

ANALYSIS OF THE CIS-ACTING ELEMENTS OF CORONAVIRUS TRANSCRIPTION

Myungsoo Joo and Shinji Makino

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

INTRODUCTION

Mouse hepatitis virus (MHV), a coronavirus, is an enveloped virus containing a single-stranded, positive-sense RNA genome of approximately 31 kb (11, 13, 22). There are seven to eight species of virus-specific subgenomic mRNAs in MHV-infected cells. These subgenomic mRNAs comprise a 3'-coterminal nested-set (9, 14). In decreasing order of size, they are mRNAs 1 through 7 (9, 14). The 5' end of the MHV genomic RNA contains a 72- to 77-nucleotide-long leader sequence (8, 10, 26). Within the 3'-region of the leader sequence there is a pentanucleotide sequence, UCUAA. This sequence repeats two to four times in different MHV strains (19). The MHV-specific genes are downstream from the leader, and each gene is separated by a special short stretch of sequence, the intergenic sequence. The intergenic sequences include the unique consensus sequence UCUAAAC or a sequence very similar sequence (25). All MHV mRNA species have a sequence identical to the 5'-end genomic leader sequence. These leader sequences are fused to the intergenic consensus sequence, which marks the start of each gene (8, 10, 25, 26). In most MHV genes the degree of intergenic sequence nucleotide homology with the leader sequence correlates with the amount of mRNA transcribed (25). This correlation is not observed in infectious bronchitis virus mRNA transcription (4). The site where the leader fuses with the mRNA is somewhere within the repeated pentanucleotide (UCUAA). The number of repeats in each given mRNA varies (19). The pentanucleotide repeats at the genomic leader sequence and at the intergenic region are identical, making identification of the fusion site of these two sequences difficult.

Several models try to explain subgenomic RNA synthesis. One model is leader RNA-primed transcription; it 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, 7). Another model, was put forth by Sawicki and Sawicki (23). This model suggests that subgenomic negative-strand RNAs are initially synthesized from the input genomic RNA followed by synthesis of the positive-strand subgenomic RNA on the subgenomic-sized, negative-stranded RNA. Another possible mechanism may be that leader RNA joins the subgenomic RNA body by a mechanism similar to RNA splicing (8, 26). None of these models has been exclusively proven. The mechanism of coronavirus transcription is unknown.

Defective interfering (DI) RNAs of MHV are vital for understanding the mechanisms of coronavirus mRNA transcription (16). An artificial subgenomic mRNA was made using the DI system; the intergenic region, derived from between genes 6 and 7 of the genome was inserted into a complete DI cDNA clone. This clone replicated and transcribed a subgenomic DI RNA after transfection of its RNA into MHV-infected cells.

In the present study, how nucleotide substitution in the consensus sequence influences mRNA transcription and the location of the leader-body fusion site were studied with mutant

DI cDNAs. The role of the sequences flanking the intergenic sequence in subgenomic DI RNA transcription was also examined.

MATERIALS AND METHODS

Viruses and cells. The plaque-cloned A59 strain of MHV (MHV-A59) (9) was used as a helper virus. Mouse DBT cells were used for growth of viruses.

DNA construction. A procedure based on recombinant polymerase chain reaction (PCR) was employed for site-directed mutagenesis (5). Construction of various plasmids will be described elsewhere.

RNA transcription and transfection. Plasmid DNAs were linearized by Xba I digestion and transcribed with T7 RNA polymerase as previously described (18). Lipofection was used for RNA transfection (16).

Preparation of virus-specific intracellular RNA and Northern (RNA) blotting. Virus-specific RNAs in virus-infected cells were extracted as previously described (20). Northern blotting of intracellular RNA species were described previously (6).

Direct sequencing of the PCR product. Direct PCR sequencing was performed according to the procedure established by Winship (27).

RESULTS

The effect of nucleotide substitutions within the consensus sequence on subgenomic DI RNA transcription

The effect of nucleotide substitutions within the conserved UCUAAC consensus sequence on MHV subgenomic DI RNA synthesis was observed in a series of constructions. The "parent" construct, MJWT, had a TCTAAAC sequence in the middle of the intergenic region located between genes 6 and 7. 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 were made; each had a specific nucleotide substitution within the UCUAAC sequence. These mutants were named according to the site of mutation at the UCUAAC consensus sequence. For example, MJU3G contained a specific nucleotide substitution from U to G at the third nucleotide of the consensus sequence. In vitro-synthesized DI RNAs were transfected by lipofection into monolayers of DBT cells preinfected with MHV-A59 helper virus(18). Virus samples obtained after overnight incubation were further passaged to generate the 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 using a probe that specifically hybridizes with all MHV RNAs (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. This observation was particularly unexpected for those mutants with G substitutions, because none of the naturally occurring transcriptionally functional intergenic consensus sequences contained G. During isolation of mutant DI cDNAs, a double mutant, MJA5GC7U, with nucleotide substitutions at both nucleotides 5 and 7 was isolated. MJA5G and MJC7U supported efficient subgenomic DI RNA synthesis, whereas no subgenomic DI RNA synthesis was observed in the MJA5GC7U-replicating cells (see Fig. 1). This would seem to indicate that although MHV transcription regulation is sufficiently flexible to recognize a one-nucleotide alteration in the consensus sequence, more than one mutation impedes such recognition.

MHV undergoes high-frequency RNA recombination (2, 15). 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. Sequences at the intergenic sites confirmed that there was subgenomic DI RNA synthesis from the mutated intergenic consensus sequences. No recombinants were detected in the mutant DIs (data not shown).

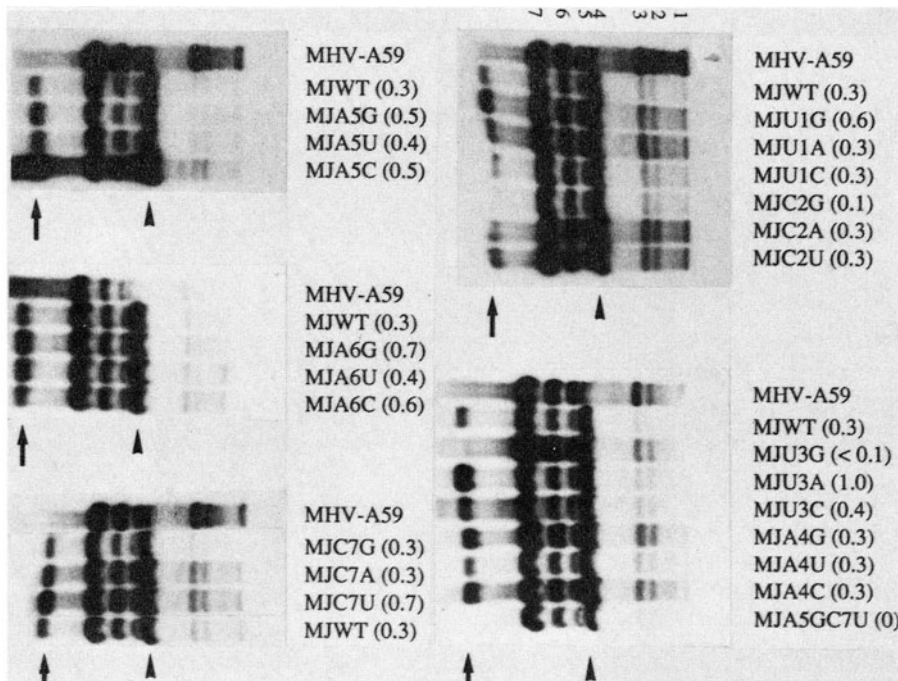


Fig. 1. Northern blot analysis of MJWT-derived mutant subgenomic DI RNAs. Passage 1 virus samples were used for virus inoculum. Intracellular RNAs were extracted 7 h postinfection, separated by 1% formaldehyde-agarose gel electrophoresis and transferred to nylon membrane. The probe was prepared by random-primed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

Identification of the leader-body fusion site on subgenomic DI RNAs

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, a subgenomic DI RNA-specific PCR product was prepared. Leader-fusion site PCR products were examined first by direct sequencing. Those PCR products with ambiguous direct-sequencing results were further cloned into a plasmid vector for dideoxy sequencing. Figure 2 shows the results of these sequence analyses. All subgenomic DI RNAs were shown to contain two pentanucleotide sets at the leader-body fusion region. All of them had the UCUAA sequence as a first pentanucleotide. The sequence at the second set differed among the mutants. Mutants with nucleotide substitutions in intergenic site positions three through seven maintained the substituted nucleotides in the second pentanucleotide. 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 set. Some maintained the substituted nucleotide, while others contained UCUAA. The leader and body sequences appeared to fuse either at the first or the second nucleotide of the intergenic consensus sequence.

Besides the region examined above, two As upstream of the intergenic consensus sequence also represent an area of sequence homology where leader-body fusion may occur. The possibility that leader-fusion might occur at those upstream As was investigated by creating an additional four mutant DI cDNAs. Each mutant had the upstream two A nucleotides of the consensus sequence deleted and each contained a different substituted

Leader Sequence --- UGUAGUUUAAA UC^UAAUC^UAAAC UUUAAUAAACGGCAC - -

MJWT Intergenic Sequence --- AUUGUUGAGAA UC^UAAAC UUUAAAGGAUGUCUU - -

Subgenomic DI RNA --- UGUAGUUUAAA UC^UAAUC^UAAAC UUUAAAGGAUGUCUU - -^{1,2}

	Intergenic Sequence	Subgenomic DI RNA		Intergenic Sequence	Subgenomic DI RNA
MJU1G	<u>GC</u> UAAAC	UCUAAUCUAAAC ^{1,2} UCUAA <u>AG</u> C ^U AAAC ¹	MJU3A	UC <u>AA</u> AAC	UCUAAUC <u>CA</u> AAAC ²
MJU1A	<u>AC</u> UAAAC	UCUAAUCUAAAC ² UCUAA <u>A</u> C ^U AAAC ²	MJU3C	UCC <u>AA</u> AAC	UCUAAUCC <u>AA</u> AAAC ¹
MJU1C	<u>CC</u> UAAAC	UCUAAUCUAAAC ^{1,2} UCUAA <u>C</u> C ^U AAAC ^{1,2}	MJA4G	UCU <u>G</u> AAC	UCUAAUC <u>UG</u> AAC ²
MJC2G	<u>UG</u> UAAAC	UCUAAUCUAAAC ²	MJA4U	UCU <u>U</u> AAC	UCUAAUC <u>U</u> AAAC ¹
MJC2U	<u>UU</u> UAAAC	UCUAAUCUAAAC ¹ UCUAA <u>U</u> <u>U</u> AAAC ^{1,2}	MJA4C	UCU <u>C</u> AAC	UCUAAUC <u>C</u> AAAC ²
MJC2A	<u>UA</u> UAAAC	UCUAAUCUAAAC ¹ UCUAA <u>A</u> <u>U</u> AAAC ^{1,2}	MJA5U	UCU <u>A</u> UAC	UCUAAUC <u>A</u> UAC ¹
			MJA5C	UCU <u>A</u> CAC	UCUAAUC <u>A</u> CAC ²
			MJA6G	UCUAA <u>GC</u>	UCUAAUCUAA <u>GC</u> ²
			MJA6U	UCUAA <u>UC</u>	UCUAAUCUAA <u>UC</u> ¹
			MJC7G	UCUAA <u>AG</u>	UCUAAUCUAA <u>AG</u> ²
			MJC7A	UCUAA <u>AA</u>	UCUAAUCUAA <u>AA</u> ²

Fig. 2. Sequence comparison of the 5' end genomic leader sequence, intergenic regions and leader-body fusion sites of MJWT-derived mutants. The intergenic sequence and the leader fusion site on the subgenomic MJWT DI RNA sequences are surrounded by boxes and the corresponding regions of mutants are listed below. The subgenomic DI RNA sequences obtained by direct PCR sequencing of consistent and clear sequencing data are shown by a number 1, while those obtained from cloned PCR products are indicated by number 2. At least three cDNA clones were analyzed.

nucleotide at the first position of the consensus sequence. All the subgenomic DI RNAs had two repeats of the UC^UAA sequence and demonstrated no sequence heterogeneity. When the intergenic region contained one consensus sequence and when most of its flanking sequences were conserved, the leader and body fused at the first or second nucleotide of the consensus sequence.

The role of the intergenic region flanking sequences in subgenomic DI RNA synthesis

Two series of mutants were constructed to test the effect of intergenic site flanking sequences on MHV transcription. One set of mutants was altered upstream of the 18 nucleotide intergenic sequence and the other was altered downstream. Four mutants, each with a 435 nucleotide-long downstream sequence, were tested in the study of the upstream intergenic effect on transcription; the altered upstream flanking sequences contained 1440, 283, 89 and 0 nucleotides. An A nucleotide, two nucleotides upstream of the UC^UAA repeats, is referred to as nucleotide 1. The sequence from nucleotides 1 through 18 represents the intergenic region sharing complete sequence homology with the 3'-region of the leader sequence (25). Analysis of the intracellular virus-specific RNA species from these mutants showed that the amount of subgenomic DI RNA was about 80% that of the genomic DI RNA (data not shown). For the downstream flanking sequence study, ten mutants with downstream intergenic flanking sequences ranging in size from 1671 nucleotides to zero nucleotides were examined. Each of these mutants lacked an upstream flanking sequence. Like the effect of the upstream flanking sequence, there seemed to be no appreciable differences in transcription levels of these mutants. These analyses suggested that upstream and downstream sequences flanking the intergenic sequence were not important in the regulation of subgenomic DI RNA transcription.

DISCUSSION

The present study showed that subgenomic DI RNAs were synthesized from almost all the intergenic consensus site mutants. The data indicated that sequence homology between the genomic RNA leader sequence and the intergenic consensus sequence was not the sole determinant of subgenomic DI RNA quantity. Instead, the amount of subgenomic DI RNA was affected by the site and species of the substituted nucleotide. It seemed that sequence integrity within the consensus sequence was more flexible for MHV transcription than was previously thought. A recent description, that sequences differing from the UC_UAAAC consensus sequence are also used for MHV transcription (12, 17, 24, 28), is consistent with the data presented here. Therefore, not only the sequence but also the secondary or tertiary structure dictated by the intergenic consensus sequence regulated MHV transcription. Moreover, this structure interacted with a leader sequence structure that was also affected by its primary sequence. The intergenic sequence flanking sequences did not alter the regulation of subgenomic DI RNA synthesis. Although, there is the possibility that other genomic region(s) may regulate the efficiency of MHV subgenomic mRNA synthesis (see below).

If most of the mutants with 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 UC_UAAAC sequence and had flanking sequences derived from the gene 6-7 intergenic sequence. Regulation of transcription at this intergenic region, like any intergenic region, could be an interrelated process; the number of intergenic consensus sequence repeats, the type of flanking sequences and the leader sequence could be involved. It will be interesting to see how nucleotide substitutions at wild type virus-derived intergenic regions affect transcription and how flanking sequences influence the effect 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 encoding essential replication proteins (20). Therefore, it is possible that an MHV variant producing a different molar ratio of genomic RNA to subgenomic RNA may replicate less efficiently than wild type MHV. Such a variant would be easily eliminated from the virus population after several rounds of virus replication. Consequently, although MHV transcription regulation mechanisms are flexible enough to recognize single nucleotide substitutions in the consensus sequence, it is probably selectively advantageous for the wild type MHV to contain the UC_UAAAC consensus sequence *in vivo*.

Another question raised from the present study is why sequences which are very similar to the UC_UAAAC consensus sequence are not recognized for transcription. Analysis of entire MHV-JHM sequences demonstrated that MHV-JHM genomic RNA contains 19 regions where only one nucleotide is different from the UC_UAAAC sequence and the UC_UAAAC sequence is only present at the transcriptionally functional intergenic regions. The intergenic region between genes 1 and 2 is UC_UAUAC, and that between genes 5 and 6 is UC_CAAAC. Both of these different consensus sequences seemed to be at least equally as active as the UC_UAAAC consensus sequence based on 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 (24). 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" exhibit 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-complimentary flanking sequences are necessary for RNA transcription. Alternatively, it is possible that transcription at the nonfunctional "intergenic regions" is prevented by the presence of flanking sequences which down-regulate transcription.

Sequence analysis of subgenomic DI RNA indicated that the leader-body fusion site on the mRNA was most probably the first or second nucleotide of the second consensus sequence repeat. This was consistent with the previous observation that the UC_CAAAC sequence present at the intergenic region between genes 5 and 6 is conserved in mRNA 6 (19). All these data indicate that the site of the leader-body fusion was also affected by the nucleotide sequence or the structure of the intergenic sequence or both. This differs from the proposed MHV transcription model which hypothesizes that leader RNA is cleaved by viral

RNA polymerase at the mismatch site between the leader sequence and the intergenic sequence (1).

If MHV subgenomic mRNAs are synthesized by the leader-primed transcription mechanism then how does leader-body fusion take place? One possibility is that only small-sized 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 to three nucleotides and the leader RNAs would not need to undergo RNA cleavage. Alternatively, relatively long free leader RNA species may bind to the intergenic region along with virus-derived proteins and host cell factors. After RNA-protein complex formation at the intergenic site, one of the viral proteins might cleave the leader sequence with the resulting leader RNA binding to the intergenic region and priming RNA synthesis (7). If this model is correct, then the leader RNA cleavage site may be determined by nucleotide sequence or RNA structure, and not simply by a nucleotide mismatch between free leader RNA and the intergenic sequence. It is possible that a ribozyme activity could cleave leader RNA, although this activity has not been found in MHV RNA.

An unanswered question from these studies (16) is why the ratio of subgenomic DI RNA to the genomic DI RNA was significantly different from that of the mRNA 7 to mRNA 1 ratio. It was found that the amount of mRNA 1 is only 1.5% of that of mRNA 7 (14), whereas the amount of subgenomic DI RNA was about 80% that of the genomic DI RNA. One possibility is that the amount of helper virus-derived MHV transcription function in DI RNA-replicating cells is too low for efficient subgenomic DI RNA transcription. MHV DI RNA replication strongly inhibits helper MHV RNA replication and transcription (21) leading to the speculation that significantly decreased amounts of MHV-specific transcription factors are present in DI RNA-replicating cells. A limited amount of MHV transcriptional function might not be enough to drive recognition of the template DI RNA molecules and subsequent subgenomic DI RNA transcription. Another possibility is that host derived-factor(s) limit the efficiency of MHV RNA transcription. If MHV RNA transcription takes place at specific sites within the cells or requires host-derived factors(s) or both, then the maximum number of RNA molecules involved in MHV RNA transcription may be saturated after accumulation of large amounts of template RNA molecules. Although genomic DI RNA accumulates more efficiently than MHV genomic RNA, only a small fraction of the DI RNA would be used as a transcription template because of the limited availability of host functions. Alternatively, it is possible that an enhancer sequence(s) for subgenomic RNA synthesis is present somewhere within gene 1 to gene 4 of MHV genomic RNA. The lack of such a sequence(s) in the DI RNA might result in a lower efficiency of subgenomic DI RNA synthesis.

ACKNOWLEDGMENTS

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. *EMBO J.* 9:4173-4179. (1990).
2. Baric, R. S., K. Fu, M. C. Schaad, and S. A. Stohlman. *Virology* 177:646-656. (1990).
3. Baric, R. S., S. A. Stohlman, and M. M. C. Lai. *J. Virol.* 48:633 (1983).
4. Brown, T. D. K., M. E. G. Bournsnel, M. M. Binns, and F. M. Tomley. *J. Gen. Virol.* 67:221 (1986).
5. Higuchi, R. Recombinant PCR, in "PCR Protocols," M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White eds. Academic Press, San Diego (1990).
6. Jeong Y. S., and S. Makino. *J. Virol.* 66: 3339 (1992).
7. Lai, M. M. C. *Annu. Rev. Microbiol.* 44:303 (1990).
8. Lai, M. M. C., R. S. Baric, P. R. Brayton, and S. A. Stohlman. *Proc. Natl. Acad. Sci. USA* 81: 3626 (1984).
9. Lai, M. M. C., P. R. Brayton, R. C. Armen, C. D. Patton, C. Pugh, and S. A. Stohlman. *J. Virol.* 39: 823 (1981).

10. Lai, M. M. C., C. D. Patton, R. S. Baric, and S. A. Stohlman. *J. Virol.* 46: 1027 (1983).
11. Lai, M. M. C., and S. A. Stohlman. *J. Virol.* 26: 236 (1978).
12. La Monica, N., K. Yokomori, and M. M. C. Lai. *Virology* 188:402 (1992).
13. Lee, H.-J., C.-K. Shieh, A. E. Gorbalenya, E. V. Eugene, N. La Monica, J. Tuler, A. Bagdzhadzhyan, and M. M. C. Lai. *Virology* 180:567 (1991).
14. Leibowitz, J. L., K. C. Wilhelmsen, and C. W. Bond. *Virology* 114: 39 (1981).
15. Makino, S., J. G. Keck, S. A. Stohlman, and M. M. C. Lai. *J. Virol.* 57:729 (1986).
16. Makino, S., M. Joo, and J. K. Makino. *J. Virol.* 65.:6031 (1991).
17. Makino, S., and M. M. C. Lai. *Virology* 169:227 (1989).
18. Makino, S., and M. M. C. Lai. *J. Virol.* 63: 5285 (1989).
19. Makino, S., L. H. Soe, C.-K. Shieh, and M. M. C. Lai. *J. Virol.* 62:3870 (1988).
20. Makino, S., F. Taguchi, N. Hirano, and K. Fujiwara. *Virology* 139: 138 (1984).
21. Makino, S., K. Yokomori, and M. M. C. Lai. *J. Virol.* 64:6045 (1990).
22. Pachuk, C. J., P. J. Bredenbeek, P. W. Zoltick, W. J. M. Spaan, and S. R. Weiss. *Virology* 171:141(1989).
23. Sawicki, S. G., and D. L. Sawicki. *J. Virol.* 64:1050 (1990).
24. Shieh, C.-K., H.-J. Lee, K. Yokomori, N. La Monica, S. Makino, and M. M. C. Lai. *J. Virol.* 63:3729 (1989).
25. Shieh, C.-K., L. H. Soe, S. Makino, M.-F. Chang, S. A. Stohlman, and M. M. C. Lai. *Virology* 156: 321 (1987).
26. Spaan, W., H. Delius, M. Skinner, J. Armstrong, P. Rottier, S. Smeekens, B. A. M. van der Zeijst, and S. G. Siddell. *EMBO J.* 2: 1939 (1983).
27. Winship P. R. 1989. *Nucleic Acids Res.* 17:1266 (1989).
28. Yokomori, K., L. R. Banner, and M. M. C. Lai. *Virology* 183:647 (1991).