VIROLOGY 156, 321–330 (1987)

The 5′-End Sequence of the Murine Coronavirus Genome: Implications for Multiple Fusion Sites in Leader-Primed Transcription

CHIEN-KOU SHIEH, LISA H. SOE, SHINJI MAKINO, MING-FU CHANG, STEPHEN A. STOHLMAN, AND MICHAEL M. C. LAI[1]

Departments of Microbiology and Neurology, University of Southern California School of Medicine, Los Angeles, California 90033

Received July 28, 1986; accepted October 31, 1986

The coronavirus leader-primed transcription model proposes that free leader RNA species derived from the 5′-end of the genomic RNA are utilized as a primer for the transcription of subgenomic mRNAs. To elucidate the precise mechanism of leader-priming, we cloned and sequenced the 5′-end of the mouse hepatitis virus genomic RNA. The 5′-terminal sequences are identical to the leader sequences present at the 5′-end of the subgenomic mRNAs. Two possible hairpin loop structures and an AU-rich region around the 3′-end of the leader sequence may provide the termination site for leader RNA synthesis. The comparison of 5′-end genomic sequences and the intergenic start sites for mRNA transcription revealed that there are homologous regions of 7–18 nucleotides at the putative leader/body junction sites. Some intergenic regions contain a mismatching nucleotide within this homologous region. We propose that free leader RNA binds to the intergenic region due to this homology and is cleaved at the mismatching nucleotide before serving as a primer. Thus, the free leader RNA species may be longer than the leader sequences in the subgenomic mRNAs and different mRNAs may have different leader/body junction sites. © 1987 Academic Press, Inc.

INTRODUCTION

Murine coronaviruses, or mouse hepatitis viruses (MHV), are a group of enveloped viruses with a nonsegmented positive-stranded RNA genome of mol wt 5.4×10^6 (Lai and Stohlman, 1978), which replicates exclusively in the cytoplasm of several established mouse cell lines (Brayton *et al.*, 1981; Wilhelmsen *et al.*, 1981). Upon entry into cells, the virion RNA encodes an "early" RNA polymerase which transcribes the RNA genome into a full-length negative-stranded RNA (Brayton *et al.*, 1982; Lai *et al.*, 1982). The latter is then transcribed by separate "late" RNA polymerases into a positive-sense genomic RNA and six subgenomic mRNA species (Brayton *et al.*, 1981; Lai *et al.*, 1981). These mRNAs have a nested-set structure, containing sequences from the 3′-end of the genomic RNA that extend for various distances toward the 5′-end (Lai *et al.*, 1982). Each mRNA and genomic RNA also contain an identical leader sequence of approximately 72 nucleotides at the 5′-end (Lai *et al.*, 1984; Spaan *et al.*, 1983). The uv transcriptional mapping suggested that MHV subgenomic mRNAs are not derived by splicing of larger precursor RNAs (Jacobs *et al.*, 1981). Since MHV replication does not involve a nuclear phase, the joining of the leader sequences to the mRNA body sequences does not utilize conventional eukaryotic RNA splicing mechanisms. Studies of replicative-intermediate RNA and double-stranded replicative form RNA further suggest that leader RNA is joined to the mRNA during transcription, but not post-transcriptionally, and that the leader RNA joining mechanism most likely involves a free leader RNA species participating in mRNA transcription (Baric *et al.*, 1983). This model has recently been supported by the detection of free leader RNAs in the cytoplasm of MHV-infected cells and also by the isolation of a temperature-sensitive mutant which synthesizes only leader RNA but not mRNAs at nonpermissive temperature (Baric *et al.*, 1985). Furthermore, during mixed infections the leader sequence of different MHVs can be freely reassorted among the mRNAs of the coinfecting viruses, suggesting that the leader sequence functions as a separate transcriptional unit (Makino *et al.*, 1986). Finally, sequence analysis of the leader RNA and the intergenic regions of various mRNAs has revealed sequence homology of 6–10 nucleotides between the 3′-end of the leader RNA and the initiation sites of various subgenomic mRNAs, thus providing a possible mechanism for leader RNA to bind to the initiation sites of various subgenomic mRNAs (Budzilowicz *et al.*, 1985). These studies support the model of leader-primed transcription, in which free leader is synthesized initially from the 5′-end of genomic RNA, dissociates from the negative-strand template, and rebinds to the template at the initiation sites of the various mRNAs, thereby serving as a primer for transcription (Baric *et al.*, 1983, 1985).

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

0042-6822/87 $3.00

To understand the precise mechanism of leader RNA synthesis and leader priming, we cloned the 5′-end of MHV genomic RNA by using synthetic oligodeoxyribonucleotides specific for sequences close to the 5′-end of the genome as primers for cDNA cloning. Clones containing the leader sequences were identified and the 5′ leader/body junction site was characterized by sequence analysis. These sequences reveal several interesting features which provide significant insights into the mechanism of MHV RNA transcription.

MATERIALS AND METHODS

Viruses and cells

The plaque-cloned JHM strain of MHV (Makino *et al.*, 1984a, 1985) was used throughout. Viruses were propagated on DBT cells at low multiplicities of infection (m.o.i.) as previously described (Makino *et al.*, 1984a).

Preparation of virion RNA

The JHM-infected cell culture fluid was harvested at 14 hr postinfection and clarified at 15,000 *g* for 30 min at 4°. Pooled viruses were precipitated with 50% ammonium sulfate according to published procedures (Makino *et al.*, 1984a). The pellet was resuspended in NTE buffer (0.1 *M* NaCl, 0.01 *M* Tris–hydrochloride (pH 7.2), and 1 m*M* EDTA) and was placed on a discontinuous sucrose gradient consisting of 60, 50, 30, and 20% (w/w) sucrose in NTE buffer and centrifuged at 27,000 rpm for 4 hr at 4° in a Beckman SW 28.1 rotor. A virus band at the interface between 50 and 30% sucrose was collected and diluted threefold with NTE buffer. The diluted virus suspension was centrifuged on a linear 20 to 60% sucrose gradient at 27,000 rpm in an SW 28.1 rotor for 18 hr at 4°. Virus band was collected, diluted with NTE buffer, and then pelleted at 40,000 rpm in an SW41 rotor for 1.5 hr at 4°. Pellets of purified virus were resuspended in NTE buffer, and genomic RNA was extracted with phenol/chloroform as described (Makino *et al.*, 1984a).

RNA sequencing

Genomic RNA was digested with RNase T1 and separated by two-dimensional polyacrylamide gel electrophoresis as described (Lai *et al.*, 1981). An RNase T1-resistant oligonucleotide (oligonucleotide 3) which had previously been mapped to the 5′-end of the genome (Makino *et al.*, 1984b) was eluted from the gel in 0.5 *M* NaCl. The oligonucleotide was treated with bacterial alkaline phosphatase and 5′-end-labeled with [γ-^{32}P]ATP according to the published protocol (Pedersen and Haseltine, 1980). The 5′-end-labeled oligonucleotide was then partially cleaved at each of the four bases using ribonucleases T1, U2, Phy M, and *Bacillus cereus* (Pharmacia) under the conditions described by D'alessio (1982). The products were separated on a 20% polyacrylamide/8.3 *M* urea sequencing gel (Maxam and Gilbert, 1977).

Primer extension

Synthetic oligodeoxyribonucleotides were 5′-end-labeled with [γ-^{32}P]ATP by polynucleotide kinase (Pedersen and Haseltine, 1980). Thirty nanograms of 5′-end-labeled oligonucleotide was mixed with 3 μg of genomic RNA in 10 μl of a solution containing 5 m*M* sodium phosphate (pH 7.0) and 5 m*M* EDTA. The RNA and oligonucleotide mixture was heated at 90° for 5 min, and then 1 μl of 1 *M* KCl preheated to 90° was added. The reaction mixture was allowed to cool to room temperature over a period of 20 min, and 25 units of RNasin (Promega Biotec) was added. The reaction was transferred to an ice water bath. The extension reaction was subsequently carried out as described previously (Lai *et al.*, 1984). Reaction products were analyzed by electrophoresis on 1.4% agarose gels in the presence of 30 m*M* NaOH and 1 m*M* EDTA.

cDNA cloning of the 5′-end of the MHV genome

cDNA cloning followed the general method of Gubler and Hoffman (1983). This procedure was modified by the use of a synthetic oligodeoxyribonucleotide (5′-ATAATGGGTTTTGTATAATA-3′), which is complementary to a T1-oligonucleotide close to the 5′-end of the genome (Makino *et al.*, 1984b) (see Results), to prime reverse transcription.

To specifically clone 5′-end of the genomic RNA, we employed a cloning strategy different from that described above: First-strand cDNA was prepared according to Maniatis *et al.* (1982), using another synthetic oligomer (see Results) as a primer. Reverse transcription was carried out at 42° for 1 hr, and the RNA–DNA hybrid was separated by boiling for 90 sec and quick chilled on ice. Two units of RNase T1 was added to the reaction mixture and incubated at 37° for 30 min to remove the RNA template. Single-stranded cDNA was extracted with phenol/chloroform and precipitated with ethanol. The second-strand DNA was synthesized with 5 units of Klenow fragment of DNA polymerase I, using a synthetic oligomer (5′-AGAGTGATTGGCGT-3′), which corresponds to the 5′-end of the leader sequences (Lai *et al.*, 1984), as a primer. The reaction was carried out in 50 μl of 10 m*M* $MgCl_2$, 70 m*M* KCl, 50 m*M* Tris–hydrochloride (pH 7.5), 0.2 m*M* DTT, and 1 m*M* dNTPs at 37° for 1 hr. Double-stranded DNA was dC-tailed in a 12-μl reaction mixture containing 10 units of terminal deoxynucleotide transferase, 200 m*M*

potassium cacodylate, 0.5 m*M* $CoCl_2$, 25 m*M* Tris–hydrochloride (pH 6.9), 2 m*M* DTT, 250 μg/ml BSA, and 50 μ*M* dCTP at 37° for 5 min. The dC-tailed double-stranded DNA was annealed to 200 ng of dG-tailed *Pst*I-cut PBR322 plasmid in 20 μl of a buffer containing 10 m*M* Tris–hydrochloride (pH 7.4), 100 m*M* NaCl, and 0.25 m*M* EDTA. The DNA mixture was heated at 68° for 5 min and then cooled slowly overnight for annealing. The annealed molecules were used to transform *Escherichia coli* MC1061 as described (Dagert and Ehrlich, 1979).

Colony hybridization

Replica plating of the colonies was performed by using transfer pads (FMC Corp.). Colonies were incubated at 37° for 12 hr and transferred to Colony/Plaque Screen disks (New England Nuclear). Bacteria lysis and DNA fixation were carried out according to the procedures previously described (Grunstein and Hogness, 1975). The disks were prehybridized in a solution containing 0.2% polyvinyl–pyrrolidone (M_r 40,000), 0.2% Ficoll (M_r 400,000), 0.2% bovine serum albumin, 0.05 *M* Tris–hydrochloride (pH 7.5), 1% SDS, 1 *M* NaCl, 10% dextran sulfate, and 100 μg/ml denatured salmon sperm DNA at 65° for 6 hr. 5′-end-^{32}P-labeled leader-specific 72-mer derived from the leader sequences of mRNA 7 (Lai *et al.*, 1984) was added and incubated at 65° for 20 hr. After hybridization, disks were washed twice in 2× SSC (0.3 *M* NaCl and 30 m*M* sodium citrate) at room temperature, twice in 2× SSC containing 1% SDS at 65° for 30 min, and twice in 0.1× SSC at room temperature. The disks were air-dried and exposed to X-ray film at −70°.

Northern hybridization

Intracellular RNA from virus-infected cells were denatured by glyoxal treatment and was separated by electrophoresis on 1% agarose gels containing 10 m*M* sodium phosphate (pH 7.0) (McMaster and Carmichael, 1977). After electrophoresis, the gel was treated with 50 m*M* NaOH for 30 min and neutralized with 100 m*M* Tris–hydrochloride (pH 7.2) for 30 min. The gel was then soaked in 20× SSC for 30 min. RNA transfer to Biodyne nylon filters (ICN Radiochemicals) and subsequent hybridization were done according to the method of Thomas (1980).

DNA sequencing

Sequencing was carried out by Sanger's dideoxyribonucleotide chain termination method (Sanger *et al.*, 1977). Sequence analysis and predicted RNA secondary structures were obtained using the Intelligenetics Sequencing Program.

RESULTS

RNA sequencing of RNase T1-resistant oligonucleotides

The leader-primed transcription model for coronavirus subgenomic mRNA synthesis predicts that the mRNAs of MHV are derived by the fusion of the leader sequences from the 5′-end of the viral genome to the body sequences of the mRNAs. The sequence of the 5′-end of the RNA genome is, therefore, crucial for understanding the mechanism of leader RNA synthesis and leader priming. To clone the 5′-end region of the genomic RNA, we utilized the RNase T1-resistant oligonucleotide 3 of JHM, which had previously been mapped to gene A (Makino *et al.*, 1984b). This oligonucleotide was eluted from a two-dimensional fingerprinting gel, was ^{32}P-labeled at 5′-end, and was partially digested with RNases T1, U2, Phy M, and *B. cereus* in separate reactions. The digested products were then separated by electrophoresis on a 20% polyacrylamide gel (Fig. 1). The sequence obtained by this procedure was confirmed by wandering spots analysis (data not shown) (Nomoto and Imura, 1979). The complementary sequence of this oligonucleotide (5′-ATAAT-GGGTTTTGTA-3′) was synthesized (designated as oligo 13) and used as the primer for cDNA synthesis.

Localization of the synthetic oligomer on the viral genome

To determine the binding site of the oligo 13 on the viral genomic template, we carried out primer extension studies. Briefly, first-strand synthesis was carried out in the presence of [α-^{32}P]dATP, and the products were resolved by electrophoresis on a 1.4% alkaline agarose gel. This reaction generated a product of 6.2 kb in size, (Fig. 2), suggesting that the oligomer is localized roughly 6.2 kb from the 5′-end of the genomic RNA. This analysis enabled us to map the positions of cDNA clones subsequently generated from this primer to gene A, which encompasses more than 8 kb of the 5′-end of the MHV genome (Lai *et al.*, 1981). The double-stranded DNA made from this primer in an independent reaction utilized a 3′-terminal hairpin loop of the first strand and was roughly twice the size of the single-stranded DNA (Fig. 2). This result confirms that this oligomer is mapped approximately 6 kb from the 5′-end of the genome.

Molecular cloning of the 5′-end of the MHV genome

cDNA clones were generated according to the general method of Gubler and Hoffman (1983), using the synthetic oligodeoxyribonucleotide (oligo 13) to prime

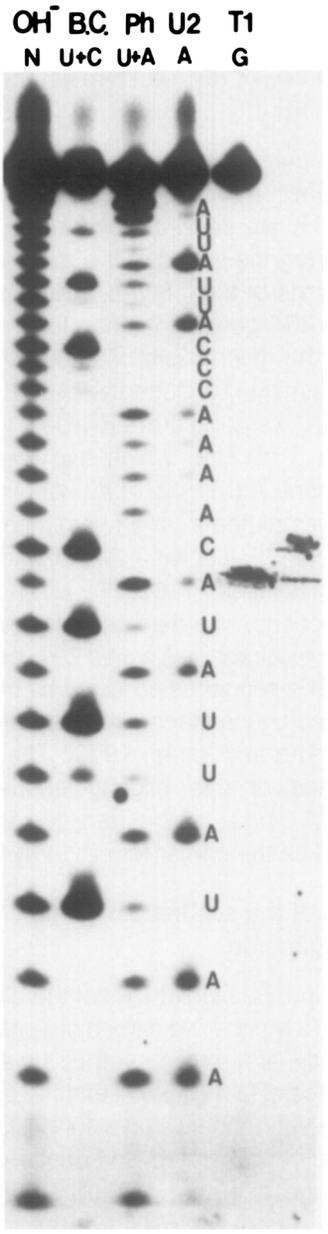

FIG. 1. RNA sequence analysis of RNase T1-resistant oligonucleotide 3 isolated from two-dimensional polyacrylamide fingerprinting gels. The oligonucleotide, which was previously mapped to gene A (Makino *et al.*, 1984b), was 5′-end labeled and partially cleaved with RNases T1, U2, Phy M (ph), and *B cereus* (B.C.). Alkali digestion products in the far left lane (OH^-) denote nucleotide positions in the sequencing ladder. The products were separated on a 20% polyacrylamide/8.3 *M* urea sequencing gel.

reverse transcription. Using this approach, we obtained cDNA clones ranging in size from 0.5 to 4.5 kb. These clones (group II) were characterized by restriction mapping (Fig. 3) and hybridization with intracellular RNA from MHV-infected cells. None of the clones hybridized to a leader-specific probe (data not shown), suggesting that these clones did not extend completely to the 5′-end of the viral genome.

To obtain a primer to synthesize cDNA clones spanning the 5′-end of the genome RNA, we subcloned an internal *Pst*I fragment from clone B27 (see Fig. 3) into M13 and performed dideoxy sequence analysis. The sequence of the first 300 nucleotides of this fragment revealed the presence of a single open reading frame (data not shown), enabling us to determine the positive-sense sequence. We utilized these sequences to generate another oligodeoxyribonucleotide (5′-TGTCATCTCGCACTCCAA-3′) (designated as oligo 14) complementary to this open reading frame and used it as a primer for first-strand cDNA synthesis. In this way, we were able to "walk" toward the 5′-end of the genomic RNA.

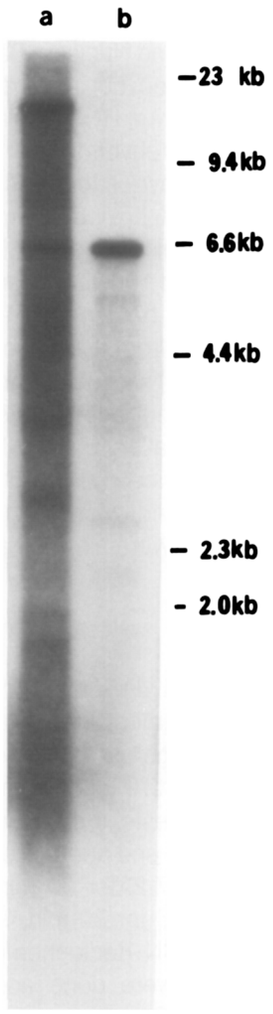

FIG. 2. Electrophoretic analysis of MHV genomic RNA primer extension products. Primer extensions were carried out using oligo 13 as a primer for first-strand cDNA synthesis. First-strand synthesis (lane b) was done using reverse transcriptase in the presence of [α-^{32}P]dATP. Second-strand synthesis (lane a) was carried out in the presence of [α-^{32}P]dATP on unlabeled first-strand cDNA products. The products were analyzed on a 1.4% alkaline agarose gel. The molecular weights of *Hin*dIII digestion products of wild-type λ DNA are indicated in the far right lane.

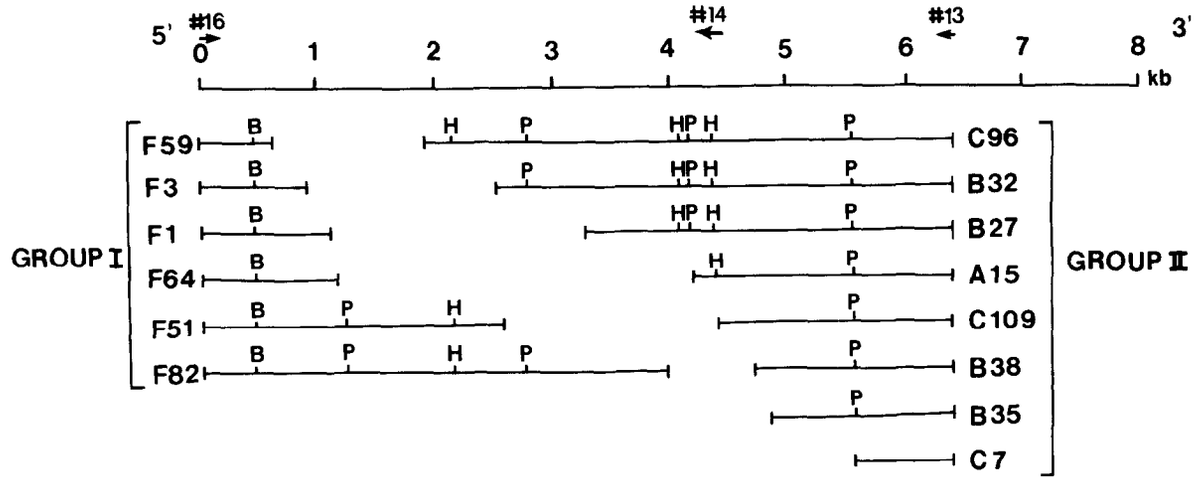

FIG. 3. Restriction endonuclease map of the 5'-end cDNA clones and location of the synthetic oligodeoxyribonucleotides used as primers for cDNA synthesis. Group II clones were derived using oligo 13 for first-strand cDNA synthesis and the method of Gubler and Hoffman (1983) for second-strand synthesis. Group I clones were obtained using oligo 14 and 16 as primers for first- and second-strand synthesis, respectively. P, *Pst*I; H, *Hin*dIII; B, *Bam*HI.

To ensure cloning of the 5'-terminus of the viral genome, we used a modified procedure for second-strand synthesis. A synthetic oligodeoxyribonucleotide corresponding to the 5'-end of the leader sequence was derived from the sequence of A59 mRNA 7 (Lai *et al.*, 1984). This oligodeoxyribonucleotide (oligo 16) (5'-AGAGTGATTGGCGT-3') was used to prime second-strand DNA synthesis employing the Klenow fragment of DNA polymerase I. cDNA clones obtained by this strategy were screened by colony hybridization using 5'-end-labeled leader-specific 72 mer. Several positive colonies were identified and characterized further (Fig. 3). These clones (group I) range in size from 0.5 to 3.8 kb. Restriction analysis and cross-hybridization of cDNA clones confirmed that the largest clone (F82) overlaps with the cDNA clones (group II) obtained using oligo 13 primer.

Characterization of the cDNA clones containing leader sequences

To eliminate the possibility that the cDNA clones containing leader sequences might have been derived from subgenomic mRNAs or leader-homologous sequences present at intergenic sites, we characterized the clones by Northern blot analysis. We utilized a unique *Bam*HI site at the 5'-end of the genome (Fig. 3) to separate the leader-containing clones into two fragments. The two fragments were used to generate ^{32}P-labeled probes by nick-translation and hybridized with intracellular RNAs from MHV-infected cells (Fig. 4). The unique 0.4-kb *Pst*I–*Bam*HI fragment of clone F1 hybridized to all of the mRNAs (Fig. 4, lane A), suggesting that this fragment contains the leader sequence since leader-specific sequences are present in all subgenomic mRNAs (Lai *et al.*, 1984). This probe also detected an extra RNA species, RNA b, which is consistently observed in this strain of MHV (Makino *et al.*, 1984b, 1985). The remaining 0.6-kb fragment hybridized exclusively to mRNA 1 (Fig. 4, lane B). This latter probe did not hybridize to subgenomic mRNAs nor to uninfected cellular RNA. As deduced from the 3'-nested set structure of the MHV mRNAs (Lai *et al.*, 1984), this hybridization pattern locates the fragment to gene A. These results confirm that all the leader-containing cDNA clones were indeed derived from the 5'-end of the genomic RNA.

Sequencing of the cDNA clones containing the 5'-end of the MHV genome

To determine the sequence of these 5'-end cDNA clones, we subcloned the 0.4-kb *Pst*I–*Bam*HI fragment into an M13 vector for dideoxy chain termination sequencing. Sequence data derived from four of the leader-containing clones are shown in Fig. 5. The leader sequences of the genomic RNA are identical to the leader sequences present at the 5'-end of the subgenomic mRNAs (Spaan *et al.*, 1983; Lai *et al.*, 1984). Thus, the leader RNA in the subgenomic mRNAs are most likely derived from the 5'-end of the genome. However, we cannot rule out the possibility of sequence variation in the 5'-ultimate dodecameric region of the genomic RNA since this sequence in our clones was derived from the synthetic primer used for second-strand DNA synthesis. We consider this possibility unlikely since the leader sequences at the 5'-ends of the genome and mRNAs have the same size (Lai *et al.*,

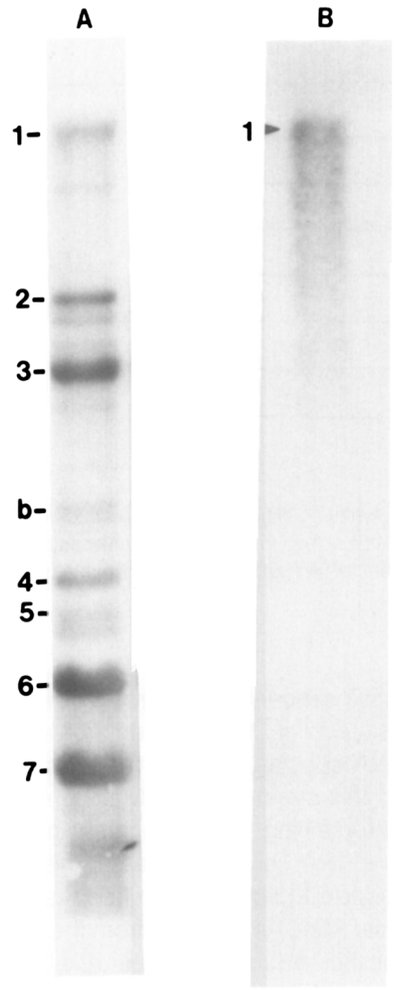

FIG. 4. Northern blot analysis of MHV-specific intracellular RNA hybridized to 5′-end genomic cDNA clones. MHV-JHM intracellular RNA was denatured by glyoxal treatment, separated on a 1% agarose gel, and transferred to Biodyne nylon filters as described under Materials and Methods. Probes used were: Lane A, nick-translated probe derived from the 5′ 0.4-kb *Pst*I–*Bam*HI fragment of cDNA clone F1; lane B, probe derived from the 3′ 0.6-kb *Bam*HI–*Pst*I fragment of cDNA clone F1. The positions and designations of the MHV-specific RNAs are indicated by the numbers on the left side of the figure.

1984), and in a different coronavirus, avian infectious bronchitis virus (IBV), the leader sequence of the genome has also been shown to be identical to that of the subgenomic mRNAs (M. Boursnell, personal communication). The leader sequence at the 5′-end of the genomic RNA also contains several additional structural features which are highlighted in Fig. 5. We note the presence of a mirror-image direct repeat (AAAUCUAAU) at the 3′-end of the leader sequence from position 57 to 75 (Fig. 5A). A similar dyad symmetry has previously been described for leader sequences of the subgenomic mRNAs of the A59 strain of MHV (Lai *et al.*, 1984). A stretch of inverted repeats (UUUA) follows this dyad symmetry.

The sequence also reveals the presence of two possible hairpin loop structures formed in the regions of nucleotides 52–77 ($\Delta G = -1.6$ kcal/mol) and 80–140 ($\Delta G = -18.7$ kcal/mol), which are separated by an AU-rich region (UUUAUAAA) at position 78–84 (Fig. 5B). In addition, we note that the leader sequence at the 5′-end of the genomic RNA contains three repeats (UCUAA). These repeats are also present on the leader sequences of subgenomic mRNAs (Spaan *et al.*, 1983). One of the cDNA clones (F82) contains four of these repeats (data not shown), suggesting the presence of heterogeneity at the 5′-end of the MHV genome. The significance of these structural features and the presence of homology between the 5′-end of the genome and intergenic start sites for transcription are elaborated under Discussion in relation to the leader-primed transcription model.

DISCUSSION

Several lines of evidence support the leader-primed model of MHV transcription. In this model, leader RNA functions as a free and separate transcriptional unit, which binds to intergenic start sites on the negative-stranded genome RNA and is then utilized as a primer for subgenomic mRNA transcription (Baric *et al.*, 1983). In this paper, we report the sequence at the 5′-end of the genomic RNA of the JHM strain of MHV. These sequences reveal several important features which are helpful in understanding the probable size of the free leader RNA and suggest the mechanism by which the leader is fused to the body sequences of the subgenomic mRNAs. One of the interesting features is the presence of two possible hairpin loop structures separated by an AU-rich region (Fig. 5B). Hairpin loop structures have been shown to be transcriptional pausing sites in QB phage RNA (Mills *et al.*, 1978). AT-rich regions, such as TTTTTATA or derivatives thereof, have also been shown to be the transcriptional termination signals in certain eukaryotic cells (Zaret and Sherman, 1982; Henikoff *et al.*, 1983). We postulate that the hairpin loop structures and the intervening AU-rich region at the 3′-end of the MHV leader sequence may serve as the termination site of leader RNA synthesis. The instability of AU hybrids (Martin and Tinoco, 1980) may further facilitate the dissociation of the leader RNA from the template. Indeed, several free leader-containing RNA species have been detected in MHV-infected cells (Baric *et al.*, 1987), which could have resulted from termination within this region.

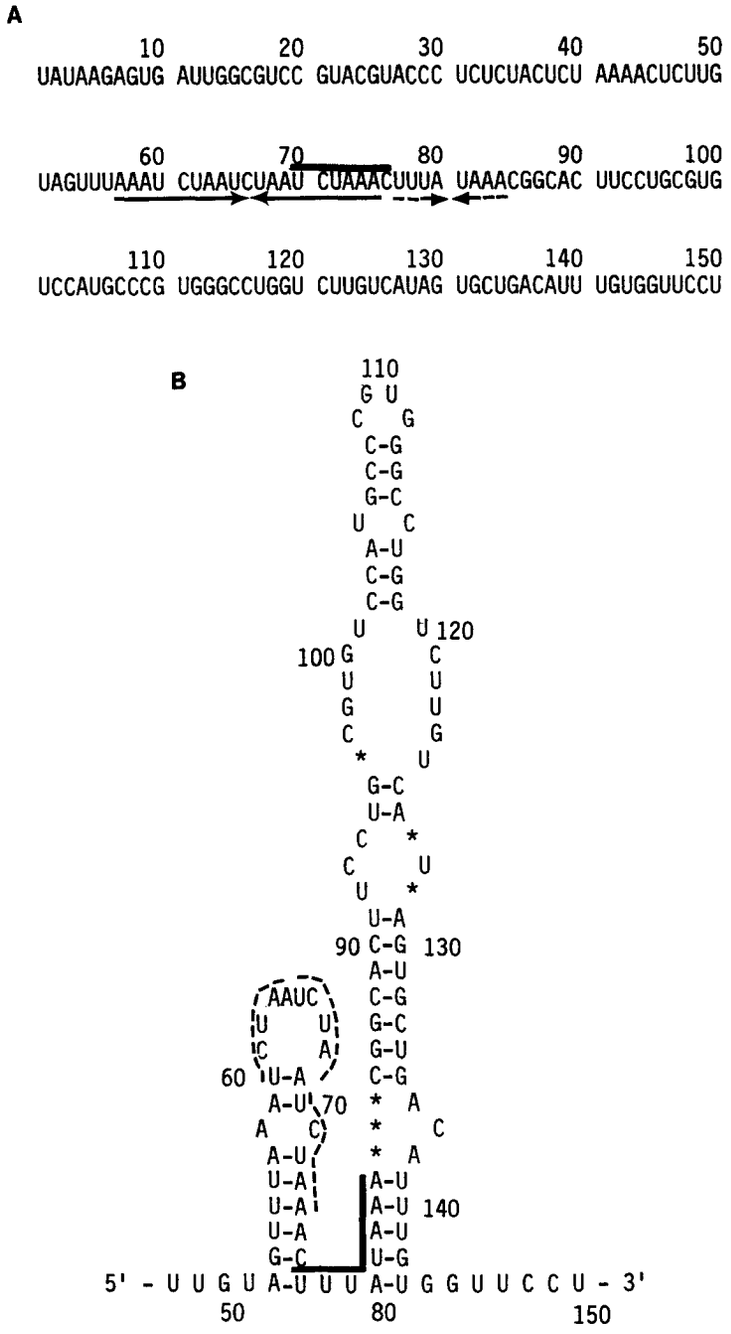

FIG. 5. Sequence of the 5'-end of the genomic RNA of MHV. (A) The primary nucleotide sequence of clones F1, F64, F51, and F59. A 9-base sequence present as a mirror-image direct repeat is denoted by the solid arrows. A 4-base inverted repeat is delineated by dashed arrows. The 7-base sequence highly conserved among the intergenic start sites for mRNA transcription is overscored by a bold solid line. (B) The possible leader termination site is illustrated as two hairpin loops at the 3' end of the leader sequence. An 8-base AU-rich sequence is overscored beginning at base 77. The UCUAA repeats are denoted by dashed lines.

Comparison of 5'-end sequences of the genomic RNA with the intergenic sequences of various mRNAs reveals a possible mechanism of leader-primed transcription and length of the leader RNA (Fig. 6). Previously, the length of the leader in the mRNAs was defined by the leader/body fusion site at first base mismatch between the leader and intergenic region (Spaan *et al.*, 1983; Budzilowicz *et al.*, 1985). It has also been noted that homology exists between the 3'-end of the leader sequence and the intergenic regions upstream of this leader/body fusion site (Budzilowicz *et al.*, 1985). Accordingly, the length of the JHM leader RNA was

Intergenic site		No. bases of homology	Ratio of RNA amount
3-4	.....*AGAA*AAUCUAAAC*AAUUUAUAGC*.....	9	1.69
4-5	.....*ACUAGU*UCUAAAC*CUCAUCUUAA*.....	7	0.86
5-6	...*AUGA*UAAUCUAAUCCAAAC*AUUAUG*.....	10 + 4	31.5
6-7	....*UGAG*AAUCUAAUCUAAACUUUA*AGGAUG*.....	18	100

Leader ....AUCUAAUCUAAUCUAAACUUUAUAAACG.....
60 70 80

FIG. 6. Comparison of the leader sequences at the 5′ end of the genomic RNA with intergenic sequences. The sequences at the intergenic sites are presented in plus-sense (genomic RNA) (Skinner *et al.*, 1985; Skinner and Siddell, 1983, 1985). The regions of homology with the 3′-end of the leader RNA are underscored by a thin line. The 7-base sequence UC$^{U}_{C}$AAAC underscored by the bold line is the sequence common to the initiation sites of all the mRNAs. The number of bases of homology between the leader RNA and intergenic start sites and the relative ratio of the mRNAs (Leibowitz *et al.*, 1981) are included for comparison.

estimated to be 72 nucleotides on the basis of a mismatch at position 72 of mRNA 6. Our sequence analysis of the 5′-end of the genome reveals that, in addition to the homology present upstream of the putative leader/body junction sites, there is homology downstream of these sites which extends for an additional five nucleotides (UAAAC). This 3′-homology would expand the potential base pairing between the leader RNA and the intergenic start sites and increase the overall binding stability of leader RNA to these intergenic start sites. The presence of homology 3′ to the leader/body fusion sites has also been noted in another coronavirus, IBV (Brown *et al.*, 1983). Hence, we postulate that the free leader RNA of JHM may be 77 nucleotides or longer to utilize the 3′-homology. The significance of the 3′-homologous sequences has previously been referred to by Skinner and Siddell (1985) who noted the common sequence UC$^{U}_{C}$AAAC at the intergenic start sites of several mRNAs. Significantly, a similar sequence CUAAAC is also found in the intergenic region between the genes M and N of porcine transmissible gastroenteritis virus (Kapke and Brian, 1986). The presence of these sequences at the 5′-end of the MHV genome suggests that these sequences are part of free leader RNA and important for transcriptional priming.

Based upon our sequence analysis of the 5′-end of the genome, we propose a model for leader-primed transcription (Fig. 7). In this model, a free leader RNA longer than 72 nucleotides binds to the intergenic start site of each gene. An endonuclease activity then cleaves the leader RNA at the position of a base mismatch. Transcription proceeds utilizing the nicked leader as the primer and the negative-strand RNA as the template for the synthesis of the body sequences. The leader/body junction sites may be different for each mRNA, depending on the exact site of the base mismatch. The proposed endonuclease activity allows the use of leader RNA species of variable lengths for transcription. Furthermore, the extent of homology between the leader and intergenic start sites of the subgenomic mRNAs may regulate the expression of different mRNA species. The sequence analysis presented in this paper reveals a strong correlation between the extent of homology to leader RNA at the intergenic start sites and the level of RNA expression (Fig. 6). However, we cannot rule out the possibility that other regulatory factors are also involved.

The significance of the UCUAA repeats at the 3′-end of the leader sequence is suggested by the structure of clone F82. In contrast to the other 5′-end clones which have three repeats, F82 contains an additional fourth repeat. It is conceivable that, when the free leader RNA rebinds to the original site of transcription, it might bind to an alternative repeat. As a result, genomic RNA with an extra repeat might be generated. This additional repeat may explain the mRNA 7 structure of a small-plaque mutant of JHM, which has a leader/body junction sequence approximately five nucleotides longer than that of the wild-type virus (Makino *et al.*, 1984b). A similar variant cDNA clone containing an additional copy of the UCUAA repeat in mRNA 6 has previously been demonstrated (Armstrong *et al.*, 1984). These mutants could have arisen from variable rebinding of the leader to the negative-strand RNA template, resulting in RNA containing different numbers

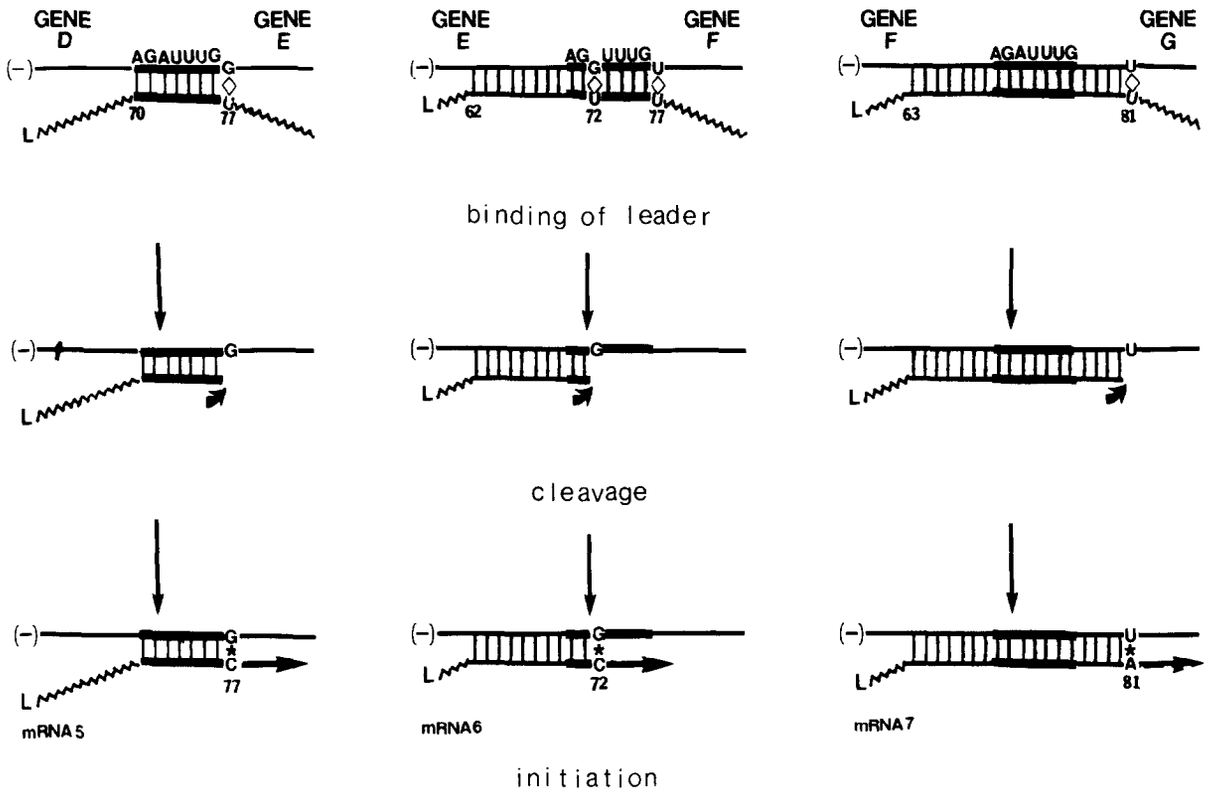

FIG. 7. The proposed model of leader RNA priming. The leader RNA derived from the 5'-end of the genomic RNA is represented by the lower line of each diagram denoted by the letter L. The upper line of each diagram represents the negative-strand RNA template at the various intergenic start sites for mRNAs 5 (left column), 6 (middle column), and 7 (right column). Vertical bars denote bases of complementarity and the open rhombuses delineate the first base of mismatch. The position of the proposed recognition sequence $AG^{A}_{G}UUUG$ is highlighted by the bold-lined region and referenced by the numbers below corresponding to the position from the 5'-end of the leader RNA. In this model, the leader RNA binds to the minus-strand RNA template by utilizing homologous sequences present at the intergenic start sites. The polymerase recognizes the first base mismatch and cleaves the free leader at this position. Transcription then initiates from the 3'-end of the cleaved primer.

of this repeated sequence. This observation further supports the discontinuous and nonprocessive model of coronavirus RNA replication (Baric *et al.*, 1985, 1987).

ACKNOWLEDGMENTS

We acknowledge the excellent technical assistance of Victoria Aguinaldo and Monica Mueller and the editorial help of Carol Flores in the preparation of this manuscript. This work was supported by the U.S. Public Health Research Grants AI 19244 and NS 18146, National Multiple Sclerosis Society Research Grant RG 1449, and National Science Foundation Grant PCM-4507.

REFERENCES

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.

BARIC, R. S., SHIEH, C. K., STOHLMAN, S. A., and LAI, M. M. C. (1987). Analysis of intracellular small RNAs of mouse hepatitis virus: Evidence for discontinuous transcription. *Virology* **156,** 342–354.

BARIC, R. S., STOHLMAN, S. A., and LAI, M. M. C. (1983). Characterization of replicative intermediate RNA of mouse hepatitis virus: Presence of leader RNA sequences on nascent chains. *J. Virol.* **48,** 633–640.

BARIC, R. S., STOHLMAN, S. A., RAZAVI, M. K., and LAI, M. M. C. (1985). Characterization of leader-related small RNAs in coronavirus-infected cells: Further evidence for leader-primed mechanism of transcription. *Virus Res.* **3,** 19–33.

BRAYTON, P. R., GANGES, R. G., and STOHLMAN, S. A. (1981). Host cell nuclear function and murine hepatitis virus replication. *J. Gen. Virol.* **56,** 457–460.

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.

BROWN, T. D. K., BOURSNELL, M. E. G., BINNS, M. M., and TOMELY, F. M. (1986). Cloning and sequencing of 5' terminal sequences from avian infectious bronchitis virus genomic RNA. *J. Gen. Virol.* **67,** 221–228.

BUDZILOWICZ, C. J., WILCZYNSKI, S. P., and WEISS, S. R. (1985). Three intergenic regions of coronavirus mouse hepatitis virus strain A59 genome RNA contain a common nucleotide sequence that is homologous to the 3' end of the viral mRNA leader sequence. *J. Virol.* **53,** 834–840.

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

D'ALESSIO, J. M. (1982). RNA sequencing. *In* "Gel Electrophoresis of Nucleic Acids" (D. Rickwood and B. D. Hames, Eds.), pp. 173–197. IRL Press, Eynsham.

GRUNSTEIN, M., and HOGNESS, D. S. (1975). Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* **72,** 3961–3965.

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

HENIKOFF, S., KELLY, J. D., and COHEN, E. H. (1983). Transcription terminates in yeast distal to a control sequence. *Cell* **33,** 607–614.

JACOBS, L., SPAAN, W. J. M., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M. (1981). Synthesis of subgenomic mRNAs of mouse hepatitis virus is initiated independently: Evidence from U.V. transcriptional mapping. *J. Virol.* **34,** 401–406.

KAPKE, P. A., and BRIAN, D. A. (1986). Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* **151,** 41–49.

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 sequence divergence from hepatotropic strain MHV-3 *J. Virol.* **39,** 823–834.

LAI, M. M. C., PATTON, C. D., and STOHLMAN, S. A. (1982). Replication of mouse hepatitis virus: Negative-stranded RNA and replicative form RNA are of genomic length. *J. Virol.* **44,** 487–492.

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

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.

MAKINO, S., FUJIOKA, N., and FUJIWARA, K. (1985). Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus. *J. Virol.* **54,** 329–336.

MAKINO, S., STOHLMAN, S. A., and LAI, M. M. C. (1986). Leader sequences of murine coronavirus mRNAs can be freely reassorted: Evidence for the role of free leader RNA in transcription. *Proc. Natl. Acad. Sci. USA* **83,** 4204–4208.

MAKINO, S., TAGUCHI, F., and FUJIWARA, K. (1984a). Defective interfering particles of mouse hepatitis virus. *Virology* **133,** 9–17.

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

MANIATIS, T., FRITSCH, E. F., and SAMBROOK, J. (1982). "Molecular Cloning—A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

MARTIN, F. H., and TINOCO, I., JR. (1980). DNA-RNA hybrid duplexes containing oligo (dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res.* **8,** 2295–2299.

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

MCMASTER, G. K., and CARMICHAEL, G. G. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA* **74,** 4835–4838.

MILLS, D. R., DABKIN, C., and KRAMER, F. R. (1978). Template-determined, variable rate of RNA chain elongation. *Cell* **15,** 541–550.

NOMOTO, A., and IMURA, N. (1979). A convenient sequencing method for 5′ protein-linked RNAs. *Nucleic Acids Res.* **7,** 1233–1246.

PEDERSEN, F. S., and HASELTINE, W. A. (1980). A micromethod for detailed characterization of high molecular weight RNA. *In* "Methods in Enzymology" (L. Grossman and K. Moldave, Eds.), Vol. 65, pp. 680–687. Academic Press, Orlando, FL.

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

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.* **15,** 5045–5054.

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

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.

THOMAS, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **77,** 5201–5205.

WILHELMSEN, K. C., LEIBOWITZ, J. L., BOND, C. W., and ROBB, J. A. (1981). The replication of murine coronaviruses in enucleated cells. *Virology* **110,** 225–230.

ZARET, K. S., and SHERMAN, F. (1982). DNA sequence required for efficient transcription termination in yeast. *Cell* **28,** 563–573.