MURINE CORONAVIRUS 5'-END GENOMIC RNA SEQUENCE REVEALS MECHANISM OF LEADER-PRIMED TRANSCRIPTION

Lisa H. Soe, Chien-Kou Shieh, Shinji Makino, Ming-Fu Chang, Stephen A. Stohlman and Michael M.C. Lai

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

INTRODUCTION

Mouse hepatitis virus (MHV) contains a single-strand, positive-sense RNA genome which is transcribed in infected cells, first, into a full-length negative-strand RNA (Brayton et al., 1982; Lai et al., 1982) and then into a positive-sense genomic RNA and six species of subgenomic mRNA. The mRNAs consist of a 3' co-terminal nested-set structure (Lai et al., 1981), and also contain an identical leader sequence of approximately 72 nucleotides at the 5' ends (Lai et al., 1984; Spaan et al., 1983). Ultraviolet transcriptional mapping studies (Jacobs et al., 1981) and the fact that no nuclear function is required for replication (Brayton et al., 1981; Wilhelmsen et al., 1981) suggest that the joining of the leader sequences to coronavirus mRNAs does not utilize conventional eukaryotic splicing mechanisms.

Several transcriptional models have been proposed as possible mechanisms for the synthesis of such leader-containing mRNAs (Baric et al., 1983). The data obtained by our laboratory favor the model of leader-primed transcription, in which the leader RNA is synthesized independently, dissociates from the negative-strand RNA template, and then rebinds to the template at the initiation sites for different mRNAs, thereby serving as a primer for transcription.

Several lines of evidence support this model:

(1) Several free leader RNA species of 50-90 nucleotides have been detected in the cytoplasm of MHV-infected cells. These RNAs represent discrete RNA species, some of which have dissociated from the membrane-bound transcriptional complex (Baric et al., 1985 and in this volume).

(2) A temperature-sensitive mutant has been isolated which synthesizes only the small leader RNAs but not mRNAs at the nonpermissive temperature. These data suggested that leader RNAs and mRNAs are synthesized discontinuously and require different proteins (Baric et al., 1985).

(3) During mixed infection with two different strains of MHV, the leader RNA sequences from one virus could be detected in the mRNAs of the co-infecting virus at very high frequency. This result suggests that the leader RNA represents a separate transcriptional unit which can be freely reassorted between the mRNAs of two co-infecting viruses (Makino et al., 1986).

M. M. C. Lai et al. (eds.), *Coronaviruses*
© Plenum Press, New York 1987

(4) Sequence analysis of the leader RNA and intergenic regions of the various mRNA species indicated that there is sequence homology of approximately 10 nucleotides between the 3' end of the leader RNA and the initiation sites of several mRNAs. Hence, the leader RNA could possibly bind to the negative-strand RNA template at these complementary sequences (Budzilowicz et al., 1985).

To elucidate the precise mechanism of leader RNA synthesis and priming, we cloned and sequenced the 5' end of the MHV genomic RNA, from which the leader RNA is derived. These sequences suggest a detailed mechanism for leader-primed transcription.

MATERIALS AND METHODS

Viruses and cells

The plaque-cloned JHM strain of MHV (Makino et al, 1984a, 1985) was used throughout. Viruses were propagated on DBT cells at low multiplicities of infection. Virus was harvested, purified from media, and viral RNA was prepared as previously described (Makino et al., 1984a).

cDNA cloning

cDNA cloning followed the general method of Gubler and Hoffman (1983). This procedure was modified by the use of a synthetic oligodeoxyribonucleotide #13 which is complementary to a Tl-oligonucleotide close to the 5'-end of the genome (Makino et al., 1984b) (see Results), to prime reverse transcription.

To specifically clone the 5'-end of the genomic RNA, we employed a different cloning strategy. Briefly, first-strand cDNA was prepared according to Maniatis et al. (1982), with reverse transcriptase using synthetic oligomer #14 (see Fig. 1) as a primer. The second-strand DNA was synthesized with Klenow fragment of DNA polymerase I, using as a primer, a synthetic oligomer #16 (5'-AGAGTGATTGGCGT-3') which corresponds to the 5'-end of the leader sequences (Lai et al., 1984). Double-stranded DNA was dC-tailed with terminal deoxynucleotide transferase and annealed to dG-tailed PstI-cut pBR322 plasmid. The annealed molecules were used to transform *Escherichia coli* MC 1061 as described (Dagert and Ehrlich, 1979).

DNA sequencing

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

RESULTS

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

To obtain cDNA clones covering the 5'-end of the genomic RNA, we used synthetic oligonucleotides complementary to sequences near the 5' end of the genome as primers for reverse transcription. For this purpose, we sequenced an RNase Tl-resistant oligonucleotide #3 of JHM previously mapped to gene A (Makino et al., 1984b), and generated a complementary synthetic oligodeoxyribonucleotide (5'-ATAATGGGTTTTGTATAATA-3') to serve as a primer (#13) for first strand cDNA synthesis. The binding site for this synthetic oligomer was localized to 6.2 kb from th 5'-end of the viral genomic template by

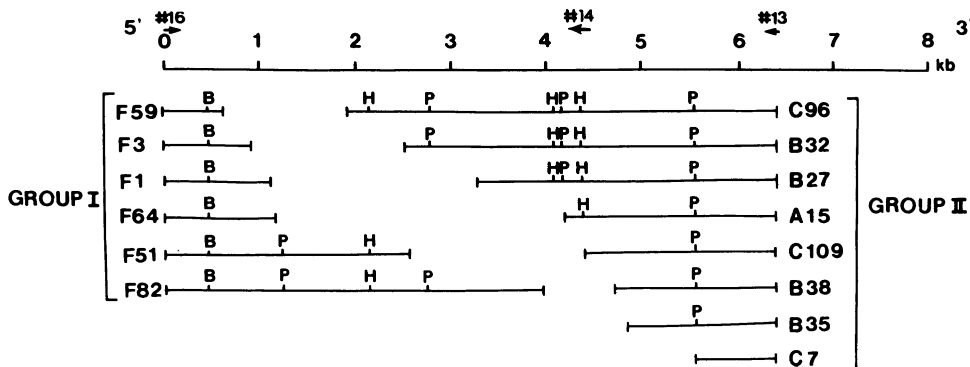

Fig. 1. Map of the 5'-end cDNA clones and location of the synthetic oligodeoxyribonucleotides used as primers for cDNA synthesis. P, PstI; H, HindIII; B, BamHI.

primer extension studies (data not shown). This analysis allowed us to map the positions of cDNA clones subsequently generated to the 5' end of gene A.

The cDNA clones obtained, designated Group II, ranged from 0.5 to 4.5 kb in length (Fig.1). None of these clones hybridized to a leader-specific probe, suggesting that they did not completely extend to the 5' end of the viral genome. To obtain clones of the 5'-most region, we generated a second primer (oligo #14) by sequencing an internal Pst I fragment from clone B27 (see Fig. 1). In addition, a third primer, a 17-mer, which represents the extreme 5'-end of the leader region, was used to initiate second strand cDNA synthesis. The resulting clones, designated Group I, ranged in size from 0.5 to 3.8 kb, and hybridized to the leader-specific probe. The largest clone (F82) overlaps with the Group II cDNA clones (Fig. 1).

The map positions of these cDNA clones were confirmed by Northern blot analysis. The unique BamHI site at the 5' end of the genome on several Group I clones (Fig. 1) was used to separate these clones into 5' and 3' fragments, which were then hybridized separately to intracellular RNA from MHV-infected cells (data not shown). These studies showed that the 5' fragment hybridized to all 7 major mRNAs, suggesting that this fragment contains the leader sequence which is present on all the mRNAs. In contrast, the 3' fragment hybridized exclusively to mRNA 1 and, thus, verified the location of these clones to gene A.

Sequence analysis and predicted secondary structure of the MHV cDNA clones

Sequence data derived from several of the leader-containing cDNA clones were obtained to provide insight into the mechanism of leader RNA synthesis and priming. These sequences are shown in Fig. 2. The leader sequences present on the genomic RNA are identical to the leader sequences present at the 5'-end of the subgenomic mRNAs (Lai et al., 1984; Spaan et al., 1983). To rule out the possibility of sequence variation in the 5'-ultimate dodecameric region of the cDNA clones derived from synthetic primer #16, we carried out primer extension reactions using another synthetic primer #19A (5'-AATGTCAGCAGCACTATGACA-3') derived from nucleotides 123-140 (see Fig. 2). The sequence of these primer-extended products was identical to that obtained from the 5' cDNA clones. In addition, we note the presence of a mirror-image direct repeat (AAAUCUAAU) at the 3'-end of the leader sequence

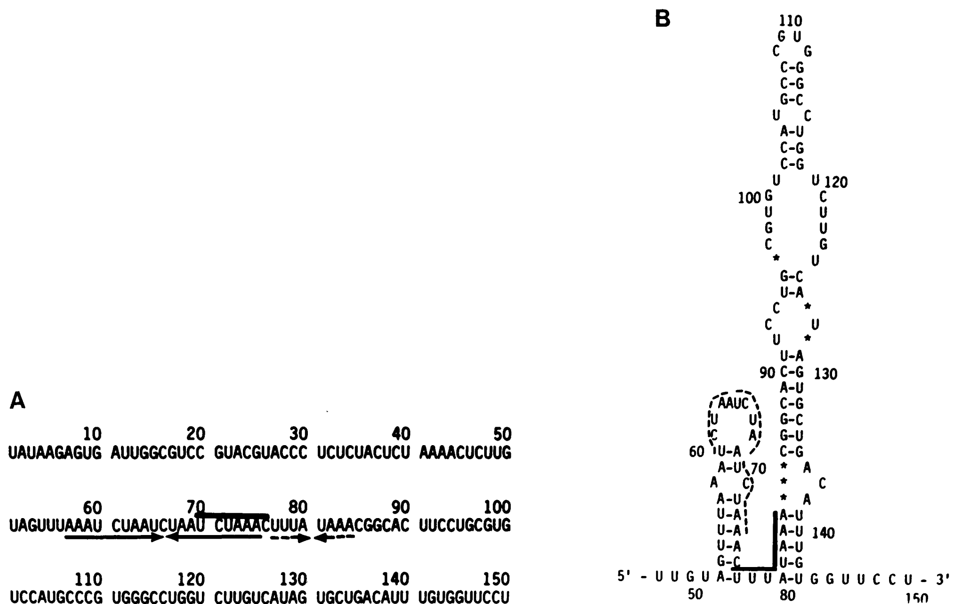

Fig. 2. (A) Sequence of the 5'-end of the MHV genomic RNA. The possible leader termination site (B) is illustrated as two hairpin loops separated by an AU-rich sequence. The UCUAA repeats are denoted by dashed lines.

from position 57 to 75 and a stretch of inverted repeats (UUUA) which follows this dyad symmetry (Fig. 2A). The significance of these structures is still not clear. The sequence also reveals the presence of two possible hairpin loop structures formed in the region of nucleotides 52-77 (G = -1.6 kcal/mole) and 80-143 (G = -18.7 kcal/mole), which are separated by an AU-rich region (UUUAUAAA) at position 78-84 (Fig. 2B). This region may serve as the termination site of leader RNA synthesis (see Discussion).

To further examine the potential secondary structure at the 5' region of the RNA genome, we extended our sequence analysis to the coding region for gene A. Figure 3 illustrates the predicted secondary structure for the MHV genome, which was compared to the published sequence of IBV (Brown et al., 1986). Both these RNAs show an extensive and stable structure preceding the first AUG start codon at position 215 (G = -42.5 kcal/mole) for MHV and G = -103.4 kcal/mole for IBV (Brown et al., 1986). The 5'-end sequence also contains three pentameric repeats (UCUAA) which have previously been described in the leader sequences of the subgenomic mRNAs (Spaan et al., 1983). One of the cDNA clones (F82) contains four of these repeats, suggesting the presence of heterogeneity at the 5'-end of the MHV genome.

Predicted amino acid sequence of the 5'-end of the MHV genome

The 5'-end gene has been suggested to code for RNA-dependent RNA polymerases. A three-frame translation of the 5' 1.2 kb sequence reveals a single long open reading frame beginning at position 215 from the 5' end of the genomic RNA (Fig. 4). The two alternative reading frames are closed by numerous termination codons. The first AUG at position 215 is at an optimal context for translation as defined by Kozak's rules (Kozak, 1983) and is followed by several additional in-frame methionine residues. These internal methionine residues may provide alternative initiation sites for translation and result in the overlapping proteins observed by in vitro translation studies (Leibowitz et al., 1982). The amino terminal 150 amino acids of this predicted protein are primarily

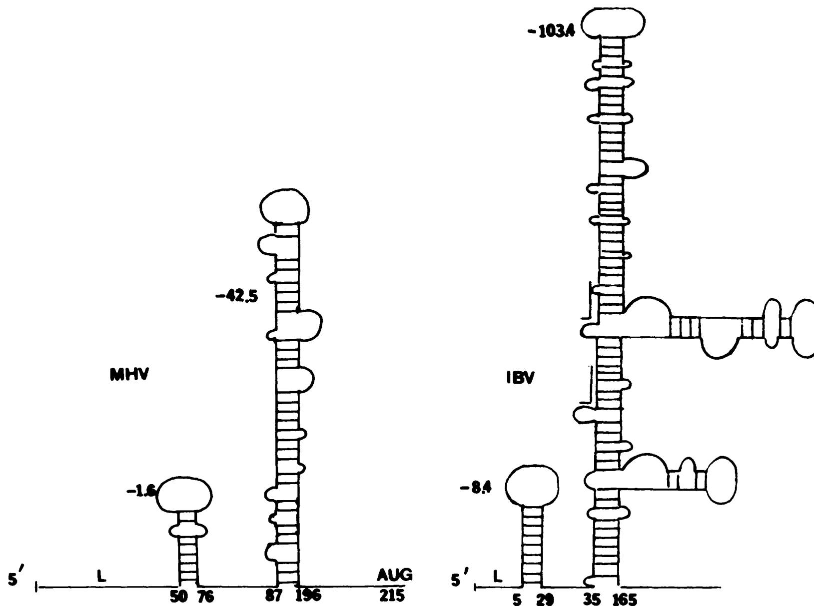

Fig. 3. Predicted secondary structure at the 5' region of the MHV-JHM and IBV (Brown et al., 1986) RNA genomes.

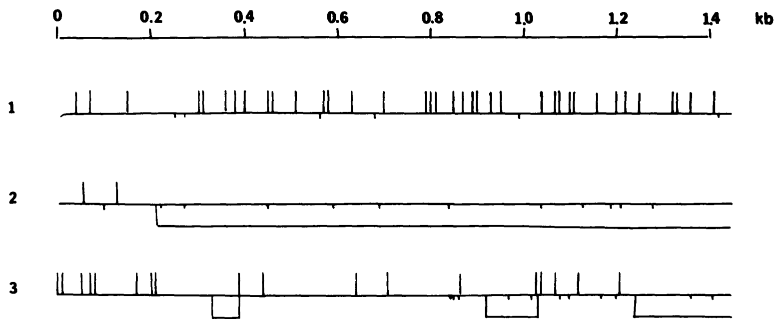

Fig. 4. Three-frame translation of the 5' 1.4 kb sequence. Long vertical lines above the baseline represent termination codons, and short vertical lines below the baseline are potential start codons. A single long open reading frame beginning at position 215 is observed in frame 2.

hydrophobic and is consistent with the membrane-associated nature of the RNA polymerase (Brayton et al., 1984). In addition, we note the presence of a potential zinc-binding domain at amino acids 196 to 213, formed by the specific juxtaposition of cysteine and histidine residues. Zinc-binding domains have been shown to be a common feature of proteins involved in nucleic acid binding or gene regulation (Berg, 1986). The significance of this structure in MHV transcription remains to be elucidated.

DISCUSSION

The coronavirus leader-primed transcription model proposes a novel mechanism for fusing two noncontiguous RNA segments in the cytoplasm. In this model, free leader RNA derived from the 5'-end of the genomic RNA dissociates from the negative-strand template and then rebinds at the initiation sites for the mRNAs, thereby serving as a primer for transcription (Baric et al., 1983). Several issues concerning this transcription process remained to be resolved. What is the structure and mechanism of synthesis of free leader RNA? How does the free leader recognize the various initiation sites on the RNA template and what is the mechanism of fusion? What regulates the transcription of the various mRNA species which are synthesized at different rates? The sequences at the extreme 5'-end of the genome provided answers to some of these questions.

The 5'-end sequences of the viral genome reveal the presence of two possible hairpin loop structures separated by an AU-rich region near the 3'-end of the leader sequences. Since both regions of secondary structure and AU-rich areas are involved in transcriptional pausing and termination (Henikoff et al., 1983; Mills et al., 1978; Zaret and Sherman, 1982), we postulate that these structures may serve as the termination site of leader RNA synthesis. This model is supported by the detection of several free leader-containing RNA species in MHV-infected cells (Baric et al., this volume) which may have resulted from termination within this region.

Comparison of 5'-end sequences of the genomic RNA with the intergenic regions of various mRNAs reveals a higher degree of homology than previously recognized (Fig. 5). Previously, the leader-body junction site for the different subgenomic mRNAs was determined by the 3'-most point of the intergenic regions where the homology between the 3'-end of the leader RNA and the intergenic regions end (Budzilowicz et al., 1985; Spaan et al., 1983). The 5'-end sequence reveals that, in addition to this homology, there is downstream homology which extends for an additional five nucleotides (UAAAC). Thus, the potential base pairing between the leader RNA and intergenic start sites varies from 7 to 18 nucleotides. The presence of homology 3' to the leader/body fusion sites has also been noted in another coronavirus, IBV (Brown et al., 1986). Hence, the free leader RNA of JHM may be 77 nucleotides or longer to utilize the 3'-homology.

Based upon these data, we propose a model for leader-primed transcription (Fig. 6). In this model, a free leader RNA is synthesized from the 3'-end of the negative-strand template and binds to the intergenic start site of each gene. The leader RNA undergoes endonucleolytic cleavage at the position of a base mismatch. Transcription then proceeds utilizing the nicked leader as the primer and the negative-strand RNA as the template for the synthesis of mRNA body

Intergenic site		No. bases of homology	Ratio of RNA amount
3-4	.........*AGAA*AAUCUAAAC*AAUUUAUAGC*......	9	1.69
4-5	.........*ACUAGU*UCUAAAC*CUCAUCUUAA*......	7	0.86
5-6	...*AUGA*UAAUCUAAUCCAAAC*AUUAUG*..........	10 + 4	31.5
6-7	....*UGAG*AAUCUAAUCUAAACUUUA*AGGAUG*......	18	100
Leader	....AUCUAAUCUAAUCUAAACUUUAUAAACG...... (60, 70, 80)		

Fig. 5. Comparison of the leader sequence at the 5' end of the genomic RNA with intergenic sequences for mRNAs 4 (Skinner and Siddell, 1985), 5 (Skinner et al., 1985), 6 and 7 (Skinner and Siddell, 1983). The number of bases of homology and the relative ratio of the mRNAs (Leibowitz et al., 1981) are included.

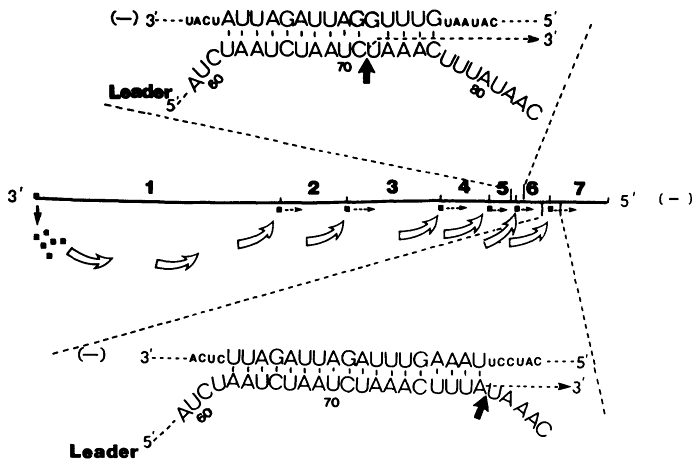

Fig. 6. Postulated model of leader-primed transcription. The solid squares represent free leader RNA which binds to the initiation points of mRNAs on the negative-strand RNA template. The expanded regions show the sequences of the intergenic start sites of mRNAs 6 and 7 and the mechanism of leader RNA binding. The solid arrows represent the cleavage point of the leader RNA at the first base mismatch. Transcription then initiates from the 3'-end of the cleaved primer.

sequences. This model predicts that the leader-body junction sites may be different for each mRNA, depending upon the position of the mismatched nucleotide. The degree of homology between the leader and intergenic start sites of the subgenomic mRNAs strongly correlates with the abundance of individual mRNA species (Fig. 5). Thus, the rate of mRNA transcription may be regulated by the degree of homology, which, in turn, influences the strength or stability of binding of the leader RNA to the intergenic sites.

We also note that the canonical sequence (UCUAAAC), which is present at position 70-76 from the 5' end of the RNA genome, is nearly repeated two nucleotides downstream at position 79 to 85 (UAUAAAC) (see Fig. 2). This sequence is the consensus sequence within each of the intergenic start sites. Thus, it may also represent an intergenic start site for mRNA 1 transcription. If this is indeed the case, mRNA 1 of MHV may be nine nucleotides shorter than the virion genomic RNA. Similarly, we observe a region of homology between the 3'-end of the leader and the 5' noncoding region of IBV (Brown et al., 1986) in which the sequnce ACUUAA is repeated eight nucleotides downstream from the 3'-end of the IBV leader. This mechanism for mRNA 1 transcription of MHV is strongly supported by the recent finding that the sequence of an MHV defective-interfering RNA, DIssA, contains the wild-type MHV leader/body junction site and 5' gene A sequences, but is missing the precise nine nucleotides which are predicted by our model (Makino, S., unpublished observations). This utilization of alternate consensus sequences during transcription and replication may distinguish the fates of mRNA 1 and genomic RNA.

ACKNOWLEDGMENTS

We are grateful to David Vannier and Victoria Aguinaldo for excellent technical assistance, Carol Flores for editorial help and Susan Baker for critical reading of the manuscript. This work was supported by U.S. Public Health Research Grants AI 19244 and NS 18146, National Multiple Sclerosis Society Research Grant RG 1449, and National Science Foundation Grant PCM-4507.

REFERENCES

Baric, R.S., Stohlman, S.A., and Lai, M.M.C. (1983). Characterization of replicative intermediate RNA of mouse hepatitis virus: Presence of leader RNA sequences on nascent chains. J. Virol. 48, 633-640.

Baric, R.S., Stohlman, S.A., Razavi, M.K., and Lai, M.M.C. (1985). Characterization of leader-related small RNAs in coronavirus-infected cells: Further evidence for leader-primed mechanism of transcription. Virus Res. 3, 19-33.

Berg, J.M. (1986). Potential metal-binding domains in nucleic acid binding proteins. Science 232, 485-487.

Brayton, P.R., Ganges, R.G., and Stohlman, S.A. (1981). Host cell nuclear function and murine hepatitis virus replication. J. Gen. Virol. 56, 457-460.

Brayton, P.R., Lai, M.M.C., Patton, C.D., and Stohlman, S.A. (1982). Characterization of two RNA polymerase activities induced by mouse hepatitis virus. J. Virol. 42, 847-853.

Brown, T.D.K., Boursnell, M.E.G., Binns, M.M., and Tomley, F.M. (1986). Cloning and sequencing of 5' terminal sequences from avian infectious bronchitis virus genomic RNA. J. Gen. Virol. 67, 221-228.

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

Dagert, M., and Ehrlich, S.D. (1979). Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. Gene 6, 23-29.

Gubler, U., and Hoffman, B.J. (1983). A simple and very efficient method for generating cDNA libraries. Gene 25, 263-269.

Henikoff, S., Kelly, J.D., and Cohen, E.H. (1983). Transcription terminates in yeast distal to a control sequence. Cell 33, 607-614.

Jacobs, L., Spaan, W.J.M., Horzinek, M.C., and Van der Zeijst, B.A.M. (1981). Synthesis of subgenomic mRNAs of mouse hepatitis virus is initiated independently: Evidence from U.V. transcriptional mapping. J. Virol. 34, 401-406.

Kozak, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Micro. Rev. 47, 1-45.

Lai, M.M.C., Baric, R.S., Brayton, P.R., and Stohlman, S.A. (1984). Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus - A cytoplasmic RNA virus. Proc. Natl. Acad. Sci. USA 81, 3626-3630.

Lai, M.M.C., Brayton, P.R., Armen, R.C., Patton, C.D., Pugh, C., and Stohlman, S.A. (1981). Mouse hepatitis virus A59: mRNA structure and genetic localization of the sequence divergence from hepatotropic strain MHV-3. J. Virol. 39, 823-834.

Lai, M.M.C., Patton, C.D., and Stohlman, S.A. (1982). Replication of mouse hepatitis virus: Negative-stranded RNA and replicative form RNA are of genomic length. J. Virol. 44, 487-492.

Leibowitz, J.L., Wilhemsen, K.C., and Bond, C.W. (1981). The virus-specific intracellular RNA species of two murine coronaviruses: MHV-A59 and MHV-JHM. Virology 114, 39-51.

Makino, S., Fujioka, N. and Fujiwara, K. (1985). Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus. J. Virol. 54, 329-336.

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

Makino, S., Taguchi, F., and Fujiwara, K. (1984a). Defective interfering particles of mouse hepatitis virus. Virology 133, 9-17.

Makino, S., Taguchi, R., Hirano, N., and Fujiwara, K. (1984b). Analysis of genomic and intracellular viral RNAs of small plaque mutants of mouse hepatitis virus, JHM strain. Virology 139, 138-151.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). "Molecular Cloning-A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Mills, D.R., Dabkin, C. and Kramer, F.R. (1978). Template-determined, variable rate of RNA chain elongation. Cell 15, 541-550.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Skinner, M.A., Ebner, D., and Siddell, S.C. (1985). Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. J. Gen. Virol. 66, 581-592.

Skinner, M.A., and Siddell, S.G. (1983). Coronavirus JHM: Nucleotide sequence of the mRNA that encodes nucleocapsid protein. Nucleic Acids Res. 15, 5045-5054.

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

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., Van der Zeijst, B.A.M., and Siddell, S.G. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. EMBO J. 2, 1839-1844.

Wilhelmsen, K.C., Leiboweitz, J.L., Bond, C.W., and Robb, J.A. (1981). The replication of murine coronaviruses in enucleated cells. Virology 110, 225-230.

Zaret, K.S., and Sherman, F. (1982). DNA sequence required for efficient transcription termination in yeast. Cell 28, 563-573.