JOURNAL OF VIROLOGY, Nov. 1992, p. 6330–6337
0022-538X/92/116330-08$02.00/0

Vol. 66, No. 11

Mutagenic Analysis of the Coronavirus Intergenic Consensus Sequence

MYUNGSOO JOO AND SHINJI MAKINO*

Department of Microbiology, The University of Texas at Austin, Austin, Texas 78712

Received 19 May 1992/Accepted 28 July 1992

Previously, a system in which an intergenic region from mouse hepatitis virus (MHV) inserted into an MHV defective interfering (DI) RNA led to transcription of a subgenomic DI RNA in helper virus-infected cells was established. In the present study, a DI cDNA containing one UCUAAAC consensus sequence in the middle of the 0.3-kb-long intergenic region located between genes 6 and 7 was constructed. From this DI cDNA clone, 21 mutant DI RNAs were constructed so that each of the seven consensus sequence nucleotides was changed individually to the three alternative bases. These mutants were used to define how changes in the integrity of MHV transcription consensus sequence UCUAAAC affected mRNA transcription. Except for two mutants with the sequences UGUAAAC and UCGAAAC, all of the mutants supported efficient subgenomic DI RNA transcription. This indicated that MHV transcription regulation was sufficiently flexible to recognize altered consensus sequences. Next, these and other mutants were used to examine the leader-body fusion site on the subgenomic DI RNAs. Sequence analysis demonstrated that all subgenomic DI RNAs analyzed contained two pentanucleotide sequences; the first sequence seemed to be contributed by the leader, and the leader-body fusion most likely took place at either the first or the second nucleotide of the second sequence. This observation was not consistent with the proposed coronavirus transcription model (S. C. Baker and M. M. C. Lai, EMBO J. 9:4173–4179, 1990) which states that nucleotide mismatch can be corrected by RNA polymerase proofreading activity.

Mouse hepatitis virus (MHV), a coronavirus, is an enveloped virus containing a single-stranded, positive-sense RNA genome of approximately 31 kb (14, 16, 28). In MHV-infected cells, seven or eight species of virus-specific subgenomic mRNAs constituting a 3′-coterminal nested set (12, 17) are synthesized, and these subgenomic mRNAs are named mRNAs 1 to 7, in decreasing order of size (12, 17). Among these mRNA species, only mRNA 1 contains a packaging signal and is efficiently packaged into MHV virions (6, 26, 44); the remaining mRNAs are not packaged (14, 22). The 5′ end of the MHV genomic RNA contains a 72- to 77-nucleotide-long leader sequence (11, 13, 42). The 3′ region of the leader sequence contains a pentanucleotide sequence, UCUAA, which is repeated two to four times in different MHV strains (23). Downstream of the leader sequence are the MHV-specific genes, which are separated by special short stretches of sequence termed intergenic sequences. Each of the MHV intergenic sequences, which are located upstream of genes essential for MHV replication, includes the unique consensus sequence UCUAAAC or a very similar sequence (38). A sequence identical to the 5′-end genomic leader sequence is also found at the 5′ end of every MHV mRNA species. These leader sequences are fused with the mRNA body sequence which starts from the intergenic consensus sequence (11, 13, 38, 42). In most MHV genes there is a correlation between the degree of intergenic-sequence nucleotide homology with the 5′-end genomic leader sequence and the amount of mRNA transcribed (38). This correlation is not observed in infectious bronchitis virus mRNA transcription (5). The site where the leader fuses with the mRNA is located somewhere within the repeated pentanucleotide (UCUAA), and the number of repeats in each given mRNA varies (23). Because the pentanucleotide repeats at the 5′-end genomic leader sequence and at the intergenic region are the same sequence, the precise leader-body fusion site on MHV mRNA has not been identified.

It has been clearly demonstrated that there are at least two stages in coronavirus subgenomic RNA synthesis: one is primary transcription, in which subgenomic-size RNA is synthesized from the genomic-size template RNA, and the other is secondary transcription, in which the subgenomic-size RNA serves as a template (9, 18). The polarity of the template RNA has not been conclusively demonstrated for either transcription mechanism. Several models explaining how subgenomic RNA is synthesized have been proposed. One model is leader RNA-primed transcription, which proposes that a free leader RNA is transcribed from the 3′ end of the genomic-size, negative-strand template RNA, dissociates from the template, and then rejoins the template RNA at downstream intergenic regions to serve as the primer for mRNA transcription (3, 10). Another model, put forth by Sawicki and Sawicki (33), suggests that subgenomic negative-strand RNAs are initially synthesized from the input genomic RNA. Then, the positive-strand subgenomic RNA is synthesized on the subgenomic-size, negative-strand RNA during secondary transcription. The leader sequence on the subgenomic-size RNA may be acquired during primary transcription or during secondary transcription. Another possible mechanism may be that leader RNA joins the subgenomic RNA body during primary transcription by a mechanism similar to RNA splicing (11, 42). To date, none of these models has been conclusively proven or disproven. The mechanism of coronavirus transcription remains to be elucidated.

Previously, a system that exploits defective interfering (DI) RNAs of MHV for studying the mechanisms of coronavirus mRNA transcription was established (18). In this

* Corresponding author.

system, a complete MHV DI cDNA clone containing an inserted intergenic region, derived from between genes 6 and 7 of the genome, was constructed. Replication of genomic DI RNA as well as transcription of subgenomic DI RNA was observed after transfection of in vitro-synthesized DI RNA into MHV-infected cells. Analysis of a series of mutants with deletions in the intergenic region demonstrated that the sequences flanking the consensus sequence of UCUAAAC affected the efficiency of subgenomic DI RNA transcription and that the consensus sequence was necessary though not sufficient for the synthesis of the subgenomic DI RNA (18). Some MHV intergenic regions do not contain the UCU AAAC consensus sequence but contain a slightly different sequence, UCUAUAC or UCCAAAC. It is not known how plastic the consensus sequence is with respect to transcription.

In the present study, a series of mutant DI RNAs with mutated nucleotides in the consensus sequence was constructed in order to examine how nucleotide substitution in the consensus sequence affects mRNA transcription. Surprisingly, the MHV transcription mechanism is sufficiently flexible to recognize a mutated consensus sequence for mRNA transcription. Furthermore, by using these mutant DI RNAs, the leader-body fusion site on subgenomic DI RNA was examined. Possible coronavirus transcription regulation mechanisms are discussed.

MATERIALS AND METHODS

Viruses and cells. The plaque-cloned A59 strain of MHV (MHV-A59) (12) was used as a helper virus. Mouse DBT cells (8) were used for growth of viruses.

DNA construction. A procedure based on recombinant polymerase chain reaction (PCR) was employed for site-directed mutagenesis (7). MHV DI cDNA clone PR6 mutant 9, which contains one repeat of TCTAAAC, was used as template DNA for the construction of mutant DI cDNAs (18). All mutants were constructed under the following PCR conditions: plasmid DNA was incubated with two oligonucleotides in PCR buffer (0.05 M KCl, 0.01 M Tris-hydrochloride [pH 8.0], 0.0025 M $MgCl_2$, 0.01% gelatin, 0.17 mM each deoxynucleoside triphosphate, and 5 U of *Taq* polymerase [Perkin-Elmer Cetus]) at 93°C for 30 s, 37°C for 45 s, and 72°C for 100 s for a total of 25 cycles. The first round of PCR consisted of two separate reactions; the resultant two sets of products were then mixed for the second round of PCR. Oligonucleotide 1189 (5′-GTTGGATATCTGCTTGGGC-3′), which contains an *Eco*RV site and hybridizes 732 to 750 nucleotides from the 3′ end of PR6 mutant 9, and oligonucleotide 1115 (5′-TCTAGCACGTGGCACTA-3′), which hybridizes 1,060 to 1,077 nucleotides from the 3′ end of PR6 mutant 9, were used as the outside oligonucleotides for the second round of PCR. For the initial constructions, the first PCR products were obtained with oligonucleotide 1178 (5′-GTTTAGATTCTCAACAAT-3′), which hybridizes 902 to 918 nucleotides from the 3′ end of PR6 mutant 9, and oligonucleotide 1115. Oligonucleotides used in combination with oligonucleotide 1189 were synthesized to contain degenerated bases at the target position in order to obtain an array of mutants. All these oligonucleotides hybridize to a region 902 to 916 nucleotides from the 3′ end of PR6 mutant 9. For the construction of MJΔWT, MJΔU1A, MJΔU1G, and MJΔU1C, oligonucleotide 1241 (5′-CGCATTGTTG AGTCTAA-3′), which hybridizes 903 to 921 nucleotides from the 3′ end of PR6 mutant 9, was used in combination with oligonucleotide 1189, and oligonucleotide 1115 was used with the oligonucleotides containing degenerated bases at the target position. Each final PCR product was then digested with *Kpn*I and *Eco*RV, and the 0.27-kb *Kpn*I-*Eco*RV PCR fragment was cloned into the 6.0-kb *Kpn*I-*Eco*RV fragment of PR6 mutant 9 (18). The *Kpn*I-*Eco*RV region of each clone was sequenced to confirm the presence of the specific mutation and the absence of other mutation sites.

RNA transcription and transfection. Plasmid DNAs were linearized by *Xba*I digestion and transcribed with T7 RNA polymerase as previously described (21). The lipofection procedure was used for RNA transfection as previously described (18).

Preparation of virus-specific intracellular RNA and Northern (RNA) blotting. Virus-specific RNAs in virus-infected cells were extracted as previously described (25). For each sample, 1.5 μg of intracellular RNA was denatured and electrophoresed through a 1% agarose gel containing formaldehyde and separated RNA was blotted onto nylon filters as described previously (18). The nylon filter was soaked in a prehybridization buffer, and Northern blot hybridization was performed (6, 9). The gel-purified 0.25-kb *Nru*I-*Msc*I fragment of DF1-2 (26), which corresponds to a region 18 to 262 nucleotides from the 3′ end of the MHV DI cDNA, was labeled with ^{32}P according to the random-priming procedure (31) and used as a probe.

Direct sequencing of the PCR product. The PCR products were separated by agarose gel electrophoresis and recovered from the gel slices by using GeneClean II (Bio 101, La Jolla, Calif.). Direct PCR sequencing was performed according to the procedure established by Winship (46). Briefly, the gel-purified PCR products were incubated at 100°C in 10 μl of Sequenase buffer (U.S. Biochemicals) containing a primer and 10% dimethyl sulfoxide. After 3 min of incubation, the samples were quickly placed on dry ice. The sequencing reactions were done with Sequenase (U.S. Biochemicals); 0.5 μl of dimethyl sulfoxide and 2.5 μl of each termination mix were included in the termination reaction mixture.

RESULTS

Effects of nucleotide substitutions within the consensus sequence on subgenomic DI RNA transcription. The effects of nucleotide substitutions within the conserved UCUAAAC consensus sequence on MHV subgenomic DI RNA synthesis were studied. An MHV DI cDNA clone, MJWT, containing a TCTAAAC sequence in the middle of the intergenic region located between genes 6 and 7 was constructed from MHV cDNA clone PR6 mutant 9 (18). MJWT also contained the wild-type regions flanking the intergenic consensus sequence, 0.1 kb upstream and 0.17 kb downstream. Twenty-one MJWT-derived mutants, each with a specific nucleotide substitution within the UCUAAAC sequence, were constructed. These mutants were named according to the site of mutation in the UCUAAAC consensus sequence; e.g., in MJU3G, G replaced U at the third nucleotide of the consensus sequence. Plasmid DNAs were linearized with *Xba*I and transcribed by T7 RNA polymerase in the presence of a cap analog [$m^7G(5')ppp(5')G$] (21), and the resulting RNAs were transfected by lipofection into monolayers of DBT cells infected with MHV-A59 helper virus 1 h prior to transfection (21). After incubation of virus-infected cells at 37°C for 16 h, the culture fluid was harvested and the cell debris was removed by low-speed centrifugation. This sample was named passage 0. Normally, subsequent passage of virus samples obtained from DI RNA-transfected cells results in

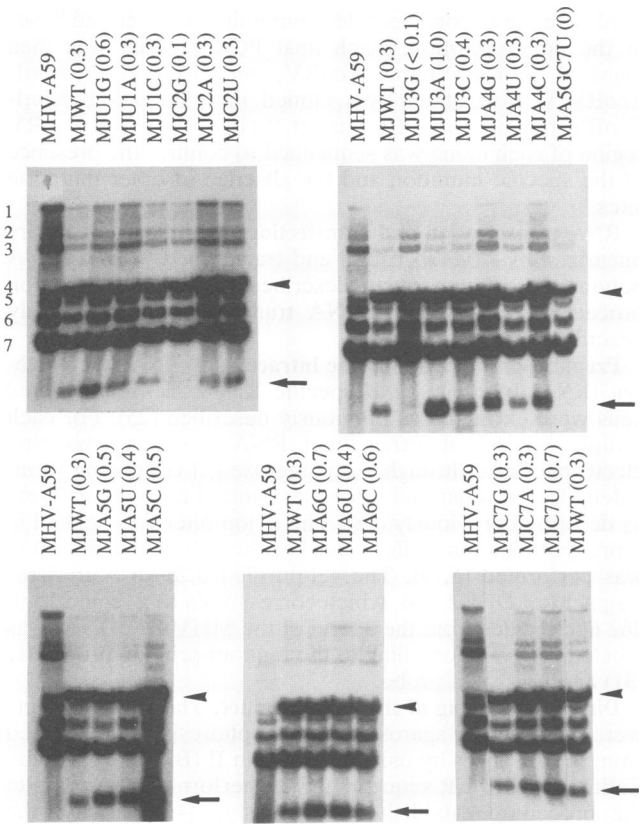

FIG. 1. Northern blot analysis of MJWT-derived mutant subgenomic DI RNAs. Passage 1 virus samples were used as inocula. Intracellular RNAs were extracted 7 h postinfection, separated by formaldehyde–1% agarose gel electrophoresis, and transferred to a nylon membrane. The probe was prepared by random-primed ^{32}P labeling of the MHV-specific cDNA fragment corresponding to the 3′ region of MHV genomic RNA. 1 to 7, MHV-A59-specific mRNA species. Arrowheads and arrows, genomic and subgenomic DI RNAs, respectively. The molar ratios of genomic DI RNA to subgenomic DI RNA are shown in parentheses.

efficient amplification of DI particles (26); for this reason, the virus was further passaged to generate passage 1 virus samples. The passage 1 virus sample harvested after 16 h of culture was used as the inoculum for the analysis of intracellular RNA species.

Virus-specific intracellular RNA was extracted at 7 h postinfection and analyzed by Northern blotting with a probe which specifically hybridizes with all MHV RNAs (Fig. 1). The number of negative-strand coronavirus RNA species is much lower than that of positive-strand coronavirus RNA species (32); therefore, the signal obtained from this Northern blot analysis represented mostly positive-strand RNA species. The molar ratio of subgenomic DI RNA to the genomic DI RNA was determined by densitometric scanning of the autoradiogram (Fig. 1).

Northern blot analysis demonstrated that many of the single-nucleotide substitutions in the consensus sequence did not abolish subgenomic DI RNA transcription. Furthermore, except for MJC2G and MJU3G, the mutants supported at least the same level of subgenomic DI RNA synthesis as MJWT. Only a minute amount of subgenomic DI RNA was synthesized from MJU3G genomic DI RNA, and sometimes subgenomic DI RNA was not detectable even after amplification of subgenomic DI RNA by PCR (data not shown). This analysis indicated that the substitution of G for wild-type nucleotides at the second and third positions of the consensus sequence had a negative effect on MHV RNA transcription. The following mutants demonstrated increased subgenomic DI RNA synthesis compared with that of MJWT: MJU1G, MJU3A, MJU3C, MJA5G, MJA5U, MJA5C, MJA6G, MJA6U, MJA6C, and MJC7U. This observation was particularly unexpected for those mutants with G substitutions, because none of the naturally occurring transcriptionally functional intergenic consensus sequences contained G (Fig. 2). During isolation of mutant DI cDNAs, a double mutant, MJA5GC7U, with nucleotide substitutions at nucleotides 5 and 7 was isolated. As shown in Fig. 1, MJA5G and MJC7U supported efficient subgenomic DI RNA synthesis, whereas no subgenomic DI RNA synthesis was observed in the MJA5GC7U-replicating cells.

Leader sequence	UUAAAUCUAA	UCUAAAC	UUUAUAAACG		
1	AGGAGGUUCA	UCUUAAC	CCGAGGGGUA	Gene 1	895-902
2	ACUGGCGUUG	UCUAAAG	UGUGGCAUGG	Gene 1	3,846-3,872
3	GCAUGUGAAA	UGUAAAC	CCAAGUACCA	Gene 1	5,887-5,913
4	GUUGAGGGUG	UCUACAC	UAAUUUUAAA	Gene 1	6,143-6,169
5	AGCGGACCUC	UCUAAAG	AGUUGAAACG	Gene 1	7,906-7,932
6	UCUGCUGCAU	UCUAAAG	UUAAAGGUGU	Gene 1	8,084-8,100
7	AGGCCGGUUU	UCUUAAC	GCUGCUGUUU	Gene 1	8,142-8,168
8	GGCUGCUUGU	UCUCAAC	UAGCAAAGGC	Gene 1	10,162-10,188
9	AUGCUGCUAU	UCUUAAC	AGGUGCAAUU	Gene 1	10,890-10,916
10	GGUUGUUGGG	UAUAAAC	AGAGUUUUAG	Gene 1	11,422-11,448
11	ACACGAAUUU	UUUAAAC	GGGUUCGGGG	Gene 1	13,644-13,670
[12]	GUUAAAUAAA	UCUAUAC	UUGUCAUGGC	Gene 1	21,790- Gene 2 17

Leader sequence	UUAAAUCUAA	UCUAAAC	UUUAUAAACG		
13	CUAAUCAUUU	UAUAAAC	UUUCCUCUAG	Gene 2	44-70
14	AAAGCUUCAG	UCUCAAC	UUGUGGAAAU	Gene 2	102- 128
15	UCAAUGAACC	UCUUAAC	AUCGUUUCACA	Gene 2	929- 955
[16]	UGAGGCAUAA	UCUAAAC	AUGCUGUUCG	Gene 2	2,166- Gene 3 17
17	CCUUACGUGG	UUUAAAC	CACCCUUUCU	Gene 3	271- 297
[18]	AGACAGAAAA	UCUAAAC	AAUUUAUAGC	Gene 3	3,737- Gene 4 17
[19]	UAUUACUAGU	UCUAAAC	CUCAUCUUAA	Gene 4	372- 404 (Gene 5 17)
[20]	GAUAAUAUAA	UCCAAAC	AUUAUGAGUA	Gene 5	634- Gene 6 17
[21]	GAGAAUCUAA	UCUAAAC	UUUAAGGAUG	Gene 6	693- Gene 7 17
22	GGCAGAAAAU	UUUAAAC	AAGCCUCGUC	Gene 7	860- 886
23	AAAAGAGGAC	UCCAAAC	AAGCAGUGCC	Gene 7	887- 913
24	CUUCUUUGGA	UCUAAAU	UAGAAUUGGU	Gene 7	1,041- 1,067
25	AGGUCCUACG	UCUAACC	AUAAGAACGG	Gene 7	1,615- 1,641

FIG. 2. Sequence comparison of the MHV-JHM intergenic regions, 5′-end genomic leader sequence, and the genomic regions in which only one nucleotide differs from the UCUAAAC consensus sequence. The regions homologous with the 3′-end genomic leader RNA and the intergenic region between genes 6 and 7 are underlined on the leader sequence. The regions homologous with the 3′ end of the leader RNA are underlined throughout. The first nucleotide in the intergenic consensus sequence of each of the seven MHV genes was considered the first nucleotide of the gene. Numbers denoting the functional intergenic regions are boxed. Sequence data were obtained from published MHV-JHM sequences (16, 29, 34, 37, 39–41).

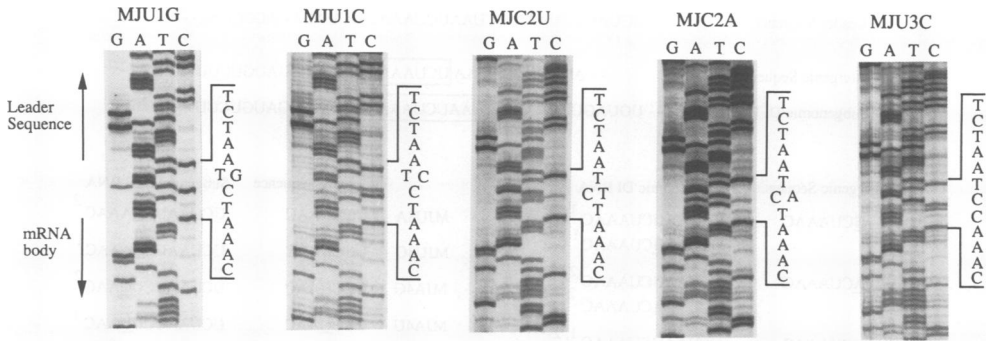

FIG. 3. Direct PCR sequencing of the leader-body fusion sites of subgenomic DI RNAs. Oligonucleotide 1189 was used as a sequencing primer.

This seems to indicate that although MHV transcription regulation is sufficiently flexible to recognize a one-nucleotide alteration in the consensus sequence, mutations of more than one nucleotide impede such recognition.

MHV undergoes high-frequency RNA recombination (2, 19). Possibly, the increase in subgenomic RNA transcription efficiency of the mutants was due to recombination between the mutant DI RNA and the helper virus-derived intergenic sites. To confirm that the subgenomic DI RNAs were in fact synthesized from the mutated intergenic consensus sequences, the intergenic-region sequences of the genomic DI RNAs were examined. For this analysis, MHV-specific cDNAs were first synthesized from intracellular RNA extracted from passage 1 virus-infected cells with oligonucleotide 1189, which specifically hybridizes to positive-sense DI RNA 732 to 750 nucleotides from the 3′ end of DI RNA, used as a primer. The cDNAs were mixed with oligonucleotide 1115, which hybridizes to negative-sense genomic DI RNA 1060 to 1077 nucleotides from the 3′ end of genomic DI RNA, and PCR was performed. The genomic DI RNA-specific PCR products were gel purified. The sequences were examined by direct PCR sequencing, and it was found that none of the mutants had altered intergenic sequences (data not shown). Therefore, it was concluded that there was no detectable RNA recombination in the intergenic region of any mutant DI RNA and that the subgenomic DI RNAs were indeed transcribed from mutated intergenic sequences.

Identification of the leader-body fusion site on subgenomic DI RNAs. Because the 3′ region of the genomic RNA leader sequence and the intergenic consensus sequences contain identical or closely related sequences, the exact site of the leader-body fusion on MHV mRNAs has not been identified. The mutant DI RNAs created in the present study were used to locate the exact leader-body fusion site on subgenomic DI RNA. To analyze this region, cDNA was made from the intracellular RNA species in DI RNA-replicating cells by priming the first strand with oligonucleotide 1189 and the second strand with oligonucleotide 78 (5′-AGCTTTACG TACCCTCTCTACTCTAAAACTCTTGTAGTTT-3′) (21), which specifically hybridizes to negative-strand MHV RNA at the leader sequence. After 25 cycles of PCR, one subgenomic DI RNA-specific PCR product of the predicted size was obtained from each RNA sample. The leader-body fusion site was initially examined by direct sequencing of the PCR products (Fig. 3). The PCR products with ambiguous direct-sequencing results were further cloned into a plasmid vector for dideoxy sequencing. The results of these sequence analyses are summarized in Fig. 4. All subgenomic DI RNAs were shown to contain two pentanucleotides at the leader-body fusion region. All of them contained the UC UAA sequence as the first pentanucleotide. The sequences at the second pentanucleotide differed among the mutants. All mutants with nucleotide substitutions in the third through seventh positions of the intergenic consensus sequence maintained the substituted nucleotides in the second pentanucleotide of the subgenomic DI RNA. Most of the subgenomic DI RNAs transcribed from the genomic DI RNAs with a substituted nucleotide at the first or second position of the consensus sequence demonstrated sequence heterogeneity in the second pentanucleotide: some maintained the substituted nucleotide, while others contained UCUAA. In the case of the MJC2G-derived PCR product, sequence analysis of six cDNA clones demonstrated that the second pentanucleotide was always changed to UCUAA. These analyses indicated that for this region, the leader-body fusion took place at either the first or the second nucleotide of the intergenic consensus sequence.

In addition to the region examined above, two A's upstream of the intergenic consensus sequence also represent an area of sequence homology where leader-body fusion may occur. The possibility that leader-body fusion might occur at these upstream A's was investigated by creating an additional four mutant DI cDNAs, each with the two A nucleotides upstream of the consensus sequence deleted and each containing a different substituted nucleotide at the first position of the consensus sequence. The intergenic consensus sequences of these mutants, MJΔWT, MJΔU1G, MJΔU1A, and MJΔU1C, are shown in Fig. 5. All four mutants replicated efficiently in MHV-infected cells and synthesized subgenomic DI RNA species (data not shown). The subgenomic DI RNA-specific PCR products were isolated and sequenced directly. For mutants MJΔU1A and MJΔU1C, the PCR products were further cloned into a plasmid vector and sequenced. These sequencing analyses demonstrated that all subgenomic DI RNAs had two repeats of the UCUAA sequence, and no sequence heterogeneity was observed (Fig. 5). These data demonstrated that in this series of mutants, the first UCUAA sequence on the subgenomic DI RNA was derived from the genomic RNA leader, the DI leader, or both. Sequence mismatches present at the first nucleotide of the second pentanucleotide sequence were substituted with the correct nucleotides, which could only have come from the leader sequence. These analyses indicated that when the intergenic region contains one consensus sequence and most of its flanking sequences are conserved, the leader-body fusion takes place at the first or second nucleotide of the consensus sequence.

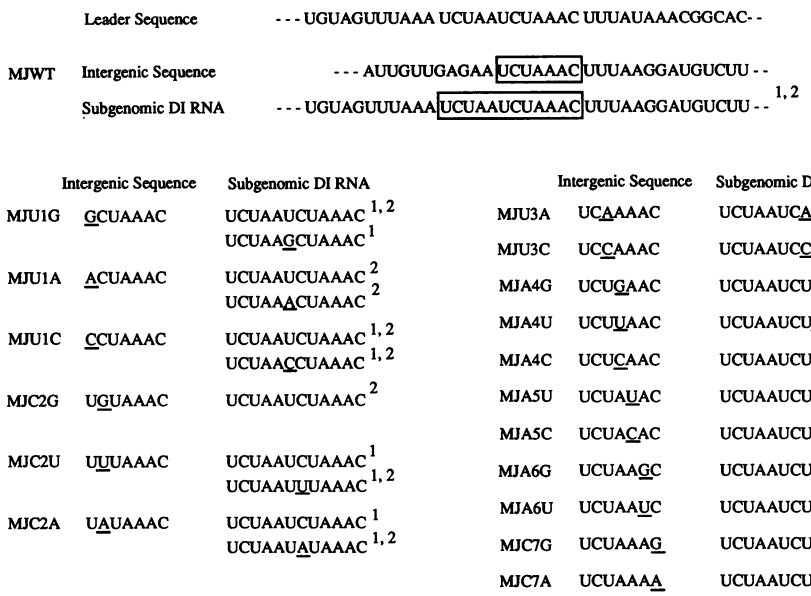

FIG. 4. Sequence comparison of the 5′-end genomic leader sequence and intergenic regions and leader-body fusion sites of MJWT-derived mutants. The intergenic sequence and the leader-body fusion site on the subgenomic MJWT DI RNA sequences are boxed, and the corresponding regions of mutants are listed. The subgenomic DI RNA sequences obtained by direct PCR sequencing of consistent and clear sequencing data and those obtained from cloned PCR products are indicated (1 and 2, respectively). At least three cDNA clones were analyzed.

DISCUSSION

Regulation of MHV transcription by the intergenic consensus sequence. In the present study, it was shown that subgenomic DI RNAs were synthesized from almost all mutant MHV DI RNAs with single-nucleotide substitutions in the inserted intergenic consensus sequence. The following conclusion can be made on the basis of this site-directed mutagenic analysis: sequence homology between the genomic RNA leader sequence and the intergenic consensus sequence was not the sole determinant of the quantity of subgenomic DI RNA; instead, the amount of the subgenomic DI RNA was affected by the site and the species of the substituted nucleotide. Only two mutants, MJC2G and MJU3G, demonstrated significantly decreased levels of subgenomic DI RNA transcription. The other mutants demonstrated transcriptional efficiency essentially the same as or even greater than that of MJWT. None of the naturally occurring transcriptionally functional intergenic regions contain a G within the consensus sequence, and the first, fourth, sixth, and seventh nucleotides of the UCUAAAC consensus sequence are completely conserved in all the functional intergenic sequences of essential MHV genes (Fig. 2). Therefore, these results were particularly surprising, because they indicated that sequence integrity within the

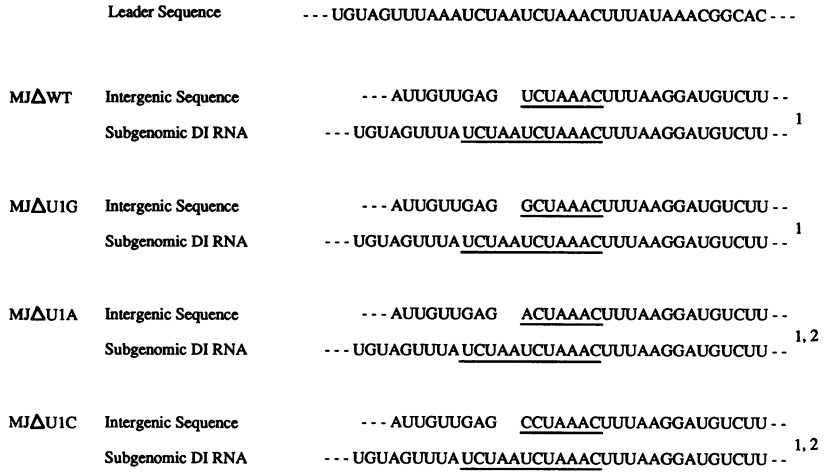

FIG. 5. Sequence comparison of the 5′-end genomic leader sequence and intergenic regions and leader-body fusion sites of MJΔWT-derived mutants. The intergenic consensus sequence and the leader-body fusion sites are underlined. Subgenomic DI RNA sequences obtained by direct PCR sequencing and those obtained from cloned PCR products are indicated (1 and 2, respectively). At least three cDNA clones were analyzed.

consensus sequence is more flexible for MHV transcription than was previously thought.

The recent observation that unique sequences quite different from the UCUAAAC consensus sequence are also utilized for MHV transcription (15, 20, 37, 47) is consistent with data presented here. MHV mRNAs are synthesized by utilizing these unique consensus sequences only if the leader sequence contains a characteristic number of the UCUAA repeats at the 5′ end of the genome (20) and only if the expression of these mRNAs is not essential for MHV replication (15, 37, 47). Together, these data indicate that not only the sequence but also the secondary or tertiary structure dictated by the intergenic consensus sequence regulate MHV transcription. Moreover, this structure interacts with a leader sequence structure that is also affected by its primary sequence.

As demonstrated in the present study, most of the mutants containing single-nucleotide substitutions within the consensus sequence supported subgenomic RNA transcription. Then why is this experimentally plastic sequence so highly conserved in the virus? For reasons of simplicity, the intergenic region used for this study contained only one UCUAAAC sequence and had flanking sequences derived from the intergenic sequence between genes 6 and 7. Regulation of transcription via this intergenic region, as with any intergenic region, could conceivably be an interrelated process that is also affected by the number of intergenic consensus sequence repeats, the type of intergenic-region-flanking sequences, and the nature of the leader sequence. It will be interesting to see how nucleotide substitutions at wild-type-virus-derived intergenic regions affect transcription and how flanking sequences influence the effects of those substitutions. It may be that the molar ratios of subgenomic mRNAs and those of the mRNA translation products are important for efficient MHV replication and that alterations of these ratios may result in less efficient MHV replication. So far, none of the MHV mutants demonstrated significantly altered ratios of any subgenomic RNA species which encodes proteins essential for MHV replication (25). Therefore, it is possible that an MHV variant which produces a different molar ratio of subgenomic mRNA because of a nucleotide substitution at the consensus sequence may replicate less efficiently than wild-type MHV. Such a variant may be easily eliminated from the virus population after several rounds of virus replication. Consequently, although MHV transcription regulation mechanisms are flexible enough to recognize the single-nucleotide substitution at the consensus sequence, it is probably selectively advantageous for the wild-type MHV to contain the UCUAAAC consensus sequence in vivo.

Another question raised by the present study is why sequences which are very similar to the UCUAAAC consensus sequence are not recognized for transcription. As shown in Fig. 2, MHV-JHM genomic RNA contains 19 regions in which only one nucleotide is different from the UCUAAAC sequence. It is evident that the UCUAAAC sequence is present only at the transcriptionally functional intergenic regions. The intergenic region between genes 1 and 2 is UCUAUAC, and that between genes 5 and 6 is UCCAAAC. Both of these consensus sequences seem to be at least as active as the UCUAAAC consensus sequence, as shown by analysis of MJA5U and MJU3C (Fig. 1). The intergenic consensus sequence between genes 1 and 2 has two adjacent upstream and downstream nucleotides which are complementary to the leader sequence (Fig. 2). The altered intergenic consensus sequence between genes 5 and 6 is also flanked by a number of nucleotides which are complementary with the leader sequence. None of the nonfunctional "intergenic regions" contain this flanking sequence homology. This may indicate that if a single-nucleotide substitution is present in the consensus sequence, then the presence of adjacent leader-complementary flanking sequences is necessary for RNA transcription. Alternatively, it is possible that transcription at the nonfunctional intergenic regions is prevented by the presence of flanking sequences which downregulate transcription.

Specific sequence alteration at the leader-body junction site. The intergenic consensus sequences and the 5′-end genomic leader sequence contain identical or very similar sequences; therefore, the exact site of leader-body fusion on MHV mRNA has not been identified. In this study, all the subgenomic DI RNAs analyzed contained two pentanucleotide sequences: the first sequence was always UCUAA, while the second sequence varied among the mutant DI RNAs. When subgenomic DI RNAs were transcribed from the genomic DI RNAs containing a nucleotide substitution in the first or second nucleotide of the intergenic consensus sequence, some RNA molecules kept the mutations in the second pentanucleotide sequence and some had this sequence changed to UCUAA. It was found that leader-body fusion did not occur in the third through seventh nucleotides of the intergenic consensus sequence. Analysis of MJΔU1G, MJΔU1A, and MJΔU1C indicated that the leader-body fusion took place most likely downstream of the first nucleotide of the intergenic consensus sequence (Fig. 5). These data indicated that the leader-body fusion site on the mRNA was most probably the first or second nucleotide of the second pentanucleotide of the consensus sequence. The data obtained from the present study are consistent with our previous observation that the UCCAAAC sequence present at the intergenic region between genes 5 and 6 is conserved in mRNA 6 (23). All these data indicate that the site of the leader-body fusion was also affected by the sequence and/or structure of the intergenic sequence. This model is different from the proposed MHV transcription model in which leader RNA is cleaved at the mismatch site when mismatched sequence between the leader sequence and the intergenic sequence is recognized by viral polymerase (1).

If MHV subgenomic mRNAs are synthesized by the leader-primed transcription mechanism, how does the leader-body fusion take place? One possibility is that only small free leader RNAs with one UCUAA sequence plus one or two additional U and C nucleotides at the 3′ end are used for transcription. In this case, sequence complementarity between the free leader RNA and the intergenic sequence would be only two or three nucleotides and the leader RNAs would not need to undergo RNA cleavage. This mechanism is possible, as it was shown previously that influenza virus transcription does not require sequence homology between the template RNA and the primer RNA (30). Alternatively, relatively long free leader RNA species may bind to the intergenic region along with some virus-derived proteins and host cell factors. Once an RNA-protein complex is formed at the intergenic site, one of the viral proteins may cleave the leader sequence, with the resulting leader RNA binding to the intergenic region and priming RNA synthesis (10). If this model is correct, then the leader RNA cleavage site may be determined by the nucleotide sequence or RNA structure of the intergenic sequence and not simply by a nucleotide mismatch between free leader RNA and the intergenic sequence. It is possible that ribozyme activity could cleave

leader RNA, although this activity has not been found in MHV RNA.

Recent coronavirus transcription studies indicate that the possibility of splicing-type transcription during primary transcription has not been excluded (35, 36). It has been shown that the leader sequence of MHV mRNA reassorts with a high level of efficiency (24). If coronavirus primary transcription undergoes such a splicing-type transcription, it may be more similar to *trans* than to *cis* splicing (27, 43). However, there are a number of differences between coronavirus primary transcription and eukaryotic splicing. It has been shown that coronavirus transcription takes place in the cytoplasm (4, 45). Splicing acceptor and donor sites are not found in the intergenic sequences or the leader sequence. The number of pentanucleotide repeats on a given mRNA species varies (23), and the present study showed that the leader-body fusion site varies if the sequence at the intergenic region changes. Therefore, even if coronavirus primary transcription involves a *trans* splicing-like event, its mechanism should be quite different from that of eukaryotic splicing.

ACKNOWLEDGMENTS

We thank Akio Yamada for valuable suggestions about direct sequencing of the PCR products. We thank Jennifer Fosmire for helpful comments on the manuscript.

This work was supported by Public Health Service grant AI29984 from the National Institutes of Health.

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. **Baric, R. S., S. A. Stohlman, and M. M. C. Lai.** 1983. Characterization of replicative intermediate RNA of mouse hepatitis virus: presence of leader RNA sequences on nascent chains. J. Virol. **48**:633–640.
4. **Brayton, P. R., R. G. Ganges, and S. A. Stohlman.** 1981. Host cell nuclear function and murine hepatitis virus replication. J. Gen. Virol. **50**:457–460.
5. **Brown, T. D. K., M. E. G. Boursnell, M. M. Binns, and F. M. Tomley.** 1986. Cloning and sequencing of 5′ terminal sequences from avian infectious bronchitis virus genomic RNA. J. Gen. Virol. **67**:221–228.
6. **Fosmire, J. A., K. Hwang, and S. Makino.** 1992. Identification and characterization of a coronavirus packaging signal. J. Virol. **66**:3522–3530.
7. **Higuchi, R.** 1990. Recombinant PCR, p. 177–183. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols. Academic Press, San Diego, Calif.
8. **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.
9. **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.
10. **Lai, M. M. C.** 1990. Coronavirus: organization, replication and expression of genome. Annu. Rev. Microbiol. **44**:303–333.
11. **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.
12. **Lai, M. M. C., P. R. Brayton, R. C. Armen, C. D. Patton, C. Pugh, and S. A. Stohlman.** 1981. Mouse hepatitis virus A59: mRNA structure and genetic localization of the sequence divergence from hepatotropic strain MHV-3. J. Virol. **39**:823–834.
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. **Lai, M. M. C., and S. A. Stohlman.** 1978. RNA of mouse hepatitis virus. J. Virol. **26**:236–242.
15. **La Monica, N., K. Yokomori, and M. M. C. Lai.** 1992. Coronavirus mRNA synthesis: identification of novel transcription initiation signals which are differentially regulated by different leader sequences. Virology **188**:402–407.
16. **Lee, H.-J., C.-K. Shieh, A. E. Gorbalenya, E. V. Eugene, 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. Wilhelmsen, 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. **Makino, S., M. Joo, and J. K. Makino.** 1991. A system for study of coronavirus mRNA synthesis: a regulated, expressed subgenomic defective interfering RNA results from intergenic site insertion. J. Virol. **65**:6031–6041.
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**:227–232.
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., C.-K. Shieh, J. G. Keck, and M. M. C. Lai.** 1988. Defective-interfering particles of murine coronaviruses: mechanism of synthesis of defective viral RNAs. Virology **163**:104–111.
23. **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.
24. **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.
25. **Makino, S., F. Taguchi, N. Hirano, and K. Fujiwara.** 1984. Analysis of genomic and intracellular viral RNAs of small plaque mutants of mouse hepatitis virus, JHM strain. Virology **139**:138–151.
26. **Makino, S., K. Yokomori, and M. M. C. Lai.** 1990. Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. J. Virol. **64**:6045–6053.
27. **Murphy, W. J., K. P. Watkins, and N. Agabian.** 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for trans splicing. Cell **47**:517–525.
28. **Pachuk, C. J., P. J. Bredenbeek, P. W. Zoltick, W. J. M. Spaan, and S. R. Weiss.** 1989. Molecular cloning of the gene encoding the putative polymerase of mouse hepatitis virus, strain A59. Virology **171**:141–148.
29. **Pfleiderer, M., M. A. Skinner, and S. G. Siddell.** 1986. Coronavirus MHV-JHM: nucleotide sequence of the mRNA that encodes the membrane protein. Nucleic Acids Res. **14**:6338.
30. **Plotch, S. J., M. Bouloy, I. Ulmanen, and R. M. Krug.** 1981. A unique cap (m^7GppX^m)-dependent influenza virion endonuclease cleaves capped mRNAs to generate the primers that initiate viral RNA transcription. Cell **23**:847–858.
31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Sawicki, S. G., and D. L. Sawicki.** 1986. Coronavirus minus-strand RNA synthesis and effect of cycloheximide on coronavirus RNA synthesis. J. Virol. **57**:328–334.
33. **Sawicki, S. G., and D. L. Sawicki.** 1990. Coronavirus transcrip-

tion: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. J. Virol. **64**:1050–1056.
34. **Schmidt, I., M. Skinner, and S. Siddell.** 1987. Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus MHV-JHM. J. Gen. Virol. **68**:47–56.
35. **Sethna, P. B., M. A. Hofmann, and D. A. Brian.** 1991. Minus-strand copies of replicating coronavirus mRNAs contain antileaders. J. Virol. **65**:320–325.
36. **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.
37. **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.
38. **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.
39. **Skinner, M. A., D. Ebner, and S. G. Siddell.** 1985. Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. J. Gen. Virol. **66**:581–592.
40. **Skinner, M. A., and S. G. Siddell.** 1983. Coronavirus JHM: nucleotide sequence of the mRNA that encodes nucleocapsid protein. Nucleic Acids Res. **15**:5045–5054.
41. **Skinner, M. A., and S. G. Siddell.** 1985. Coding sequence of coronavirus MHV-JHM mRNA 4. J. Gen. Virol. **66**:593–596.
42. **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 noncontiguous sequences. EMBO J. **2**:1939–1944.
43. **Sutton, R. E., and J. C. Boothroyd.** 1986. Evidence for trans splicing in trypanosomes. Cell **47**:527–535.
44. **van der Most, R. G., P. J. Bredenbeek, and W. J. M. Spaan.** 1991. A domain at the 3′ end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs. J. Virol. **65**:3219–3226.
45. **Wilhelmsen, K. C., J. L. Leibowitz, C. W. Bond, and J. A. Robb.** 1981. The replication of murine coronaviruses in enucleated cells. Virology **110**:225–230.
46. **Winship, P. R.** 1989. An improved method for directly sequencing PCR material using dimethyl sulfoxide. Nucleic Acids Res. **17**:1266.
47. **Yokomori, K., L. R. Banner, and M. M. C. Lai.** 1991. Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. Virology **183**:647–657.