JOURNAL OF VIROLOGY, Feb. 1996, p. 705–711 Vol. 70, No. 2
0022-538X/96/$04.00+0

A 5′-Proximal RNA Sequence of Murine Coronavirus as a Potential Initiation Site for Genomic-Length mRNA Transcription

XUMING ZHANG[1]* AND MICHAEL M. C. LAI[1,2,3]

Departments of Neurology[1] and Molecular Microbiology and Immunology[2] and Howard Hughes Medical Institute,[3] University of Southern California School of Medicine, Los Angeles, California 90033-1054

Received 26 May 1995/Accepted 9 October 1995

Coronavirus transcription is a discontinuous process, involving interactions between a *trans*-acting leader and the intergenic transcription initiation sequences. A 9-nucleotide (nt) sequence (UUUAUAAAC), which is located immediately downstream of the leader at the 5′ terminus of the mouse hepatitis virus (MHV) genomic RNA, contains a sequence resembling the consensus intergenic sequence (UCUAAAC). It has been shown previously that the presence of the 9-nt sequence facilitates leader RNA switching and may enhance subgenomic mRNA transcription. It is unclear how the 9-nt sequence exerts these functions. In this study, we inserted the 9-nt sequence into a defective interfering (DI) RNA reporter system and demonstrated that mRNA transcription could be initiated from the 9-nt sequence almost as efficiently as from the intergenic sequence between genes 6 and 7. Sequence analysis of the mRNAs showed that the 9-nt sequence served as a site of fusion between the leaders and mRNA. The transcription initiation function of the 9-nt sequence could not be substituted by other 5′-terminal sequences. When the entire 5′-terminal sequence, including four copies of the UCUAA sequence plus the 9-nt sequence, was present, transcription could be initiated from any of the UCUAA copies or the 9-nt sequence, resulting in different copy numbers of the UCUAA sequence and the deletion of the 9-nt sequence in some mRNAs. All of these heterogeneous RNA species were also detected from the 5′-terminal region of the viral genomic-length RNA in MHV-infected cells. These results thus suggest that the heterogeneity of the copy number of UCUAA sequences at the 5′ end, the deletion of the 9-nt sequence in viral and DI RNAs, and the leader RNA switching are the results of transcriptional initiation from the 9-nt site. They also show that an mRNA species (mRNA 1) that lacks the 9-nt sequence can be synthesized during MHV infection. Therefore, MHV genomic RNA replication and mRNA 1 transcription may be distinguishable.

Mouse hepatitis virus (MHV), a prototype coronavirus, contains a single-strand, positive-sense RNA genome of 31 kb (11, 16). Upon infection into susceptible cells, the viral genomic RNA serves as a template for the synthesis of the negative-strand RNA, which, in turn, is used for the synthesis of the genomic and six to seven subgenomic mRNAs (mRNAs 1 to 7) that are conested at the 3′ ends (11, 13, 17). Each mRNA has a leader RNA of approximately 70 nucleotides (nt) at the 5′ end, which is identical to the leader sequence of genomic RNA (12, 15, 31). It is generally accepted that at the beginning of the viral replication cycle, viral genomic RNA first serves as an mRNA for translation of RNA polymerases from gene 1, which are required for subsequent RNA transcription and replication. The genomic-length RNA (mRNA 1) is eventually packaged into virions to become virion RNA. Thus, viral genomic RNA and mRNA 1 are functionally and probably structurally indistinguishable, and their names are used interchangeably in the literature. However, it has been shown that transcription of all coronavirus subgenomic mRNAs is a discontinuous process (1, 7, 11, 14, 31, 36), involving the interaction and fusion between a *trans*-acting leader and the intergenic (IG) regions, whereas genomic RNA replication is likely a continuous process. It is not clear whether mRNA 1 and viral genomic RNA are synthesized by a mechanism similar to that for subgenomic mRNA synthesis (i.e., discontinuous transcription) or by a continuous replication mechanism or both.

The leader RNA of coronaviruses undergoes rapid exchanges between two RNA molecules. This has been shown to occur between mRNAs of two MHV strains (20, 25) or between MHV RNA and defective interfering (DI) RNA (23). Recently, it has also been demonstrated between the genomic RNA of a bovine coronavirus and its DI RNA (3). Recombinant MHVs with leader RNAs exchanged between two MHV strains have also been isolated (10). Thus, leader exchange appears to be a common phenomenon during coronavirus replication. Studies have revealed that the leader exchange occurs only in the presence of a 9-nt sequence (UUUAUAAAC), which is located at the 5′-proximal untranslated region (UTR) immediately downstream of the leader of the genomic RNA (Fig. 1 and 2) (23). In MHV-infected and DI RNA-transfected cells, the majority of the DI RNA leaders were switched to that of the helper virus after a single cycle of replication if and only if this DI RNA contained the 9-nt sequence (23). Thus, the 9-nt sequence is required for leader RNA exchange during coronavirus replication. How the 9-nt sequence affects leader exchange is unknown. Recently, we have found that the 9-nt sequence also plays a role in coronavirus subgenomic mRNA transcription (36). When a DI RNA contained the 9-nt sequence, mRNAs transcribed from the DI RNA contained the leader sequence derived from the 5′ end of the same DI RNA; however, when the DI RNA lacked the 9-nt sequence, the leaders of subgenomic DI RNAs were derived exclusively in *trans* from the helper virus. Furthermore, the DI RNAs containing the 9-nt sequence have a higher transcription efficiency than the corresponding DI RNA without this sequence (35). Under certain circumstances, the 9-nt sequence could also become part of the leader RNA and be utilized for leader-body

* Corresponding author. Phone: (213) 342-3504. Fax: (213) 342-9555. Electronic mail address: Xumingzh@hsc.usc.edu.

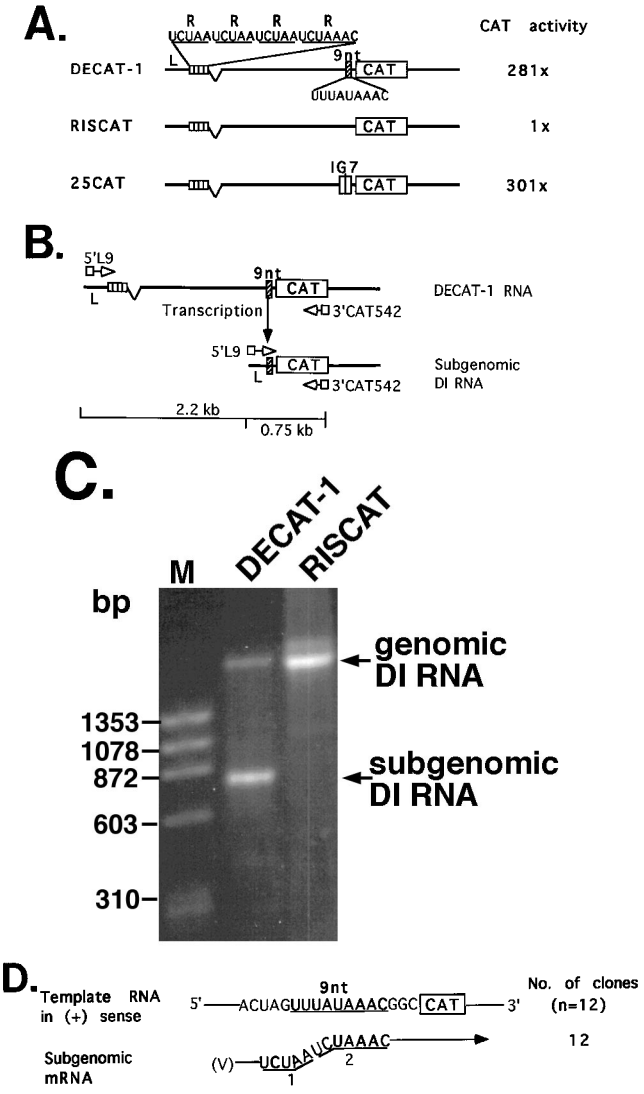

FIG. 1. The 5′-proximal 9-nt sequence of MHV functions as a transcription initiation site. (A) Structures and transcription efficiencies of the DI RNA reporter constructs. All three constructs have the same structure and sequence except for the IG region. The IG sequence in pDECAT-1 is the 9-nt sequence (UUUAUAAAC), which is shown as a small patterned box upstream of the CAT reporter gene. Leader sequence (L) denotes the 5′-end sequence common to all subgenomic mRNAs and upstream of the UCUAA repeats. The four UCUAA repeats (R) are indicated by four narrow boxes, and their sequences are shown at the top and underlined. The V-shaped breaks in the lines represent a 9-nt deletion. p25CAT, which contains a known transcription initiation site (IG7), and pRISCAT, which lacks an IG sequence, are used as positive and negative controls, respectively (18). CAT activities expressed from these constructs, shown on the right, represent fold increase against the background, which is set as onefold (1×) (36). The CAT activities are representative of five independent experiments. (B) Positions and names of the primers (5′L9 and 3′CAT542) used in RT-PCR to amplify genomic and subgenomic mRNAs (shown on the right) in MHV A59-infected and pDECAT-1 RNA-transfected cells. The expected sizes of the PCR products are indicated in kilobases below the sequences. The solid arrow indicates the transcription start site from the 9-nt sequence. (C) PCR products were analyzed by electrophoresis on a 1% agarose gel. The gel was stained with ethidium bromide, photographed with Polaroid film, and scanned with a ScanJet IIc (Hewlett-Packard), and the computer imaging was saved in a tagged image file format (TIFF) with Adobe Photoshop 2.5.1 LE software. Lanes DECAT-1 and RISCAT are the RT-PCR products from helper virus A59-infected and DECAT-1 RNA- and RISCAT RNA-transfected cells, respectively. Bands corresponding to the genomic or subgenomic RNAs are indicated on the right with arrows. On the left are positions of molecular size markers (lane M). (D) Summary of sequences of cDNA clones representing subgenomic CAT-containing RNAs as shown in panel C. Only the IG sequences of the template RNA and the leader fusion site of the subgenomic mRNA are shown. The number of the UCUAA copies at the leader-body fusion site is underlined and indicated. (V), helper viral origin of the leader RNA. The number of cDNA clones specifying the mRNA species is shown on the right. The total number (n) of cDNA clones sequenced is shown in parentheses. The arrow indicates the direction of transcription.

fusion at the IG region during subgenomic mRNA synthesis (33). These studies suggest that the 9-nt sequence is part of the acceptor and donor sequences for leader RNA in subgenomic mRNA synthesis and that the leader switching is a result of discontinuous transcription involving the 9-nt sequence. This may explain how DI RNAs and some variant MHV strains (e.g., JHM2c) lacking the 9-nt sequence were derived (14, 26). It also suggests that transcription may be initiated from the 9-nt sequence in the wild-type viral RNA, thus generating heterogeneous mRNA 1 species, some of which will lack the 9-nt sequence.

The 9-nt sequence in the MHV genome is preceded by a stretch of pentanucleotide (UCUAA) sequences, which are repeated two to four times in all MHVs (22, 23). Part of the 9-nt sequence resembles the consensus IG sequences or transcription initiation sites (UCUAAAC or a similar sequence) for subgenomic mRNAs (2, 29, 30, 32). This fact suggests the potential capacity of the 9-nt sequence to serve as a transcription initiation site. Indeed, previous site-specific mutagenesis studies of MHV RNA have shown that UAUAAAC can serve as a transcription initiation site when inserted in the DI RNAs (8, 21, 32). However, it has not been demonstrated that the 9-nt sequence in the context of the viral genome can initiate mRNA transcription. In this study, we show that the 9-nt sequence in the context of 5′ UTR flanking sequence could serve as a transcription initiation site, resulting in the synthesis of an mRNA 1 species lacking this 9-nt sequence. These results suggest a possible distinction between genomic RNA replication and mRNA 1 transcription and may explain the possible mechanism of leader RNA exchange and the frequent generation of MHV DI RNAs with the 9-nt deletion.

MATERIALS AND METHODS

Viruses and cells. MHV strains A59 (27) and JHM(2) and JHM(3) (22) were used for infection. DBT cells (6) were used for all experiments throughout this study.

Plasmid constructions. To construct a DI RNA reporter plasmid containing the 9-nt sequence (UUUAUAAAC) upstream of the chloramphenicol acetyltransferase (CAT) gene, we performed PCR with a sense primer, 5′Spe9ntCAT (5′-TAACTAG**TTTATAAAC**GGCATGGAGAAAAAAAT-3′), which contains an *Spe*I site (underlined), the 9-nt sequence (boldface), and the first 14 nt of the CAT open reading frame (ORF), and an antisense primer, 3′CAT542 (36), which is complementary to the 3′ end of the CAT ORF. Plasmid DNA p25CAT (18) was used as a template. The PCR fragments were digested with *Spe*I and *Bsp*EI, and the *Spe*I-*Bsp*EI fragment was directionally cloned into the *Spe*I-*Bsp*EI sites of p25CAT, resulting in pDECAT-1 (Fig. 1).

For construction of pDECAT-1U, three sets of PCRs were performed. In the first PCR, a fragment of the 5′ UTR was generated with a sense primer (5′Spe9ntUTR; 5′-TAACTAG**TTTATAAAC**GGCACTTCC-3′, containing an *Spe*I site, the 9-nt sequence, and the 5′ UTR), and an antisense primer (3′UTRCAT; 5′-TTTTCTCCATTATGCAACCTATG-3′, complementary to the 3′ end of the MHV 5′ UTR and the first 10 nt of the CAT ORF). In the second PCR, the CAT gene was synthesized with a sense primer (5′UTRCAT; 5′-GTTGCATAATGGAGAAAAAAATC-3′, containing the last 8 nt of the 5′ UTR and the first 15 nt of the CAT ORF) and an antisense primer, 3′CAT542. p25CAT+9nt DNA (36) was used as a template for both PCRs. The PCR fragments were purified by electrophoresis on a low-melting-point agarose gel. In the third PCR, the two PCR fragments described above were used as templates, and 5′Spe9ntUTR and 3′CAT542 were used as a primer pair. The resulting PCR products were digested with *Spe*I and *Bsp*EI and directionally cloned into p25CAT+9nt (36) as described above.

The construction of pDECAT-4RU and pDECAT-4R1U was similar to that of pDECAT-1U except that in the first PCR, the 5′ primer was 5′SpeI4R (5′-TAA CTAGTCTAATCTAATCTAATCTAAAC-3′), and p25CAT and p25CAT+9nt

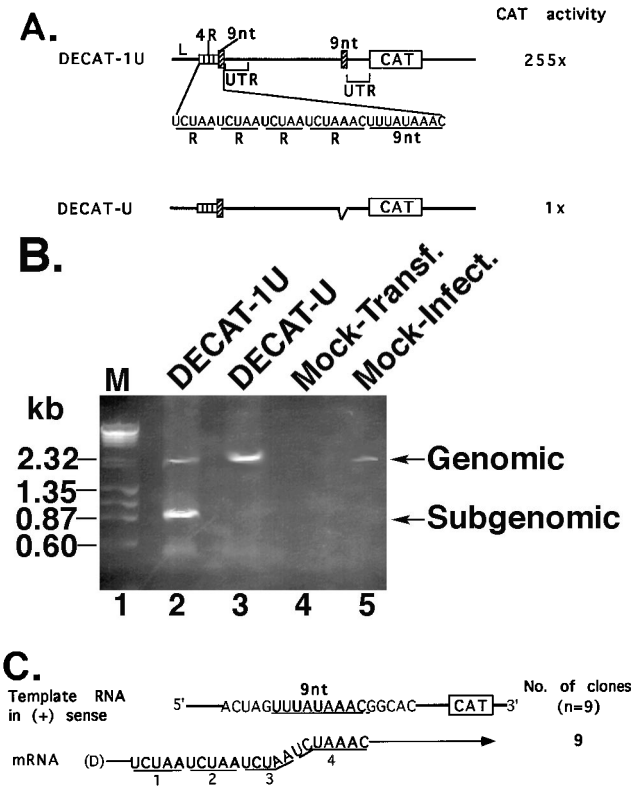

FIG. 2. The 9-nt sequence functions as a transcription initiation site in the presence of the 5′ UTR sequence. (A) The structures of pDECAT-1U and pDECAT-U differ only in the presence (pDECAT-1U) and absence (pDECAT-U) of the 9-nt sequence. CAT activities expressed from these DI RNA constructs are shown on the right and are representative of five independent experiments. UTR, 5′ UTR of MHV genomic RNA (nt 86 to 214 from the 5′ end). (B) Detection of DI RNA transcripts by RT-PCR. RT-PCR products were generated by using primers as shown in Fig. 1B and analyzed by electrophoresis on a 1% agarose gel. Lanes 2 and 3, RT-PCR products of RNA samples from helper virus A59-infected and pDECAT-1U- and pDECAT-U RNA-transfected cells, respectively; lane 4, RNAs from A59-infected and mock-transfected cells; lane 5, RNAs from mock-infected and DECAT-1U-transfected cells. Bands corresponding to genomic and subgenomic RNAs are indicated on the right with arrows. Positions of molecular size markers (M; lane 1) are shown on the left. (C) Structures of the subgenomic mRNA species. Only the sequences of the IG site of the template and the leader fusion site of the subgenomic mRNA are shown. (D) indicates that the leader RNA in subgenomic RNA was derived from the DI RNA. Other depictions are the same as in the legend to Fig. 1.

DNAs were used as templates for pDECAT-4RU and pDECAT-4R1U, respectively.

To construct pDECAT-U, PCR was performed with a pair of primers (5′SpeIUTR [5′-TAACTAGTGGCACTTCCTGCGTG-3′, containing an *Spe*I site and the 5′ UTR lacking the 9-nt sequence] and 3′CAT542) and pDECAT-1U DNA as a template. The PCR fragments were digested with *Spe*I and *Bsp*EI and cloned into the *Spe*I-*Bsp*EI sites of p25CAT+9nt (36).

In vitro transcription and RNA transfection. All plasmid DNAs were linearized with *Xba*I and transcribed in vitro with T7 RNA polymerase as recommended by the manufacturer (Promega). RNA transfection was carried out by the DOTAP method (Boehringer Mannheim) as described previously (36).

Extraction of cell lysates and CAT assay. Infected and transfected DBT cells from a 60-mm-diameter petri dish were harvested at 8 h postinfection in most experiments, resuspended in 150 μl of 0.25 M Tris-HCl (pH 8.0), and lysed by freezing and thawing three times. Cellular lysates were incubated at 60°C for 10 min. Following a brief centrifugation, 50 μl of each sample was assayed for CAT activity, using an assay kit as instructed by the manufacturer (Promega). The CAT reaction was carried out for 12 h as described previously (36).

RT-coupled PCR (RT-PCR) and cloning of subgenomic mRNAs and viral genomic RNAs. For detection of subgenomic DI mRNAs, total RNAs were isolated from MHV-infected and RNA-transfected DBT cells at 7 h postinfection by the Nonidet P-40 method (19) and used for cDNA synthesis by reverse transcription (RT) with the CAT-specific antisense primer 3′CAT542. An additional sense primer (5′L9) representing the 5′ end of the leader was used for the subsequent PCR amplification (36). The specific PCR fragment, which represents the 5′ terminus of the subgenomic CAT-containing mRNA, was analyzed either by agarose (1%) gel electrophoresis or by sequencing. For the latter, the PCR products were digested with *Sna*BI and *Eco*RI, and the *Sna*BI-*Eco*RI fragment was cloned into the *Sma*I and *Eco*RI sites of pBluescript (Stratagene). For detection of viral genomic RNAs, total intracellular RNAs were isolated from DBT cells infected with JHM(3) at 9 h postinfection. The 3′ primer used in RT and PCR was 3′499 (5′-CATCATAGTCGAGGCCTCCAC-3′, complementary to the nt 470 to 499 of gene 1, including an endogenous *Stu*I sequence) (16) to amplify the 5′-terminal 499-nt sequence. The 499-nt PCR fragment was digested with *Sna*BI and *Stu*I and cloned into the *Sma*I site of pBluescript.

DNA sequencing. The sequences of all cDNA clones were determined by the dideoxynucleotide chain termination method (28). Primers 3′-A140 (5′-AAT GTCAGCACTATGACAAG-3′) (36) and 3′-CAT56 (5′-TTACGATGCCATT GGGATA T-3′, complementary to nt 37 to 56 from the beginning of the CAT ORF) were used to determine the 5′-terminal sequences of genomic and subgenomic RNAs, respectively.

RESULTS

The 5′-proximal 9-nt sequence (UUUAUAAAC) immediately downstream of the leader serves as a transcription initiation site in the DI RNA reporter. Since part of the 9-nt sequence (UAUAAAC) resembles the consensus IG sequence (UCUAAAC), and UAUAAAC has been shown to serve as a transcription initiation signal for mRNA synthesis when inserted into DI RNAs (21, 32), we expected that the 9-nt sequence may also act as a transcription initiation site for mRNA synthesis. However, the transcription initiation potential of various IG-related sequences as determined by site-specific mutagenesis in DI RNA and those in viral genome did not correlate precisely (21, 32), suggesting that neighboring sequence context could affect transcriptional initiation. Therefore, to determine the transcriptional potential of the 9-nt sequence, we used a previously characterized DI-CAT construct, p25CAT (18, 36), and replaced the IG sequence with the 9-nt sequence, so that the CAT gene was under the control of this 9-nt sequence (pDECAT-1) (Fig. 1). Previous studies have shown that the CAT activity from this construct (p25CAT) could be only detected when a CAT-containing subgenomic mRNA is transcribed from the IG site (18). Following infection of cells with a helper MHV (A59) and transfection with the in vitro-transcribed DI-CAT RNA, CAT activities in the cell lysates were assayed. Since the leader sequences of helper virus (A59) and DI RNA (derived from JHM) are different, the origin of the leader sequence in the subgenomic mRNA could be identified (7, 18, 36). As shown in Fig. 1A, the CAT activity from pDECAT-1 was expressed at a level (281-fold increase) comparable to that from p25CAT (301-fold increase). In contrast, when the construct pRISCAT (18), which lacks the consensus IG sequence, was used, no CAT activity was detected. These results were reproducible in five independent experiments. The results suggest that the 9-nt sequence alone is sufficient to promote the transcription of subgenomic mRNAs.

We next examined the structure of the CAT-containing subgenomic mRNA by RT-PCR using a CAT-specific antisense 3′ primer and a leader-specific sense 5′ primer (Fig. 1B). Figure 1C shows that CAT-containing subgenomic mRNAs were detected in cells transfected with pDECAT-1 RNAs, but not in pRISCAT RNA-transfected cells, consistent with the results obtained from the CAT assay (Fig. 1A). Sequence analysis of the 12 cDNA clones showed that they represent CAT subgenomic mRNAs containing a leader RNA which is joined via two UCUAA copies (without the 9-nt sequence) to the CAT gene sequence. These results suggest that the leader-mRNA fusion occurred within the 9-nt sequence (Fig. 1D). All of the leader sequences in these clones were derived in *trans* from the helper virus RNA, which contains a trinucleotide sequence

CUC at positions 30 to 32 (36) (data not shown). This result is consistent with the previous finding that when DI RNA does not contain the 9-nt sequence, the leader of subgenomic RNAs is derived exclusively in *trans* (18, 36), inasmuch as this DI RNA does not contain the 9-nt sequence at the 5′ end.

The 9-nt sequence could initiate transcription in the context of the 5′-terminal sequence. The experiment described above suggested that the 9-nt sequence could serve as a potential transcription initiation site. Previous studies have also suggested that some neighboring sequences of the IG region could affect transcription efficiencies (8, 9, 21). Therefore, to determine whether the 9-nt sequence located at the 5′ end of the genomic RNA can initiate mRNA transcription, we examined the effect of the neighboring 5′-terminal sequence on the transcriptional initiation from the 9-nt sequence. We first studied the effects of the 5′ UTR (nt 86 to 214 from the 5′ end), which is downstream of the 9-nt sequence and immediately precedes the ORF for gene 1 in the viral genomic RNA (Fig. 2A). We used a construct, pDECAT-1U, which contains the MHV 5′ UTR inserted between the 9-nt sequence and the CAT gene (Fig. 2A). The subgenomic mRNA transcribed from this DI RNA would mimic the structure of a potential mRNA 1 transcribed from the 9-nt sequence in the viral genomic RNA. As shown in Fig. 2A, CAT activity was expressed from pDECAT-1U RNA at a high level (255-fold increase), similar to that from pDECAT-1 RNA (Fig. 1A), suggesting that the 9-nt sequence initiated the transcription of a CAT mRNA, even in the presence of the 5′ UTR sequence. No CAT activity was detected from pDECAT-U RNA, which contains the 5′ UTR but lacks the 9-nt sequence (Fig. 2A), again indicating the requirement of the 9-nt sequence in initiating CAT mRNA transcription. To confirm these results, we performed RT-PCR using a CAT-specific antisense primer and a leader-specific sense primer to detect possible subgenomic RNA transcripts in DI RNA-transfected and MHV-infected cells. A subgenomic CAT-containing mRNA species was detected in pDECAT-1U-transfected and A59-infected cells by RT-PCR (Fig. 2B, lane 2). No specific PCR products derived from subgenomic CAT mRNAs were detected in pDECAT-U-transfected, mock-transfected, or mock-infected cells (Fig. 2B, lanes 3 to 5). As expected, all of the cells transfected with DI RNAs yielded a PCR product representing the full-length DI RNA. Since these PCR primers would have favored shorter RNA (subgenomic mRNA, ca. 0.86 kb) over the longer RNA (full-length DI RNA, ca. 2.36 kb), the total absence of PCR products representing subgenomic mRNAs suggested that very little, if any, subgenomic mRNA was synthesized by pDECAT-U. These results thus indicate that the 9-nt sequence is the only sequence at the 5′ UTR (downstream of the leader) required for mRNA transcription.

The structures of CAT-containing subgenomic mRNAs transcribed from pDECAT-1U RNA were studied by sequencing of RT-PCR clones. The sequences of these clones suggest that the leader fusion sites in these subgenomic mRNAs were likely at the 9-nt sequence (Fig. 2C), similar to those transcribed from pDECAT-1 RNAs (Fig. 1D), further suggesting that the 9-nt sequence could serve as a transcriptional start sequence. Interestingly, the leaders of all nine cDNA clones were of DI RNA origin, which contains the trinucleotide sequence UAA at positions 30 to 32 from the 5′ end, in contrast to those from pDECAT-1 RNA. This result is consistent with the previous finding that the presence of the 9-nt sequence enables the DI RNA-derived leader to contribute to subgenomic mRNA (36), since pDECAT-1U contains the 9-nt sequence immediately downstream of the leader.

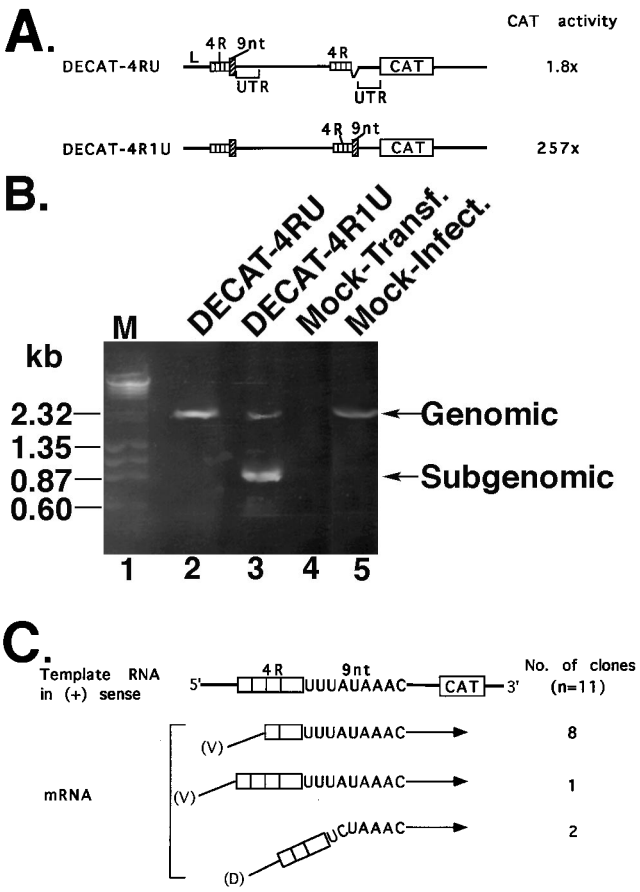

FIG. 3. The 9-nt sequence is the only transcription initiation site at the 5′-terminal region of MHV RNA. (A) Structures of pDECAT-4RU and pDECAT-4R1U. They differ only in the presence (pDECAT-4R1U) and absence (pDECAT-4RU) of the 9-nt sequence. CAT activities expressed from these DI RNA constructs are shown on the right and are representative of five independent experiments. UTR, 5′ UTR of MHV genomic RNA. The V-shaped break indicates the deletion of the 9-nt sequence. (B) Detection of RNA transcripts by RT-PCR. RT-PCR products were generated by using primers as shown in Fig. 1B and analyzed by electrophoresis on a 1% agarose gel. Lanes 2 and 3, RT-PCR products of RNA samples from helper virus A59-infected and pDECAT-4RU- and pDECAT-4R1U RNA-transfected cells, respectively; lane 4, RNAs from A59-infected and mock-transfected cells; lane 5, RNAs from mock-infected and pDECAT-4R1U-transfected cells. Bands corresponding to genomic and subgenomic RNAs are indicated on the right with arrows. Positions of molecular size markers (M; lane 1) are shown on the left. (C) Structures of the subgenomic mRNA species transcribed from pDECAT-4R1U. Only the IG sequence of the template and the leader fusion site of the subgenomic mRNAs are shown. Each box represents one copy of UCUAA. The larger box represents UCUAAAC. (V) and (D) indicate that the leaders of subgenomic mRNAs were derived from the helper viral and DI RNAs, respectively. Other depictions are the same as explained in the legend to Fig. 1.

The transcription initiation function of the 9-nt sequence could not be replaced by the 5′-UCUAA repeat sequences. Viral and DI genomic RNAs contain several copies of the UCUAA sequence immediately upstream of the 9-nt sequence. Since the pentanucleotide repeat sequence is similar to the 9-nt sequence and part of it is identical to the consensus IG sequence, it was predicted that this repeat sequence may also function as a transcription initiation site. To test this possibility, we constructed a DI RNA reporter plasmid (pDECAT-4RU) which contains the CAT gene behind the four copies of the UCUAA sequence and the 5′ UTR (nt 86 to 214 from the 5′ end) (16) but lacks the 9-nt sequence (Fig. 3A). The result surprisingly showed that no CAT activity was de-

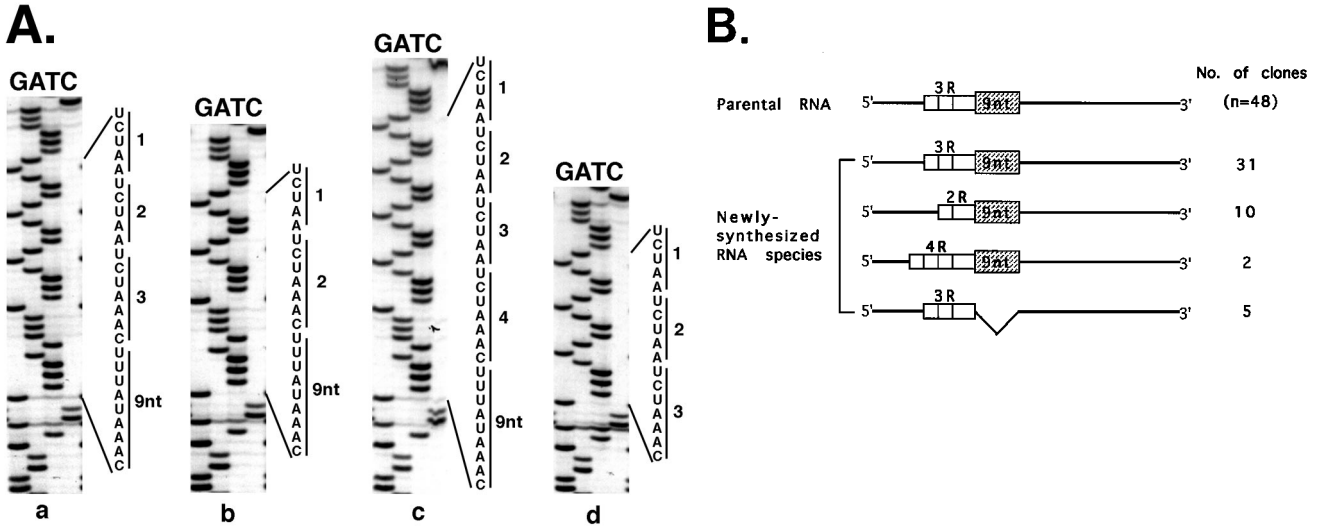

FIG. 4. Heterogeneity of the 5′-terminal sequence and structure of JHM(3) genomic RNA. (A) DNA sequences of the RT-PCR products, showing the sequences at the 3′ end of the leader of JHM(3) genomic-length RNAs. The copy number of the UCUAA sequence and the 9-nt sequence are indicated on the right for each RNA species (a through d). (B) Structural diagram of the 5′ terminus of the newly synthesized RNA species and the parental viral genomic RNA. The last repeat in the leader is a heptanucleotide sequence (UCUAAAC) instead of a pentanucleotide sequence (UCUAA); it is represented by a longer rectangle. The V-shaped break indicates the 9-nt deletion. The numbers of cDNA clones representing each RNA species are shown on the right. The total number (n) of cDNA clones sequenced is shown in parentheses.

tected (Fig. 3A), indicating that the pentanucleotide sequences at the 5′-terminal viral genomic RNA could not substitute for the 9-nt sequence to initiate mRNA transcription. When the 9-nt sequence was inserted between the four copies of the UCUAA sequence and the 5′ UTR (pDECAT-4R1U), the CAT activity was restored (Fig. 3A). These data were confirmed by five independent experiments. This result was unexpected since the four copies of UCUAA include the consensus sequence found in most of the IG sequences. To rule out the possibility that pDECAT-4RU had transcribed a subgenomic mRNA which could not be translated, we performed RT-PCR to detect possible subgenomic CAT RNA transcripts in MHV-infected and pDECAT-4RU- or pDECAT-4R1U-transfected cells. Figure 3B shows that a specific CAT-containing mRNA was transcribed from pDECAT-4R1U RNA (lane 3) but not from DECAT-4RU RNA (lane 2), whereas the full-length DI RNAs were detected in all of the samples. No CAT-specific subgenomic mRNAs were detected in mock-transfected or mock-infected cells (lanes 4 and 5). These results are consistent with those of the CAT assay (compare Fig. 3A and B), thus confirming that the lack of CAT activity from pDECAT-4RU was due to failure of mRNA transcription. The combined results suggest that the 9-nt sequence is the only signal at the 5′ terminus of the genomic RNA capable of transcriptional initiation. The failure of the four UCUAA repeats to serve as a transcription initiation signal may be due to RNA conformation, as the neighboring sequence has been shown to affect the transcription initiation (8, 9).

The leader-body joining regions of the CAT-containing mRNAs transcribed from pDECAT-4R1U RNA were examined by RT-PCR cloning and DNA sequencing. Eight of the eleven cDNA clones contained two UCUAA copies and the 9-nt sequence at the junction, and one clone contained four UCUAA copies and the 9-nt sequence (Fig. 3C). These RNAs represent the fusion of the leader RNA at the UCUAA repeat region. The leaders of these mRNAs were derived from the helper virus, which has a CUC at nt 30 to 32 from the 5′ end of the genome (reference 36 and data not shown). Two additional clones contained four UCUAA copies but lacked the 9-nt sequence at the junction site (Fig. 3C). These mRNAs represent the fusion of the leader RNA at the 9-nt sequence, resulting in the loss of the 9-nt sequence in the mRNA.

Detection of the 9-nt deletion in MHV genomic-length RNA in vivo. As demonstrated above, the 9-nt sequence in the context of 5′-end genomic sequence functions as a transcription initiation site in the DI RNA system. If it also functions as a transcription initiation site in the viral genomic RNA, then the genomic-length RNAs in the virus-infected cells should contain some mRNA species that have lost the 9-nt sequence at the 5′ terminus. Such an RNA species has not been detected previously by primer extension and direct RNA sequencing (22). We reexamined this possibility by RT-PCR. MHV strain JHM(3), which has three copies of the UCUAA sequence and the 9-nt sequence, was plaque purified and then propagated in DBT cells for fewer than five passages. Cells were infected with JHM(3) at a multiplicity of infection of 5, and total intracellular RNAs were isolated at 9 h postinfection. The 5′-terminal 499-nt sequence of the viral genomic RNA was reverse transcribed and monitored by PCR amplification for 10 cycles. Individual cDNA clones were sequenced (Fig. 4). Of the total of 48 clones sequenced, 31 clones contained three UCUAA copies and the 9-nt sequence at the 5′-terminal end (Fig. 4A, panel a), identical to the parental genomic RNA. Ten other clones had two UCUAA copies and the 9-nt sequence (panel b), consistent with the previous finding that viruses with three UCUAA copies rapidly lost one copy during virus evolution (22). Interestingly, two other clones contained four UCUAA copies and the 9-nt sequence (panel c); these clones thus contained one more copy of UCUAA than the parental virus. This type of RNA has not been observed previously in JHM(3) virus. Additionally, five clones had three UCUAA copies but were missing the 9-nt sequence (panel d), consistent with an mRNA 1 species initiated from the 9-nt sequence (Fig. 4A and B). These data suggest that the 9-nt sequence in the viral genomic-length RNA could serve as a transcriptional initiation site, resulting in the loss of the 9-nt sequence. These results

combined suggest that a discontinuous transcription process may occur during coronavirus genomic-length RNA synthesis, generating an mRNA 1 species which is distinct from the viral genomic RNA.

DISCUSSION

The 9-nt sequence as a transcription initiation site. In this study, we demonstrated by using a DI RNA reporter system that a 5′-proximal 9-nt sequence (UUUAUAAAC) of MHV genomic RNA functioned as a transcription initiation site for mRNA synthesis. The 9-nt sequence either alone or in the presence of other neighboring sequences was able to initiate the synthesis of a subgenomic mRNA. This finding is consistent with the fact that part of the 9-nt sequence resembles the consensus transcriptional initiation sequence at various IG sites. It is also supported by the previous findings that two mutant IG sequences containing the motif UAUAAAC in DI RNA constructs had a transcription efficiency similar to that of the wild-type consensus IG sequence (UCUAAAC) (8, 32). However, the UAUAAAC sequence alone may not necessarily be a functional promoter; its activity may be modulated by other flanking sequences. In fact, the UAUAAAC sequence is present at two other locations in MHV genomic RNA (at nt 11432 to 11438 of gene 1 and 54 to 60 of gene 2), but no corresponding mRNA species were synthesized from these sites. Our present finding directly demonstrated that the 9-nt sequence serves as a transcription initiation site in the context of the 5′ UTR sequence. It is surprising, however, that the four copies of UCUAA sequence at the 3′ end of the leader plus the 5′ UTR, but in the absence of the 9-nt sequence (construct pDECAT-4RU in Fig. 3), did not serve as a transcription initiation site for mRNA synthesis, despite the fact that the last 14 nt of the UCUAA repeat sequence (AAUCUAAUC UAAAC) are identical to the consensus sequence IG7, which is the strongest transcription initiation signal of MHV. One possible interpretation is that multiple copies of the UCUAA sequence form a stem-loop structure (data not shown), which is unfavorable for interacting with a *trans*-acting leader. The inability of the UCUAA repeat sequences in initiating mRNA transcription is consistent with the finding that the UCUAA sequence alone is not sufficient to cause leader RNA switching (23) (see below). More recently, it was reported that the UC UAA sequence also was not required for leader RNA exchange of bovine coronavirus (4), again consistent with this concept. In contrast, the 9-nt sequence resides at the end of a possible stem-loop structure (data not shown), thus probably allowing the binding of *trans*-acting leaders. Alternatively, the 9-nt sequence may bind a protein, which is required for transcription, in a mechanism similar to the proposed model for subgenomic mRNA transcription (32, 34, 36). It will be interesting to determine whether the 5′-terminal sequences of MHV genomic RNA containing or lacking the 9-nt sequence have different protein-binding properties.

Leader RNA switching. Previously, it was found that the presence of the 9-nt sequence promoted leader RNA exchange (23). It was unclear how the 9-nt sequence exerts such a function. Data from the current report and previous studies (22, 23) suggest a possible mechanism (Fig. 5). When the 9-nt sequence is absent at the 5′ terminus of the genomic RNA, no discontinuous transcription (mRNA 1) occurs at the 5′ terminus. Thus, the leader RNA cannot be switched to another RNA molecule (Fig. 5A). In this case, the genomic-length RNA results from faithful copying of the template strand through a continuous replication mechanism. When the 9-nt sequence is present between the UCUAA repeat sequence of the leader and the 5′ UTR sequence, the negative-strand repetitive sequence (including the UCUAA repeats and the 9-nt sequence) provides an initiation site for a *trans*-acting leader to transcribe an mRNA 1, resulting in leader RNA switching (Fig. 5B). Because of the imprecise nature of leader binding during the discontinuous transcription process, the leader-body fusion site of mRNA 1 is heterogeneous [Fig. 5B, (a) to (c)], containing different numbers of UCUAA copies [Fig. 3 and 4 and Fig. 5B, (a′) and (b′)]. When the leader binding takes place on the 9-nt sequence, an mRNA species lacking the 9-nt sequence will result [Fig. 5B, (c′)].

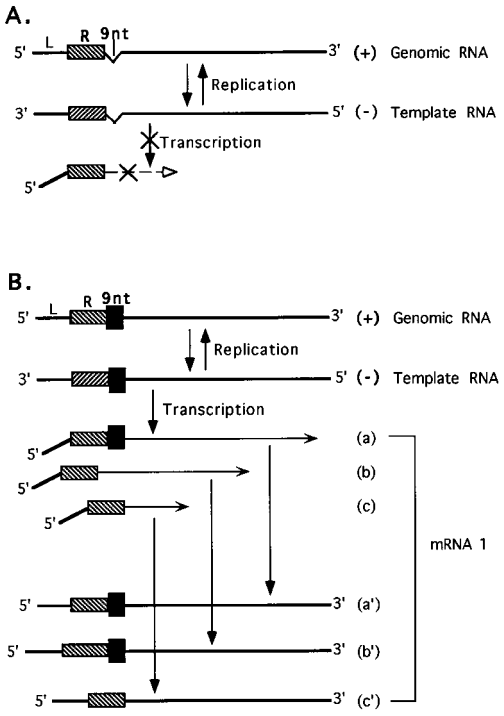

FIG. 5. Proposed mode of leader switching and evolutionary pathway of MHV variants and DI RNAs. (A) In the absence of the 9-nt sequence, the *trans*-acting leader cannot initiate transcription. Hence, there is no mRNA 1 transcription and consequently no leader RNA switching, although RNA replication still takes place. In this case, the synthesis of genomic-length RNA results from a continuous replication process. The patterned boxes and breaks indicate the UCUAA repeats (R) in the leader (L) and the 9-nt sequence deletion, respectively. (B) In the presence of the 9-nt sequence in the genomic RNA, the *trans*-acting leader primes the mRNA 1 transcription, and as a result, leader RNA switching takes place. (a) to (c), the *trans*-acting leaders bind to the different positions of the template, resulting in the synthesis of mRNA 1 species differing in sequences at the leader-body joining site [(a′) to (c′)]. Thus, the newly synthesized RNA species are heterogeneous at the 5′-terminal repetitive sequence, including the UCUAA repeats and the 9-nt sequence.

Evolution of MHV variants and DI RNAs. It was shown previously that JHM(3), which contains three UCUAA copies in the leader, undergoes rapid conversion to JHM(2), which contains only two UCUAA copies (22). In this study, an RNA species containing four UCUAA copies in the leader was found in JHM(3)-infected cells (Fig. 4A, panel c). This increase of the copy number of UCUAA sequence in the viral genome during virus replication has not been observed previously. The current finding that the 9-nt sequence serves as a transcription initiation site for mRNA synthesis suggests that the heterogeneities of UCUAA copy number may have resulted from transcriptional start from the UCUAA repeats or the 9-nt sequence at the 5′ end of the viral genome, similar to the heterogeneity of the leader fusion in MHV subgenomic

mRNAs (24). Since these genomic-length mRNAs contain the packaging signal (5), they are expected to be packaged into virions and function as viral genomic RNAs. This interpretation may explain why MHV genomic RNA evolves so rapidly in this region. In contrast, if the heterogeneity of viral RNA were due to RNA polymerase slippage during continuous RNA replication, similar heterogeneity would have been detected in the IG regions of viral genomic RNA. Such a heterogeneity did not occur. Likewise, this finding may also explain why many recombinant viruses have recombination sites at the 5′ end of the viral genome, exchanging only the leader sequences (10). It should be noted that although the experiments performed here and previously (33) used plaque-cloned viruses, it cannot be rigorously ruled out that the heterogeneity of viral genome existed in the virus population and did not arise de novo during viral RNA replication. A definite experiment would require an infectious recombinant viral RNA. Nevertheless, the heterogeneity of RNA sequence in this region, in contrast to other regions, suggests the unusual property of the 5′-proximal sequence in the viral genome.

Possible distinction between genomic RNA and mRNA 1. Our findings also suggest that the synthesis of coronavirus genomic-length RNAs may include both a discontinuous process (i.e., transcription of mRNA 1) and a continuous, faithful replication mechanism. This raises an interesting question: are mRNA 1 and genomic RNA functionally distinguishable? Conceivably, the mRNA 1 species, which lacks the 9-nt sequence, detected in this study may be translated more efficiently than the viral genomic RNA, which contains the 9-nt sequence, in virus-infected cells. If this is the case, the viral gene 1 products would be more abundantly synthesized during the peak of viral mRNA synthesis but not of RNA replication. This possibility can be experimentally tested.

ACKNOWLEDGMENTS

This work was supported by Public Health Service Research grants AI 19244 and NS 18146 from the National Institutes of Health. M.M.C.L. is an Investigator of the Howard Hughes Medical Institute.

REFERENCES

1. **Baker, S. C., and M. M. C. Lai.** 1990. An *in vitro* system for the leader-primed transcription of coronavirus mRNAs. EMBO J. **9:**4173–4179.
2. **Budzilowicz, C. J., S. P. Wilczynski, and S. R. Weiss.** 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.
3. **Chang, R.-Y., M. A. Hofmann, P. B. Sethna, and D. A. Brian.** 1994. A *cis*-acting function for the coronavirus leader in defective interfering RNA replication. J. Virol. **68:**8223–8231.
4. **Chang, R.-Y., R. Krishnan, and D. A. Brian.** 1995. The leader UCUAAAC sequence element on the bovine coronavirus defective-interfering RNA replicon is not required for rapid reversion of a mutated leader, abstr. W32-7, p. 181. *In* Abstracts of the 14th Scientific Meeting of the American Society for Virology 1995. American Society for Virology, Austin, Tex.
5. **Fosmire, J. A., K. Hwang, and S. Makino.** 1992. Identification and characterization of a coronavirus packaging signal. J. Virol. **66:**3522–3530.
6. **Hirano, N., K. Fujiwara, S. Hino, and M. Matsumoto.** 1974. Replication and plaque formation of mouse hepatitis virus (MHV-2) in mouse cell line DBT culture. Arch. Gesamte Virusforsch. **44:**298–302.
7. **Jeong, Y. S., and S. Makino.** 1994. Evidence for coronavirus discontinuous transcription. J. Virol. **68:**2615–2623.
8. **Joo, M., and S. Makino.** 1992. Mutagenic analysis of the coronavirus intergenic consensus sequence. J. Virol. **66:**6330–6337.
9. **Joo, M., and S. Makino.** 1995. The effect of two closely inserted transcription consensus sequences on coronavirus transcription. J. Virol. **69:**272–280.
10. **Keck, J. G., S. A. Stohlman, L. H. Soe, S. Makino, and M. M. C. Lai.** 1987. Multiple recombination sites at the 5′-end of murine coronavirus RNA. Virology **156:**331–341.
11. **Lai, M. M. C.** 1990. Coronavirus: organization, replication and expression of genome. Annu. Rev. Microbiol. **44:**303–333.
12. **Lai, M. M. C., R. S. Baric, P. R. Brayton, and S. A. Stohlman.** 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.
13. **Lai, M. M. C., P. R. Brayton, R. C. Armen, C. D. Patton, C. Pugh, and S. A. Stohlman.** 1981. Mouse hepatitis virus A59: messenger RNA structure and genetic localization of the sequence divergence from the hepatotropic strain MHV3. J. Virol. **39:**823–834.
14. **Lai, M. M. C., S. Makino, L. H. Soe, C.-K. Shieh, J. G. Keck, and J. O. Fleming.** 1987. Coronavirus: a jumping RNA transcription. Cold Spring Harbor Symp. Quant. Biol. **52:**359–365.
15. **Lai, M. M. C., C. D. Patton, R. S. Baric, and S. A. Stohlman.** 1983. Presence of leader sequences in the mRNA of mouse hepatitis virus. J. Virol. **46:**1027–1033.
16. **Lee, H. J., C. K. Shieh, A. E. Gorbalenya, E. V. Koonin, N. La Monica, J. Tuler, A. Bagdzhadzhyan, and M. M. C. Lai.** 1991. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology **180:**567–582.
17. **Leibowitz, J. L., K. C. Wilhemsen, and C. W. Bond.** 1981. The virus-specific intracellular RNA species of two murine coronaviruses: MHV-A59 and MHV-JHM. Virology **114:**39–51.
18. **Liao, C.-L., and M. M. C. Lai.** 1994. Requirement of the 5′-end genomic sequence as an upstream *cis*-acting element for coronavirus subgenomic mRNA transcription. J. Virol. **68:**4727–4737.
19. **Makino, S., N. Fujioka, and K. Fujiwara.** 1985. Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus. J. Virol. **54:**329–336.
20. **Makino, S., J. G. Keck, S. A. Stohlman, and M. M. C. Lai.** 1986. High-frequency RNA recombination of murine coronaviruses. J. Virol. **57:**729–737.
21. **Makino, S., and M. Joo.** 1993. Effect of intergenic consensus sequence flanking sequences on coronavirus transcription. J. Virol. **67:**3304–3311.
22. **Makino, S., and M. M. C. Lai.** 1989. Evolution of the 5′-end of genomic RNA of murine coronaviruses during passages in vitro. Virology **169:**227–232.
23. **Makino, S., and M. M. C. Lai.** 1989. High-frequency leader sequence switching during coronavirus defective interfering RNA replication. J. Virol. **63:**5285–5292.
24. **Makino, S., L. H. Soe, C. K. Shieh, and M. M. C. Lai.** 1988. Discontinuous transcription generates heterogeneity at the leader fusion sites of coronavirus mRNAs. J. Virol. **62:**3870–3873.
25. **Makino, S., S. A. Stohlman, and M. M. C. Lai.** 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.
26. **Makino, S., F. Taguchi, N. Hirano, and K. Fujiwara.** 1984. Analysis of genomic and intercellular viral RNAs of small plaque mutants of mouse hepatitis virus, JHM strain. Virology **39:**138–151.
27. **Manaker, R. A., C. V. Piczak, A. A. Miller, and M. F. Stanton.** 1961. A hepatitis virus complicating studies with mouse leukemia. J. Natl. Cancer Inst. **27:**29–51.
28. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
29. **Shieh, C. K., H. J. Lee, K. Yokomori, N. La Monica, S. Makino, and M. M. C. Lai.** 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.
30. **Shieh, C. K., L. H. Soe, S. Makino, M. F. Chang, S. A. Stohlman, and M. M. C. Lai.** 1987. The 5′-end sequence of the murine coronavirus genome: implications for multiple fusion sites in leader-primed transcription. Virology **156:**321–330.
31. **Spaan, W., H. Delius, M. Skinner, J. Armstrong, P. Rottier, S. Smeekens, B. A. M. van der Zeijst, and S. G. Siddell.** 1983. Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. EMBO J. **2:**1839–1844.
32. **Van der Most, R. G., R. J. de Groot, and W. J. M. Spaan.** 1994. Subgenomic RNA synthesis directed by a synthetic defective interfering RNA of mouse hepatitis virus: a study of coronavirus transcription initiation. J. Virol. **68:**3656–3666.
33. **Zhang, X. M., and M. M. C. Lai.** 1994. Unusual heterogeneity of leader-mRNA fusion in a murine coronavirus: implications for the mechanism of RNA transcription and recombination. J. Virol. **68:**6626–6633.
34. **Zhang, X. M., and M. M. C. Lai.** 1995. Interactions between the cytoplasmic proteins and the intergenic (promoter) sequence of mouse hepatitis virus RNA: correlation with the amounts of subgenomic mRNA transcribed. J. Virol. **69:**1637–1644.
35. **Zhang, X. M., and M. M. C. Lai.** 1995. Regulation of coronavirus RNA transcription is likely mediated by protein-RNA interactions, p. 515–521. *In* P. J. Talbot and G. A. Levy (ed.), Corona and related viruses, Plenum Press, New York.
36. **Zhang, X. M., C.-L. Liao, and M. M. C. Lai.** 1994. Coronavirus leader RNA regulates and initiates subgenomic mRNA transcription both in *trans* and in *cis*. J. Virol. **68:**4738–4746.