JOURNAL OF VIROLOGY, Oct. 1994, p. 6626–6633
0022-538X/94/$04.00+0

Vol. 68, No. 10

# Unusual Heterogeneity of Leader-mRNA Fusion in a Murine Coronavirus: Implications for the Mechanism of RNA Transcription and Recombination

XUMING ZHANG AND MICHAEL M. C. LAI*

*Howard Hughes Medical Institute and Department of Microbiology, University of Southern California School of Medicine, Los Angeles, California 90033-1054*

Received 18 April 1994/Accepted 30 June 1994

**Coronavirus mRNA transcription was thought to be regulated by the interaction between the leader RNA and the intergenic sequence (IS), probably involving direct RNA-RNA interactions between complementary sequences. In this study, we found that a particular strain of mouse hepatitis virus, JHM2c, which has a deletion of a 9-nucleotide (nt) sequence (UUUAUAAAC) immediately downstream of the leader RNA, transcribed subgenomic mRNA species containing a whole array of heterogeneous leader fusion sites. Using a transfected defective interfering RNA which contains an IS and a reporter (chloramphenicol acetyltransferase) gene and JHM2c as a helper virus, we demonstrated that subgenomic mRNAs transcribed from the defective interfering RNAs were extremely heterogeneous. The leader-mRNA fusion sites in this virus can be grouped into five types. In type I, the leader is fused with the consensus IS of the template RNA at a site within the UCUAA repeats, consistent with the classical model of discontinuous transcription. In type II, the leader is fused with the consensus IS as in type I, but the leader of mRNA contains some nucleotide substitutions within the UCUAA repeats. In type III, the leader is fused with mRNAs at a site either upstream or downstream of the consensus IS. The sequences around the fusion sites bear little or no homology to the leader. As a result, mRNAs contain sequences complementary to the template sequences upstream of the IS or have sequence deletions downstream of the IS. In type IV, the leader is fused to the IS at the 9-nt sequence immediately downstream of the UCUAA repeats. In type V, the leader-mRNA fusion site contains a duplication of a portion of the leader sequence or an insertion of nontemplated sequences which are not present in either leader or template RNA. These patterns of leader-mRNA fusion resemble the aberrant homologous recombination frequently seen in other RNA viruses. The degree of heterogeneity of leader fusion sites is dependent on the sequences of both the leader RNA and IS. These results suggest that leader-mRNA fusion in coronavirus transcription does not require direct RNA-RNA interaction between complementary sequences. A modified model of RNA transcription and recombination based on protein-RNA and protein-protein interactions is proposed. This study also provides a paradigm for aberrant homologous recombination.**

Coronavirus mRNA synthesis involves a unique mechanism of discontinuous transcription, generating subgenomic mRNAs which contain a leader RNA fused to a distant RNA sequence (8). The precise mechanism of this RNA synthesis is not yet fully understood. At least two alternative, though not necessarily mutually exclusive, models have been proposed. One is the leader-primed transcription mechanism, in which a leader RNA is transcribed from the 3′ end of the genome-size, negative-strand template RNA, dissociates, and rejoins the same or a different template RNA at downstream intergenic regions to serve as the primer for subgenomic mRNA synthesis (8). An abundance of experimental data are compatible with this model (for a review, see reference 8). More recently, by using a defective interfering (DI) RNA containing a reporter gene, it has been demonstrated that most of the leader sequence of the subgenomic mRNAs is derived in *trans* from a separate RNA molecule (16, 33), further substantiating the *trans*-acting nature of coronavirus leader RNA. However, this model does not account for the possible roles of the subgenomic, negative-strand RNAs (27) and subgenomic replicative form RNAs (26) in coronavirus-infected cells, although conceivably the former can serve as templates for the leader-primed transcription. Thus, an alternative model has been proposed; i.e., the discontinuous transcription step occurs during negative-strand RNA synthesis, and subgenomic mRNAs are transcribed continuously from subgenomic, negative-strand RNA templates (26). Although neither model has been firmly established, our recent findings that mRNA transcription is regulated by three RNA elements, i.e., intergenic (promoter) sequence, *trans*-acting leader, and *cis*-acting leader sequence, and that the leader RNA initiates (primes) subgenomic mRNA transcription both in *trans* and in *cis* are more consistent with the leader-primed transcription model (16, 33). How leader RNA is tethered to mRNAs has been the subject of numerous studies.

Mouse hepatitis virus (MHV) contains a single-stranded, positive-sense RNA genome of 31 kb (8, 14). All of the subgenomic mRNAs of MHV are conested at the 3′ end of the RNA genome (11, 15) and share a 5′-end leader sequence of approximately 70 nucleotides (nt), which is derived from the 5′ end of genomic RNA (10, 13, 30). The 3′ end of the leader on the genomic RNA contains two to four copies of a pentanucleotide sequence (UCUAA) (20, 21). At the transcription start site for each mRNA is a stretch of intergenic sequence (IS), which contains a consensus UCUAAAC or similar sequence (29). The UCUAAAC sequence is necessary and sufficient for the initiation of mRNA synthesis (6). Fusion of the leader sequence with the mRNA body sequence usually occurs between the UCUAA repeats of the leader and this intergenic

* Corresponding author. Phone: (213) 342-1748. Fax: (213) 342-9555.

consensus sequence (1, 22). It thus has been hypothesized that mRNA initiation is mediated by direct RNA-RNA interaction between the leader RNA and intergenic consensus sequence (7, 8). The copy number of UCUAA at the 3′ end of the subgenomic mRNA leader may differ from that of the genomic RNA leader, presumably because of imprecise alignment between the leader and the intergenic regions (8, 22). The 5′ end of genomic RNA, including the leader, also acts as an enhancer-like element for mRNA synthesis (16). Furthermore, the sequence of IS significantly affects the pattern of transcriptional regulation. For instance, the efficiency of transcription initiation from the IS of mRNA2-1, which encodes the hemagglutinin-esterase (HE) protein, varies with the leader RNA, containing either two or three UCUAA copies (20, 28), while the IS of mRNA7, which encodes N protein, is constitutively transcribed. In addition, a 9-nt sequence (UUUAUAAAC) immediately downstream of the leader influenced the ability of the leader RNA to act as an initiator for mRNA transcription (33). Deletion of this sequence also prevented leader switching during RNA replication (21). Thus, coronavirus mRNA transcription is regulated by multiple regions of the genomic RNA sequences.

Coronavirus transcription can be viewed as a site-specific RNA recombination, since it involves interaction and fusion between homologous sequences in two distant and distinct RNA regions during transcription, similar to the copy choice mechanism of RNA recombination (9). High-frequency exchange of leader sequences between mRNAs of coinfecting MHVs (23) or between MHV and transfected DI RNA (21) suggested that recombination involving leader RNA occurred during both viral RNA transcription and replication. However, despite the high frequency of coronavirus RNA recombination, which approaches almost 25% for the entire genome (2, 9), only homologous and, in a rare case, nonhomologous recombinations (between coronavirus and possibly influenza C virus) (17) have been observed in coronaviruses so far (9). In contrast, many other RNA viruses undergo a special type (type II) of recombination (9), i.e., aberrant homologous recombination, in which two RNA molecules with similar sequences recombine, but crossovers occur not at the homologous or comparable sites but at unrelated sites on each parental RNA molecule. As a result, recombinant RNA contains sequence duplication, deletion, mutation, and, in some cases, even insertion of nucleotides of unknown origin (9 and references therein). This type of recombination has not been observed in coronaviruses so far.

During the course of studying MHV mRNA transcription with a DI RNA-reporter gene system (16, 33), we unexpectedly discovered that a particular MHV strain, which has a deletion of the 9-nt sequence (UUUAUAAAC) immediately downstream of the leader, generated very heterogeneous mRNAs from this DI RNA. In this study, we have now shown that these mRNAs have unusually heterogeneous leader fusion sites, which have never been seen in any other coronaviruses. The patterns of leader-mRNA fusion resemble aberrant homologous recombination seen in other RNA viruses, such as turnip crinkle virus (3). The occurrence of heterogeneity of leader-mRNA fusion is contingent on the sequences of both the leader RNA and IS. This study provides a unique paradigm for aberrant homologous recombination. The unexpected occurrence of leader-mRNA fusion at sites where there is no sequence homology also defies the previous transcription models, which imply direct interaction between complementary RNA sequences in the leader and IS. These results thus suggest that leader-mRNA fusion in coronavirus transcription does not require direct RNA-RNA interactions between complementary sequences. A modified model of RNA transcription and recombination based on protein-RNA and protein-protein interaction is proposed.

## MATERIALS AND METHODS

**Viruses and cells.** MHV strains JHM(3) and JHM2c(3) were used throughout this study. JHM(3) contains three UCUAA repeats at the 3′ end of the leader RNA (20). JHM2c(3) was derived from JHM2c(4), which is a small-plaque isolate of JHM derived from persistent infection of JHM in DBT cells (24) and contains four UCUAA repeats (12), by serial passages of JHM2c(4) in DBT cells (20); the plaque-purified JHM2c(3) contains three UCUAA repeats at the 3′ end of the leader of the genomic RNAs. Both JHM2c(4) and JHM2c(3) have a deletion of a 9-nt sequence (UUUAUAAAC) immediately downstream of the leader. The properties of these viruses have been described previously (20, 28). DBT cells (5) were used for infections and transfections.

**Plasmid construction.** Constructions of all plasmids used in this study were described previously (33). Their structures are illustrated in Fig. 1.

**RNA transcription and transfection.** Plasmid constructs were linearized with *Xba*I and transcribed in vitro with T7 RNA polymerase according to the manufacturer's recommended procedure (Promega). RNA transfection was carried out by the DOTAP method (Boehringer Mannheim) as described before (33).

**RT-PCR and cloning of subgenomic mRNAs.** Cytoplasmic RNAs were isolated from MHV-infected and RNA-transfected DBT cells at 8 h posttransfection by the Nonidet P-40 method (18) and used for cDNA synthesis by reverse transcription (RT) using a chloramphenicol acetyltransferase (CAT)-specific antisense primer, 3′CAT542 (33). For synthesis of cDNA specific for viral mRNA2-1, cytoplasmic RNAs were isolated from JHM2c(3)-infected DBT cells at 9 h postinfection by the same method and subjected to RT with an mRNA2-1-specific antisense primer, 3′HE74 (5′-ATGGAAT TCGATGTTAAGAGGTTCATTGA-3′, complementary to nt 54 to 74 of the HE open reading frame) (28). An additional sense primer specific to the leader (5′L9: nt 9 to 29) was used for subsequent PCR amplification of these cDNAs (33). The PCR fragments, which represent the 5′ portion of either CAT-containing subgenomic mRNA or viral mRNA2-1, were digested with *Sna*BI and *Eco*RI, and the *Sna*BI-*Eco*RI fragment (0.3 kb for CAT-containing mRNA and 0.15 kb for viral mRNA2-1) was cloned into the *Sma*I and *Eco*RI sites of pBluescript vector (Stratagene).

**Restriction endonuclease digestion analysis and agarose gel electrophoresis.** All pBluescript plasmid DNAs containing cloned cDNA fragments were digested with *Xba*I and *Eco*RI and analyzed directly by 1% agarose gel electrophoresis.

**DNA sequencing.** The sequences of the leader and the leader-body fusion sites of individual clones were determined by dideoxynucleotide chain termination sequencing (25) with primers 3′-CAT56 (5′-TTACGATGCCATTGGGATAT-3′, complementary to nt 37 to 56 from the beginning of the CAT open reading frame) for CAT-containing mRNAs and 3′HE74 for viral mRNA2-1.

## RESULTS

**Heterogeneity in the length of DI-derived subgenomic mRNA in JHM2c(3)-infected cells.** By using an MHV DI RNA containing the CAT gene behind an IS, we have previously shown that the leader sequences of the subgenomic mRNAs

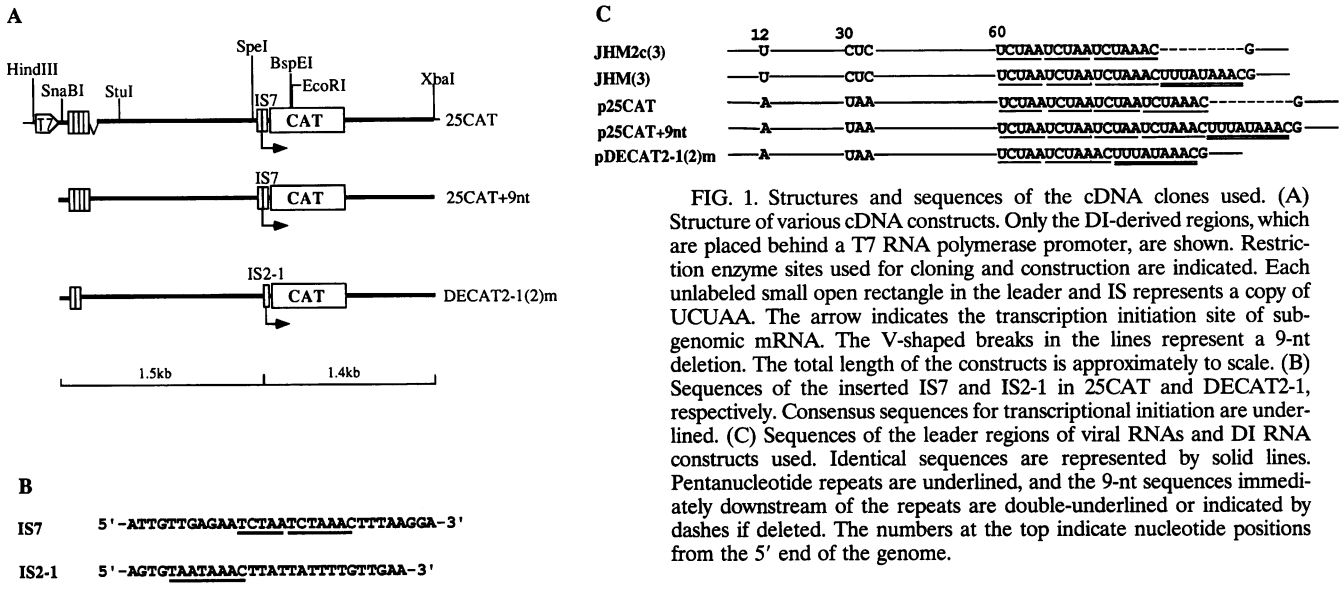


FIG. 1. Structures and sequences of the cDNA clones used. (A) Structure of various cDNA constructs. Only the DI-derived regions, which are placed behind a T7 RNA polymerase promoter, are shown. Restriction enzyme sites used for cloning and construction are indicated. Each unlabeled small open rectangle in the leader and IS represents a copy of UCUAA. The arrow indicates the transcription initiation site of subgenomic mRNA. The V-shaped breaks in the lines represent a 9-nt deletion. The total length of the constructs is approximately to scale. (B) Sequences of the inserted IS7 and IS2-1 in 25CAT and DECAT2-1, respectively. Consensus sequences for transcriptional initiation are underlined. (C) Sequences of the leader regions of viral RNAs and DI RNA constructs used. Identical sequences are represented by solid lines. Pentanucleotide repeats are underlined, and the 9-nt sequences immediately downstream of the repeats are double-underlined or indicated by dashes if deleted. The numbers at the top indicate nucleotide positions from the 5′ end of the genome.

transcribed from the DI RNA were derived both in *trans* from the helper viral RNA and in *cis* from the 5′-end leader of DI RNA. However, when the DI RNA contains a deletion of the 9-nt sequence (UUUAUAAAC) immediately downstream of the leader sequence, the leader RNA of the subgenomic mRNA was derived exclusively in *trans* from the helper viral RNA, suggesting that the DI leader in this sequence context is defective (33). We suggested the possibility that if the 9-nt sequence is absent, leader RNA synthesis could not be terminated properly, interfering with the ability of the leader RNA to initiate mRNA transcription (33). If this is the case, the structure of the leader RNA in the DI-derived subgenomic mRNAs would be of considerable interest when both the DI and helper virus RNAs lack the 9-nt sequence. We therefore studied the structure of the DI-derived, CAT-containing mRNAs in cells infected with a helper virus lacking the 9-nt sequence in its genomic RNA. Such a virus, JHM2c, has previously been described (12, 24). In this study, we used a derivative [JHM2c(3)] of the original JHM2c (20); the JHM2c(3) virus contains three copies of the UCUAA pentanucleotide but lacks the 9-nt sequence immediately downstream of the UCUAA repeats.

Cells were infected with either JHM2c(3) or JHM(3), the latter of which contains the 9-nt sequence, and transfected with various DI RNAs containing or lacking the 9-nt sequence (Fig. 1). Intracellular RNAs were isolated, and cDNAs were synthesized by RT with a CAT gene-specific antisense primer and amplified by PCR with a leader RNA-specific sense primer. The PCR products were cloned, and the inserts, which represent the 5′ end of the DI-derived, CAT-containing subgenomic mRNAs, including the leader sequence, were analyzed by restriction endonuclease digestion. As shown in Fig. 2, when JHM(3) was used as a helper virus, the 5′ ends of the CAT-containing subgenomic mRNAs transcribed from all DI RNA constructs were homogeneous, regardless of the presence or absence of the 9-nt sequence and the nature of IS in DI RNA (Fig. 2A, C, and E). Surprisingly, when JHM2c(3) was used as a helper virus, the sizes of the inserts were extremely heterogeneous (Fig. 2B, D, and F), varying as much as 200 nt in some cases, particularly when DI RNA contained IS2-1. The size heterogeneity of mRNAs was specific for JHM2c(3), which lacks the 9-nt sequence, as a helper virus and was not related to the presence or absence of the 9-nt sequence in DI RNAs. The extent of heterogeneity was influenced by IS sequences (either IS2-1 or IS7), i.e., it was more pronounced when IS2-1 was used. It should be noted that the size heterogeneity of subgenomic mRNAs has been reported previously for both the A59 and JHM strains of MHV (22). However, the previously observed mRNA heterogeneity represents only a variation in

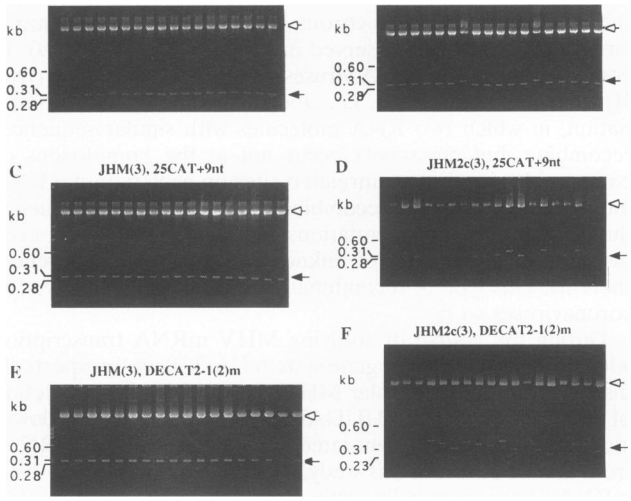


FIG. 2. Polymorphism of CAT-containing subgenomic mRNAs. cDNAs were amplified by RT-PCR with CAT- and leader-specific primers of subgenomic mRNAs from MHV-infected and DI RNA-transfected cells and cloned into the *Sma*I and *Eco*RI sites of pBluescript. Eighteen clones in each experiment were digested with *Eco*RI and *Xba*I and analyzed by agarose gel electrophoresis (1% agarose gel). DNA bands were visualized by staining with ethidium bromide. The left lane in each panel shows ϕX174 RF DNA *Hae*III fragments as size markers, and the sizes of some fragments are indicated (in kilobases). Solid arrows indicate bands of cDNA inserts, and open arrows indicate bands of the vector DNA. The helper virus and DI RNA used for infection and transfection, respectively, are indicated above each panel.

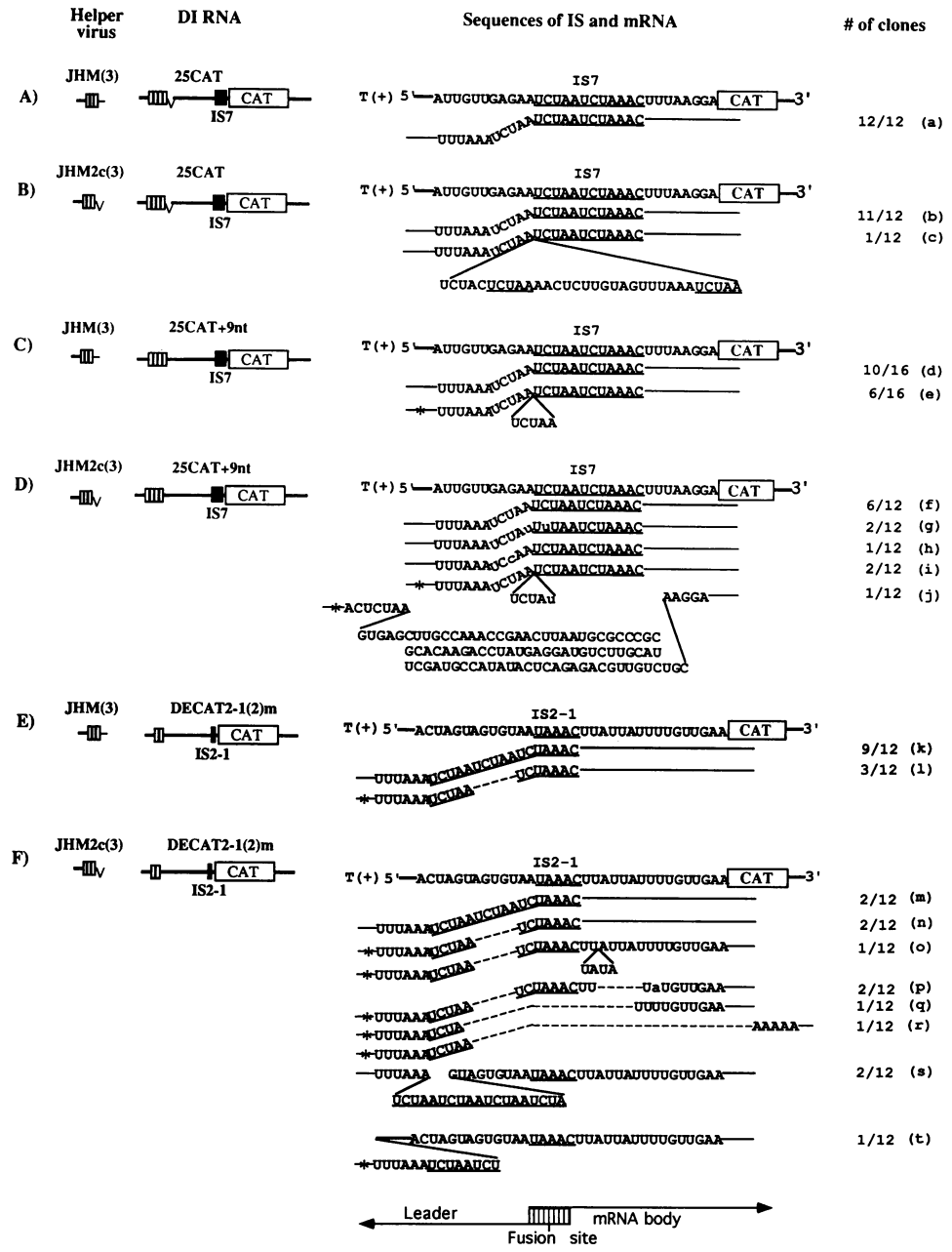


FIG. 3. Summary of sequences of cDNA clones as shown in Fig. 2. The structures of the helper virus and DI RNAs used are summarized on the left. The symbols are the same as in Fig. 1. The top line in each group represents the IS (IS7 or IS2-1) of template RNA in the plus sense [T(+)] and the CAT open reading frame. Below the template sequence are subgenomic mRNAs. Only the sequences around the leader-body fusion sites (underlined) are shown. The solid lines upstream and downstream of the fusion sites indicate sequences identical to the leader and template, respectively. Dashes indicate deletions. Nucleotide substitutions are indicated by lowercase letters, and insertions are shown by extra sequences. Asterisks indicate that the leader of the subgenomic mRNA was derived from DI RNA. Fractions of cDNA clones with the indicated sequence are shown on the right. The letters in parentheses identify individual cDNA clones.

the copy number of UCUAA, which would not have been detected in the present assay because of the large sizes of the PCR fragments analyzed here. The size differences among different cDNA clones in this report are considerably bigger. Thus, the heterogeneity seen here is a novel type of mRNA heterogeneity.

**Identification of the heterogeneity of leader-mRNA fusion sites.** Since transcription of coronavirus subgenomic mRNAs has been proposed to result from interactions between the leader and ISs (8), the size heterogeneity of the 5′ ends of subgenomic mRNAs was most likely due to the heterogeneity of the leader-mRNA fusion sites, possibly caused by imprecise leader-mRNA fusion. Therefore, to examine the nature of subgenomic mRNA heterogeneity, we determined the sequence of each individual cDNA clone, which represents the 5′-end sequence of subgenomic CAT-containing mRNAs tran-

scribed from IS. At least 12 representative clones from each combination of DI RNA and helper virus were sequenced (Fig. 3). The results showed that in JHM(3)-infected, 25CAT-transfected cells, all of the subgenomic mRNAs were identical, containing a leader RNA derived exclusively from the helper virus RNA (Fig. 3A). In JHM(3)-infected cells which had been transfected with the other two DI constructs, mRNAs were slightly heterogeneous despite the apparent homogeneity of their size (Fig. 2C and E). These mRNAs contained leader RNAs derived from both DI and helper viral RNAs. Since the number of UCUAA repeats in the leaders of the subgenomic mRNAs was identical to that of the leaders of the origin, the fusion between the leader and IS probably occurred between the last UCUAA copy of the leader and the consensus IS of the template (Fig. 3A, C, and E). This result is consistent with the previous finding (33) that, when DI RNA has a deletion of the 9-nt sequence, all of the leader sequences of the subgenomic mRNAs were derived from the helper virus (in *trans*) (Fig. 3A); in contrast, when DI RNA contains the 9-nt sequence, both the helper virus and DI RNA contributed leader sequences to subgenomic mRNAs (Fig. 3C and E). These results together suggest that the 9-nt sequence is essential for the ability of the leader RNA to initiate mRNA synthesis.

In contrast to the JHM(3)-infected cells, the DI-derived subgenomic mRNAs in JHM2c(3)-infected cells showed various novel types of leader-mRNA fusion sites (Fig. 3B, D, and F). These were detected in all three different DI RNA constructs but were particularly evident in the DI containing an IS2-1 (Fig. 3F). For example, three clones from JHM2c(3)-infected and DECAT2-1(2)m-transfected cells contained extensive deletions (Fig. 3F, clones p, q, and r), suggesting that the leader was fused to the sequence downstream of the consensus IS, whereas two other clones (clones s and t) contained the template sequence upstream of the IS, indicating that fusions occurred upstream of the consensus IS. There was no obvious sequence homology between the leader and template around these fusion sites. Another clone from JHM2c(3)-infected and 25CAT-transfected cells had a duplication of 32 nt of the leader sequence (nt 33 to 64) around the fusion site (Fig. 3B, clone c), and another contained both a deletion of 35 nt from the 3′ end of the leader sequence and an insertion of 93 nt, which is derived from a sequence approximately 300 nt upstream of the IS (Fig. 3D, clone j). In two clones (clones o and p), leader-mRNA fusion occurred not between the UCUAA repeats of the leader and the consensus IS but rather between the 9-nt sequence and the sequence downstream of the consensus IS. Thus, in these cases, the 9-nt sequence (UUUAUAAAC) or part of it was incorporated into subgenomic mRNAs. Most surprisingly, three other clones (clones g, h, and i) (Fig. 3D) had one or two mutations within the pentanucleotide sequence (UCUAU, UUUAA, or UCC AA instead of UCUAA), which is different from either the IS or leader RNA of the genomic RNA, suggesting that these sequences were derived by misincorporation of nucleotides or by nontemplated transcription. Taken together, these sequence data showed that the observed size heterogeneity of subgenomic mRNAs was the consequence of aberrant binding of the leader to template RNA. It should be noted that the leader RNA of subgenomic mRNAs was derived exclusively from the helper viral RNA (in *trans*) when neither RNA contained the 9-nt sequence (Fig. 3B).

**JHM2c(3) subgenomic mRNA2-1 also has heterogeneous leader-mRNA fusion sites.** The above results demonstrated that when JHM2c(3) was used as a helper virus, subgenomic mRNAs, particularly those transcribed from IS2-1 in DI RNA, were extremely heterogeneous in leader-mRNA fusion site. To rule out the possibility that this heterogeneity was the artifact of DI RNA, we examined whether virus-specific subgenomic mRNAs in JHM2c(3)-infected cells are also heterogeneous. JHM2c has been shown to synthesize all seven MHV-specific mRNAs plus several minor mRNAs, one of which corresponds to mRNA2-1 (24). We chose mRNA2-1 for this study because the results above demonstrated that transcription from IS2-1 was much more heterogeneous than that from IS7. Cytoplasmic RNA was isolated from JHM2c(3)-infected cells. cDNAs representing the 5′ end of subgenomic mRNA2-1 were synthesized by RT with an antisense primer specific to mRNA2-1 and amplified by PCR with an additional sense primer specific to the leader. Although mRNA2-1 is a minor RNA species in JHM2c(3)-infected cells (24) (data not shown), RT-PCR did detect a PCR product derived from mRNA2-1. The PCR products were cloned, and the size of the inserts in individual clones was analyzed after restriction endonuclease digestion. Restriction fragment polymorphism of the PCR products, similar to that of subgenomic CAT-containing mRNAs, was detected (Fig. 4), indicating that mRNA2-1 of JHM2c(3) is heterogeneous. Thus, the heterogeneity seen among the CAT-containing mRNAs transcribed from the IS2-1 of DI RNA most likely represents the heterogeneity of viral mRNAs during natural infections.

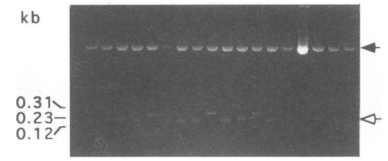


FIG. 4. Polymorphism of subgenomic mRNA2-1 of JHM2c(3). cDNAs of mRNA2-1 were amplified by RT-PCR of RNA from JHM2c(3)-infected cells and cloned into pBluescript. Eighteen clones from miniprep plasmid DNA were digested with *Eco*RI and *Xba*I and analyzed by agarose gel electrophoresis. In the left lane are φX174 RF DNA *Hae*III fragments, which are shown as size markers. The solid arrow indicates the band of vector DNA, and the open arrow indicates the bands of cDNA inserts.

The nature of the heterogeneity was then determined by sequencing. Seventeen clones were sequenced. Sequence data indicated that the leader-body fusion sites in mRNA2-1 were highly heterogeneous (Fig. 5). Nine of the clones have sequences suggesting a correct binding between the leader and template RNA within the UCUAA repeats; however, they contained two copies of UCUAA at the 3′ end of the leader, in contrast to three copies in the viral genomic RNA leader. One of the clones had a mutation (CCUAA instead of UCUAA) which is not present in either the IS or leader RNA on the genomic RNA (clone v). In other clones, leader RNA appears to fuse at distant sites either upstream (clone z) or downstream (clones w and x) of the consensus IS. Some of the clones contained extensive insertions or deletions of the template sequences around the normal fusion sites (clones z′, w, x, and y), suggesting that leader-mRNA fusions occur either downstream of the IS or upstream of the UCUAA repeat within the leader. Two clones contained an insertion of exogenous nontemplated sequences (CAG in clone x and ----GUGGUC CCC in clone z′) near the leader fusion site. These results together indicate that mRNA2-1 of JHM2c(3) has very heterogeneous leader-mRNA fusion sites.

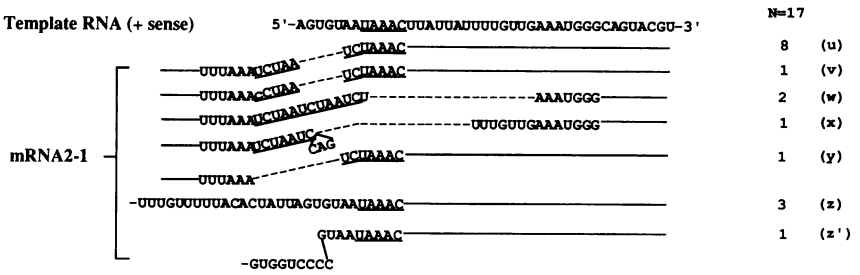


FIG. 5. Summary of leader fusion sequences of cDNA clones of JHM2c(3) mRNA2-1. IS2-1 is shown in the plus sense at the top. Depictions of different cDNA clones are the same as explained in the legend to Fig. 3. The total number of clones sequenced is 17; the numbers on the right represent the number of clones with the indicated sequence. The leader fusion site of clone z is approximately 70 nt upstream of the IS. Clone z' contains 15 nt of unknown origin upstream of the IS.

## DISCUSSION

In this study, we have demonstrated a very unusual case of heterogeneity of leader-mRNA fusion during coronavirus mRNA transcription. Previous studies have shown that the leader RNA always fuses with the subgenomic mRNA within the region homologous between the 3' end of the leader (UCUAA repeats) and the consensus sequence (UCUAAAC) of the IS region. This is the basis for the proposed leader-primed transcription mechanism and other discontinuous-transcription models, all of which putatively involve direct RNA-RNA interactions between these homologous or complementary sequences (8). In this study, we showed that an MHV strain, JHM2c(3), synthesized mRNA species with a whole array of heterogeneous leader-mRNA fusion sites (Fig. 6), which can be grouped into five types. The first type (type I) is the classical leader-mRNA fusion, in which the leader is fused at a site within the pentanucleotide repeats with the consensus IS of the template RNA. Any of the several UCUAA repeats in the leader RNA could be utilized for fusion, resulting in different numbers of UCUAA copies in the leaders of subgenomic mRNAs. In type II, the leader is fused precisely with the consensus IS of the template RNA, but the leader of the mRNA contains some nucleotide alterations within the UCUAA repeats, suggesting some modifications of the leader sequence during subgenomic mRNA transcription. In type III, the leader is fused with mRNAs at a site either upstream or downstream of the consensus IS. The sequences at or around the fusion sites bear little or no homology with the leader. As a result, mRNAs contain additional sequences upstream or have a sequence deletion downstream of the consensus UC UAAAC. In type IV, the leader is fused to the IS not by its UCUAA repeats but rather by the 9-nt sequence (UUUAU AAAC) immediately downstream, indicating that the 9-nt sequence could also provide fusion sites. These clones thus contain a longer leader sequence. In type V, the leader-mRNA fusion site contains a duplication of a portion of the leader sequence or an insertion of nontemplated sequences which are not present in either the leader or template RNA. The origin of these sequences is unknown.

Type I leader fusion conforms to the classical model of the leader-primed transcription of coronavirus (7, 8). However, other types of leader fusion involve either fusion outside the homologous regions (types III and IV) or addition of or substitution with nontemplated nucleotides (types II and V). These unusual leader fusion events thus defy the basic premises of the previously proposed leader-primed transcription mechanism and other discontinuous-transcription models, which postulate direct RNA-RNA interactions between complementary sequences. These results suggest that the leader-mRNA fusion may not be the result of direct RNA-RNA annealing between the complementary sequences in the leader and template RNAs. Nevertheless, since most of the fusion sites are located near the IS, binding between the leader RNA and IS likely represents the first step of mRNA transcriptional initiation. Subsequent RNA realignment likely occurs before

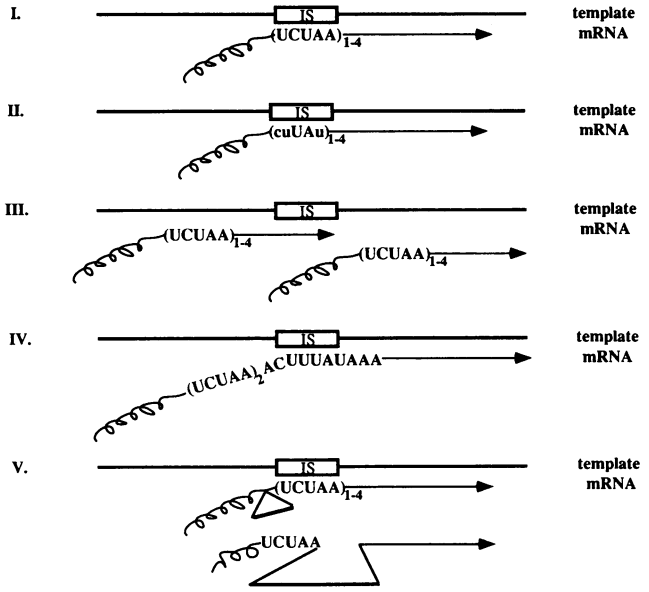


FIG. 6. Diagrammatic representation of the types of leader-mRNA fusion observed. IS indicates the consensus IS in the template RNA (thick line). The helix represents the leader of subgenomic mRNAs. The copy number of the UCUAA sequence at the 3' end of the leader varies from 1 to 4. The thin-lined arrow indicates the direction of mRNA transcription, and the body of subgenomic mRNA, which is complementary to the template RNA sequence. Type I: The leader is fused at the UCUAA repeats precisely with the consensus IS. The copy number of the UCUAA sequence at the 3' end of subgenomic mRNA varies from 1 to 4. Type II: The leader is fused at the UCUAA repeats with the consensus IS, but some nucleotide substitutions occur within the UCUAA sequence of the leader. Lowercase letters indicate nucleotide substitutions. Type III: The leader is fused at the UCUAA repeats with sequences upstream or downstream of the consensus IS. Type IV: The leader is fused at the 9-nt sequence downstream of the UCUAA repeats with the consensus IS or nearby sequences. Type V: The leader is fused with the IS at the normal UCUAA repeats or at an internal UCUAA sequence (nt 45 to 50) within the leader. Insertions or substitutions of sequences of unknown origin occur near the leader-mRNA fusion site.

transcriptional initiation. Alternatively, the IS may only provide a recognition signal for RNA polymerase. For example, polymerase and other viral or cellular transcription factors may form a transcription complex which interacts with the IS (promoter). Depending on the nature of RNA polymerase or other viral proteins in the transcription complex, interaction between the transcription complex and RNA may not be precise. The leader RNA then initiates RNA transcription inside this transcription complex. The presence of nontemplated and substituted nucleotides in the mRNA around the leader fusion sites further suggests that mRNA initiation did not involve tight binding of the leader RNA to the template RNA. Thus, the driving force for coronavirus mRNA initiation may not be direct RNA-RNA interaction between the leader and IS. Instead, it may be the RNA-protein interactions involving the recognition of the promoter sequence (IS) by cellular proteins or the putative RNA polymerase (33). A similar model has recently been suggested (31).

This kind of leader-mRNA fusion event is reminiscent of the type II aberrant homologous recombination (9) seen in turnip crinkle virus (3). In this virus, most recombinations occur near the replicase recognition site of the viral RNA, suggesting that recombination is a replicase-driven event (3). This type of recombination also causes insertion of nontemplated nucleotides at crossover sites. Mechanistically, coronavirus RNA transcription is similar to the copy choice mechanism of RNA recombination, which involves the jumping of a nascent RNA transcript to a new RNA template. Similarly, in mRNA transcription, a leader RNA jumps in *trans* from the 3′ end of template RNA to an IS located at a distant site or on a different RNA molecule. This process may be driven by the recognition of the IS by RNA polymerase in either the transcription or recombination process. Therefore, coronavirus mRNA transcription can be considered a site-specific recombination. The current study demonstrated that the occurrence of aberrant leader-mRNA fusion is dependent on the sequences of the leader and the IS (template) and is also influenced by the virus strain. Thus, coronavirus mRNA transcription may serve as a model for studying RNA recombination.

Why is leader fusion so aberrant in JHM2c(3)? One possible mechanism is that the 9-nt sequence forms part of the transcription termination signal for the leader RNA. Thus, when this sequence is missing, the amount of free leader RNA is decreased or leader RNAs of aberrant sizes are generated, leading to aberrant leader-IS interactions. Indeed, previous studies have shown that when DI RNA lacked this 9-nt sequence, the leader RNA from the DI RNA was not used to initiate mRNA synthesis; instead, most of the leader RNAs in the mRNAs were derived in *trans* from the helper viral RNA (33). In this study, in which more DI constructs were studied, the same trend of leader utilization was observed, although there were some exceptions to this general rule. Another possible mechanism is that the transcription of subgenomic mRNA may be regulated by this 9-nt sequence as well as the leader RNA at the 5′ end of the genome and that the lack of this 9-nt sequence in JHM2c(3) RNA might have contributed to the aberrant leader-mRNA fusion. We propose that the 9-nt sequence forms a secondary structure which is involved in the interaction with proteins of the transcription initiation complex. Deletion of the 9-nt sequence may result in alteration of its binding affinity for components of the transcription complex, such as RNA polymerase or cellular proteins, and consequently a decrease in the efficiency of subgenomic mRNA transcription (33). It is significant that the 9-nt sequence is not part of the leader RNA which is eventually incorporated into the mRNAs (7, 8). Therefore, the effect of the 9-nt sequence on mRNA transcription is not due to its priming function. Furthermore, the heterogeneity of leader-mRNA fusion was most pronounced at the IS2-1 site. Thus, the 9-nt sequence most likely interacts with IS either directly or indirectly through protein-protein or protein-RNA interactions. Several cellular proteins have recently been shown to bind to the 5′ end of MHV genomic RNA (4). Some of the same proteins also bind to the IS (32). It is tempting to suggest that these proteins are part of the transcription complex, which is responsible for bringing the promoter (IS) and leader RNA together and initiates mRNA transcription. Perhaps because of the altered protein-binding properties of the JHM2c(3) leader RNA, the transcription complex may not properly bring the leader RNA to the correct transcriptional initiation site.

Recently, we have suggested that protein-RNA and protein-protein interactions rather than direct RNA-RNA interactions are the driving force for the initiation of transcription (33). The identification of the unusual heterogeneity in leader-body fusion in the present report strongly supports this notion. However, since complete sequencing of the JHM2c(3) RNA genome has not been performed, we cannot rule out the possibility that the JHM2c(3) RNA polymerase may have special properties which affect the properties of the transcription complex. Elucidation of the RNA structure and nature of polymerase of JHM2c(3) may thus shed further light on the mechanism of coronavirus RNA transcription and recombination.

## ACKNOWLEDGMENTS

We thank Ching-Len Liao for providing construct 25CAT.

This work was supported by Public Health Service Research grant AI16144 from the National Institutes of Health. X.Z. is supported by a postdoctoral training fellowship from NIH. 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. **Baric, R. S., K. Fu, M. C. Schaad, and S. A. Stohlman.** 1990. Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. Virology **177**:646–656.
3. **Cascone, P. J., C. D. Carpenter, X. H. Li, and A. E. Simon.** 1990. Recombination between satellite RNAs of turnip crinkle virus. EMBO J. **9**:1709–1715.
4. **Furuya, T., and M. M. C. Lai.** 1993. Three different cellular proteins bind to complementary sites on the 5′-end-positive and 3′-end-negative strands of mouse hepatitis virus RNA. J. Virol. **67**:7215–7222.
5. **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.
6. **Jeong, Y. S., and S. Makino.** 1992. Mechanism of coronavirus transcription: duration of primary transcription initiation activity and effects of subgenomic RNA transcription on RNA replication. J. Virol. **66**:3339–3346.
7. **Lai, M. M. C.** 1986. Coronavirus leader-primed transcription: an alternative mechanism to RNA splicing. BioEssays **5**:257–260.
8. **Lai, M. M. C.** 1990. Coronavirus: organization, replication and expression of genome. Annu. Rev. Microbiol. **44**:303–333.
9. **Lai, M. M. C.** 1992. RNA recombination in animal and plant viruses. Microbiol. Rev. **56**:61–79.
10. **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.
11. **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.

12. **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.
13. **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.
14. **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.
15. **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.
16. **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.
17. **Luytjes, W., P. J. Bredenbeek, A. F. H. Noten, M. C. Horzinek, and W. J. M. Spaan.** 1988. Sequence of mouse hepatitis virus A59 mRNAs: indications for RNA recombination between coronaviruses and influenza C virus. Virology **166:**415–422.
18. **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.
19. **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.
20. **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:**227–32.
21. **Makino, S., and M. M. C. Lai.** 1989. High-frequency leader sequence switching during coronavirus defective interfering RNA replication. J. Virol. **63:**5285–5292.
22. **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.
23. **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.
24. **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.
25. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
26. **Sawicki, S. G., and D. L. Sawicki.** 1990. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. J. Virol. **64:**1050–1056.
27. **Sethna, P. B., S. L. Hung, and D. A. Brian.** 1989. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. Proc. Natl. Acad. Sci. USA **86:**5626–5630.
28. **Shieh, C. K., H. J. Lee, K. Yokomori, M. N. La, 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.
29. **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.
30. **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.
31. **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.
32. **Zhang, X. M., and M. M. C. Lai.** 1994. Unpublished observation.
33. **Zhang, X., 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.