JOURNAL OF VIROLOGY, Dec. 1987, p. 3968–3976
0022-538X/87/123968-09$02.00/0

Vol. 61, No. 12

Sequence and Translation of the Murine Coronavirus 5′-End Genomic RNA Reveals the N-Terminal Structure of the Putative RNA Polymerase

LISA H. SOE, CHIEN-KOU SHIEH, SUSAN C. BAKER, MING-FU CHANG, AND MICHAEL M. C. LAI*

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

Received 30 June 1987/Accepted 8 September 1987

A 28-kilodalton protein has been suggested to be the amino-terminal protein cleavage product of the putative coronavirus RNA polymerase (gene *A*) (M. R. Denison and S. Perlman, Virology 157:565–568, 1987). To elucidate the structure and mechanism of synthesis of this protein, the nucleotide sequence of the 5′ 2.0 kilobases of the coronavirus mouse hepatitis virus strain JHM genome was determined. This sequence contains a single, long open reading frame and predicts a highly basic amino-terminal region. Cell-free translation of RNAs transcribed in vitro from DNAs containing gene *A* sequences in pT7 vectors yielded proteins initiated from the 5′-most optimal initiation codon at position 215 from the 5′ end of the genome. The sequence preceding this initiation codon predicts the presence of a stable hairpin loop structure. The presence of an RNA secondary structure at the 5′ end of the RNA genome is supported by the observation that gene *A* sequences were more efficiently translated in vitro when upstream noncoding sequences were removed. By comparing the translation products of virion genomic RNA and in vitro transcribed RNAs, we established that our clones encompassing the 5′-end mouse hepatitis virus genomic RNA encode the 28-kilodalton N-terminal cleavage product of the gene *A* protein. Possible cleavage sites for this protein are proposed.

Mouse hepatitis virus (MHV), a murine coronavirus, contains a nonsegmented, single-stranded, positive-sense RNA of approximate M_r 6×10^6 (28, 49). In infected cells, MHV replicates exclusively in the cytoplasm and synthesizes a single genomic-length RNA and six subgenomic mRNAs, which are transcribed from a full-length negative-strand RNA template (7, 27). These mRNAs form a nested-set structure and have a common 3′ terminus but extend for various lengths in the 5′ direction (25). In addition, each mRNA contains an identical leader sequence of approximately 72 nucleotides which is derived from the 5′ end of the genomic RNA (24, 26, 43). UV transcriptional mapping studies (17) and the finding that no nuclear function is required for replication (6, 50) suggest that the joining of the leader sequences to coronavirus mRNAs does not utilize conventional eucaryotic splicing mechanisms. Instead, several lines of evidence support a 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 (2, 3, 10, 32).

In vitro translation studies in reticulocyte lysates and translations in *Xenopus* oocytes have demonstrated that MHV mRNAs are functionally monocistronic, each synthesizing only the protein encoded from the 5′-unique region of each mRNA, which does not overlap with the next smaller RNA (29, 37, 41). Hence, the MHV genome is thought to consist of at least seven genes, *A* to *G*, from 5′ to 3′(25). Genes *C*, *F*, and *G* encode three structural proteins: the nucleocapsid phosphoprotein (pp60) (44) and envelope glycoproteins (gp90/180 and gp25) (47). The remaining genes, *A*, *B*, *D*, and *E*, encode at least four nonstructural proteins (41, 42). One of the nonstructural proteins is RNA-dependent RNA polymerase, which is responsible for the synthesis of negative-strand RNA initially and positive-strand genomic and subgenomic mRNAs later in infection. Two distinct polymerase activities have been detected in MHV-infected cells (7, 8), but the polypeptide components have not been identified.

Several lines of evidence suggest that the 5′-most gene of MHV RNA encodes the RNA polymerase. (i) Since the MHV virion does not carry an RNA polymerase (7), the enzyme has to be translated directly from the incoming MHV genome, so that subsequent RNA transcription and replication can occur. This translation is only possible if the gene is localized at the 5′ end of the RNA. (ii) RNA recombination studies involving temperature-sensitive mutants which do not synthesize mRNAs at the nonpermissive temperature suggested that most of the temperature sensitive lesions in these mutants are localized to the 5′-most gene (18). (iii) The 5′-most gene is the largest of the viral genes, with a coding capacity for a protein of more than 200 kilodaltons (kDa) in the RNA genome (29). In the cases of other RNA viruses (39, 46), the largest nonstructural protein is usually RNA polymerase. Therefore, by analogy, the 5′-most gene of the MHV genome probably encodes RNA polymerase. Indeed, translation of MHV virion RNA in reticulocyte lysates produced three structurally related polypeptides of molecular masses greater than 200 kDa (29). More recently, using a similar in vitro translation system, Denison and Perlman (12) demonstrated a 250-kDa primary translation product from MHV genomic RNA. In addition, a 28-kDa protein was generated from the N terminus of this primary product by proteolytic cleavage. The 28-kDa protein has been detected in MHV-infected cells (13). Thus, this

* Corresponding author.

protein represents the first component of the potential MHV polymerase identified. Whether these proteins represent different functional polymerases which carry out the synthesis of negative-strand RNA, leader RNA, genomic RNA, and mRNAs is unknown.

To elucidate the structure and mechanism of synthesis of the potential RNA polymerase of MHV, we sequenced the 5'-end 2.0-kilobase (kb) region of genomic RNA which covers the genetic region likely to encode the p28 protein. The sequence obtained from cDNA clones predicts a single, long open reading frame (ORF). In vitro-synthesized RNAs derived from these cDNA clones were translated in a reticulocyte lysate system, and these data confirmed the position of the initiation codon and the identity of the p28 protein.

MATERIALS AND METHODS

Viruses and cells. The plaque-cloned JHM strain of MHV (31, 33) was used throughout and propagated on DBT cells, a mouse astrocytoma cell line (16), at low multiplicities of infection. Virus was harvested and purified from the medium, and viral RNA was prepared as previously described (32).

cDNA clones. The cDNA clones representing the 5'-terminal 4 kb of the MHV genome were derived from RNA extracted from sucrose gradient-purified virus. These clones were generated by priming for first-strand cDNA synthesis with a specific oligodeoxyribonucleotide which is complementary to a sequence 3.8 kb from the 5' end of the MHV genome (40). Figure 1 shows the positions of cDNA clones used to obtain the sequence of the 5'-end 2.0 kb of the MHV JHM genome. Clones F1 and F64 were shown by subsequent sequence analysis to be derived by binding of the primer oligodeoxyribonucleotide to an alternative site about 1.1 kb from the 5' end, which has closely homologous sequences. The sequences of the three clones F1, F64, and F82 were identical in the 5'-terminal 1.1-kb region except for this primer-binding region and an additional five-nucleotide repeat (TCTAA) in clone F82 (40).

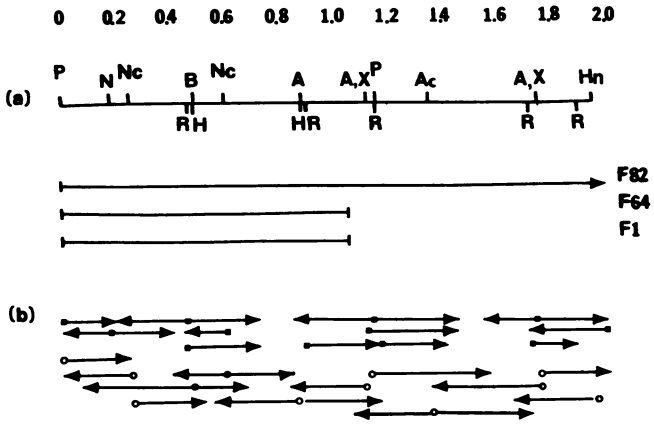

FIG. 1. Restriction map and sequencing strategy for 5'-end cDNA clones of JHM genomic RNA. (a) A map of restriction sites used in sequencing and the positions of the three cDNA clones sequenced. (b) Direction and extent of sequence information obtained from individual restriction sites. Arrows starting with solid squares indicate DNA sequenced by the dideoxy method. Arrows starting with open circles indicate Maxam-and-Gilbert sequencing with 3'-end-labeled DNA. Abbreviations: A, *Ava*I; Ac, *Acc*I; B, *Bam*HI; H, *Hae*III; Hn, *Hin*dIII; N, *Nar*I; Nc, *Nco*I; P, *Pst*I; R, *Rsa*I; X, *Xho*I. Only *Hae*III and *Rsa*I sites used in subcloning are indicated. Lengths are expressed in kilobase pairs.

DNA sequencing. Sequencing was carried out by two methods. The dideoxy chain termination method (38) was used with [α-^{35}S]dATP (5) on fragments of cDNA inserts, generated by a variety of restriction enzymes, which were cloned into the M13 vectors mp18 and mp19 (36). dITP (1) was used in place of dGTP in some cases to eliminate sequence ambiguity caused by nucleotide band compressions as a result of the DNA secondary structure. Sequence data were also obtained by chemical modification (34) of various cDNA fragments subcloned into pT7 vectors (48). DNA fragments were 3' end labeled with Klenow fragment at internal restriction sites or, alternatively, at the polylinker cloning sites of pT7. End-labeled cDNA restriction fragments were separated by electrophoresis on preparative acrylamide gels (35) and purified as described by Hansen et al. (14, 15). Sequence analysis and predicted RNA secondary structures were obtained with the Intelligenetics sequencing program.

Construction of pT7 recombinant vectors. Clone F82 DNA representing the 5'-end 4 kb of the MHV genome (40) was excised from pBR322 with *Pst*I, and the 5' 1.2-kb *Pst*I fragment was inserted into the *Pst*I site of the transcription vector pT7-5 containing the bacteriophage T7 promoter (48) (see Fig. 6). This construct was designated pT7F82P. To remove the GC tail and the potential secondary structure in the leader region of this subclone, the 1.15-kb *Dra*I-*Hin*dIII fragment of pT7F82P was subcloned into the *Bam*HI and *Hin*dIII sites of pT7-5 and designated pT7F82D. Finally, to completely remove the secondary structure in the 5' noncoding region of gene *A*, we ligated the *Nar*I-*Hin*dIII fragment of F82 (nucleotides 187 to 2,000) into the *Acc*I-*Hin*dIII sites of vector pT7-3, and the resulting clone was named pT7F82N.

In vitro transcription. Recombinant pT7 plasmids were transcribed in vitro with a T7 RNA polymerase specific for the T7 promoter as described by Tabor and Richardson (48). Transcription was carried out in 50 μl of reaction buffer containing 2 μg of linearized plasmid DNA, 40 mM Tris hydrochloride (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM dithiothreitol, 1 mM each ATP, CTP, and UTP, 0.05 mM GTP, 0.5 mM 7mGpppG cap analog (Bethesda Research Laboratories), 10 U of T7 RNA polymerase (US Biochemicals), and 40 U of RNasin ribonuclease inhibitor (Promega Biotec). Reaction mixtures were incubated at 37°C for 45 min. After incubation, the GTP concentration was increased to 1 mM, a second sample of T7 RNA polymerase was added, and the reaction was incubated for an additional 45 min. DNA templates were removed by digestion with 2 U of RNase-free DNase I (New England Nuclear Corp.) for 15 min at 37°C. RNA products were extracted twice with phenol-chloroform (1:1) and precipitated with ethanol.

In vitro translation. An mRNA-dependent rabbit reticulocyte lysate (New England Nuclear) was used for all translation reactions. Reactions were carried out in 25 μl. The conditions for translation were the same as those recommended by the manufacturer except that potassium acetate and magnesium acetate were optimized to 180 and 1.5 mM, respectively (12). RNA (200 ng) synthesized in vitro or isolated from purified MHV virions was used in each reaction. Translation mixtures were incubated for 90 min at 30°C with [^{35}S]methionine (2,400 μCi/ml; New England Nuclear) as the radiolabel.

Polyacrylamide gel electrophoresis. Polyacrylamide gel electrophoresis of translation products was performed as described by Maizel (30), with 7 to 20% gradient polyacrylamide gels containing 1% sodium dodecyl sulfate. Gels were

fixed for 30 min in 50% methanol–7% acetic acid, dried, and exposed to Kodak X-ray film at −70°C.

One-dimensional peptide mapping. Partial peptide mapping followed the method of Cleveland et al. (11). Briefly, in vitro translated proteins were separated by polyacrylamide gel electrophoresis, and their positions were located relative to a stained molecular weight marker. The [^{35}S]methionine-labeled proteins of interest were excised from gels, equilibrated for 30 min in soaking buffer (0.1% [wt/vol] sodium dodecyl sulfate, 1 mM EDTA, 0.125 M Tris hydrochloride [pH 6.8]) and loaded onto the stacking portion of a 17.5% polyacrylamide gel. The gel slices were overlaid with 20 μl of buffer containing 20% glycerol, 1% (wt/vol) dithiothreitol and bromophenol blue, followed by 20 μl of buffer containing 10% glycerol and 1 ng of *Staphylococcus aureus* V-8 protease (Miles Laboratories). The gel was run as usual except that the voltage was turned off for 30 min when the sample was at the interface between the stacking and separating gels to allow proteolytic digestion of the proteins. Following electrophoresis, the gel was dried and exposed to X-ray film at −70°C.

RESULTS

Strategy of DNA sequencing. To determine the sequence of the 5′ 2.0 kb of the MHV genome, various subclones were constructed, containing restriction fragments of each cDNA clone in both M13 and pT7 vectors. Figure 1 shows the sequencing scheme of the cDNA fragments used in determining the sequence. Each region was verified by sequencing on both strands and by sequencing with both the dideoxy chain termination (38) and chemical modification (34) methods.

Analysis of nucleotide sequence and predicted amino acid sequence. Figure 2 shows the DNA sequence obtained from

```
5'-TATAAGAGTGATTGGCGTCCGTACGTACCCTCTCTACTCTAAAACTCTTGTAGTTTAAATCTAATCTAATCTAAACTTTATAAACGGC   88

                  M  P  V  G  L  V  L  S  Z
 89 ACTTCCTGCGTGTCCATGCCCGTGGGCCTGGTCTTGTCATAGTGCTGACATTTGTGGTTCCTTGACTTTCGTTCTCTGCCAGTGACGTGT  178

                                                M  A  K  M  G  K  Y  G  L  G  F  K  W  A  P  E  F  P   18
179 CCATTCGGCGCCAGCAGCCCACCCATAGGTTGCATAATGGCAAAGATGGGCAAATACGGTCTCGGCTTCAAATGGGCCCCAGAATTTCCA  268

 19  W  M  L  P  N  A  S  E  K  L  G  N  P  E  R  S  E  E  D  G  F  C  P  S  A  A  Q  E  P  K   48
269 TGGATGCTTCCGAACGCATCGGAGAAGTTGGGTAACCCTGAGAGGTCAGAGGAGGATGGGTTTTGCCCCTCTGCTGCGCAAGAACCGAAA  358

 49  V  K  G  K  T  L  V  N  H  V  R  V  D  C  S  R  L  P  A  L  E  C  C  V  Q  S  A  I  I  R   78
359 GTTAAAGGAAAAACTTTGGTTAATCACGTGAGGGTGGATTGTAGCCGGCTTCCAGCCTTGGAGTGCTGTGTTCAGTCCGCCATAATCCGT  448

 79  D  I  F  V  D  E  D  P  Q  K  V  E  A  S  T  M  M  A  L  Q  F  G  S  A  V  L  V  K  P  S  108
449 GATATTTTTGTTGACGAGGATCCCCAGAAGGTGGAGGCCTCGACTATGATGGCATTGCAGTTCGGTAGTGCTGTCTTGGTCAAGCCATCC  538

109  K  R  L  S  V  Q  A  W  A  K  L  G  V  L  P  K  T  P  A  M  G  L  F  K  R  F  C  L  C  N  138
539 AAGCGCTTGTCTGTTCAGGCATGGGCTAAGTTGGGTGTGCTGCCTAAAACTCCGGCCATGGGGTTGTTCAAGCGCTTCTGCCTGTGTAAC  628

139  T  R  E  C  V  C  D  A  H  V  A  F  Q  L  F  T  V  Q  P  D  G  V  C  L  G  N  G  R  F  I  168
629 ACCAGGGAGTGCGTTTGTGACGCCCACGTGGCCTTTCAACTTTTTACGGTCCAGCCCGATGGTGTATGCCTGGGTAACGGCCGTTTTATA  718

169  G  W  F  V  P  V  T  A  I  P  E  Y  A  K  Q  W  L  Q  P  W  S  I  L  L  R  K  G  G  N  K  198
719 GGCTGGTTCGTTCCAGTCACAGCCATACCGGAGTATGCGAAGCAGTGGTTGCAACCCTGGTCCATCCTTCTTCGTAAGGGTGGTAACAAA  808

199  G  S  V  T  S  G  H  F  R  R  A  V  T  M  P  V  Y  D  F  N  V  E  D  A  C  E  E  V  H  L  228
809 GGGTCTGTGACATCCGGCCATTTCCGCCGCGCTGTTACCATGCCTGTGTATGACTTTAATGTAGAGGATGCTTGTGAGGAGGTTCATCTT  898
                                                                                    ▼
229  N  P  R  G  K  Y  S  C  K  A  Y  A  L  L  R  G  Y  R  G  V  K  P  I  L  F  V  D  Q  Y  G  258
899 AACCCGAGGGGTAAGTACTCCTGCAAGGCGTATGCTCTTCTTAGGGGCTATCGCGGTGTTAAGCCCATCCTGTTTGTGGACCAGTATGGT  988
                                    ▼
259  C  D  Y  T  G  C  L  A  K  G  L  E  D  Y  G  D  L  T  L  S  E  M  K  E  L  S  P  V  W  R  288
998 TGCGACTATACTGGATGTCTCGCCAAGGGTCTCGAAGACTATGGCGACCTTACTTTGAGTGAGATGAAGGAGTTGTCCCCTGTGTGGCGT 1078

289  D  S  L  D  N  E  V  V  V  A  W  H  V  D  R  D  P  R  A  V  M  R  L  Q  T  L  A  T  V  R  318
1079 GACTCCTTAGATAATGAAGTTGTTGTGGCTTGGCATGTTGATCGTGACCCTCGAGCTGTGATGCGCCTGCAGACTCTGGCTACTGTACGT 1168

319  S  I  E  Y  V  G  Q  P  I  E  D  M  V  D  G  D  V  V  M  R  E  P  A  H  L  L  A  P  N  A  348
1169 AGCATTGAGTATGTGGGCCAACCGATCGAGGATATGGTTGATGGGGATGTGGTGATGCGTGAACCCGCTCACCTCCTCGCGCCCAATGCC 1258
                                                                    ◆                 ◆
349  I  V  K  R  L  P  R  L  V  E  T  M  L  Y  T  D  S  S  V  T  E  F  C  Y  K  T  K  L  C  D  378
1259 ATTGTTAAAAGGCTTCCCCGTCTGGTGGAGACTATGCTGTATACGGATTCTTCCGTTACAGAGTTTTGTTATAAAACTAAGTTGTGTGAT 1348
                                    ◆                 ◆
379  C  G  F  I  T  Q  F  G  Y  V  D  C  C  G  D  T  C  G  F  R  G  W  V  P  G  N  M  M  D  G  408
1349 TGTGGTTTTATCACACAGTTTGGCTATGTTGATTGCTGTGGTGACACCTGTGGTTTTCGTGGGTGGGTTCCTGGCAATATGATGGATGGC 1438

409  F  P  C  P  G  C  C  K  S  Y  M  P  W  E  L  E  A  Q  S  S  G  V  I  P  E  G  G  V  L  F  438
1439 TTTCCATGCCCTGGGTGTTGTAAAAGTTACATGCCTTGGGAGTTGGAAGCTCAATCATCGGGTGTTATACCAGAAGGAGGTGTTCTATTT 1528

439  T  Q  S  T  D  T  V  N  R  E  S  F  K  L  Y  G  H  A  V  V  P  F  G  G  A  A  Y  W  S  P  468
1529 ACGCAGAGCACTGACACAGTGAATCGTGAGTCTTTTAAGCTCTACGGTCATGCTGTTGTGCCTTTTGGTGGTGCTGCGTATTGGAGTCCA 1618

469  Y  P  G  M  W  L  P  V  I  W  S  S  V  K  S  Y  S  Y  L  T  Y  T  G  V  V  G  C  K  A  I  498
1619 TATCCTGGTATGTGGCTTCCAGTAATTTGGTCTTCTGTTAAGTCATATTCCTATTTGACTTATACAGGAGTTGTTGGTTGTAAGGCAATT 1708

499  V  Q  E  T  D  A  I  C  R  F  L  Y  M  D  Y  V  Q  H  K  C  G  N  L  E  Q  R  A  I  L  G  528
1709 GTCCAGGAGACAGACGCTATATGTCGTTTTCTGTACATGGACTATGTCCAGCACAAGTGTGGCAATCTCGAGCAGAGGGCTATCCTTGGA 1798

529  L  D  D  V  Y  H  R  Q  L  L  V  N  R  G  D  Y  S  L  L  L  E  N  V  D  L  F  V  K  R  R  558
1799 TTGGACGATGTCTATCATAGACAGTTGCTTGTTAACAGGGGTGACTATAGTCTCCTCCTTGAAAATGTGGATTTGTTTGTTAAGCGGCGC 1888

559  A  E  F  A  C  K  F  A  T  C  G  D  G  L  V  P  L  L  L  D  G  L  V  P  R  S  Y  Y  L  I  588
1889 GCTGAATTTGCTTGCAAATTCGCCACCTGTGGAGATGGCCTTGTACCCCTTCTACTAGATGGTTTAGTGCCCCGCAGTTATTATTTAATT 1978

589  K  S  G  Q  A  F  T                                                                        595
1979 AAGAGTGGTCAAGCTTTTACCT                                                                    2000
```

FIG. 2. DNA sequence of the 5′-end 2 kb of MHV JHM genomic cDNA clones. A translation of the main ORF is shown in single-letter amino acid code. Potential cleavage sites for p28 are indicated by arrowheads. The amino acids underlined indicate the potential zinc-binding domain in the gene *A* polypeptide, with diamonds above the essential Cys residues.

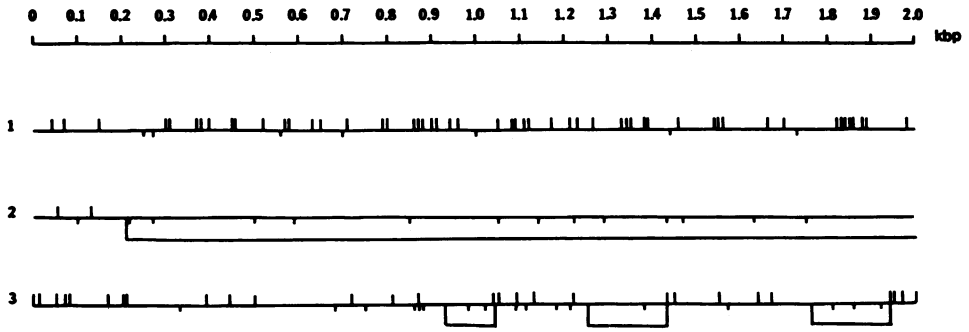

FIG. 3. Three-frame translation of the 5′-end 2.0-kb sequences. The locations of termination codons (long vertical lines above the base lines) and potential initiation codons (short vertical lines below the base lines) are indicated. ORFs of greater than 35 amino acids are indicated by open boxes. A single, long ORF beginning at position 215 is in frame 2.

the cDNA clones, with a translation in one-letter amino acid code of the main ORF. The leader sequence on the 5′ end of the genomic RNA has previously been presented (40). Potential cleavage sites for the generation of a 28-kDa N-terminal protein are indicated. We also noted a potential zinc-binding domain at amino acids 371 to 395, formed by the specific juxtaposition of cysteine residues. Zinc-binding domains are a common feature of proteins involved in nucleic acid binding or gene regulation (4).

A three-frame translation of the 5′ 2.0-kb sequence revealed a single, long ORF beginning at position 215 from the 5′ end of the genomic RNA (Fig. 3). The two alternative reading frames are closed by numerous termination codons. The 5′-proximal AUG at position 104 is in a suboptimal context for initiation of translation (20), and the reading frame is closed after only eight amino acids. The sequence context around the next AUG at position 215 is similar to that used by most eucaryotic mRNAs (20), with a purine at −3 and a G at position 4. The context around the AUG at position 224 conforms equally well to the consensus for functional initiation codons. These two initiation codons are followed by several additional in-frame methionine residues. These methionine residues may provide alternative initiation sites for translation and result in the overlapping proteins observed in in vitro translation studies (29).

To obtain more information about this putative protein, the amino acid sequences were used to determine hydropathy plots (23) (Fig. 4). The product of the extreme 5′ end of the ORF, the p28 protein, is predicted to be highly basic. In addition, the presence of a central hydrophobic domain in this protein is consistent with the membrane-associated nature of the RNA polymerase (7, 8).

Predicted secondary structure of the 5′-terminal region of MHV RNA. To examine the potential secondary structure at the 5′ region of the RNA genome, the nucleotide sequences were analyzed by the RNA secondary structure program of Zuker and Stiegler (51). Figure 5 illustrates the predicted secondary structure of the MHV genome, which was compared with the published 5′ sequence of an avian coronavirus, infectious bronchitis virus (IBV) (9). Both of these

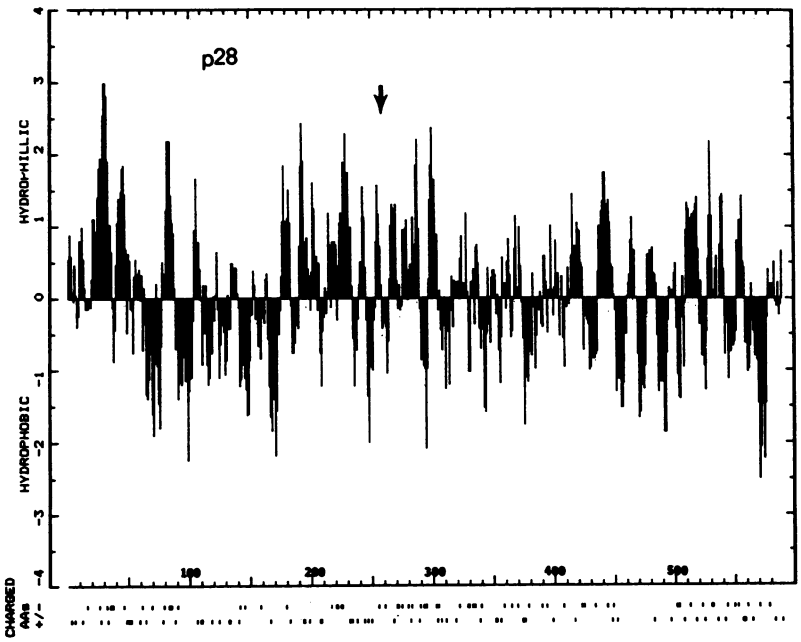

FIG. 4. Hydropathy profile of the predicted amino acid (AA) sequence of the main ORF in the 5′ 2.0 kb of the MHV JHM genome. Positive values indicate hydrophilic regions, and negative values show hydrophobic regions. Each point is the mean hydropathicity of a span of seven residues plotted at the first residue of the span (23). The potential cleavage site (at amino acid 257) for generation of the p28 protein is indicated by the arrow.

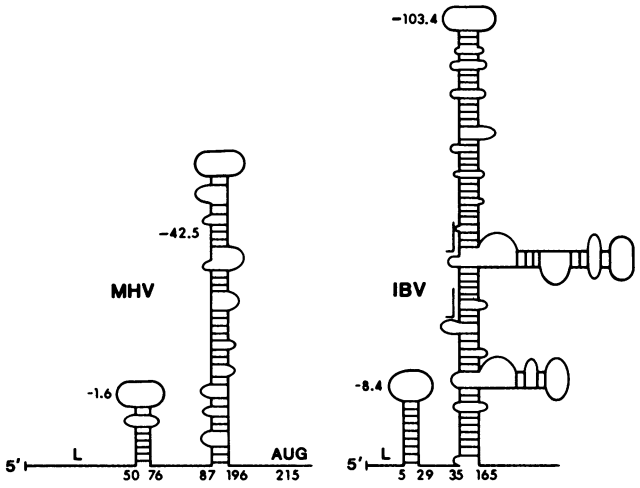

FIG. 5. Predicted secondary structure at the 5' noncoding region of the MHV JHM and IBV RNA genomes. The 5' leader sequence is indicated by the letter L, and free-energy values are given in kilocalories per mole. The IBV sequence was obtained from previously published data (9).

RNAs show an extensive and stable secondary structure preceding the first AUG for MHV (ΔG, −42.5 kcal/mol [1 cal = 4.184 J]) and for IBV (ΔG, −103.4 kcal/mol). The presence of the secondary structure at the 5' end of the genome may affect the translational efficiency of the incoming viral RNA and, hence, regulate the kinetics of viral replication and transcription.

Determination of the initiation codon for the gene A product. The predicted amino acid sequence of the gene *A* ORF contains several in-frame methionine residues. To determine which initiation codon is used to translate the gene *A* ORF, we performed in vitro translation in rabbit reticulocyte lysates of the 5'-terminal RNA of MHV. For this purpose, the pT7 construct, pT7F82D, containing the 5'-terminal 1.8 kb of the MHV genome, was linearized with various restriction enzymes (Fig. 6) and transcribed with T7 RNA polymerase as described in Materials and Methods. The predicted sizes of the RNAs generated for three restriction sites, *Bst*EII, *Ava*II, and *Sca*I, within gene *A* were 242, 617, and 945 bases, respectively, and were confirmed by electrophoresis of the transcripts on 1.5% agarose gels (Fig. 7a). Each of these RNAs was then translated in a cell-free rabbit reticulocyte lysate system, and the products were resolved on a 7.5 to 20% polyacrylamide gradient gel (Fig. 7b).

RNAs truncated at restriction sites *Bst*EII, *Ava*II, and *Sca*I yielded proteins of approximately 3.2, 18, and 26 kDa, respectively. This is consistent with the predicted sizes of the protein products if translation was initiated from either of the 5'-most optimal AUGs at positions 215 and 224. Translation of the RNA generated from the *Sca*I-cut plasmid also yielded an additional protein product of approximately 34.5 kDa, corresponding to the size of the predicted full-length ORF of pT7F82D initiated at position 215. This protein probably was made by the RNA transcribed from a small amount of recircularized plasmid. These data establish that translation initiation occurs at the 5'-most optimal start codon. However, we could not distinguish between initiation from AUGs at positions 215 and 224.

Effect of the 5'-terminal RNA secondary structure on translation. The 5'-end sequences of the MHV genome predict the presence of an extensive secondary structure preceding the initiator AUG (Fig. 5). To determine whether these 5'-terminal sequences affect the translational efficiency of the viral genomic RNA, we studied the translation of the 5'-end RNAs with different lengths of 5' untranslated regions. For this purpose, several cDNA fragments from clone F82 were inserted behind the T7 promoter, such that various lengths of 5' noncoding sequences preceded the first optimal AUG. The resulting plasmids are shown in Fig. 6.

RNAs were transcribed from pT7 recombinant plasmids pT7F82P, pT7F82D, and pT7F82N, linearized with *Ava*II, and translated in a rabbit reticulocyte lysate system. After translation, products were resolved on a 7.5 to 20% polyacrylamide gradient gel (Fig. 8). The RNAs transcribed from all three recombinant plasmids directed the synthesis of an approximately 18-kDa protein product, corresponding to the size of the polypeptide initiated from the AUG at position 215. The sequential truncation of 5' noncoding sequences of gene *A* resulted in increased yields of translational products. Hence, little or no product was translated from RNA synthesized from plasmid pT7F82P containing the complete 5'-end sequences. Removal of the poly(G) region and the 5' 58 nucleotides, as in the pT7F82D construct, allowed effi-

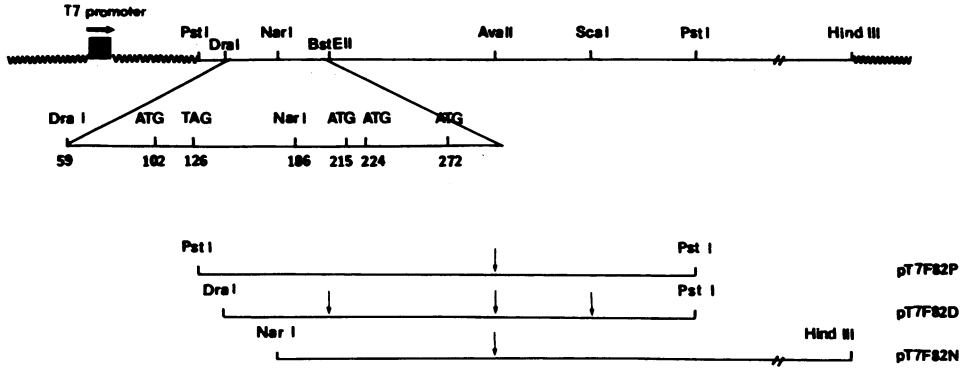

FIG. 6. Restriction map of recombinant plasmids used to transcribe RNA. Clone F82 DNA was excised from pBR322 by *Pst*I digestion and ligated into pT7-5, and the resulting clone was designated pT7F82P. pT7F82P was digested with *Dra*I and *Hin*dIII to remove the poly(G) region and the 5' 59 nucleotides of the leader sequence and then inserted into the *Bam*HI-*Hin*dIII sites of pT7-5 (pT7F82D). Alternatively, clone F82 was digested with *Nar*I and *Hin*dIII to remove the 5'-end 186 nucleotides and then inserted into *Acc*I-*Hin*dIII-cleaved pT7-3 (pT7F82N). These recombinant plasmids were linearized with *Bst*EII, *Ava*II, or *Sca*I (vertical arrows), and RNA was transcribed with T7 RNA polymerase. Jagged lines represent pT7 vector sequences, and the horizontal arrow designates the direction of transcription.

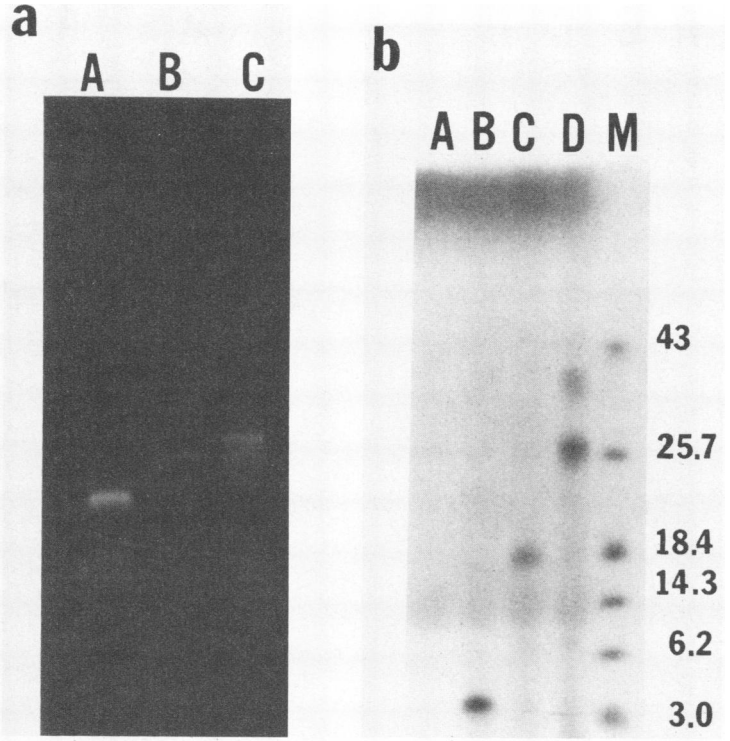

FIG. 7. In vitro transcription and translation of recombinant plasmids containing 5′ gene *A* sequences. Capped RNA was synthesized from linearized plasmid pT7F82D with T7 RNA polymerase. The RNA was analyzed on a 1.5% agarose gel and visualized by ethidium bromide staining. (a) RNA synthesized from pT7F82D cut with (lanes): A, *Bst*EII; B, *Ava*II; or C, *Sca*I. (b) Cell-free translation of in vitro synthesized RNAs (200 ng per reaction) in rabbit reticulocyte lysates in the presence of [^{35}S]methionine. The products were analyzed on a 7.5 to 20% gradient sodium dodecyl sulfate-polyacrylamide gel. Lanes A, B, C, and D, respectively, show translation products of no RNA, *Bst*EII fragment-generated RNA, *Ava*II fragment-generated RNA, and *Sca*I fragment-generated RNA. Lane M contained ^{14}C-labeled marker polypeptides; molecular masses are given in kilodaltons.

cient translation of the RNA. Further removal of the 5′ 186 nucleotides, as in pT7F82N, eliminated a large region of the secondary structure and a suboptimal AUG at position 102 and resulted in a fourfold increase in the translational product over that obtained with pT7F82D. Thus, the 5′ noncoding sequences of gene *A* could modulate the translational efficiency of the viral genome. Translation of uncapped in vitro synthesized RNA from these constructs yielded no detectable protein product (data not shown), indicating that 5′ capped ends are required for proper translation of MHV RNAs.

Identification of the N-terminal peptide of the gene *A* product. The 5′-end gene of the MHV genome has been demonstrated to encode a large (>200 kDa) protein product, which is a presumptive RNA polymerase (29). This large polypeptide was shown to be cleaved during in vitro translation to yield 28- and 220-kDa proteins (12). Preliminary evidence suggests that p28 was derived from the N terminus of the primary translation product. To establish that the p28 product derived from the translation of virion RNA is indeed the N-terminal cleavage product of a larger translation product, protein made by an RNA representing the 5′-end 1.1-kb region was compared with the p28 protein translated from virion genomic RNA.

Full-length genomic RNA extracted from purified virions yielded two major proteins of 220 and 28 kDa in rabbit reticulocyte lysate (Fig. 9a, lane A), in agreement with previously published results (12). In addition, we observed the synthesis of p60, which could be precipitated with the monoclonal antibody specific for the nucleocapsid protein of MHV (data not shown). This protein is probably the product of contaminating mRNA 7 or degraded genomic RNA. In vitro synthesized RNA generated from plasmid pT7F82N linearized by *Sca*I, representing the 5′-end 1.1-kb sequence (Fig. 6), generated a 26-kDa protein (Fig. 9a, lane B). The p28, p220, and p26 protein bands were excised, treated with V-8 protease, and fractionated on a 17.5% polyacrylamide gel (Fig. 9b). Peptide mapping demonstrated that p28 and p26 were structurally related to each other, sharing five of six peptides. In contrast, the peptide map of p220 was significantly different. These results confirm that the p28 protein translated from the virion genomic RNA is indeed derived from the 5′ end of gene *A*, i.e., it is the N-terminal cleavage product of the putative RNA polymerase.

DISCUSSION

Nucleotide sequence analysis of the 5′ 2.0 kb of the MHV JHM genome demonstrated the presence of a single, long ORF that potentially encodes a portion of the viral RNA-dependent RNA polymerase. This protein has a highly basic amino terminus and contains several hydrophobic regions, which is consistent with the membrane-associated nature of the RNA polymerase (7, 8). A potential zinc-binding domain was also identified. The latter feature suggests that the sequence may encode a nucleic acid-binding protein which is involved in viral RNA transcription.

Using RNA transcripts derived from pT7 recombinant

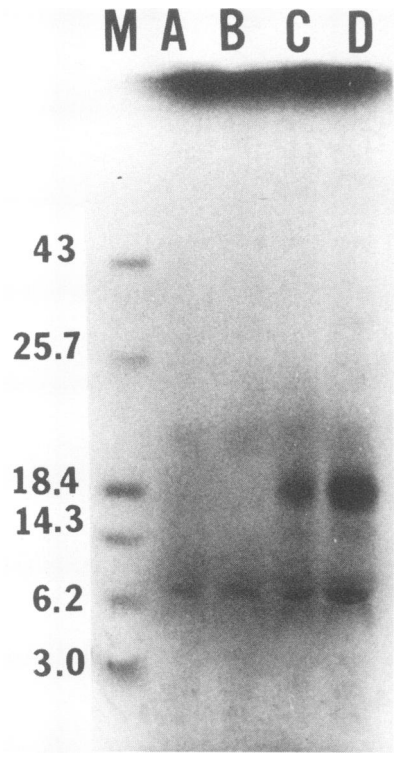

FIG. 8. Effect of truncation of 5′ noncoding sequences on the translation of gene *A* RNA. Plasmids pT7F82P, pT7F82D, and pT7F82N (Fig. 6) were linearized with *Ava*II and used as templates for T7 polymerase-directed RNA synthesis. The RNAs were translated in rabbit reticulocyte lysates, and the products were analyzed on 7.5 to 20% gradient polyacrylamide gels. Lanes A, B, C, and D, respectively, show translation products of no RNA, pT7F82P RNA, pT7F82D RNA, and pT7F82N RNA. Lane M contained ^{14}C-labeled marker proteins; molecular masses are given in kilodaltons.

plasmids containing the 5′ gene *A* sequences, we demonstrated a single translation product which is initiated from the AUG at nucleotide 215 or 224. Both potential initiation codons are in an optimal context for translation, containing a purine at position −3 and a G at position +4 (20). Because of the proximity of these two AUGs, we could not determine which AUG was actually utilized for initiation of translation. We further showed that the ORF at the 5′ end of the genome was translated more efficiently when the upstream noncoding sequences were removed. This could be due to removal of the secondary structure in the noncoding region or stabilization of the RNA by shortening of the RNA. RNA secondary structure profiles suggest that this noncoding region may contain a stable hairpin loop structure preceding the first initiation codon. A similar secondary structure is predicted for another coronavirus, IBV (9) (Fig. 5). These secondary structures may make translation inefficient, since the preinitiation complex would have to move through this extensive RNA-RNA hybrid to reach the optimal 5′-proximal AUG (22). In addition, the 5′-end sequences of both viruses contain a first AUG (nucleotide 102 in MHV) in suboptimal context which can potentially encode a very small polypeptide (8 amino acids for MHV and 11 amino acids for IBV). These small ORFs may be nonfunctional or may reduce the number of ribosomes reaching the downstream optimal translation start site (21). These data are

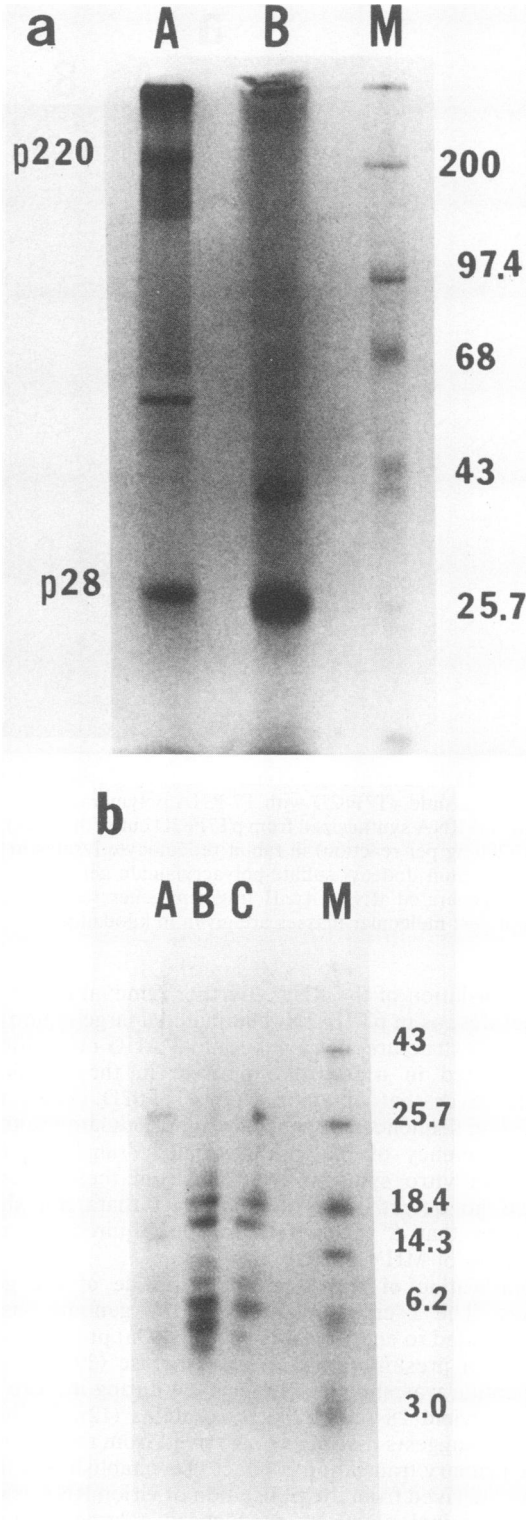

FIG. 9. Partial peptide maps of the translation products of JHM genomic RNA and pT7F82D RNA. (a) Translations in rabbit reticulocyte lysates of purified JHM virion RNA (lane A) and RNA synthesized from pT7F82D cut with *Sca*I (lane B). Products were analyzed on a 5 to 15% gradient polyacrylamide gel. The major genomic products p28 and p220 and the major pT7F82D-*Sca*I translation product of approximately 26 kDa were excised from the wet gel and treated with V-8 protease. (b) Proteolytic digestion products were analyzed on a 17.5% polyacrylamide gel. Lanes: A, p220; B, p28; C, p26. Lane M contained ^{14}C-labeled marker proteins; molecular masses are given in kilodaltons.

consistent with the very low amount of gene *A* proteins present in MHV-infected cells (29).

In vitro translation of MHV virion RNA yielded a predominant protein of greater than 200 kDa, the N terminus of which was cleaved to a 28-kDa protein (12). p28 is a basic protein which was detected in infected cells at late times postinfection, thus representing the first protein component of a putative RNA polymerase detected in coronavirus-infected cells (13). Peptide mapping experiments presented in this study, comparing the translation products of virion RNA with those of pT7 recombinant plasmid-derived RNA, confirmed that p28 is indeed the N-terminal protein of the gene *A* product(s). Furthermore, the predicted amino acid sequence of the N-terminal portion of gene *A* suggests a highly basic protein, in agreement with the basic nature of the p28 protein (13). Thus, the predicted amino acid sequence for the p28 protein presented in this study most likely represents the correct translational product of gene *A*. It is not clear whether p28 is a component of RNA polymerase. Since MHV RNA polymerase is expected to perform many different transcription functions, including transcription of negative-strand RNA, leader RNA, and mRNAs, the RNA polymerase might have multiple components, one of which may be p28. Such a multicomponent polymerase has been demonstrated in other RNA viruses, such as Sindbis virus (45). Alternatively, cleavage of p28 from a polymerase protein may alter the structure of the polymerase to enable the polymerase to carry out different transcriptional functions. Thus, cleavage of p28 may be a regulatory event. It has previously been shown that several different RNA polymerases could be detected in MHV-infected cells at different times during infection (7, 8).

The exact cleavage site for generation of the p28 protein has not been determined. Several Tyr-Gly pairs, which have been shown to be one of the cleavage sites in poliovirus polyprotein (19), are present in the vicinity of the potential cleavage site for p28. Other types of protein cleavage sites, Gln-Gly and Asn-Ser observed in picornaviruses, are not present in this area. It is also not clear whether this cleavage is carried out by a viral or a cellular proteolytic enzyme. It is likely that an autoproteolytic process is involved. However, we have not observed cleavage of the p28 protein in in vitro translations of pT7-generated RNAs (data not shown). Further studies will be needed to determine the exact site and mechanism of this cleavage.

ACKNOWLEDGMENTS

We thank David Vannier and Lee Macario Abaigar for excellent technical assistance, Carol Flores for preparation of the manuscript, and Stephen Stohlman and Stanley Tahara for helpful discussions.

This work was supported by Public Health Service research grants AI 19244 and NS 18146 from the National Institutes of Health, by National Multiple Sclerosis Society research grant 1449, and by National Science Foundation grant PCM-4507. L.H.S. is a postdoctoral fellow of the Bank of America-Giannini Foundation, and S.C.B. a postdoctoral fellow of the Arthritis Foundation.

LITERATURE CITED

1. **Bankier, A., and B. G. Barrell.** 1983. Shotgun DNA sequencing, p. 1–34. *In* R. A. Flavell (ed.), Techniques in the life sciences (biochemistry), vol. B5: techniques in nucleic acid biochemistry. Elsevier Biomedical Press, Amsterdam.
2. **Baric, R. S., S. A. Stohlman, and M. M. C. Lai.** 1983. Characterization of replicative intermediate RNA of mouse hepatitis virus: presence of leader RNA sequences on nascent chains. J. Virol. **48**:633–640.
3. **Baric, R. S., S. A. Stohlman, M. K. Razavi, and M. M. C. Lai.** 1985. Characterization of leader-related small RNAs in coronavirus-infected cells: further evidence for leader-primed mechanism of transcription. Virus Res. **3**:19–33.
4. **Berg, J. M.** 1986. Potential metal-binding domains in nucleic acid binding proteins. Science **232**:485–487.
5. **Biggins, M. C., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
6. **Brayton, P. R., R. G. Ganges, and S. A. Stohlman.** 1981. Host cell nuclear function and murine hepatitis virus replication. J. Gen. Virol. **56**:457–460.
7. **Brayton, P. R., M. M. C. Lai, C. D. Patton, and S. A. Stohlman.** 1982. Characterization of two RNA polymerase activities induced by mouse hepatitis virus. J. Virol. **42**:847–853.
8. **Brayton, P. R., S. A. Stohlman, and M. M. C. Lai.** 1984. Further characterization of mouse hepatitis virus RNA dependent RNA polymerases. Virology **133**:197–201.
9. **Brown, J. D. K., M. E. G. Boursnell, M. M. Binns, and F. M. Tomley.** 1986. Cloning and sequencing of 5' terminal sequences from avian infectious bronchitis virus genomic RNA. J. Gen. Virol. **67**:221–228.
10. **Budzilowicz, C. J., S. P. Wilczynski, and S. R. Weiss.** 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.
11. **Cleveland, D. W., S. G. Fischer, M. W. Kirschner, and U. K. Laemmli.** 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. J. Biol. Chem. **252**:1102–1106.
12. **Denison, M. R., and S. Perlman.** 1986. Translation and processing of mouse hepatitis virus virion RNA in a cell-free system. J. Virol. **60**:12–18.
13. **Denison, M. R., and S. Perlman.** 1987. Identification of a putative polymerase gene product in cells infected with murine coronavirus A59. Virology **157**:565–568.
14. **Hansen, J. N.** 1981. Use of solubilizable acrylamide disulfide gels for isolation of DNA fragments suitable for sequence analysis. Anal. Biochem. **116**:146–151.
15. **Hansen, J. N., B. H. Pheiffer, and J. A. Boehnert.** 1980. Chemical and electrophoretic properties of solubilizable disulfide gels. Anal. Biochem. **105**:192–201.
16. **Hirano, N., K. Fujiwara, S. Hino, and M. Matusumoto.** 1974. Replication and plaque formation of mouse hepatitis virus (MHV-2) in mouse cell line DBT culture. Arch. Gesamte Virusforsch. **44**:298–302.
17. **Jacobs, L., W. J. M. Spaan, M. C. Horzinek, and B. A. M. van der Zeijst.** 1981. Synthesis of subgenomic mRNA's of mouse hepatitis virus is initiated independently: evidence from UV transcription mapping. J. Virol. **39**:401–406.
18. **Keck, J. G., S. A. Stohlman, L. H. Soe, S. Makino, and M. M. C. Lai.** 1987. Multiple recombination sites at the 5'-end of the murine coronavirus RNA. Virology **156**:331–341.
19. **Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanacek, J. J. Lee, S. Van der Werf, C. W. Anderson, and E. Wimmer.** 1981. Primary structure, gene organization and polypeptide expression of poliovirus RNA. Nature (London) **291**:547–553.
20. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
21. **Kozak, M.** 1986. Bifunctional messenger RNAs in eukaryotes. Cell **47**:481–483.
22. **Kozak, M.** 1986. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA **83**:2850–2854.
23. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
24. **Lai, M. M. C., R. S. Baric, P. R. Brayton, and S. A. Stohlman.** 1984. Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus—a cytoplasmic RNA virus. Proc. Natl. Acad. Sci. USA **81**:3626–3630.

25. **Lai, M. M. C., P. R. Brayton, C. Armen, C. D. Patton, C. Pugh, and S. A. Stohlman.** 1981. Mouse hepatitis virus A59: mRNA structure and genetic localization of the sequence divergence from hepatotropic strain MHV-3. J. Virol. **39**:823–834.
26. **Lai, M. M. C., C. D. Patton, R. S. Baric, and S. A. Stohlman.** 1983. Presence of leader sequences in the mRNA of mouse hepatitis virus. J. Virol. **46**:1027–1033.
27. **Lai, M. M. C., C. D. Patton, and S. A. Stohlman.** 1982. Replication of mouse hepatitis virus: negative-stranded RNA and replicative form RNA are of genome length. J. Virol. **44**:487–492.
28. **Lai, M. M. C., and S. A. Stohlman.** 1978. RNA of mouse hepatitis virus. J. Virol. **26**:236–242.
29. **Leibowitz, J. L., S. R. Weiss, E. Paavola, and C. W. Bond.** 1982. Cell-free translation of murine coronavirus RNA. J. Virol. **43**: 905–913.
30. **Maizel, J.** 1971. Polyacrylamide gel electrophoresis of viral proteins. Methods Virol. **5**:176–246.
31. **Makino, S., N. Fujioka, and K. Fujiwara.** 1985. Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus. J. Virol. **54**:329–336.
32. **Makino, S., S. A. Stohlman, and M. M. C. Lai.** 1986. Leader sequences of murine coronavirus mRNAs can be freely reassorted: evidence for the role of free leader RNA in transcription. Proc. Natl. Acad. Sci. USA **83**:4204–4208.
33. **Makino, S., F. Taguchi, and K. Fujiwara.** 1984. Defective interfering particles of mouse hepatitis virus. Virology **133**:9–17.
34. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
35. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
36. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
37. **Rottier, P. J. M., W. J. M. Spaan, M. C. Horzinek, and B. A. M. van der Zeijst.** 1981. Translation of three mouse hepatitis virus strain A59 subgenomic RNAs in *Xenopus laevis* oocytes. J. Virol. **38**:20–26.
38. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
39. **Schubert, M., G. G. Harmison, C. D. Richardson, and E. Meier.** 1985. Expression of a cDNA encoding a functional 241-kilodalton vesicular stomatitis virus RNA polymerase. Proc. Natl. Acad. Sci. USA **82**:7984–7988.
40. **Shieh, C. K., L. H. Soe, S. Makino, M. F. Chang, S. A. Stohlman, and M. M. C. Lai.** 1987. The 5′-end sequence of the murine coronavirus genome: implications for multiple fusion sites in leader-primed transcription. Virology **156**:321–330.
41. **Siddell, S.** 1983. Coronavirus JHM: coding assignments of subgenomic mRNAs. J. Gen. Virol. **64**:113–125.
42. **Siddell, S., H. Wege, A. Barthel, and V. ter Meulen.** 1981. Coronavirus JHM: intracellular protein synthesis. J. Gen. Virol. **53**:145–155.
43. **Spaan, W., H. Delius, M. Skinner, J. Armstrong, P. Rottier, S. Smeekens, B. A. M. van der Zeijst, and S. G. Siddell.** 1983. Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. EMBO J. **2**:1839–1844.
44. **Stohlman, S. A., and M. M. C. Lai.** 1979. Phosphoproteins of murine hepatitis viruses. J. Virol. **32**:672–675.
45. **Strauss, E. G., C. M. Rice, and J. H. Strauss.** 1984. Complete nucleotide sequence of the genomic RNA of Sindbis virus. Virology **133**:92–110.
46. **Strauss, E. G., and J. H. Strauss.** 1983. Replication strategies of the single-stranded RNA viruses of eukaryotes. Curr. Top. Microbiol. Immunol. **105**:1–98.
47. **Sturman, L. S.** 1977. Characterization of a coronavirus. I. Structural proteins: effects of preparative conditions on the migration of protein in polyacrylamide gels. Virology **77**:637–649.
48. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**: **82**:1074–1078.
49. **Wege, H., A. Muller, and V. ter Muelen.** 1978. Genomic RNA of the murine coronavirus JHM. J. Gen. Virol. **41**:217–228.
50. **Wilhelmsen, K. C., J. L. Leibowitz, C. W. Bond, and J. A. Robb.** 1981. The replication of murine coronavirus in enucleated cells. Virology **110**:225–230.
51. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.