Primary Structure and Translation of a Defective Interfering RNA of Murine Coronavirus

SHINJI MAKINO,[1] CHIEN-KOU SHIEH, LISA H. SOE, SUSAN C. BAKER, AND MICHAEL M. C. LAI

Department of Microbiology, University of Southern California, School of Medicine, Los Angeles, California 90033

An intracellular defective-interfering (DI) RNA, DIssE, of mouse hepatitis virus (MHV) obtained after serial high multiplicity passage of the virus was cloned and sequenced. DIssE RNA is composed of three noncontiguous genomic regions, representing the first 864 nucleotides of the 5' end, an internal 748 nucleotides of the polymerase gene, and 601 nucleotides from the 3' end of the parental MHV genome. The DIssE sequence contains one large continuous open reading frame. Two protein products from this open reading frame were identified both by *in vitro* translation and in DI-infected cells. Sequence comparison of DIssE and the corresponding parts of the parental virus genome revealed that DIssE had three base substitutions within the leader sequence and also a deletion of nine nucleotides located at the junction of the leader and the remaining genomic sequence. The 5' end of DIssE RNA was heterogeneous with respect to the number of UCUAA repeats within the leader sequence. The parental MHV genomic RNA appears to have extensive and stable secondary structures at the regions where DI RNA rearrangements occurred. These data suggest that MHV DI RNA may have been generated as a result of the discontinuous and nonprocessive manner of MHV RNA synthesis. © 1988 Academic Press, Inc.

INTRODUCTION

Mouse hepatitis virus (MHV), a member of the Coronaviridae, contains a single-stranded, positive-sense RNA of approximately 6×10^6 Da (Lai and Stohlman, 1978; Wege *et al.*, 1978). In infected cells, the genomic RNA of MHV is first translated into an RNA-dependent RNA polymerase (Brayton *et al.*, 1982, 1984; Mahy *et al.*, 1983) which is responsible for the synthesis of a genomic-sized negative-stranded RNA (Lai *et al.*, 1982b). The negative-stranded RNA then serves as the template for the synthesis of six subgenomic and a genomic-sized mRNA (Brayton *et al.*, 1984; Lai *et al.*, 1982b). These mRNAs are arranged in the form of a 3' coterminal "nested" set, i.e., the sequence of each mRNA is contained entirely within the next larger mRNA (Lai *et al.*, 1981; Leibowitz *et al.*, 1981). In addition, each mRNA has a common leader sequence, which is derived from the 5' end of the genome (Lai *et al.*, 1982a, 1983, 1984; Spaan *et al.*, 1983). Several pieces of evidence demonstrated that MHV utilizes a novel mechanism of leader RNA-primed transcription, in which a free leader RNA species derived from the 5' end of genomic RNA is utilized as a primer for the transcription of subgenomic mRNAs (Baric *et al.*, 1983, 1985; Makino *et al.*, 1986b).

Another unusual feature of coronavirus RNA synthesis is that the virus undergoes RNA–RNA recombination at a very high frequency (Makino *et al.*, 1986a). The

unusually high frequency, approaching 10% under some circumstances (Makino *et al.*, 1986a), of coronavirus RNA recombination suggests that discontinuous RNA transcripts might be generated during coronavirus RNA synthesis. These incomplete RNA intermediates may rejoin the original or different RNA template to continue RNA synthesis, resulting in RNA recombination in the latter case. The detection of such RNA intermediates in MHV-infected cells (Baric *et al.*, 1985, 1987) suggests that coronavirus genomic RNA synthesis involves a discontinuous and nonprocessive mechanism, which may account for the high frequency of recombination via a copy choice mechanism.

Defective-interfering (DI) particles are naturally occuring deletion mutants that have been described for many virus groups. Characteristically, DI particles (a) lack part of the viral genome, (b) contain normal viral structural proteins, (c) replicate only with the aid of a helper standard virus, and (d) interfere with replication of homologous standard virus. Deletion of genomic sequence can occur in various regions of the genome; however, all of the DI RNAs apparently retain signals for RNA replication since they can be replicated in the presence of helper virus. The generation of DI RNA can be viewed as the result of abnormal RNA replication or illigitimate RNA recombination. Therefore, the structure of DI RNA is of particular interest in elucidation of the mechanism of viral RNA replication and recombination.

We have previously reported the generation of DI particles during high multiplicity passages of the JHM
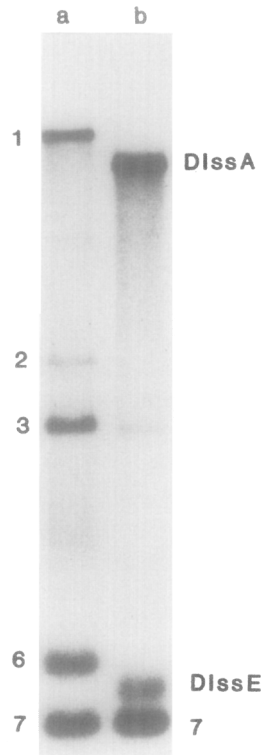

FIG. 1. Intracellular RNA species in DI-infected cells. ^{32}P-Labeled RNA from MHV-JHM-infected cells (a) and DI particles-infected cells (b) were electrophoresed in a 1% agarose gel without denaturation. Numbers 1, 2, 3, 6, and 7 represent the major MHV-JHM-specific mRNA species.

strain of MHV (MHV-JHM) (Makino et al., 1984a). In DI-infected cells, the synthesis of most of the standard viral mRNAs is inhibited. Instead, three distinct virus-specific RNA species could be detected (Makino et al., 1985) (Fig. 1). The first species, DIssA, is equivalent to DI virion RNA in length and is eventually incorporated into virus particles. This RNA differs from the standard virus genome in that it contains multiple deletions distributed throughout the genome, except for the 5' and 3' ends of the genomic RNA (Makino et al., 1985), which encode RNA polymerase (gene A) and nucleocapsid (N) protein, respectively. Surprisingly, DIssA RNA can replicate by itself in the absence of helper virus infection, suggesting that DIssA codes for functional RNA polymerases (Makino et al., 1988). Thus, DIssA is not a defective RNA in a strict sense. The second major RNA species found in DI-infected cells is indistinguishable from the mRNA 7 made by the standard virus. The synthesis of this mRNA and its product N protein is not inhibited in DI-infected cells. The third RNA species is a novel single-stranded polyadenylated DI RNA species of varying size. Oligonucleotide fingerprinting studies suggest that it represents se-

quences derived from various noncontiguous parts of the genome. The size of this RNA varies with the DI passage level (Makino et al., 1985). One of these RNAs, DIssE, which is the smallest DI RNA detected, has been analyzed in greater detail (Makino et al., 1988). In contrast to DIssA, DIssE RNA synthesis requires helper virus coinfection (Makino et al., 1988). Only a trace amount of it is incorporated into virus particles to serve as a template for RNA replication (Makino et al., 1988). Thus, it may lack packaging signals. On the other hand, since it is efficiently replicated in DI-infected cells, DIssE RNA must contain the sequences essential for viral RNA replication.

In the present study, we analyzed the primary structure of DIssE RNA. The results revealed that DIssE consists of three noncontiguous regions of MHV-JHM genomic RNA, including 5' end leader RNA and the 3' end of genomic RNA. One large open reading frame (ORF) was demonstrated and the product of this ORF was identified both in infected cells and by in vitro translation. Possible mechanisms of DI RNA generation are discussed.

MATERIALS AND METHODS

Viruses and cell culture

MHV-JHM was used as a nondefective standard virus. Serially passaged MHV-JHM stock at passage level 17 was used as the source of DI particles (Makino et al., 1985). All viruses were propagated in DBT cells as described previously (Makino et al., 1984a).

Preparation of virus-specific intracellular RNA

MHV-specific intracellular RNA was extracted by procedures described previously (Makino et al., 1984b). Poly(A)-containing RNA was obtained by oligo(dT)-cellulose column chromatography (Makino et al., 1984b).

Agarose gel electrophoresis

^{32}P-Labeled virus-specific RNA was analyzed by electrophoresis on 1% agarose gels without denaturing as described previously (Makino et al., 1988). Poly(A)-containing RNA was purified by preparative gel electrophoresis in 1% urea–agarose gels as previously described (Makino et al., 1984a). The RNA was eluted from gel slices by the methods of Langridge et al., (1980).

cDNA cloning of DIssE

cDNA cloning followed the general method of Gubler and Hoffman (1983). Five hundred nanograms of oli-

go(dT)$_{12-18}$ was mixed with 2 μg of gel-purified DIssE RNA in 10 μl of distilled water. The RNA and oligo(dT) mixture was heated at 70° for 3 min and chilled quickly. The RNA–DNA hybrid was then incubated in 50 μl of first-strand cDNA synthesis buffer containing 60 units of RNasin (Promega Biotec), 50 mM Tris–HCl (pH 8.3 at 42°),100 mM KCl, 10 mM MgCl$_2$, 10 mM DTT, 1.25 mM each of dATP, dCTP, dGTP, and TTP, and 20 units of avian myeloblastosis virus reverse transcriptase (Life Science) at 42° for 1 hr. The cDNA synthesis was stopped by adding 4.4 μl of 250 mM EDTA. Nucleic acids were extracted with phenol–chloroform–isoamyl alcohol and precipitated with ethanol.

Second-strand synthesis was carried out in a reaction volume of 100 μl containing 20 mM Tris–HCl (pH 7.4), 5 mM MgCl$_2$, 100 mM KCl, 50 μg/ml of BSA, 10 mM (NH$_4$)$_2$SO$_4$, 0.15 mM β-NAD, 100 μM dNTPs, 25 units of DNA polymerase I, 2 units of *Escherichia coli* DNA ligase, 0.8 units of RNase H, and the product from the first strand reaction. The mixture was incubated at 12° for 1 hr, and then at 22° for 1 hr. The reaction was stopped by adding 8.7 μl of 250 mM EDTA, and products were extracted with phenol–chloroform–isoamyl alcohol, and precipitated with ethanol. Double-stranded DNA was dC-tailed in a 12-μl reaction mixture containing 10 units of terminal transferase, 200 mM potassium cacodylate, 0.5 mM CoCl$_2$, 25 mM Tris–HCl (pH 6.9), 2 mM DTT, 250 μg/ml BSA, and 50 μM dCTP at 37° for 4 min. The dC-tailed double-stranded DNA was annealed to 200 ng of dG-tailed *Pst*I-cut pBR322 plasmid in 20 μl of a buffer containing 10 mM Tris–HCl (pH 7.4), 100 mM NaCl, and 0.25 mM EDTA. The DNA mixture was heated at 68° for 5 min and then cooled slowly overnight for annealing. The annealed molecules were used to transform *E. coli* MC1061 as described (Dagert and Ehrlich, 1979).

Identification of large cDNA clones containing DIssE sequence

[32]P-Labeled MHV-JHM gene A cDNA clones C96 and F82 (Shieh *et al.*, 1987) and 5' end ^{32}P-labeled leader-specific 72-mer derived from leader sequence of MHV (Lai *et al.*, 1984) were used for colony hybridization (Shieh *et al.*, 1987) to isolate DIssE-specific cDNA clones. Colonies yielding a strong signal were further analyzed by Southern hybridization (Maniatis *et al.*, 1982).

Primer extension

The gel-purified RNAs were incubated in 8 μl of distilled water containing 10 mM methyl mercury. After 10 min incubation at room temperature, RNA was incu-

bated in 50 μl of first-strand cDNA synthesis buffer with 28 mM β-mercaptoethanol and 5' end-labeled oligo-deoxyribonucleotides at 42° for 1 hr. Reaction products were extracted with phenol–chloroform–isoamyl alcohol, precipitated with ethanol, and analyzed by electrophoresis on 6% polyacrylamide gels containing 8.3 M urea and were eluted from the gels according to the published procedures (Maxam and Gilbert, 1980).

DNA sequencing

Sequencing was carried out by Sanger's dideoxyribonucleotide chain termination method (Sanger *et al.*, 1977) and Maxam–Gilbert chemical modification procedure (Maxam and Gilbert, 1980), as described previously (Soe *et al.*, 1987). Sequence analysis and predicted RNA secondary structures were obtained with the Intelligenetics sequencing program.

In vitro translation

An mRNA-dependent rabbit reticulocyte lysate (New England Nuclear) was used as previously described (Soe *et al.*, 1987).

Antisera

A monoclonal antibody, J.3.3, directed against the MHV-JHM N protein has been described (Fleming *et al.*, 1983). The anti-p28 antibody was generated in rabbits against a synthetic peptide representing a portion of the MHV-JHM p28 protein (Soe *et al.*, 1987) and will be described in detail elsewhere (S. C. Baker *et al.*, manuscript in preparation).

Labeling of intracellular proteins, immunoprecipitation, and SDS–polyacrylamide gel electrophoresis

DBT cells were infected with either wild type MHV-JHM or MHV-JHM containing DI particles at 2 PFU per cell. At 7.5 hr postinfection, cells were labeled in methionine-free medium containing 30 μCi of L-[35S]methionine/ml (ICN translabel) for 30 min. Cell extracts were prepared by treatment with lysolecithin (L-α-lysophosphatidylcholine, palmitoyl; Sigma) at 125 μg/ml for 1 min at 4°. The treated cells were scraped in 300 μl HND buffer (0.1 M HEPES, pH 8.0, 0.2 M NH$_4$Cl, 0.005 M DTT), disrupted by pipetting with a Pasteur pipet, and then centrifuged at 800 g for 5 min to remove nuclei and cell debris. The resulting supernatant was used for immunoprecipitation.

Immunoprecipitation was performed by the methods of Kessler (1981). The cell-free extracts were incubated with 3 μl of antisera for 4 hr at 4°. The antigen–antibody

complexes were collected by binding to Pansorbin (Calbiochem, La Jolla, CA) and washed three times with washing buffer (50 mM Tris–HCl, pH 7.4, 150 mM NaCl, 5 mM EDTA, and 0.5% NP-40) and eluted by boiling for 2 min in electrophoresis sample buffer (0.1 M β-mercaptoethanol, 1% SDS, 0.08 M Tris–HCl, pH 6.8, and 10% glycerol). The bacteria were removed by centrifugation and proteins were analyzed by electrophoresis on 5 to 15% SDS–polyacrylamide gels (Laemmli, 1970).

RESULTS

cDNA cloning and sequencing of DIssE RNA

To understand the primary structure of DIssE RNA, DIssE-specific cDNA clones were generated according to the general method of Gubler and Hoffman (1983), using oligo(dT) as a primer and gel-purified DIssE RNA. Since previous oligonucleotide fingerprinting analysis suggested that DIssE RNA contains the leader sequence and the 5' end region of genomic sequence (Makino et al., 1985), cDNA clones were screened by colony hybridization using 5' end-labeled, leader-specific 72-mer, and two cDNA clones F82 and C96, which correspond to the 5' end of genomic RNA of MHV-JHM (Shieh et al., 1987). Several large cDNA clones were isolated and their structure was further analyzed. A diagram representing the structure of the DIssE genome and that of MHV-JHM genomic RNA and the strategy used for sequencing the cDNA clones are shown in Fig. 2. The DIssE sequence obtained is shown in Fig. 3.

Sequence analysis of DIssE cDNA clones revealed that DIssE RNA consists of three different regions of MHV-JHM genomic RNA. The first region represents 864 nucleotides from the 5' end of the genomic RNA. The second region, 748 nucleotides in length, is a region within the polymerase gene that corresponds to the region at 3.3 to 4 kb from the 5' end of genomic RNA (Shieh, unpublished observation), and the third region contains a sequence of 601 nucleotides derived from the extreme 3' end of the genomic RNA. The entire sequence of DIssE RNA is identical to that of the corresponding regions of MHV genomic RNA (Skinner and Siddell, 1983; Soe et al., 1987; Shieh et al., unpublished data), with some exceptions in the leader sequence region (see below).

The cDNA clones obtained does not appear to have a complete sequence at its extreme 5' end. To understand the complete 5' end sequence of DIssE, we performed primer-extension studies on DIssE RNA using a specific primer (5'-AATGTCAGCACTATGACA-3') complementary to nucleotides 123–140 from the 5'

end of the genome of MHV-JHM (Shieh et al., 1987). The 5' end-labeled primer was hybridized to gel-purified DIssE RNA and extended with reverse transcriptase. Primer extension products were then analyzed by electrophoresis on 6% polyacrylamide gels containing 8 M urea. As shown in Fig. 4A, two cDNA products of 136 and 131 nucleotides were obtained, indicating heterogeneity at the 5' end sequence of DIssE. These primer-extended products were sequenced by the Maxam–Gilbert method. The sequences of both cDNA products were identical except that the faster migrating cDNA products contained three UCUAA repeats at the 3' end of the leader sequence, while the slower migrating species contained four UCUAA repeats (Fig. 4B). In addition, the 5' end sequences of DIssE and MHV-JHM genomic RNA showed several differences. Within the leader sequence, 3 bases were substituted in DIssE RNA (Fig. 4B, asterisks) and nine nucleotides (UUUAU-AAAC) were deleted in DIssE at the junction between the leader RNA and the remaining genomic sequences. The significance of the heterogeneity in the number of UCUAA repeats and of the nine-nucleotide deletion will be discussed below.

Translation of DIssE RNA in vitro and in vivo

Another significant feature of DIssE RNA is the presence of a single large ORF (Fig. 3). This ORF is expected to share amino acid sequence identity with three different regions of the standard MHV-JHM. The first 218 amino acids correspond to the N terminus of the MHV polymerase. This region represents the part of the N-terminus of the polymerase protein which is cleaved into a p28 protein (Denison and Perlman, 1986; Soe et al., 1987). The following 250 amino acids were derived from the region of the polymerase at 3.3 to 4 kb from the 5' end of the genome. The 3' end region of the ORF of DIssE RNA is the same as the ORF utilized for the N protein (Skinner and Siddell, 1983). Thus, the predicted product of this ORF should contain the N-terminus of p28 and the C-terminus of the N protein. The predicted molecular weight mass of this ORF product is 62,538.

To examine whether the ORF of DIssE RNA is utilized for translation, we first performed in vitro translation in a rabbit reticulocyte lysate of DIssE RNA purified from the DI-infected cells. Two proteins with an apparent molecular mass of approximately 88,000 (88K) and 79,000 (79K) were detected (Fig. 5A). Both proteins were immunoprecipitated with anti-N protein monoclonal antibody and anti-p28 antibody (Fig. 5A, lanes 2 and 3). Therefore, these two proteins were likely the translation products of DIssE RNA. A minor band of ap-
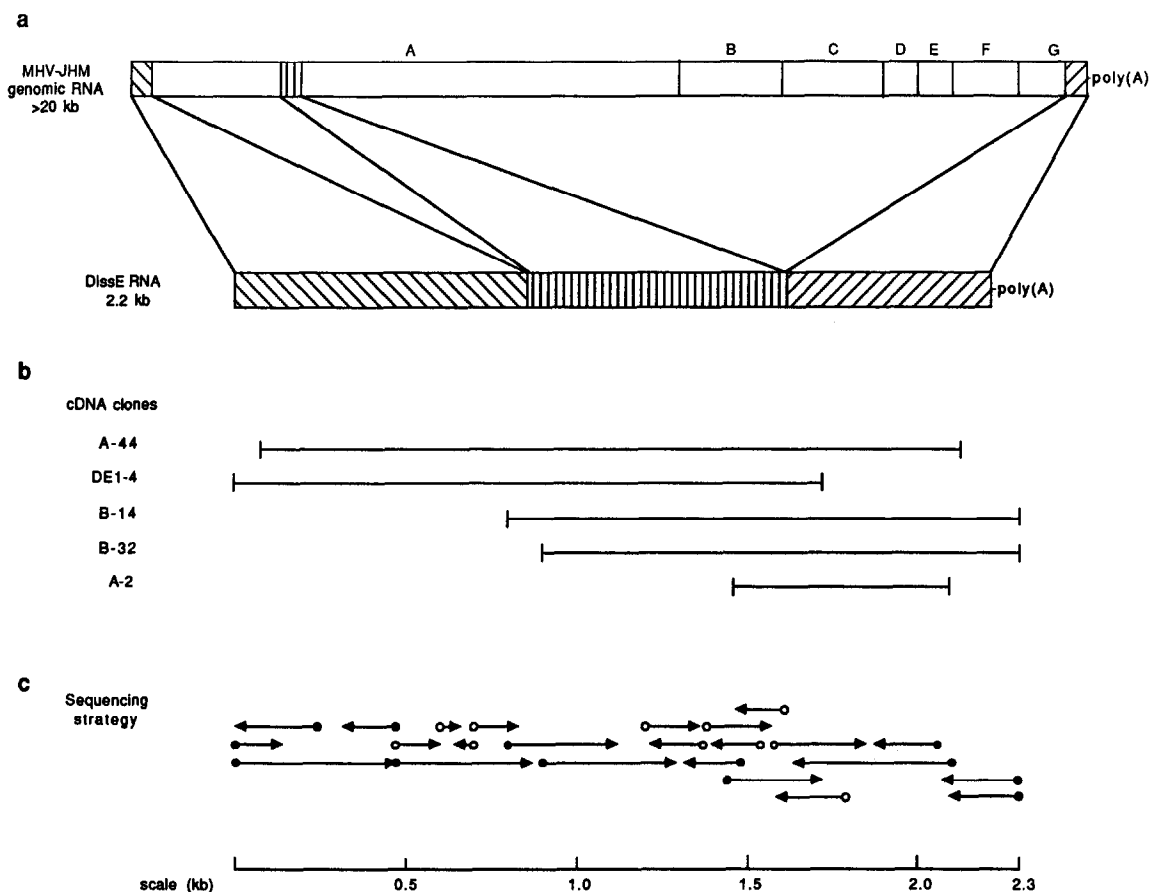
FIG. 2. Diagram of the structure of DIssE RNA and the strategy used for sequencing DIssE cDNA clones. (a) A comparison between the sequence of DIssE RNA and that of the standard MHV-JHM genomic RNA. A–G represent the seven genes of MHV (Lai *et al.*, 1981). (b) Structure of DIssE-specific cDNA clones. (c) Strategy for sequencing of DIssE. Arrows starting with solid circles indicate DNA sequenced by the Maxam–Gilbert chemical method with 3′ end-labeled DNA. Arrows starting with open circles indicate DNA sequenced by the dideoxy method.

proximately 60 kDa had the same electrophoretic mobility as the N protein of MHV-JHM, and was precipitated with anti-N monoclonal antibody, but not with anti-p28 antibody (Fig 5A, lanes 2 and 3). Thus, this protein is most likely the N protein translated from the contaminated mRNA 7 in the DIssE RNA preparation.

The synthesis of DIssE-specific protein in DI-infected cells was then examined. DBT cells were mock-infected (Fig. 5B, lanes 1 and 4), infected with MHV-JHM (Fig. 5B, lanes 2 and 5), or infected with MHV-JHM containing DI particles (Fig. 5B, lanes 3 and 6). Both 88K and 79K proteins were specifically immunoprecipitated with anti-N monoclonal antibody and anti-p28 antibody from DI-infected cells. The amount of these two proteins was low as compared to the N protein. Nevertheless, they were reproducibly detected in DI-infected cells. Thus, the DIssE RNA is a functional mRNA. The relationship between the two protein species detected is not clear. The discrepancy between the predicted and observed molecular weights of the translation

products of DIssE could be due to post-translational modification of the protein or aberrant migration of the protein. A small amount of p28 was immunoprecipitated with anti-p28 antibody in MHV-JHM-infected cells (Fig. 5B, lane 5). However, this protein was hardly detectable in DI-infected cells (Fig. 5B, lane 6). The absence of detectable amount of p28 in DI-infected cells may be due to the inhibition of MHV-JHM genomic RNA synthesis by DI particles (Makino *et al.*, 1985).

Possible secondary structure at the DI RNA rearrangment sites

Sequence analysis revealed that DIssE RNA consisted of three noncontiguous regions of MHV-JHM genomic RNA. We have previously proposed that coronavirus RNA synthesis proceeds by a discontinuous, nonprocessive mechanism, being interrupted at sites with hairpin loops (Baric *et al.*, 1987). This transcriptional interruption could account for the generation of

5'-TATAAGAGTGAATGGCGTCCGTACGTACCCAATCTACTCTAAAACTCTTGTAGTTTAAATCTAATCTAATCTAATCTAAACGGC   84

```
                MPVGLVLSZ
85  ACTTCCTGCGTGTCCATGCCCGTGGGCCTGGTCTTGTCATAGTGCTGACATTTGTGGTTCCTTGACTTTCTGTCTCTGCCAGTG    168

 1                                              MAKMGKYGLGFKWA                          14
169  ACGTGTCCATTCGGCGCCAGCAGCCCACCCATAGGTTGCATAATGGCAAAGATGGGCAAATACGGTCTCGGCTTCAAATGGGCC   252

15  PEFPWMLPNASEKLGNPERSEEDGFCPS                                                          42
253  CCAGAATTTCCATGGATGCTTCCGAACGCATCGGAGAAGTTGGGTAACCCTGAGAGGTCAGAGGAGGATGGGTTTTGCCCCTCT   336

43  AAQEPKVGKTLVNHVRVDCSRLPALEC                                                            70
337  GCTGCGCAAGAACCGAAAGTTAAAGGAAAAACTTTGGTTAATCACGTGAGGGTGGATTGTAGCCGGCTTCCAGCCTTGGAGTGC   420

71  CVQSAIIRDIFVDEDPQKVEASTMMALQ                                                           98
421  TGTGTTCAGTCCGCCATAATCCGTGATATTTTTGTTGACGAGGATCCCCAGAAGGTGGAGGCCTCGACTATGATGGCATTGCAG   504

99  FGSAVLVKPSKRLSVQAWAKLGVLPKTP                                                          126
505  TTCGGTAGTGCTGTCTTGGTCAAGCCATCCAAGCGCTTGTCTGTTCAGGCATGGGCTAAGTTGGGTGTGCTGCCTAAAACTCCG   588

127  AMGLFKRFCLCNTRECVCDAHVAFQLFT                                                          154
589  GCCATGGGGTTGTTCAAGCGCTTCTGCCTGTGTAACACCAGGGAGTGCGTTTGTGACGCCCACGTGGCCTTTCAACTTTTTACG   672

155  VQPDGVCLGNGRFIGWFVPVTAIPEYAK                                                          182
673  GTCCAGCCCGATGGTGTATGCCTGGGTAACGGCCGTTTTATAGGCTGGTTCGTTCCAGTCACAGCCATACCGGAGTATGCGAAG   756

183  QWLQPWSILLRKGGNKGSVTSGHFRRAV                                                          210
757  CAGTGGTTGCAACCCTGGTCCATCCTTCTTCGTAAGGGTGGTAACAAAGGGTCTGTGACATCCGGCCATTCCCGCCGCGCTGTT   840

211  TMPVYDFNATDVVYADENQDDDADDPVV                                                          238
841  ACCATGCCTGTGTATGACTTTAATGCAACAGATGTTGTATATGCAGATGAAAACCAAGATGATGATGCTGACGATCCTGTAGTC   924
                        ▲

239  LVADTQEEDGVAREQVDSADSEICVAHT                                                          266
925  CTTGTCGCCGATACCCAAGAAGAGGACGGCGTTGCCAGGGAGCAGGTTGATTCGGCTGATTCGGAAATTTGTGTTGCGCACACT  1008

267  GGQEMTEPDVVGSQTPIASAEETEVGEA                                                          294
1009  GTTGGTCAAGAAATGACTGAGCCTGATGTCGTCGGATCTCAAACTCCCATCGCCTCTGCTGAGGAAACCGAAGTCGGTGAGGCA  1092

295  CDREGIAEVKATVCADALDACPDQVEAF                                                          322
1093  TGCGACAGGGAAGGGATTGCTGAGGTCAAGGCAACTGTGTGTGCTGATGCTTTAGATGCCTGCCCCGATCAAGTGGAGGCATTT  1176

323  DIEKVEDSILSELQTELNAPADKTYEDV                                                          350
1177  GATATTGAAAAGGTTGAAGACAGTATCTTAAGTGAGCTTCAAACCGAACTTAATGCGCCCGCGGACAAGACCTATGAGGATGTC  1260

351  LAFDAIYSETLSAFYAVPSDETHFKVCG                                                          378
1261  TTGGCATTCGATGCCATATACTCAGAGACGTTGTCTGCATTCTATGCTGTGCCGAGTGATGAGACGCACTTTAAAGTGTGTGGA  1344

379  FYSPAIERTNCWLRSTLIVMQSLPLEFK                                                          406
1345  TTCTATTCGCCAGCTATAGAGGCTACTAATTGTTGGCTGCGTTCTACTTTGATAGTAATGCAGAGTTTACCTTTGGAATTTAAA  1428

407  DLGMQKLWLSYKAGYDQCFVDKLVKSAP                                                          434
1429  GACTTGGGGATGCAAAAGCTCTGGTTGTCTTACAAGGCTGGCTATGATCAATGCTTTGTGGACAAACTAGTTAAGAGCGCGCCC  1512

435  KSIILPQGGYVADFAYFFLSQCSFKVHA                                                          462
1513  AAGTCTATTATTCTTCCACAAGGTGGCTATGTGGCAGATTTTGCCTATTTTTTCCTAAGCCAGTGTAGCTTCAAAGTTCATGCT  1596

463  NWRCLKRFDSTLPGFETIMKVLNENLNA                                                          490
1597  AACTGGCGTTGTCTAAAGAGATTTGATAGTACTCTACCTGGTTTTGAGACTATCATGAAAGTGTTGAATGAGAATTTGAATGCC  1680
                          ▲

491  YQNQDGGADVVSPKPQRKRGTKQKAQKD                                                          518
1681  TACCAGAATCAAGATGGTGGTGCAGATGTAGTGAGCCCTAAGCCTCAGAGAAAGAGAGGGACAAAGCAAAAGGCTCAGAAAGAT  1764

519  EVDNVSVAKPKSSVQRNVSRELTPEDRS                                                          546
1765  GAAGTAGATAATGTAAGCGTTGCAAAGCCCAAAAGCTCTGTGCAGCGAAATGTAAGTAGAGAGTTAACCCCTGAGGATCGCAGC  1848

547  LLAQILDDGVVPDGLEDDSNVZ                                                                567
1849  CTTCTGGCTCAGATCCTAGATGATGGCGTAGTGCCAGATGGGTTAGAAGATGACTCTAATGTGTAAAGAGAATGAATCCTATGT  1932


1933  CGGCACTCGGTGGTAACCCCTCGCGAGAAAGTCGGGATAGGACACTCTCTATCAGAATGGATGTCTTGCTGTCATAACAGATAG  2016


2017  AGAAGGTTGTGGCAGACCCTGTATCAATTAGTTGAAAGAGATTGCAAAATAGAGAATGTGTGAGAGAAGTTAGCAAGGTCCTAC  2100


2101  GTCTAACCATAAGAACGGCGATAGGCGCCCCCCTGGGAAGAGCTCACATCAGGGTACTATTCCTGCAATGCCCTAGTAAATGAAT  2184


2185  GAAGTTGATCATGGCCAATTGGAAGAATCAC~poly(A)-3'                                            2215
```

FIG. 3. DNA sequence and deduced amino acid sequence of the DIssE cDNA clones. The extreme 5' end sequence was obtained by primer-extension studies (see Fig. 4). A translation of the main ORF is shown in single-letter amino acid code. Solid triangles indicate the sites where sequence fusion occurred.
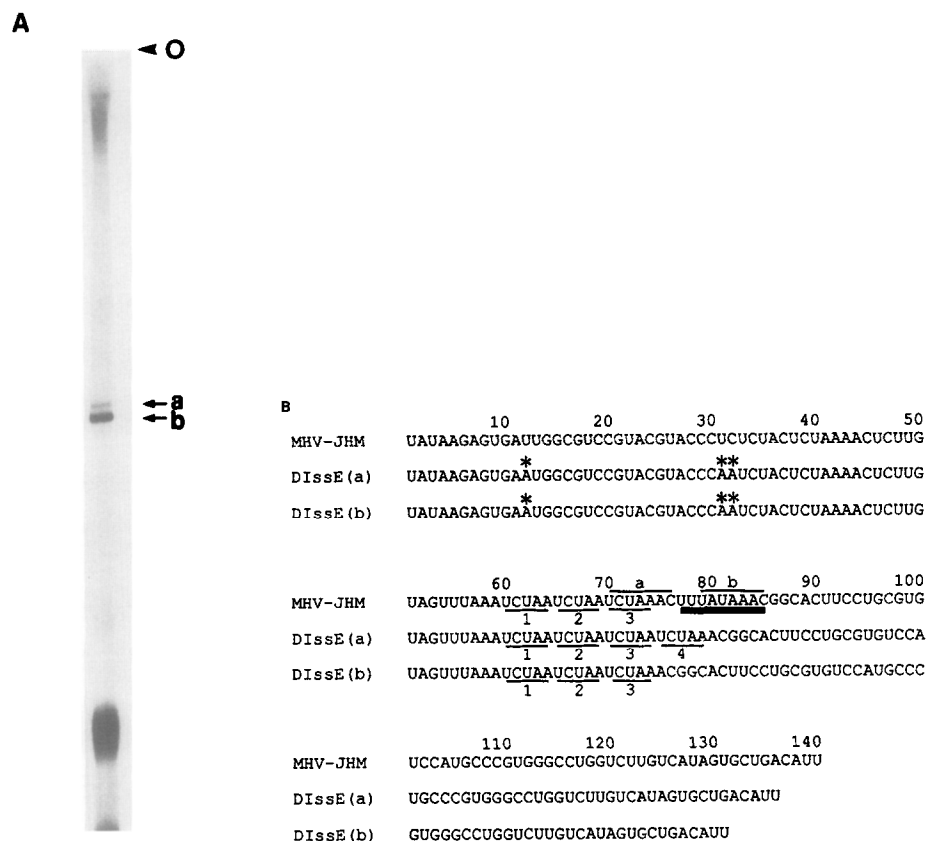
A

O ◄

a ◄
b ◄

B

```
                    10        20        30        40        50
MHV-JHM     UAUAAGAGUGAUUGGCGUCCGUACGUACCCUCUCUACUCUAAAACUCUUG
                        *                    **
DIssE(a)    UAUAAGAGUGAAUGGCGUCCGUACGUACCCAAUCUACUCUAAAACUCUUG
                        *                    **
DIssE(b)    UAUAAGAGUGAAUGGCGUCCGUACGUACCCAAUCUACUCUAAAACUCUUG


                    60        70  a      80  b    90       100
MHV-JHM     UAGUUUAAAUCUAAUCUAAUCUAAACUUUAUAAACGGCACUUCCUGCGUG
                     ─── ─── ───
                      1   2   3
DIssE(a)    UAGUUUAAAUCUAAUCUAAUCUAAUCUAAACGGCACUUCCUGCGUGUCCA
                     ─── ─── ─── ───
                      1   2   3   4
DIssE(b)    UAGUUUAAAUCUAAUCUAAUCUAAACGGCACUUCCUGCGUGUCCAUGCCC
                     ─── ─── ───
                      1   2   3


                   110       120       130       140
MHV-JHM     UCCAUGCCCGUGGGCCUGGUCUUGUCAUAGUGCUGACAUU

DIssE(a)    UGCCCGUGGGCCUGGUCUUGUCAUAGUGCUGACAUU

DIssE(b)    GUGGGCCUGGUCUUGUCAUAGUGCUGACAUU
```

FIG. 4. Primer extension analysis of the 5'-end of DIssE. (A) The synthetic oligodeoxyribonucleotides (18-mer) complementary to the nucleotides 123–140 from the 5' end of the parental MHV-JHM genomic RNA (Shieh *et al.*, 1987; Soe *et al.*, 1987) was ^{32}P-labeled at the 5' end, hybridized to the gel-purified DIssE RNA, and extended with reverse transcriptase. The products were electrophoresed on 6% polyacrylamide gels containing 8 *M* urea. O, origin of the gel. Two primer-extended products are shown as a and b. (B) The DNA sequences of these primer-extended products were determined by the Maxam–Gilbert method. The 5'-end sequence of MHV-JHM genomic sequence was obtained from previous studies (Shieh *et al.*, 1987; Soe *et al.*, 1987). The letters a and b represent the canonical seven-nucleotide sequence UCUAAAC and imperfectly repeated sequence of UAUAAAC, respectively. A bold solid line represents the nine-nucleotide sequence which is deleted in DIssE but present in MHV-JHM. DIssE (a) and DIssE (b) correspond to the sequences of primer-extended products, a and b, in Fig. 4A, respectively. Three base substitutions are indicated by asterisks.

DI RNAs. We therefore examined whether any significant secondary structure existed at rearrangement sites on MHV-JHM genomic RNA. The nucleotide sequences surrounding deleted regions of MHV-JHM genomic RNA were analyzed by an RNA secondary structure program of Zuker and Stiegler (1981). The predicted secondary structures of these rearrangement regions are shown in Fig. 6. All four genomic deletion sites have extensive and stable secondary structures. The free energies of these structures range from −73.0 to −114.2 kcal/mol. Furthermore, as previously described for the standard MHV-JHM, the sequence surrounding the junction of leader RNA and the remaining 5'-end genomic sequence also contains a stable secondary structure (Soe *et al.*, 1987). This junction region includes the nine-nucleotide deletion detected in DIssE RNA (Fig. 4B). Thus, an extensive and stable secondary structure exists at each parental MHV-JHM genomic region where deletion occurred.

DISCUSSION

The present study demonstrated that the smallest DI-specific RNA, DIssE, is composed of three discontiguous parts of the viral genome, including the 5' end and 3' end of genomic RNA. This structure is similar to many DI RNAs of other viruses, which typically retain both ends of the standard nondefective viral RNAs. Our previous study has demonstrated that DIssE is replicated from its negative template in the presence of helper virus (Makino *et al.*, 1988). Therefore, the DIssE sequence likely contains essential recognition signals for MHV RNA replication. The structure of DIssE RNA supports the likelihood that the recognition signals for the synthesis of negative-strand RNA and positive-strand RNA are localized at the 3' end and 5' end of genomic RNA, respectively.

One of the unique features of coronavirus DI RNA is that subgenomic DI RNA was poorly incorporated into
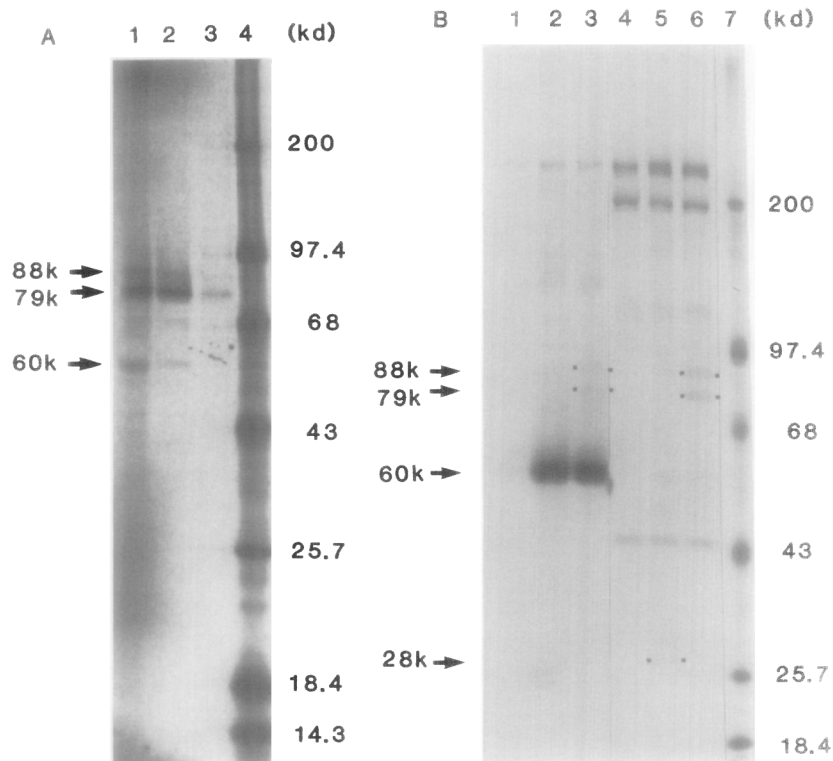
FIG. 5. Translation of DIssE-specific proteins. (A) Translation in a rabbit reticulocyte lysate of gel-purified DIssE RNA. ^{35}S-Labeled *in vitro* translation products of DIssE were analyzed by SDS–polyacrylamide gel electrophoresis directly (lane 1), and immunoprecipitated with anti-N protein monoclonal antibody (lane 2) or anti-p28 antibody (lane 3). Lane 4 contains ^{14}C-labeled marker proteins. (B) DIssE-specific proteins in DI-infected cells. DBT cells were mock-infected (lanes 1 and 4), infected with MHV-JHM (lanes 2 and 5), or infected with MHV-JHM containing DI particles (Lanes 3 and 6). At 7.5 hr postinfection, cultures were labeled with [^{35}S]methionine for 30 min, and cytoplasmic lysates were prepared, immunoprecipitated with anti-N protein monoclonal antibody (lanes 1–3) or anti-p28 antibody (lanes 4–6), and electrophoresed. Lane 7 contains ^{14}C-labeled marker proteins.
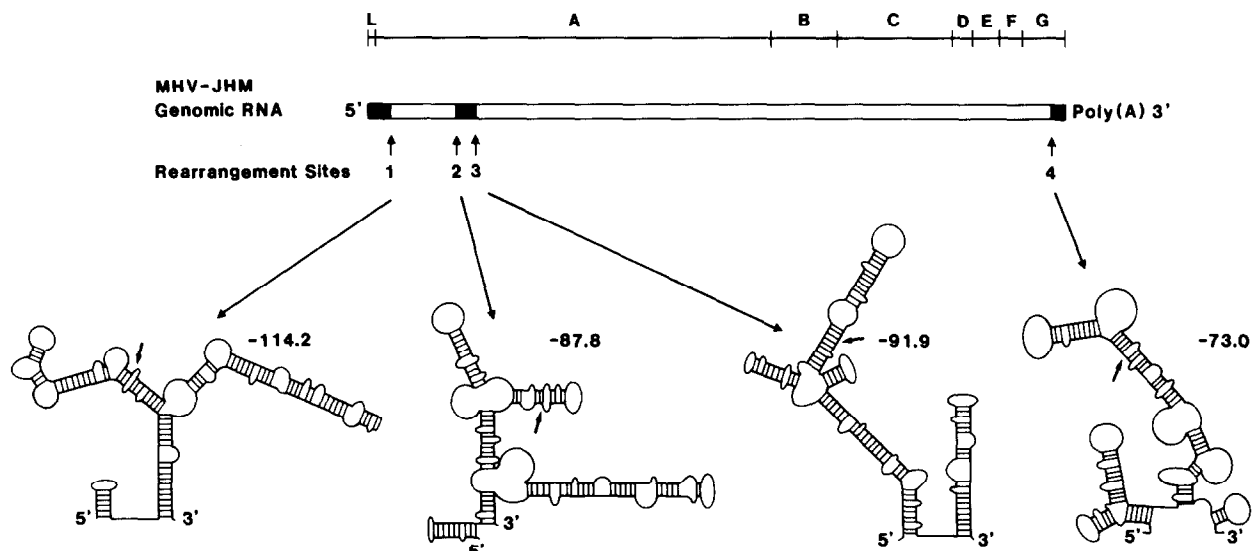
FIG. 6. Predicted secondary structure at the sequence rearrangement sites of MHV-JHM genomic RNA. The sequence of MHV-JHM genomic RNA was obtained from previously published data (Soe *et al.*, 1987) and our unpublished data (Shieh *et al.*, unpublished data). A–G represent the seven genes of MHV RNA. Solid boxes correspond to regions which share with DIssE. Free energy of the secondary structure at each rearranged site is given in kilocalories per mole. Arrows indicate the rearrangement sites.

virus particles (Makino *et al.*, 1988). One of the possible explanations is that the DI subgenomic RNAs lack a packaging signal. Since all MHV-specific subgenomic mRNAs contain the leader sequence, yet only genomic-sized RNA is efficiently packaged into virus particles, the packaging signal is probably located in gene A but not in leader sequence. The present study indicates that DIssE RNA has a nine-nucleotide (UUUAU-AAAC) deletion at the junction between the leader RNA and the remaining genomic RNA sequence. However, this deletion is not likely to account for the failure of efficient DI RNA packaging into virus particles since DIssA and the genomic RNA of a mutant MHV-JHM, both of which are packaged into virus particles, also have similar nine-nucleotide deletions (S. Makino, unpublished data). Thus, the packaging signals may be localized downstream of the 5' end 864 nucleotides. Recently we found that another intracellular DI-specific RNA, DIssF, could be packaged more efficiently than DIssE (S. Makino, unpublished data). The DIssF RNA is approximately 1.7 kb larger than DIssE and appears to contain more gene A sequences than DIssE, as determined from T1-oligonucleotide fingerprinting (Makino *et al.*, 1985). Sequence analysis of DIssF may reveal the possible reason for the poor incorporation of DIssE RNA into virus particles.

The data presented in this paper demonstrate extensive and stable secondary structures in the standard viral RNA at sites where DI RNA underwent deletions. This observation is consistent with a model of DI RNA generation, in which RNA transcription is interrupted at sites of hairpin loops on the template, and the RNA intermediates then fall off and rebind at new sites on the template to generate an RNA with extensive deletions. We have previously suggested that coronavirus RNA synthesis may utilize a discontinuous, nonprocessive mechanism, in which RNA transcription pauses at sites of secondary structures (Baric *et al.*, 1987). The incomplete RNA intermediates dissociate from templates and then rejoin the temple for subsequent RNA transcription. This mechanism is supported by the findings that MHV can undergo RNA recombination at an extremely high frequency (Makino *et al.*, 1986a), and that free incomplete RNA transcription products of various sizes were detectable in the cytoplasm of MHV-infected cells (Baric *et al.*, 1985, 1987). Furthermore, the sizes of these RNA products correspond to the lengths between the 5' end and the sites of hairpin loops (Baric *et al.*, 1987), in agreement with the notion that transcription pauses at these hairpin loops. Thus, the potential hairpin loops present in the genomic RNA at the DI RNA rearrangement sites could have interrupted RNA transcription. The incomplete RNA transcript may join the RNA template at the downstream rearrangement sites and create deleted RNA as a result. However, there is no consensus sequence at the sites of RNA deletion and reinitiation. It is not known how the reinitiation of RNA synthesis occurred.

The deletion of the nine nucleotides (UUUAUAAAC) at the 5' end where the leader RNA joins the genomic RNA may have been caused by the same discontinuous and nonprocessive transcription mechanism. It is interesting to note that the UCUAAAC, which is the consensus sequence for the leader RNA binding (Shieh *et al.*, 1987), is imperfectly repeated (UAUAAAC) at nine nucleotides downstream (Shieh *et al.*, 1987). It is these nine nucleotides which were deleted in DIssE RNA. Similar nine-nucleotide deletions have also been noted in the genomic RNA of DIssA, and that of an MHV-JHM mutant virus (S. Makino, unpublished data). This RNA structure suggests that RNA synthesis may pause at the first repeat, and then reinitiate at the second repeat because of the binding of the incomplete RNA transcript to the second repeat. Finally, the heterogeneity in the number of UCUAA repeats in DI RNAs also supports the discontinuous nature of coronavirus RNA replication. Similar heterogeneity has been noted in the genomic RNA of several different MHV strains (S. Makino and M. M. C. Lai, manuscript in preparation). Thus, DI RNA may be a product of discontinuous, nonprocessive RNA replication of coronaviruses.

There was a significant difference between the apparent molecular mass of the DIssE-specific protein products, 88K and 79K, and the predicted molecular mass of the potential product of the large ORF of DIssE RNA. This difference could be due to unusual configurations affecting electrophoretic migration, or due to the presence of phosphorylation, since the N protein is phosphorylated (Stohlman and Lai, 1979) and protein translated *in vitro* could be phosphorylated (Chattopadhyay and Banerjee, 1987). A similar difference between the predicted and actual molecular mass of the N protein has previously been noted (Skinner and Siddell, 1983). The relationship between the two protein species is not clear. The N protein has also been shown to consist of multiple species (Robbins *et al.*, 1986). It is not clear whether these proteins play any functional roles in DI-infected cells. Typically, DI RNAs do not synthesize any protein; however, in the Sindbis virus system, translation products have been detected from a DI RNA (Migliaccio *et al.*, 1985).

Although MHV genomic RNA and DIssE RNA are the major RNA species among MHV-specific mRNA species in virus-infected cells (Makino *et al.*, 1985, 1988) (Fig. 1), the gene products of these two mRNAs, RNA polymerase and both the 79K and 88K proteins, were

present in small quantities in virus-infected cells (Fig. 5B). We have previously demonstrated that the presence of stable secondary structure at the 5' end noncoding regions of the polymerase gene reduced the amount of polymerse protein synthesized in vitro (Soe et al., 1987). Also, as discussed previously, the presence of the small ORF encoding eight amino acids (Fig. 3) may reduce the number of ribosomes reaching the downstream optimal translation site (Soe et al., 1987). Since DIssE RNA has a 5' end structure similar to that of genomic RNA, the DIssE RNA may provide a tool to better our understanding of the mechanism of translational control of MHV RNAs. Furthermore, the fusion protein synthesized by DIssE RNA may be useful for understanding the functional and structural domains of the MHV polymerase and N protein.

ACKNOWLEDGMENTS

REFERENCES

BARIC, R. S., SHIEH, C.-K, STOHLMAN, S. A., and LAI, M. M. C. (1987). Analysis of intracellular small RNAs of mouse hepatitis virus: Evidence for discontinuous transcription. Virology 156, 342–354.

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.

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.

BRAYTON, P. R., STOHLMAN, S. A., and LAI, M. M. C. (1984). Further characterization of mouse hepatitis virus RNA-dependent RNA polymerases. Virology 133, 197–201.

CHATTOPADHYAY, D., and BANERJEE, A. K. (1987). Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription in vitro. Cell 49, 407–414.

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

DENISON, M. R., and PERLMAN, S. (1986). Translation and processing of mouse hepatitis virus virion RNA in a cell-free system. J. Virol. 60, 12–18.

FLEMING, J. O., STOHLMAN, S. A., HARMON, R. C., LAI, M. M. C., FRELINGER, J. A., and WEINER, L. P. (1983). Antigenic relationships of murine coronaviruses: Analysis using monoclonal antibodies to JHM (MHV-4) virus. Virology 131, 296–307.

GUBLER, U., and HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. Gene 25, 263–269.

KESSLER, S. W. (1981). Use of protein A-bearing staphylococci for the immunoprecipitation and isolation of antigens from cells. In "Methods in Enzymology" (J. J. Langone and H. Van Vunakis, Eds.), Vol. 73, pp. 442–459. Academic Press, New York.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227, 680–685.

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., BARIC, R. S., and STOHLMAN, S. A. (1983). Presence of leader sequences in the mRNA of mouse hepatitis virus. J. Virol. 46, 1027–1033.

LAI, M. M. C., PATTON, C. D., and STOHLMAN, S. A. (1982a). Further characterization of mRNAs of mouse hepatitis virus: Presence of common 5'-end nucleotides. J. Virol. 41, 557–565.

LAI, M. M. C., PATTON, C. D., and STOHLMAN, S. A. (1982b). Replication of mouse hepatitis virus: Negative-stranded RNA and replicative form RNA are of genome length. J. Virol. 44, 487–492.

LAI, M. M. C., and STOHLMAN, S. A. (1978). RNA of mouse hepatitis virus. J. Virol. 26, 236–242.

LANGRIDGE, L., LANGRIDGE, P., and BERQUIST, P. L. (1980). Extraction of nucleic acids from agarose gels. Anal. Biochem. 103, 264–271.

LEIBOWITZ, J. L., WILHELMSEN, 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.

MAHY, B. M. J., SIDDELL, S., WEGE, H., and TER MEULEN, V. (1983). RNA-dependent RNA polymerase activity in murine coronavirus-infected cells. J. Gen. Virol. 64, 103–111.

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., KECK, J. G., STOHLMAN, S. A., and LAI, M. M. C. (1986a). High-frequency RNA recombination of murine coronaviruses. J. Virol. 57, 729–737.

MAKINO, S., SHIEH, C.-K, KECK, J. G., and LAI, M. M. C. (1988). Defective-interfering particles of murine coronavirus: Mechanism of synthesis of defective viral RNAs. Virology 163, 104–111.

MAKINO, S., STOHLMAN, S. A., and LAI, M. M. C. (1986b). 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, F., 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, NY.

MAXAM, A. M., and GILBERT, W. (1980). Sequencing end-labelled DNA with base-specific chemical cleavages. In "Methods in Enzy-

mology" (L. Grossman and K. Moldave, Eds.), Vol. 65, pp. 449–560. Academic Press, New York.

MIGLIACCIO, G., CASTAGNOLA, P., LEONE, A., CERASUOLO, A., and BONATTI, S. (1985). mRNA activity of a Sindbis virus defective-interfering RNA. *J. Virol.* **55,** 877–880.

ROBBINS, S. G., FRANA, M. F., MCGOWAN, J. J., BOYLE, J. F., and HOLMES, K. V. (1986). RNA-binding proteins of coronavirus MHV: Detection of monomeric and multimeric N protein with an RNA overlay-protein blot assay. *Virology* **150,** 402–410.

SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

SHIEH, C.-K., SOE, L. H., MAKINO, S., CHANG, M.-F., STOHLMAN, S. A., and LAI, M. M. C. (1987). The 5'-end sequence of the murine coronavirus genome: Implications for multiple fusion sites in leader-primed transcription. *Virology* **156,** 321–330.

SKINNER, M. A., and SIDDELL, S. G. (1983). Coronovirus JHM: Nucleo-tide sequence of the mRNA that encodes nucleocapsid protein. *Nucleic Acids Res.* **15,** 5045–5054.

SOE, L. H., SHIEH, C.-K., BAKER, S. C., CHANG, M.-F., and LAI, M. M. C. (1987). Sequence and translation of the murine coronavirus 5'-end genomic RNA reveals the N-terminal structure of the putative RNA polymerase. *J. Virol.* **61,** 3968–3976.

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 sequence. *EMBO J.* **2,** 1939–1944.

STOHLMAN, S. A., and LAI, M. M. C. (1979). Phosphoproteins of murine hepatitis viruses. *J. Virol.* **36,** 672–675.

WEGE, H., MÜLLER, A., and TER MEULEN, V. (1978). Genomic RNA of the murine coronavirus JHM. *J. Gen. Virol.* **41,** 217–227.

ZUKER, M., and STIEGLER, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9,** 133–148.