K nonstructural protein are functionally monocistronic (Rottier *et al.*, 1981; Siddell, 1983), and sequence analyses have shown that the coding regions are located at the 5′-end of these individual mRNAs (Siddell, 1987). There is one possible exception: sequence analysis of the 5′-end of mRNA 5 (region E) revealed two open reading frames (Skinner *et al.*, 1985; Budzilowicz and Weiss, 1987). Whether both reading frames are used is not known.

The coronaviruses studied to date show an identical order of the genes encoding the structural proteins: 5′-E2-E1-N-3′ (De Groot *et al.*, 1987). Between coronaviruses these genes are highly homologous. In contrast, differences are found in the structure and number of the genes encoding the nonstructural proteins, which is reflected in the number of subgenomic mRNAs that is synthesized by each coronavirus. In infectious bronchitis virus (IBV), feline infectious peritonitis virus (FIPV), and its close relative transmissible gastroenteritis virus (TGEV), members of different antigenic clusters from MHV, the largest subgenomic mRNA encodes the peplomer protein E2 or S (Binns *et al.*, 1985; Niesters *et al.*, 1986; De Groot *et al.*, 1987; Rasschaert and Laude, 1987; Jacobs *et al.*, 1987). In contrast, in MHV-infected cells an additional, larger RNA (mRNA 2) has been identified (Spaan *et al.*, 1981; Weiss and Leibowitz, 1983). *In vitro* translation of this mRNA yields a 30K–35K protein (Leibowitz *et al.*, 1982; Siddell, 1983). In MHV-JHM-infected cells, small amounts of a 30K protein can be detected (Siddell *et al.*, 1981). However, the size of the unique region of mRNA 2, approximately 2 kb, indicates a larger coding capacity.

In order to study the function of mRNA 2 we have cloned and sequenced region B of MHV-A59. Here we

[1] To whom requests for reprints should be addressed.

0042-6822/88 $3.00
Copyright © 1988 by Academic Press, Inc.
All rights of reproduction in any form reserved.

present its primary structure and show that it contains two open reading frames (ORF). The predicted amino acid sequence of the second ORF is remarkably similar to the HA1 sequence of the hemagglutinin protein of influenza C virus. We discuss the possibility that this ORF has been acquired by a recombination event.

MATERIALS AND METHODS

cDNA synthesis and cloning

A MHV-A59-specific cDNA library was created using random primers on purified genomic RNA. Procedures were identical to those described previously (Luytjes *et al.,* 1987). Full details will be presented elsewhere (P. J. Bredenbeek *et al.,* manuscript in preparation).

Selection and analysis of cDNA clones

Recombinant cDNA clones were selected by hybridization (Meinkoth and Wahl, 1984) to oligonucleotide probes specific for the viral mRNAs (P. J. Bredenbeek *et al.,* manuscript in preparation). Plasmid DNA from recombinant clones was prepared according to Birnboim and Doly (1979). Inserts were subcloned into M13 vectors (Messing, 1983). Selection of M13 subclones specific for the unique region of mRNA 2 was performed by hybridizing phage supernatant to pentamer primed probes (Feinburg and Vogelstein, 1983; Roberts and Wilson, 1985) from previously oligonucleotide-selected cDNA clones.

DNA sequence analysis

Sequence analysis was essentially done according to Sanger *et al.* (1977). Computer assembly of sequence data was performed using the Staden program set (1986).

Similarity search of protein sequences

The predicted amino acid sequences were compared to the National Biomedical Research Foundation (NBRF) Protein Library (release 11) using the FASTP program set created by Lipman and Pearson (1985). Additional analysis of similarities was carried out with the DIAGON program of Staden (1982).

RESULTS

Isolation of region B specific cDNA clones

We have recently constructed an almost complete random-primed cDNA library of the MHV-A59 genome. A set of oligonucleotides was synthesized, based upon the sequence of previously obtained MHV-A59-specific cDNA clones which had been mapped on the viral mRNAs (P. J. Bredenbeek, manuscript in preparation). Oligonucleotides OL 4 (specific for mRNA 1), OL 6 (mRNA 2), and OL 7 (mRNA 3, see Luytjes *et al.,* 1987) were used to screen the cDNA library for clones covering region B. Two completely overlapping clones (30, 96) and several clones with partial overlaps (4D, 35, F71, 95, 918) were isolated. Clone 96 was digested with *Sau*3A and subsequently ligated into the *Bam*H1 site of M13mp9. The other selected cDNA clones were subcloned using restriction enzymes as indicated in Fig. 1. Each nucleotide of region B was determined on at least two different cDNA clones and selected regions on three or more cDNA clones.

Identification of the unique region of mRNA 2

The 3′-end of region B has already been identified at the junction sequence 5′-UAAUCUAAAC-3′, which separates it from the peplomer coding sequence (Luytjes *et al.,* 1987). The only other potential junction sequence within the consensus sequence of the region B-specific cDNA clones was found at position −9589 (Fig. 1) from the start of the poly(A)-tail of the genome: 5′-AAAUCUAUAC-3′ (Fig. 2). Immediately upstream of this sequence an ORF terminates, the primary structure of which shows a high similarity to the 3′-terminal sequence of the unique region of IBV mRNA F (Boursnell *et al.,* 1987, and data not shown). This strongly suggests that the junction sequence at position −9589 corresponds to the 5′-end of the unique region of mRNA 2.

Nucleotide and amino acid sequence

The consensus nucleotide sequence of region B is 2176 residues long (Fig. 2). It contains two open reading frames. The first open reading frame (ORF1) starts 18 nucleotides downstream from the junction sequence and is 261 amino acids (aa) long. The second ORF (ORF2) starts 903 nucleotides downstream and is 413 aa long. It terminates 23 nucleotides upstream from the junction sequence that separates regions B and C (the peplomer gene). Between ORF1 and ORF2 lies a stretch of 92 nucleotides with several termination codons in each reading frame (see Fig. 2).

Analysis of ORF1

In ORF1 three potential translation initiation codons can be found. The first AUG is in a strong context (Kozak, 1986) and is therefore most probably used. The coding capacity of ORF1 is 30K, which is in agreement with the products obtained after *in vitro* translation of mRNA 2. There are no membrane protein sequence characteristics, such as a signal sequence, a transmembrane anchor sequence, or potential N-glycosylation sites. Diagon comparison (Staden, 1982) of the

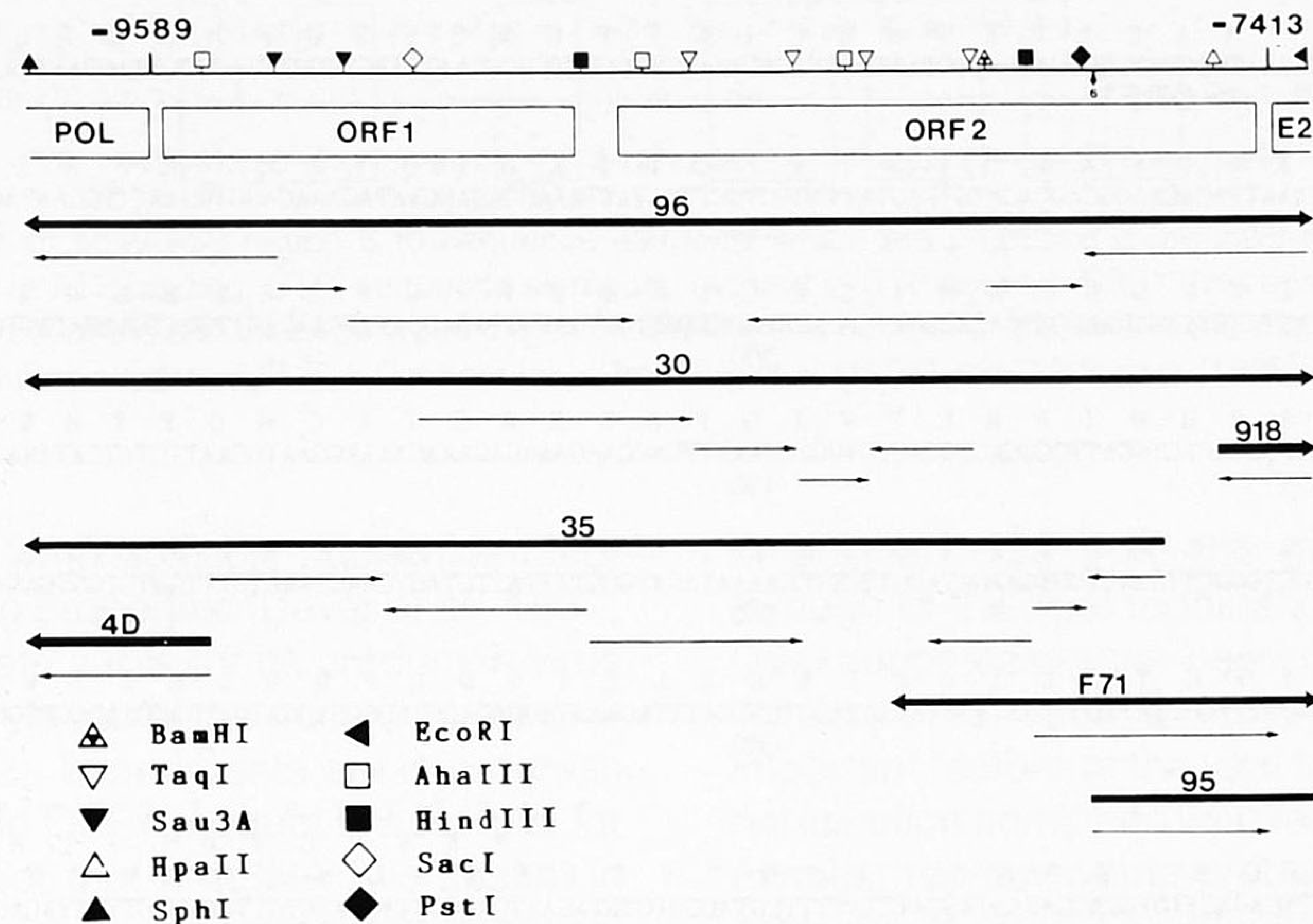

FIG. 1. Cloning and sequencing strategy of the MHV-A59 region B. The upper line represents the MHV genome. Symbols indicate the restriction enzyme recognition sites (specified in the figure) used in subcloning. Vertical bars and the negative numbers above mark the starts of the junction sequences and the distances to the start of the poly(A)-tail of the genome. The arrow points to the position of oligonucleotide 6 (OL 6). Open boxes represent open reading frames. Pol, polymerase; E2, peplomer protein; ORF1 and ORF2 are region B open reading frames. Numbered bars refer to cDNA clones; direction and extent of sequencing of subclones is indicated by the arrows below.

ORF1 amino acid sequence with available sequences of other coronaviruses did not reveal any similarities. A FASTP similarity search (Lipman and Pearson, 1985) of the NBRF protein library produced an alignment to several proteins with nucleotide binding properties (data not shown). Recently, consensus sequence elements have been published, for which an involvement in nucleotide binding is proposed (Dever *et al.*, 1987; Fry *et al.*, 1986). Three regions in the ORF1 sequence match to these elements (Fig. 3).

Analysis of ORF2

ORF2 does not start with an AUG codon; the first potential initiation codon within ORF2 is found at position 110. Interestingly, in the region upstream of ORF2 an AUG codon (position 879) is found in a favorable context, which precedes a short reading frame, separated from ORF2 by only one opal termination codon (Fig. 2). This short reading frame is 90% homologous (83% at the nucleotide level) to the N-terminus of the signal sequence of several MHC class I genes (Fig. 4; Schepart *et al.*, 1986). There is no other significant similarity between class I sequences and any MHV sequence. The region overlapping the end of ORF1 and the beginning of ORF2 has been sequenced on three independent cDNA clones. The sequences are identical, excluding the possibility that the presence of the termination codon is a cloning or sequencing artifact.

The sequence of ORF2 shows characteristics of a membrane protein sequence: the C-terminal hydrophobic residues (underlined in Fig. 2) could provide a membrane anchor and 10 potential N-glycosylation sites are present.

The most remarkable aspect of the ORF2 sequence came from FASTP analysis of the NBRF protein library: the predicted amino acid sequence encoded by ORF2 shows a 30% homology with the HA1 sequence of the hemagglutinin protein of influenza C virus (Nakada *et al.*, 1984; Pfeiffer and Compans, 1984). The alignment presented in Fig. 5 shows that several regions are completely identical and that many conservative substitutions (Dayhoff *et al.*, 1983) are present.

We could not detect similarities between the predicted ORF2 amino acid sequence and other influenza C (or A or B) virus sequences, nor was there any similarity to available coronavirus sequences.

DISCUSSION

In this paper we present the primary structure of the unique region of MHV-A59 mRNA 2. Sequence analysis revealed two ORFs. ORF1 has a coding capacity of 30K. *In vitro* translation of mRNA 2 of MHV-JHM (Siddell, 1983) and MHV-A59 (Leibowitz *et al.*, 1982) yielded a 30K protein. Also in MHV-JHM-infected cells small amounts of a 30K protein have been detected (Siddell *et al.*, 1981). This suggests that this protein is encoded by ORF1 from mRNA 2. We assume that the ORF1 translation product is initiated at the 5′-proximal AUG since this codon is in a preferred context (Kozak, 1986). The presence of three consensus elements in

```
                         M  A  F  A  D  K  P  N  H  F  I  N  F  P  L  A  Q  F  S  G  F  M  G  K  Y  L  K  L  Q  S  Q     31
[AAATCTATAC]TTGTCGTGGCTGTGAAAATGGCCTTTGCTGACAAGCCTAATCATTTCATAAACTTTCCCCTGGCCCAATTTAGTGGCTTTATGGGTAAGTATTTAAAGCTACAGTCTCAA   120
                         ->ORF1              60

L  V  E  M  G  L  D  C  K  L  Q  K  A  P  H  V  S  I  T  L  L  D  I  K  A  D  Q  Y  K  Q  V  E  F  A  I  Q  E  I  I  D     71
CTTGTGGAAATGGGTTTAGACTGTAAATTACAGAAGGCACCACATGTTAGTATTACCCTGCTTGATATTAAAGCAGACCAATACAAACAGGTGGAATTTGCAATACAAGAAATAATAGAT   240
                                                   180

D  L  A  A  Y  E  G  D  I  V  F  D  N  P  H  M  L  G  R  C  L  V  L  D  V  R  G  F  E  E  L  H  E  D  I  V  E  I  L  R    111
GATCTGGCGGCATATGAGGGAGATATTGTCTTTGACAACCCTCACATGCTTGGCAGATGCCTTGTTCTTGATGTTAGAGGATTTGAAGAGTTGCATGAAGATATTGTTGAAATTCTCCGC   360
                                                   300

R  R  G  C  T  A  D  Q  S  R  H  W  I  P  H  C  T  V  A  Q  F  D  E  E  R  E  T  K  G  M  Q  F  Y  H  K  E  P  F  Y  L    151
AGAAGGGGTTGCACGGCAGATCAATCCAGACACTGGATTCCGCACTGCACTGTGGCCCAATTTGACGAAGAAAGAGAAACAAAAGGAATGCAATTCTATCATAAAGAACCCTTCTACCTC   480
                                                   420

K  H  N  N  L  L  T  D  A  G  L  E  L  V  K  I  G  S  S  K  I  D  G  F  Y  C  S  E  L  S  V  W  C  G  E  R  L  C  Y  K    191
AAGCATAACAACCTATTAACGGATGCTGGGCTTGAGCTCGTGAAGATAGGTTCTTCCAAAATAGATGGGTTTTATTGTAGTGAACTGAGTGTTTGGTGTGGTGAGAGGCTTTGTTATAAG   600
                                                   540

P  P  T  P  K  F  S  D  I  F  G  Y  Ć  C  I  D  K  I  R  G  D  L  E  I  G  D  L  P  Q  D  D  E  E  A  W  A  E  L  S  Y    231
CCTCCAACACCCAAATTCAGTGATATATTTGGCTATTGCTGCATAGATAAAATACGTGGTGATTTAGAAATAGGAGACCTACCGCAGGATGATGAGGAAGCGTGGGCCGAGCTAAGTTAC   720
                                                   660

                                                                                  *  N  E  G  L  Y  V  L  I  C  F  Y  T  I  S  V  I
                                                                                     *  R  V  V  C  V  D  L  F  L  H  Y  *  C  N  K
H  Y  Q  R  N  T  Y  F  F  R  H  V  H  D  N  S  I  Y  F  R  T  V  C  R  M  K  G  C  M  C  *  F  V  F  T  L  L  V  *  *    261
CACTATCAAAGAAACACCTACTTCTTCAGACATGTGCACGATAATAGCATCTATTTTCGTACCGTGTGTAGAATGAAGGGTTGTATGTGTTGATTTGTTTTTACACTATTAGTGTAATAA   840
                                                   780

S  L  L  F  C  *  K  G  Q  D  V  H  S  Y  G  S  S  H  T  A  F  A  D  L  M  S  A  G  V  W  V  Q  *
 L  I  I  L  L  K  R  A  G  C  A  *  L  W  L  L  A  H  C  F  C  *  F  D  V  S  W  C  L  G  S  M  N  L  L  T  S  F  H  I
A  Y  Y  F  V  E  K  G  R  M  C  I  A  M  A  P  R  T  L  L  L  L  I  *  C  Q  L  V  F  G  F  N  E  P  L  N  I  V  S  H     16
GCTTATTATTTTGTTGAAAAGGGCAGGATGTGCATAGCTATGGCTCCTCGCACACTGCTTTTGCTGATTTGATGTCAGCTGGTGTTTGGGTTCAATGAACCTCTTAACATCGTTTCACAT   960
                                                   900              ->ORF2

*
L  N  D  D  W  F  L  F  G  D  S  R  S  D  C  T  Y  V  E  N  N  G  H  P  K  L  D  W  L  D  L  D  P  K  L  C  N  S  G  K     56
TTAAATGATGACTGGTTTCTATTTGGTGACAGTCGTTCTGACTGTACCTATGTAGAAAATAACGGTCATCCTAAATTAGATTGGCTTGACCTCGACCCAAAGTTGTGTAATTCAGGAAAG  1080
                                                  1020

I  S  A  K  S  G  N  S  L  F  R  S  F  H  F  T  D  F  Y  N  Y  T  G  E  G  D  Q  I  V  F  Y  E  G  V  N  F  S  P  S  H     96
ATTTCCGCAAAGAGTGGTAACTCTCTCTTTAGGAGTTTTCACTTCACTGATTTTTACAATTATACGGGTGAGGGAGACCAAATTGTATTTTATGAAGGAGTTAATTTTAGTCCCAGCCAT  1200
                                                  1140

G  F  K  C  L  A  H  G  D  N  K  R  W  M  G  N  K  A  R  F  Y  A  R  V  Y  E  K  M  A  Q  Y  R  S  L  S  F  V  N  V  S    136
GGCTTTAAATGCCTGGCTCATGGAGATAATAAAAGATGGATGGGCAATAAAGCTCGATTTTATGCCCGAGTGTATGAGAAGATGGCCCAATATAGGAGCCTATCGTTTGTTAATGTGTCT  1320
                                                  1260

Y  A  Y  G  G  N  A  K  P  A  S  I  C  K  D  N  T  L  T  L  N  N  P  T  F  I  S  K  E  S  N  Y  V  D  Y  Y  Y  E  S  E    176
TATGCCTATGGAGGTAATGCAAAGCCCGCCTCCATTTGCAAAGACAATACTTTAACACTCAATAACCCCACCTTCATATCGAAGGAGTCTAATTATGTTGATTATTACTATGAGAGTGAG  1440
                                                  1380

A  N  F  T  L  E  G  C  D  E  F  I  V  P  L  C  G  F  N  G  H  S  K  G  S  S  S  D  A  A  N  K  Y  Y  T  D  S  Q  S  Y    216
GCTAATTTCACACTAGAAGGTTGTGATGAATTTATAGTACCGCTCTGTGGTTTTAATGGCCATTCCAAGGGCAGCTCTTCGGATGCTGCCAATAAATATTATACTGACTCTCAGAGTTAC  1560
                                                  1500

Y  N  M  D  I  G  V  L  Y  G  F  N  S  T  L  D  V  G  N  T  A  K  D  P  G  L  D  L  T  C  R  Y  L  A  L  T  P  G  N  Y    256
TATAATATGGATATTGGTGTCTTATATGGGTTCAATTCGACCTTGGATGTTGGCAACACTGCTAAGGATCCGGGTCTTGATCTCACTTGCAGGTATCTTGCATTGACTCCTGGTAATTAT  1680
                                                  1620

K  A  V  S  L  E  Y  L  L  S  L  P  S  K  A  I  C  L  H  K  T  K  R  F  M  P  V  Q  V  V  D  S  R  W  S  S  I  R  Q  S    296
AAGGCTGTGTCCTTAGAATATTTGTTAAGCTTACCCTCAAAGGCTATTTGCCTCCATAAGACAAAGCGCTTTATGCCTGTGCAGGTAGTTGACTCAAGGTGGAGTAGCATCCGCCAGTCA  1800
                                                  1740

D  N  M  T  A  A  A  C  Q  L  P  Y  C  F  F  R  N  T  S  A  N  Y  S  G  G  T  H  D  A  H  H  G  D  F  H  F  R  Q  L  L    336
GACAATATGACCGCTGCAGCCTGTCAGCTGCCATATTGTTTCTTTCGCAACACATCTGCGAATTATAGTGGTGGCACACATGATGCGCACCATGGTGATTTTCATTTCAGGCAGTTATTG  1920
                                                  1860

S  G  L  L  Y  N  V  S  C  I  A  Q  Q  G  A  F  L  Y  N  N  V  S  S  S  W  P  A  Y  G  Y  G  H  C  P  T  A  A  N  I  G    376
TCTGGTTTGTTATATAATGTTTCCTGTATTGCCCAGCAGGGTGCATTTCTTTATAATAATGTTAGTTCCTCTTGGCCAGCCTATGGGTACGGTCATTGTCCAACGGCAGCTAACATTGGT  2040
                                                  1980

Y  M  A  P  V  C  I  Y  D  P  L  P  V  I  L  L  G  V  L  L  G  I  A  V  L  I  I  V  F  L  N  V  L  F  Y  D  G  *          413
                                     ___________________________________________________________________
TATATGGCACCTGTTTGTATCTATGACCCTCTCCCGGTCATACTGCTAGGTGTGTTATTGGGTATAGCTGTGTTGATTATTGTGTTTTTGAATGTTTTATTTTATGACGGATAGCGGTGT  2160
                                                  2100

TAGATTGCATGAGGCA[TAATCTAAAC]    2186
```

FIG. 2. Nucleotide sequence of the MHV-A59 region B and predicted amino acid sequence of the open reading frames ORF1 and ORF2. Junction sequences (see text) are boxed. The start of the open reading frames is indicated by the arrows below the sequence. The region between ORF1 and ORF2 is translated in three reading frames. The hydrophobic C-terminus of ORF2 is underlined. ORF1 is numbered 1 (M)-261 (C), ORF2 is numbered 1 (C)- 413 (G). Nucleotide numbering starts at relative position −9589 from the start of the poly(A)-tail. Single letter amino acid code is used.

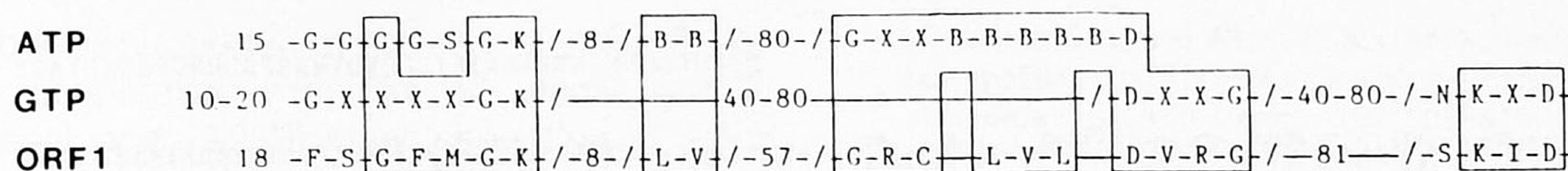

FIG. 3. Alignment of ORF1 of MHV-A59 region B to sequence elements which are proposed to be involved in nucleotide binding. ATP, sequence elements involved in ATP binding. GTP, sequence elements involved in GTP binding. ORF1, first open reading frame of region B of MHV-A59. The numbers represent the distances between the elements (the first number is the distance to the start of the sequences). X, any amino acid; B, hydrophobic amino acids L, V, F, Y, I. Data are taken from Fry *et al.* (1986) and Dever *et al.* (1987).

the sequence of ORF1 with possible nucleotide binding and phosphorylating properties (Dever *et al.*, 1987; Fry *et al.*, 1986) suggests a role for its product in virus replication or phosphorylation of the nucleocapsid protein (Siddell *et al.*, 1982). Experiments are in progress to establish whether the ORF1 product is essential for MHV, in view of the fact that a mRNA 2 is absent in cells infected with coronaviruses from other antigenic clusters.

Unexpected was the presence of a second open reading frame, ORF2, located between ORF1 and the peplomer gene, without a translation initiation codon, showing a remarkable amino acid similarity to the HA1 sequence of influenza C virus. The percentage of identity is high enough to rule out convergent evolution (Dayhoff *et al.*, 1983; Doolittle, 1981). We believe that this similarity is the result of a recombination between coronaviruses and influenza C virus. Recent studies have indicated that coronaviruses are indeed capable of recombination. Makino *et al.* (1986) described homologous recombination between coronaviruses in mixed infections; the stretch of 267 nucleotides that we have found in the MHV-A59 peplomer gene and that is absent in MHV-JHM (Luytjes *et al.*, 1987) could indicate a nonhomologous recombination.

In MHV-A59-infected cells a protein that can be assigned to ORF2 has never been detected (Siddell *et al.*, 1982). Since nonfunctional reading frames of RNA viruses show a high rate of mutation (Holland *et al.*, 1982), ORF2 must be either functional or the result of recent genetic changes. In the first case, possible ways of translating ORF2 would be either internal initiation at AUG codons in suboptimal contexts (which is unlikely) or protein initiation at an upstream AUG codon at position −33 from the start of ORF2 and read-through of the opal termination codon at position −3. Opal suppression has been reported for RNA viruses (Strauss *et al.*, 1983; Morch *et al.*, 1987) and can be an important feature of the viral translation strategy. Internal initiation combined with read-through of an opal termination codon would probably lead to undetectable amounts of protein in infected cells. The number and location of termination codons in the region between ORF1 and ORF2 excludes the possibility of frame shifting.

In the second case ORF2 could have been acquired recently by recombination between MHV and influenza C virus. However, there is considerable evolutionary distance between both viruses: the nucleotide sequences of ORF2 and the HA1 gene are not similar and the codon usage in both reading frames is different (data not shown). Therefore, recombination must have taken place between ancestors of these viruses. This means that closely related coronaviruses should exist in which ORF2 is still expressed and that ORF2 in MHV-A59 must have been recently inactivated by genetic changes. An ORF2 product would range in size from 45K (unglycosylated) to 65K (N-glycosylated) and several coronaviruses containing additional proteins in this range have been reported. MHV-JHM, which shares at least 87% homology with MHV-A59 in the nucleotide sequences from the peplomer gene down to the poly(A)-tail (Luytjes *et al.*, 1987), encodes one additional glycoprotein: gp 65 (Siddell, 1982). Sequence data indicate that the corresponding gene must be located upstream of the peplomer protein gene. Taguchi *et al.* (1985, 1986) described a JHM variant (CNS) which shows a high expression level of a 65K protein

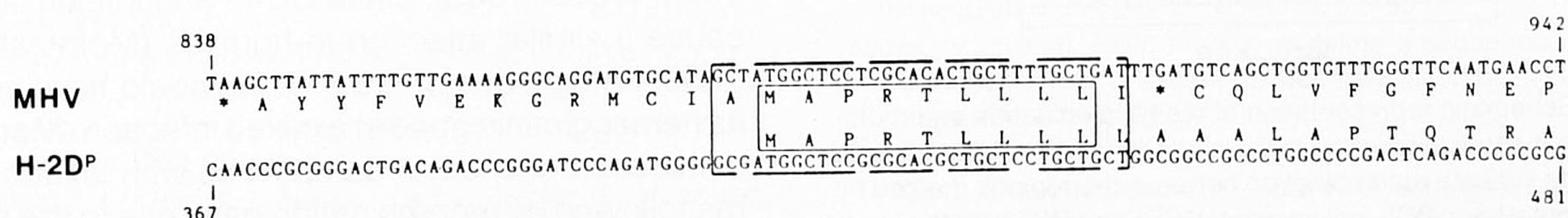

FIG. 4. Alignment of the MHV-A59 sequence around the start of ORF2 from region B and a MHC class I mRNA: H2-Dp. The MHV-A59 sequence is numbered according to Fig. 2. The H2 sequence is taken from Schepart *et al.* (1986). The H2-Dp amino acid sequence depicted represents the signal sequence. Identical nucleotides are marked with lines and identical amino acids are boxed.

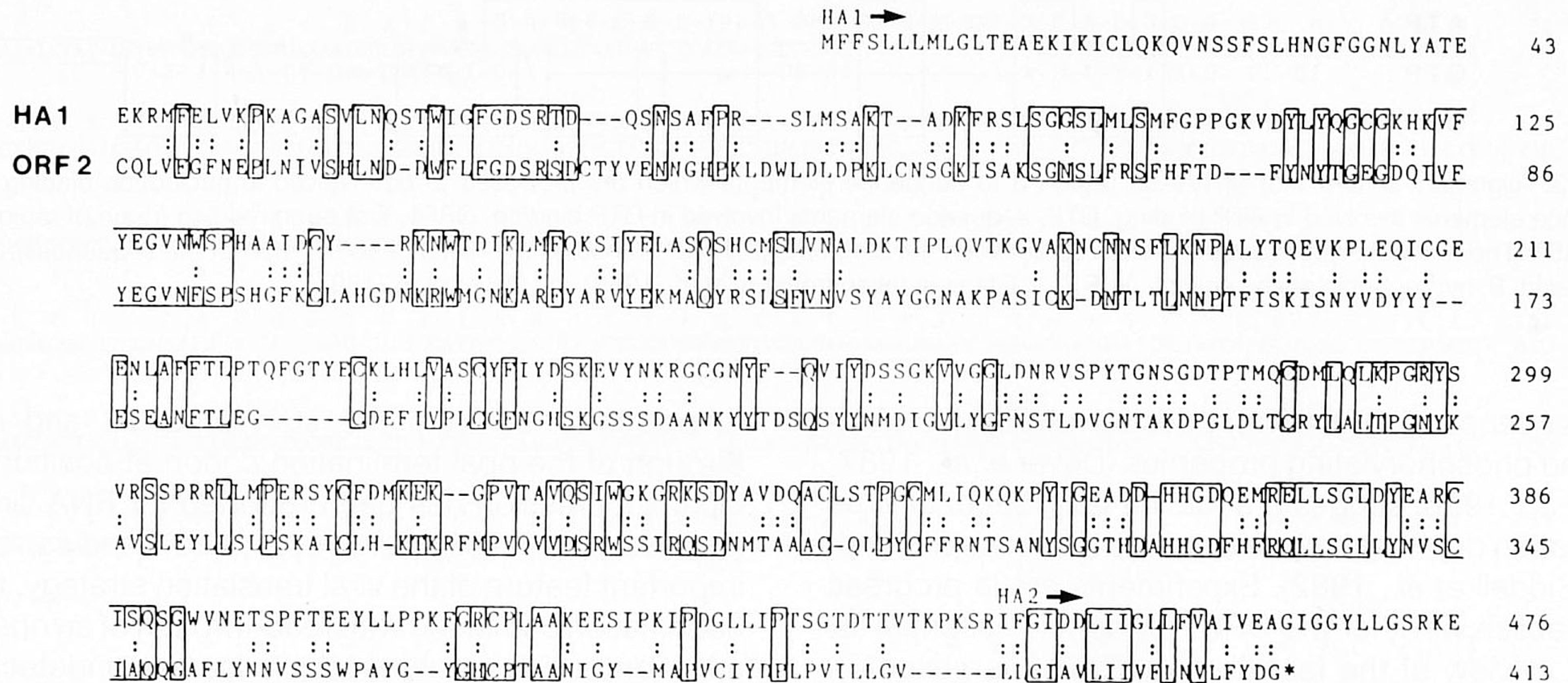

FIG. 5. Alignment of the MHV-A59 ORF2 sequence from region B and the influenza C hemagglutinin HA1 sequence and part of the HA2 sequence. Identical residues are boxed, substitutions scoring 0 or positive according to Dayhoff *et al.* (1983) are indicated by colons. Dashes represent gaps which were inserted to maximize similarity. The sequence was taken from Nakada *et al.* (1986).

and an additional mRNA 2a, intermediate in size between mRNA 2 and mRNA 3. Bovine coronavirus (BCV) shows a strong similarity to MHV-A59 in the nucleocapsid and matrix protein sequences (Lapps *et al.*, 1987) and it contains an additional spike protein E3, a hemagglutinin (King *et al.*, 1985; Deregt *et al.*, 1987). The size of the hemagglutinin monomer is 65K and BCV also encodes a mRNA 2a (Keck *et al.*, 1987). The data on these coronaviruses lead us to suggest that ORF2 in MHV-A59 corresponds to the reading frames

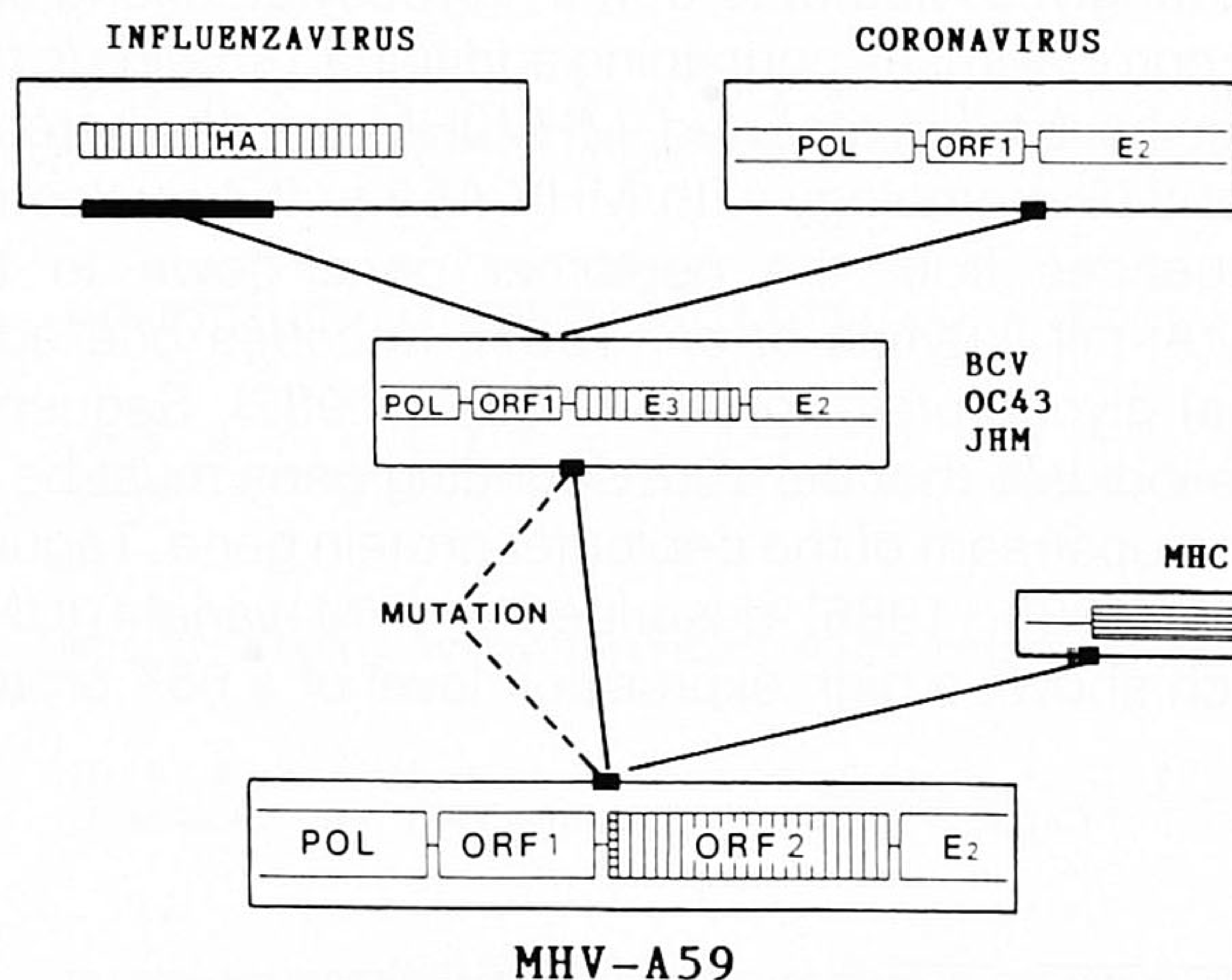

FIG. 6. Schematic representation of the recombination and mutation events that could have led to the situation in MHV-A59 region B. Drawn lines indicate recombination between the regions marked by dark horizontal bars. POL, polymerase; ORF1 and ORF2 are the reading frames of MHV-A59 region B; E2, peplomer protein gene; HA, hemagglutinin; E3, putative membrane protein (gp 65 of MHV-JHM, hemagglutinin of BCV and OC43); MHC, class I MHC mRNA.

encoding gp 65 in JHM and the 65K hemagglutinin E3 in BCV and that these genes are located on a separate mRNA 2a in the JHM CNS variant and in BCV. Junction sequences are involved in the initiation of coronavirus mRNAs. The apparent absence of a junction sequence upstream of ORF2 in MHV-A59 explains the absence of a mRNA 2a in infected cells (Spaan *et al.*, 1981; Weiss and Leibowitz, 1983). This could have been the result of an accumulation of recent point mutations. However, the strong similarity at both the amino acid and the nucleotide levels between the region immediately upstream of the opal termination codon (in front of ORF2) and the 5′-end of the coding region of several MHC class I mRNAs indicates that the initiation codon of ORF2 and the junction sequence upstream were lost because of a recent nonhomologous recombination event with MHC mRNA.

The suggested homology between ORF2 of MHV-A59 and the BCV E3 gene leads us to propose a model for the relation between several coronaviruses in the antigenic cluster of MHV. Human coronavirus OC43 is closely related to BCV (Lapps and Brian, 1985) and shows sequence similarity to MHV-A59 (Hogue *et al.*, 1984; Weiss, 1983). Since OC43 and influenza C virus cause a similar infection in humans (McIntosh *et al.*, 1969; Katagiri *et al.*, 1983) OC43 could have acquired its hemagglutinin gene in a mixed infection. More likely, coinfection of another coronavirus with influenza C virus followed by recombination gave rise to the new coronavirus OC43. The hemagglutinin gene of OC43 and BCV would then be the evolutionary intermediate between influenza C virus HA and MHV ORF2 (see Fig.

6). This model is supported by recent experiments performed in cooperation with Drs. R. Vlasak and P. Palese (Vlasak *et al.,* 1988) which show that BCV and OC43 recognize the same receptor and possess the same esterase activity as has been reported for the influenza C virus hemagglutinin protein (Vlasak *et al.,* 1987).

It has been suggested that virus evolution is a modular event, in which viral genomes are the result of the assembly of a set of primitive genes (see Goldbach, 1987). This mechanism can offer an alternative explanation for the relation between MHV and influenza C virus. However, the similarity with MHC RNA and the previously reported extra stretch of nucleotides in the A59 peplomer gene (Luytjes *et al.,* 1987) indicate that coronaviruses are probably capable of nonhomologous recombination during replication. To date nonhomologous recombination at the RNA level in animal RNA viruses has been reported only for defective interfering RNA (see King *et al.,* 1987). Coronaviruses are the first example of nontumor RNA viruses being able to take up directly into their genome genetic material from the host cell. This may be a strong force in generating strains with new host spectra and tissue tropisms and could have important implications for the prevention of coronavirus infections.

ACKNOWLEDGMENTS

The authors thank Dr. J. A. Lenstra for stimulating discussions, Drs. B. A. M. van der Zeijst and P. Rottier for critical reading of the manuscript, and Dr. A. Maagdenberg of the Duphar B. V. computer facility for setting up the computer programs. Part of the computer analyses have been performed using the CAOS-CAMM system of the University of Nijmegen, The Netherlands. W.L. was supported by a grant from Duphar B. V., Weesp, The Netherlands. P.B. was supported by a grant from the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO).

REFERENCES

Binns, M. M., Boursnell, M. E. G., Cavanagh, D., Pappin, D. J. C., and Brown, T. D. K. (1985). Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV. *J. Gen. Virol.* **66,** 719–726.

Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7,** 1513–1523.

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.

Bredenbeek, P. J., Charite, J., Noten, J. F. H., Luytjes, W., Horzinek, M. C., Van der Zeijst, B. A. M., and Spaan, W. J. M. (1987). Sequences involved in the replication of coronaviruses. *In* "Coronaviruses" (M. M. C. Lai, and S. A. Stohlman, Eds.), Vol. 218, pp. 65–72. Plenum, New York/London.

Budzilowicz, C. J., and Weiss, S. R. (1987). *In vitro* synthesis of two polypeptides from a nonstructural gene of coronavirus mouse hepatitis virus strain A59. *Virology* **157,** 509–515.

Dayhoff, M. O., Barker, W. C., and Hunt, L. T. (1983). Establishing homologies in protein sequences. *In* "Methods in Enzymology" (C. H. W. Hirs and S. N. Timasheff, Eds.), Vol. 91, pp. 524–545, Academic Press, New York.

De Groot, R. J., Maduro, J., Lenstra, J. A., Horzinek, M. C., Van der Zeijst, B. A. M., and Spaan, W. J. M. (1987). cDNA cloning and sequence analysis of the gene encoding the peplomer protein of feline infectious peritonitis virus. *J. Gen. Virol.* **68,** 2639–2646.

De Groot, R. J., Ter Haar, R. J., Horzinek, M. C., and Van der Zeijst, B. A. M. (1987). Intracellular RNAs of feline infectious peritonitis coronavirus strain 79-1146. *J. Gen. Virol.* **68,** 995–1002.

Deregt, D., Sabara, M., and Babiuk, L. A. (1987). Structural proteins of bovine coronavirus and their intracellular processing. *J. Gen. Virol.* **68,** 2863–2877.

Dever, T. E., Glynias, M. J., and Merrick, W. C. (1987). GTP-binding domain: Three consensus sequence elements with distinct spacing. *Proc. Natl. Acad. Sci. USA* **84,** 1814–1818.

Doolittle, R. F. (1981). Similar amino acid sequences: Chance or common ancestry? *Science* **214,** 149–159.

Feinberg, A. P., and Vogelstein, B. (1983). DNA probes by random priming with Klenow synthesis. *Anal. Biochem.* **132,** 6–13.

Fry, D. C., Kuby, S. A., and Mildvan, A. S. (1986). ATP-binding site of adenylate kinase: Mechanistic implications of its homology with *ras*-encoded p21, F_1-ATPase, and other nucleotide-binding proteins. *Proc. Natl. Acad. Sci. USA* **83,** 907–911.

Golbach, R. (1987). Genome similarities between plant and animal RNA viruses. *Microbiol. Sci.* **4,** 197–202.

Hogue, B. G., King, B., and Brian, D. A. (1984). Antigenic relationships among proteins of bovine coronavirus, human respiratory coronavirus OC43, and mouse hepatitis coronavirus A59. *J. Virol.* **51,** 384–388.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and Vande Pol, S. (1982). Rapid evolution of RNA genome. *Science* **215,** 1577–1585.

Jacobs, L., De Groot, R., Van der Zeijst, B. A. M., Horzinek, M. C., and Spaan, W. (1987). The nucleotide sequence of the peplomer gene of porcine transmissible gastroenteritis virus (TGEV): Comparison with the sequence of the peplomer protein of feline infectious peritonitis virus (FIPV). *Virus Res.* **8,** 363–371.

Katagiri, S., Ohizumi, A., and Homma, M. (1983). An outbreak of influenza C in a children's home. *J. Infect. Dis.* **148,** 51–56.

Keck, J. G., Hogue, B. G., Brian, D. A., and Lai, M. M. C. (1987). Temporal regulation of RNA synthesis of bovine coronavirus. *In* "Coronaviruses" (M. M. C. Lai, and S. A. Stohlman, Eds.), Vol. **218,** pp. 157–158. Plenum, New York/London.

King, A. M. Q., Ortlepp, S. A., Newman, J. W. I., and McCahon, D. (1987). Genetic recombination in RNA viruses. *In* "The Molecular Biology of the Positive Strand RNA Viruses" (D. J. Rowlands, M. A. Mayo, and B. W. J. Mahy, Eds.), pp. 129–152. Academic Press, New York/London.

King, B., Potts, B. J., and Brian, D. A. (1985). Bovine coronavirus hemagglutinin protein. *Virus Res.* **2,** 53–59.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44,** 283–292.

Lai, M. M. C., Baric, R. S., Brayton, P. R., and Stohlman, S. A. (1984). Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc. Natl. Acad. Sci. USA* **81,** 3626–3630.

Lai, M. M. C., Brayton, P. R., Armen, R. C., Patton, C. D., Pugh, C., and Stohlman, S. A. (1981). Mouse hepatitis virus A59: mRNA

structure and genetic localization of the differences from hepatotropic strain MHV-3. *J. Virol.* **39,** 823–834.

LAPPS, W., and BRIAN, D. A. (1985). Oligonucleotide fingerprints of antigenically related bovine coronavirus and human coronavirus OC43. *Arch. Virol.* **86,** 101–108.

LAPPS, W., HOGUE, B. G., and BRIAN, D. A. (1987). Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology* **157,** 47–57.

LEIBOWITZ, J. L., WEISS, S. R., PAAVOLA, E., and BOND, C. W. (1982). Cell-free translation of murine coronavirus RNA. *J. Virol.* **43,** 905–913.

LEIBOWITZ, J. L., WILHELMSEN, K. C., and BOND, C. W. (1981). The virus-specific intracellular RNA species of two murine coronaviruses: MHV-A59 and MHV-JHM. *Virology* **114,** 39–51.

LIPMAN, D. J., and PEARSON, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* **227,** 1435–1441.

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

MAKINO, S., KECK, J. G., STOHLMAN, S. A., and LAI, M. C. (1986). High-frequency RNA recombination of murine coronaviruses. *J. Virol.* **57,** 729–737.

MCINTOSH, K., KAPIKIAN, A. Z., HARDISON, K. A., HARTLEY, J. W., and CHANOCK, J. W. (1969). Antigenic relationship among the coronaviruses of man and between human and animal coronaviruses. *J. Immunol.* **102,** 1109–1118.

MEINKOTH, J., and WAHL, G. (1984). Hybridization of nucleic acids immobilized on solid supports. *Anal. Biochem.* **138,** 267–284.

MESSING, J. (1983). New M13 vectors for cloning. *In* "Methods in Enzymology" (R. Wu, L. Grossman, and K. Moldave, Eds.), Vol. 101, pp. 20–79. Academic Press, New York.

MORCH, M-D., VALLE, R. P. C., and HAENNI, A-L. (1987). Regulation of translation of viral mRNAs. *In* "The Molecular Basis of Viral Reproduction" (R. P. Bercoff, Ed.), pp. 113–162. Plenum, New York/London.

NAKADA, S., CREAGER, R. S., KRYSTAL, M., AARONSON, R. P., and PALESE, P. (1984). Influenza C virus hemagglutinin: comparison with influenza A and B virus hemagglutinins. *J. Virol.* **50,** 118–124.

NIESTERS, H. G. M., LENSTRA, J. A., SPAAN, W. J. M., ZIJDERVELD, A. J., BLEUMINK-PLUYM, N. M. C., HONG, F., VAN SCHARRENBURG, G. J. M., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M. (1986). The peplomer protein sequence of the M41 strain of coronavirus IBV and its comparison with Beaudette strains. *Virus Res.* **5,** 253–263.

PFEIFER, J. B., and COMPANS, R. W. (1984). Structure of the influenza C glycoprotein gene as determined from cloned DNA. *Virus Res.* **1,** 281–296.

RASSCHAERT, D., and LAUDE, H. (1987). The predicted primary structure of the peplomer protein E2 of the porcine coronavirus transmissible gastroenteritis virus. *J. Gen. Virol.* **68,** 1883–1890.

ROBERTS, C., and WILSON, G. L. (1985). DNA probes by random priming with Klenow synthesis. *Focus* **7-3,** 16.

ROTTIER, P. J. M., SPAAN, W. J. M., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M. (1981). Translation of three mouse hepatitis virus strain A59 subgenomic RNAs in *Xenopus laevis* oocytes. *J. Virol.* **38,** 20–26.

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

SCHEPART, B. S., TAKAHASHI, H., COZAD, K. M., MURRAY, R., OZATO, K., APPELLA, E., and FRELINGER, J. A. (1986). The nucleotide sequence and comparative analysis of the H-2D^{p} class I H-2 gene. *J. Immunol.* **136,** 3489–3495.

SIDDELL, S. (1982). Coronavirus JHM: Tryptic peptide fingerprinting of virion proteins and intracellular polypeptides. *J. Gen. Virol.* **62,** 259–269.

SIDDELL, S. (1983). Coronavirus JHM: Coding assignments of subgenomic mRNAs. *J. Gen. Virol.* **64,** 113–125.

SIDDELL, S. (1987). The organization and expression of coronavirus genomes. *In* "The Molecular Biology of the Positive Strand RNA Viruses" (D. J. Rowlands, M. A. Mayo, and B. W. J. Mahy, Eds.), pp. 117–128. Academic Press, New York/London.

SIDDELL, S., WEGE, H., BARTHEL, A., and TER MEULEN, V. (1981). Coronavirus JHM: Intracellular protein synthesis. *J. Gen. Virol.* **53,** 145–155.

SIDDELL, S., WEGE, H., and TER MEULEN, V. (1982). The structure and replication of coronaviruses. *Curr. Top. Microbiol. Immunol.* **99,** 131–163.

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

SPAAN, W., DELIUS, H., SKINNER, M., ARMSTRONG, J., ROTTIER, P., SMEEKENS, S., VAN DER ZEIJST, B. A. M., and SIDDELL, S. G. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO J.* **2,** 1839–1844.

SPAAN, W. J. M., ROTTIER, P. J. M., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M. (1981). Isolation and identification of virus-specific mRNAs in cells infected with mouse hepatitis virus (MHV-A59). *Virology* **108,** 424–434.

SPAAN, W. J. M., ROTTIER, P. J. M., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M. (1982). Sequence relationships between the genome and the intracellular RNA species 1, 3, 6, and 7 of mouse hepatitis virus strain A59. *J. Virol.* **42,** 432–439.

STADEN, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res.* **10,** 2951–2961.

STADEN, R. (1986). The current status and portability of our sequence handling software. *Nucleic Acids Res.* **14,** 217–233.

STRAUSS, E. G., RICE, C. M., and STRAUSS, J. H. (1983). Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. *Proc. Natl. Acad. Sci. USA* **80,** 5271–5275.

TAGUCHI, F., MASSA, P. T., and TER MEULEN, V. (1986). Characterization of a variant virus isolated from neural cell culture after infection of mouse coronavirus JHMV. *Virology* **155,** 267–270.

TAGUCHI, F., SIDDELL, S. G., WEGE, H., and TER MEULEN, V. (1985). Characterization of a variant virus selected in rat brains after infection by coronavirus mouse hepatitis virus JHM. *J. Virol.* **54,** 429–435.

VLASAK, R., KRYSTAL, M., NACHT, M., and PALESE, P. (1987). The influenza C virus glycoprotein (HE) exhibits receptor-binding (hemagglutinin) and receptor-destroying (esterase) activities. *Virology* **160,** 419–425.

VLASAK, R., LUYTJES, W., SPAAN, W., and PALESE, P. (1988). Human and bovine coronaviruses recognize sialic acid containing receptors similar to those of influenza C viruses. *Proc. Natl. Acad. Sci. USA* **85,** in press.

WEGE, H., SIDDELL, S., and TER MEULEN, V. (1982). The biology and pathogenesis of coronaviruses. *Curr. Top. Microbiol. Immunol.* **99,** 165–200.

WEISS, S. R. (1983). Coronaviruses SD and SK share extensive nucleotide homology with murine coronavirus MHV-A59, more than that shared between human and murine coronaviruses. *Virology* **126,** 669–677.

WEISS, S. R., and LEIBOWITZ, J. L. (1983). Characterization of murine coronavirus RNA by hybridization with virus-specific cDNA probes. *J. Gen. Virol.* **64,** 127–133.