

Molecular Cloning of the Gene Encoding the Putative Polymerase of Mouse Hepatitis Coronavirus, Strain A59

CATHERINE J. PACHUK,* PETER J. BREDENBEEK,† PHILIP W. ZOLTICK,*
WILLY J. M. SPAAN,† AND SUSAN R. WEISS*,¹

*Department of Microbiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6076, and
†Institute of Virology, Department of Infectious Diseases and Immunology, State University Utrecht, The Netherlands

Received December 13, 1988; accepted March 8, 1989

Complementary DNA (cDNA) libraries were constructed representing the genome RNA of the coronavirus mouse hepatitis virus, strain A59 (MHV-A59). From these libraries clones were selected to form a linear map across the entire gene A, the putative viral polymerase gene. This gene is approximately 23 kb in length, considerably larger than earlier estimates. Sequence analysis of the 5' terminal region of the genome indicates the presence of the 66-nucleotide leader that is found on all mRNAs. Secondary structure analysis of the 5' terminal region suggests that transcription of leader terminates in the region of nucleotide 66. The sequence of the first 2000 nucleotides is very similar to that reported for the closely related JHM strain of MHV and potentially encodes p28, a basic protein thought to be a component of the viral polymerase (L. Soe, C. K. Shieh, S. Baker, M. F. Chang, and M. M. C. Lai, 1987, *J. Virol.*, 61, 3968-3976). Gene A contains two of the consensus sequences found in intergenic regions. One is adjacent to the 5' leader sequence and the other is upstream from the initiation codon for translation of gene B. © 1989 Academic Press, Inc.

INTRODUCTION

The genome of the murine coronavirus, mouse hepatitis virus, strain A59 (MHV-A59) is a single-stranded polyadenylated RNA of positive polarity and is at least 30 kb in length (see below) (Lai and Stohlman, 1978). During infection six capped, polyadenylated subgenomic messenger RNAs are synthesized in addition to full-length positive-sense RNA. The subgenomic RNAs form a nested set, all overlapping with the 3' end of genome RNA and all containing the same 5' 66-nucleotide "leader" sequence (Lai *et al.*, 1981, 1984; Leibowitz *et al.*, 1981; Spaan *et al.*, 1982, 1983; Weiss and Leibowitz, 1983, see Fig. 1A). A full-length negative-stranded RNA serves as a template for the synthesis of all virus-specific mRNAs (Lai *et al.*, 1983). The common 5' leader sequence is transcribed from the 3' end of this template and is subsequently used as a primer for mRNA synthesis or alternatively spliced co-transcriptionally to the body of the mRNAs (Baric *et al.*, 1987a; Lai *et al.*, 1983, 1984; Spaan *et al.*, 1988).

The coronavirus polymerase needs to possess many activities to initiate and synthesize negative-strand RNA, positive-strand genome length RNA and mRNAs (both genomic and subgenomic), leader RNA, and poly(A). In addition it is likely to possess capping activities as well as the ability to switch templates during

transcription to promote recombination (Lai *et al.*, 1985). Polymerase activity is membrane associated and has been detected in extracts of infected cells (Brayton *et al.*, 1982, 1984), but has not yet been purified. Gene A, the putative polymerase gene is extremely large and like the avian coronavirus IBV polymerase gene has the capacity to encode 700,000 kDa in proteins (Bournsnell *et al.*, 1987). To understand both leader transcription and the structure of the polymerase, we have cloned the MHV-A59 gene A. We describe here the complementary DNA (cDNA) clones and the analysis of the 5' terminal end of the genome.

MATERIALS AND METHODS

Cells and viruses

MHV-A59 (Manaker *et al.*, 1961) was grown in monolayer cultures of murine fibroblast 17Cl-1 cells. For preparation of virus-specific RNA, cells were infected at an m.o.i. of 1 and harvested approximately 16 hr later.

Preparation and analysis of RNA

Viral genome RNA was prepared from purified virions as described previously (Budzilowicz *et al.*, 1985). Intracellular RNA was extracted from the cytoplasm of MHV-A59-infected or mock-infected cells as previously described (Budzilowicz *et al.*, 1985). Electro-

¹ To whom requests for reprints should be addressed.

phoresis of RNA in the presence of formaldehyde was carried out according to the method of Lehrach *et al.* (1977). After electrophoresis, RNA was transferred to nitrocellulose and hybridized with radiolabeled probes according to the method of Thomas (1980).

Complementary DNA cloning

First-strand cDNA synthesis was carried out as described previously using genome RNA that had been denatured with methyl mercury hydroxide and either random oligomers of calf thymus DNA or restriction fragments as primers (Budzilowicz *et al.*, 1985). In most cases second-strand cDNA was synthesized using *Escherichia coli* DNA polymerase I in the presence of RNase H according to the method of Gubler and Hoffman (1983). Oligo(dC) tails were added to the 3' ends of double-stranded cDNA molecules. These tailed DNAs were annealed to oligo(dG)-tailed *Pst*I-digested pBR322 and used to transform *E. coli*, strain DH5. In a second library, derived from an independent A59 stock, oligo(dG)-tailed cDNAs were inserted into pUC9 and transformed into *E. coli*, strain JM109. The cDNA clones containing the genomic leader sequence were identified using a leader-specific probe containing four tandem copies of the 72 nucleotides at the 5' end of A59 genome (pGEM1.L3/4).

Preparation and analysis of DNA

DNA was electrophoresed on agarose gels and then transferred to nitrocellulose for Southern blot analysis (Southern, 1975). DNA was labeled with ^{32}P by nick-translation to a specific activity of $>10^8$ cpm/ μg (Rigby *et al.*, 1977).

DNA sequencing

DNA restriction fragments as described under Results were ligated into the polylinker region of M13mp19 RF which was used to transform *E. coli* JM103 (Messing *et al.*, 1981). Single-stranded DNA from M13 subclones was sequenced according to a modification of the method of Sanger *et al.* (1977) using [^{35}S]dATP (Biggins *et al.*, 1983) and T7 DNA polymerase (Sequenase, U.S. Biochemicals) in place of the Klenow fragment of polymerase I. Sequencing primers were the universal primers (Promega) or synthetic oligonucleotide primers. The synthetic oligonucleotide primers which were 20 bases long were synthesized at the Wistar Institute or the Chemistry Department at the University of Pennsylvania (Philadelphia, PA). Sequencing reactions were resolved on gradient polyacrylamide gels (Biggins *et al.*, 1983). All sequence analysis was carried out using the programs of PC Gene on an Ep-

son Equity III+ computer or the Intelligenetics programs through Bionet.

RESULTS

Construction of gene A-specific cDNA clones

Complementary DNA cloning was carried out using random oligomers of calf thymus DNA as primers for transcription of cDNA from polyadenylated genome RNA. In order to maximize the size of the cDNAs, cloning was carried out according to the method of Gubler and Hoffman (1983) in which there is no S1 nuclease treatment. The average size of the clones obtained was approximately 1.5 kb with the largest being 4.1 kb.

The recombinant plasmid DNAs from 40 of the largest clones were radiolabeled by nick-translation and used as probes on Northern blots of RNA from both MHV-A59-infected and mock-infected cells. The nested set nature of the RNAs allows the mapping of clones to the unique regions of the RNAs and thus to individual viral genes. DNA from clones mapping to gene G hybridize to all seven viral mRNAs since gene G sequences are present in all seven mRNAs. Conversely, gene A-specific sequences hybridize only to mRNA 1 since this is the only viral mRNA that contains gene A sequences (Fig. 1). It was important to include uninfected cell RNA on these blots because earlier we obtained clones that appeared to hybridize to MHV-A59 RNA 1 but also hybridized to a band of similar mobility in uninfected cells. Twenty-seven of the 40 cDNA clones analyzed this way mapped to gene A.

Clone 1033 was used to orient the gene A clones. Although results from Northern blot analysis indicated that cDNA 1033 mapped to gene B (data not shown, but hybridization was to mRNAs 1 and 2 as in Fig. 1B, lane D), the size of the insert (3600 bp) suggested that 1033 must also contain sequences from the 3' end of gene A since gene B is approximately 2.1 kb (Luytjes *et al.*, 1988). In an attempt to more precisely localize cDNA 1033 within genes A and B, the two fragments derived by digestion with *Pst*I were radiolabeled and used as probes on Northern blots of infected cell RNAs (Fig. 1B). The 1.3-kb fragment hybridized to mRNA 1 only while the 2.3 kb fragment hybridized to both mRNAs 1 and 2, and must therefore extend into gene B. We have confirmed the presence of the gene A/B intergenic region by sequencing a portion of the 2.3-kb fragment (see below).

To determine overlaps among the gene A clones, the viral inserts were digested with restriction enzymes and analyzed by cross-hybridization on Southern blots (Southern, 1975). The gene A-specific clones fell into clusters separated by gaps. The remaining uncharac-

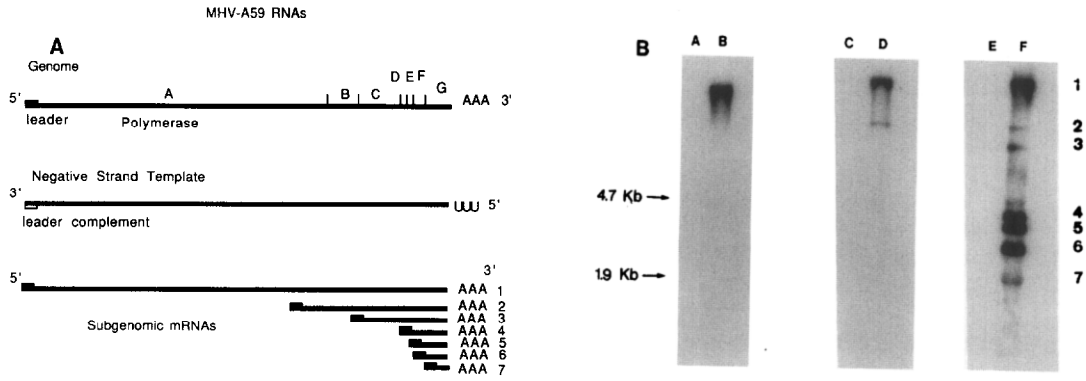


FIG. 1. Mapping of virus-specific cDNAs to viral mRNAs. (A) The MHV-A59 positive-strand genome, its negative-strand complement, and seven mRNAs are shown. (B) RNA was extracted from MHV-A59-infected and mock-infected cells, electrophoresed on agarose gels, and transferred to nitrocellulose. DNA fragments were excised from pBR322, labeled by nick-translation, and used as probes on these Northern blots. Lane A, mock-infected cell RNA, 1033 1.3-kb fragment; lane B, infected cell RNA, 1033 1.3-kb fragment; lane C, mock-infected cell RNA, 1033 2.3-kb fragment; lane D, infected cell RNA, 1033 2.3-kb fragment; lane E, mock-infected cell RNA, clone g344 (representing genes D, E, F, G); and lane F, infected cell RNA, g344.

terized clones in the cDNA library were screened by hybridization to probes made from cDNAs which mapped to regions bordering a gap. This extended the lengths of the clusters resulting in three groups of clones. These were 1033 to 920, 1220 to 917, and 136 to 1410 (Fig. 2). There remained as well a gap at the 5'

end of genome; none of the clones hybridized to a probe made from the leader sequence-specific clone pGEM1.L3/4. From a cloning using a 1200-nucleotide *EcoRI/HindIII* 917 restriction fragment as primer, we selected clone D34 which was contiguous with clone 136. Clones 67 and 98, obtained from a second ran-

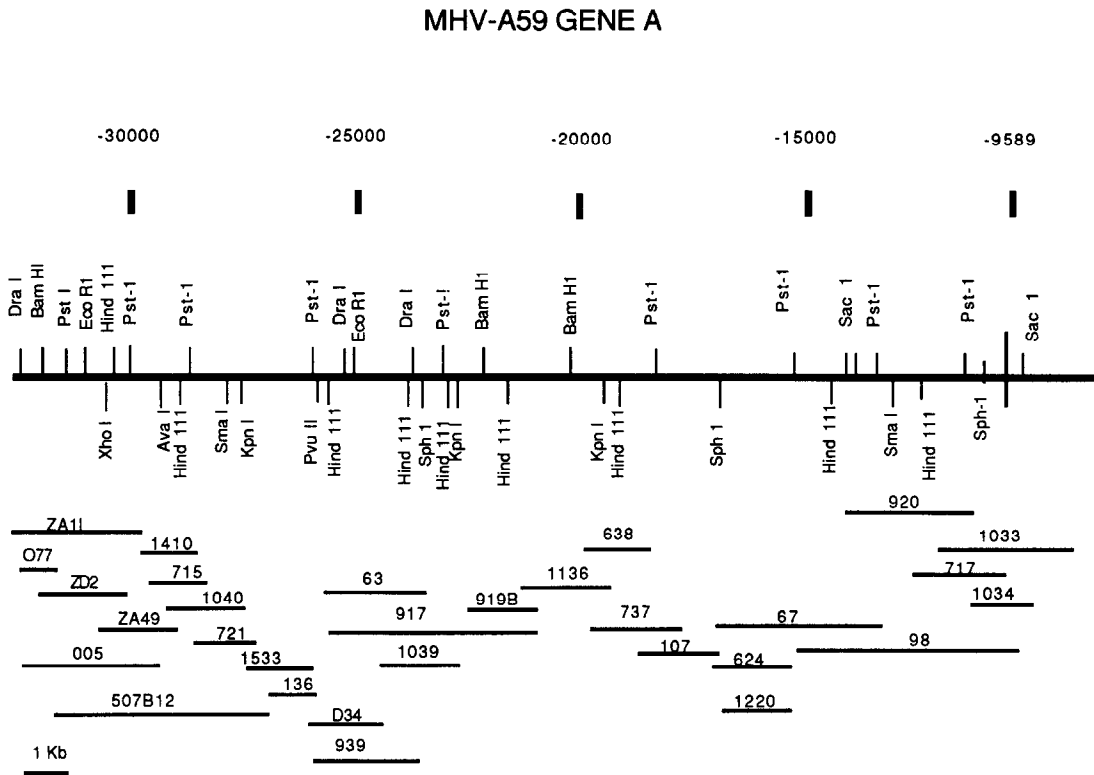


FIG. 2. Map of gene A-specific DNA clones. cDNAs were mapped by restriction analysis and Northern and Southern blots as described in the text. Only the enzymes *PstI*, *EcoRI*, *HindIII*, and *BamHI* were used across the entire genome. Nucleotide -9589 marks the beginning of the intergenic sequence between genes A and B and is 9589 nucleotides from the poly(A) tail of genome RNA.

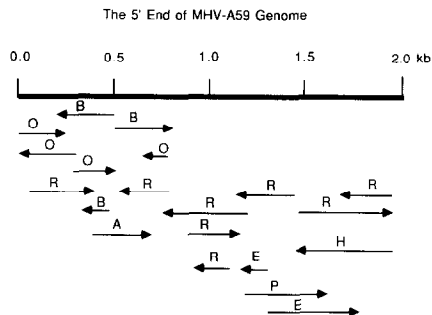


Fig. 3. Sequencing strategy for the 5' end of MHV-A59 genome RNA. The viral fragments from clones O77, ZA11, 005, and 507B12 were cleaved with restriction enzymes and subcloned into M13mp19 RF. Sequencing was carried out using the universal sequencing primer in the regions and directions indicated by the arrows. The letters above the arrows designate the enzymes used to make the subclones. *RsaI* (R), *AluI* (A), *BamHI* (B), *PstI* (P), *EcoRI* (E), *HindIII* (H). The arrows designated O were derived by sequencing using synthetic oligonucleotides as primers on O77 DNA.

domly primed library made from an independent stock of MHV-A59, closed the gap between 1220 and 920. From a cloning using a 750-nucleotide *AvaI/HindIII* fragment of 1410, we obtained many leader-positive clones, among which were clones O77 and ZA11. Clone ZA11 shared homology with 1410 and thus completed the cloning of the 5' end. Clones 005 and 507B12, both located near the 5' end of genome, were obtained from the second randomly primed library. Of all the 5' proximal clones only ZA11 contained the entire leader sequence. The contiguous series of cDNAs revealed that gene A was considerably larger than the 7 kb previously estimated (Lai and Stohman, 1978; Leibowitz *et al.*, 1981; Spaan *et al.*, 1982). The restriction digests and Southern blots indicated that the A59 gene A is approximately 23 kb, similar to the avian coronavirus (IBV) gene A (Fig. 2).

The 5' end of the genome

The sequence of the 5' 2000 nucleotides of the MHV-A59 genome RNA was obtained by sequencing of clones O77, ZA11, 005, and 507B12 (Fig. 2). The viral fragments were digested with the restriction enzymes listed in the legend to Fig. 3 and the fragments subcloned into the RF of M13mp19. Sequencing was carried out using the dideoxynucleotide chain termination technique with a universal sequencing primer and synthetic oligonucleotide primers (Fig. 3). This 5' leader sequence, obtained from ZA11, is the same as that obtained from A59 mRNA (Lai *et al.*, 1984). A typical intergenic sequence, AAUCUAAC, is observed at nucleotides 61–71 (Table 1). Downstream of the intergenic

region, the 5' A59 sequence contains a short open reading frame starting at nucleotide 99 and a longer one starting at nucleotide 210 continuing past nucleotide 2000 (Fig. 4). The sequence was translated into amino acids. The protein predicted from the sequence of the open reading frame beginning at nucleotide 210 is basic and almost identical to the predicted JHM protein which has been shown to be related to p28, the putative polymerase-related polypeptide (Soe *et al.*, 1987). The few amino acid differences between the strains, most of which are conservative, are shown below the A59 sequence. The JHM sequence contains an extra AAUCU repeat at the end of the leader RNA compared to A59 (see Fig. 5).

The secondary structure of the first 150 nucleotides of the A59 genome was predicted using the method of Zucker and Steigler (1981) and as shown in Fig. 5A contains two stems and loops. The structure of the 5' terminus is of interest because the leader RNA presumably is transcribed from this region. A similar sequence is observed for this region in the JHM genome (Fig. 5B). In this structure, nucleotides 52–69 of A59, containing the end of the leader and the leader/gene A intergenic regions, are found in the region between the loops and therefore may be a likely site for termination of leader synthesis as A/U-rich regions following areas of secondary structure have been associated with termination sites in several systems (Henikoff *et al.*, 1983; Mills *et al.*, 1980; Zaret and Sherman, 1982).

Because IBV contains terminal repeats proposed as possible polymerase recognition sites, computer analysis using the Intelligenetics programs for sequence or structural homologies were carried out comparing the 3' end of negative strand (complement of the 5' end of plus strand) with the published sequence of the 3' end of the MHV-A59 genome (Armstrong *et al.*, 1983). There were no significant homologies observed between the 350 nucleotides at the 3' ends of MHV-A59 positive and negative strands.

The 3' end of MHV gene A

The 3' end of the MHV-A59 gene A was sequenced using clone 1033. A 750-nucleotide *SphI/SacI* piece of the 2.3-kb *Pst* fragment of 1033 was subcloned into pGEM4 (Promega) and sequenced from each end using the universal sequence primers and using a synthetic oligonucleotide primer derived from the sequence 50 nucleotides upstream of the intergenic region found between genes A and B. This intergenic sequence, AAUCUAUAA, is followed by the first open reading frame of gene B (Luytjes *et al.*, 1988). This completes the list of intergenic regions for MHV-A59

TABLE 1
THE MHV-A59 INTERGENIC REGIONS

Leader/A ¹	AAUCUAAAC
Gene A/B ²	AAUCUAUAC
Gene B/C ³	AAUCUAAAC
Gene C/D ⁴	AAUCUAAAC
Gene D/E ⁵	AAUCUAAAC
Gene E/F ⁵	AAUC <u>CAAAC</u>
Gene F/G ⁵	AAUCUAAAC
Consensus	AAUCUAAAC
MHV-A59 leader sequence	
	10 20 30 40 50 60
	UAUAAGAGUG AUUGGCGUCC GUACCCUUUCA ACUCUAAAC UCUUGUAGUU UAAAUCUAAU
	70
	<u>CUAAUCUAAAC</u> ⁶

¹ Sequence obtained from clone ZA11.

² Sequence obtained from clone 1033.

³ Sequence obtained from Luytjes *et al.* (1987).

⁴ Sequence from our unpublished sequence of gene D.

⁵ Sequence from Budzilowicz *et al.* (1985).

⁶ The underlined sequence at the end of the leader is that region homologous with the intergenic regions.

(see Table 1). All intergenic regions contain the sequence AAUCUAAAC with the exception of the 1 nucleotide changes in two of the intergenic regions (A/B and E/F) indicated by the underlined nucleotides in Table 1.

DISCUSSION

We have created a restriction map of the unique portion of mRNA 1, the putative polymerase gene. This region is bounded by the leader sequence and intergenic sequence on the 5' end and an intergenic sequence on the 3' end. The data suggest that the entire gene is approximately 23 kb. This would suggest that the genome is in excess of 30 kb. We cannot know the exact size of the genome until the entire sequence is obtained.

Gene A is likely to encode the MHV polymerase because coronaviruses contain infectious genome RNA (Lomniczi, 1977; our unpublished results) and initiate infection by the translation of viral polymerase(s) from virion RNA. Furthermore, it is likely that translation of genome proceeds from the 5' end as is the case with other eukaryotic mRNAs. The 5' unique region of the avian coronavirus IBV is composed of two large ORFs (F1 and F2) which potentially encode polypeptides of 441K and 300K (Boursnell *et al.*, 1987). It is thought that a frameshifting mechanism (Jacks and Varmus, 1985; Brierly *et al.*, 1987) results in the synthesis of one

large polypeptide from the two ORFs. Although the 5' regions of MHV-A59 and IBV do not share significant homology (Fig. 4) (Boursnell *et al.*, 1987), protein sequence homology between the protein predicted by the IBV ORF2 and the polypeptide predicted from the nucleotide sequence of the 3' portions of MHV-A59 gene A has been observed (Bredenbeek *et al.*, manuscript in preparation).

A direct identification of the coronavirus polymerase is lacking and there is little information about the MHV gene A products. Purified MHV virion RNA directs the cell-free synthesis of a 250K polypeptide that is processed *in vitro* into 220K and 28K polypeptides (Denison and Perlman, 1986); the 28K polypeptide is derived from the amino terminus and has been identified in the infected cell (Denison and Perlman, 1987). A basic protein with the characteristics of the 28K polypeptide is encoded in the 5' end of the MHV-JHM and A59 genome RNAs (Soe *et al.*, 1987). We are currently using antisera directed against procaryotic/viral fusion proteins representing various portions of gene A to characterize the viral polypeptides responsible for polymerase activity in infected cells.

The 5' end sequence of the MHV-A59 genome RNA is almost identical to that of the JHM genome (Soe *et al.*, 1987). The major difference is an additional AAUCU (see Fig. 5) at the end of the JHM leader in the region thought to hybridize with template. It is not clear whether these differences are meaningful since the leader RNAs from each of these strains readily reasso-

```

UAUAAAGAGAGUUGGGCCGUCAGCCUCCUCAAUCUAAAUUCUAGUUAAAUCUAAUAAACUUUA 75
UUAAACGGCACUUCUGGUGUCCAUUGCCCGGGCCUGGUUUUUGUUAUGUGUGUGUGUAGUUCUUGGUGU 150
      H P A G L V L S
      V
UUCGUUCUUCGECACGUCACCGUUCUUGGGGCGCCACGACCCCAUAGGUUGCAUAAAGCCAAAGUGGCCA 225
      H A K M G
AAUACGGUCUGGCUCAAUUGGGCCCCGAAUUUCCAUUGGAGUUCGCGAACGCAUCGGAGAAUGGUGUAACC 300
K Y G L G P K W A P E F P W H L P H A S E K L C H
CUGAGAGUCAGAGGGAGUUGGUGUUUUGCCCCUCUGCGCAAGAACCGAAAGUUAAGGAAAUAAUUUGUUA 375
P E R S E E D G F C P S A A Q E P K V K G K T L V
AUCACGUGAGGGGAAUUGUAGCCGHHUCCAGCUUUGGAGUGUGUUCAGUCUGCAAUUAUUGCCUGAUUU 450
N H V R V N C E R X P A L E C C V Q S A I T R D I
      D L
UUGUAGUAGAGGAUCCCGAAGGUGGAGCCUCAAUGAUGGCAUUGCAGUUCGGUAGUGCCGUCUUGUUA 525
F V D E D P Q K V E A S T H M A L Q F G S A V L V
AGCCAUCCAAGCGCUUGUUAUUCAGCAGUGGACUAAUUUGGGGUGUUCUCCAAAACAGUGCCAUUGGGGUGU 600
K P S K R L S T Q A W T N L G V L P K T A A H G L
      V A K
UCAAGCGGUCUGCCUGUAAACACAGGAGUGUCUUGUGAGCCCAUGGCGUUCACCUUUUUCAGGUC 675
F X R V C L C N T R E C S C D A H V A F H L P T V
      F V Q
AAACCGAUGUGUAGUGCCGUAUUGCCGUAUUGGCGUUGUUGUCCAGUCACGACCAUACCGGAGUUG 750
Q P D G V C L G N G R F I G W F V P V T A Y P E Y
CGAAGCAGUGGUGCAAUUGCCUGGCUUCCUUCUUGUAAGGUGGUAACAAGGUGUGGACAUCCGGCCAU 825
A X Q W L Q P W S I L L R K G G H K G S V T S G H
UCCCGCGCGUGUUAUUGCCUGUUAUGCAUUUAUGAUGAGGUGUUGUGAGGAGGUGUUAUUCUUAACCGCA 900
F R R A V T H P V U D P V F N V E D A C E E V H L N P
AGGUAAGUACUCCCGCAAGGCGUAGUCUUAUUGGCGUUAUGCGGUGUAUAGCCCAUCCUUGUUGGAGCC 975
K G K Y S C K A Y A L L X G Y R G V K P I L F V D
      R
AGUAUUCUUCGACUUAUCUGGAGUCUCUCCGAAAGGUCUUCAGGACUUAUGCCGUAUUCUUCAGUGAGUA 1050
Q Y G C D Y T G C L A K G L E D Y G D L T L S E H
AGGUAUUGUCCUUGGCGGUCAGUCCUUGGUAUGGAAUGUCCUUGGUGGUGGUGGUGGUGUUGGAGUAGUCC 1125
K E L F P V C A D S L D S E V L V A W H V D R D P
      S H R N V
GGGUGUUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGG 1200
R A A M R V Q T L A T V R C I D Y V G Q P T E D V
      V L S E I M
UGGAGGAGUUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGG 1275
V D C D V V R E P A H L L A A N A I V K R L P R
      M P
UGGUGGAGUUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUG 1350
L V E T M L Y T D S S V T V E F C Y K T K L C E C G
      D
UUUAUCGCAUUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUG 1425
F I T Q F G Y V D C C G D T C D F R G W V A G N M
      G P
UGGAGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUG 1500
H D C F P C P G C T X S N Y M P W E L E A Q S S G V
      C
UACCAAGGAGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGG 1575
I A P E G G V L F T Q S T D T V N R E S F K L Y G H
CUGUUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGG 1650
A V V P P G S A V Y W S P C P G H W L P V I W S S
      G A Y
UUUAUCGCAUUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUG 1725
V K S Y S G L T Y T G V V G C K A I V Q E T D A I
      Y
GUCGUUCUUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGG 1800
C R S L Y M D Y V Q H K C G N L E Q R A I L G L D
      F
AUCUCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1875
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
AGCGCGCGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUG 1950
K R R A E P A C K P A T C G G C L V F L L L D G L
      D
UGCCCGCAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2000
V P R S Y Y L I K A G U G G U C A A C C U T S H
    
```

Fig. 4. Sequence of the 5' end of genome RNA. The sequence of the 2000 nucleotides and the amino acids encoded at the 5' end of MHV-A59 genome RNA are shown. Also shown are amino acids that differ between A59 and JHM.

ciate with templates from either strain (Makino *et al.*, 1986). The differences in the predicted amino acid sequences are mostly conservative and are indicated in

Fig. 4. Thus the p28 putative polymerase related protein is conserved between the two strains.

Since the 3' end of the viral negative strand (the complement of the 5' end of the genome) is likely to be the template for transcription of leader RNA, we analyzed the secondary structure of this region. Secondary structure analysis of the 5' ends of the A59 genome RNA showed two stem and loop structures with $\Delta G = -35.2$ kcal. This structure is similar to that obtained by a similar analysis of the 5' end sequence of the JHM genome. This structure is maintained even as more nucleotides (up to 350) are added to the analysis which suggests that this is maintained as transcription continues further. It is not clear how leader is transcribed, where its transcription is terminated, and where it is released from its template. It seems likely that nucleotide 67 marks the end of the leader transcript as it is the first nonhomologous nucleotide between leader and template in the gene E/F intergenic region (Table 1). This is supported by the observation that this is an AU-rich region and is found in a region not involved in a secondary structure loop (Henikoff *et al.*, 1983; Mills *et al.*, 1980; Zaret and Sherman, 1982). If nucleotide 66 is the end of the leader transcript, then the AAC (nucleotides 68–71) may be important not for primer hybridization but perhaps for protein interactions needed for initiation of transcription.

Secondary structure analysis of nucleotides 48–150 of MHV-JHM led to the hypothesis that the AU-rich region around nucleotides 75–76 (Fig. 5) may be a site where polymerase pauses and marks the end of leader transcription (Baric *et al.*, 1987a,b; Shieh *et al.*, 1987). This is supported by the presence of leader-related transcripts greater than 66 nucleotides in infected cells (Baric *et al.*, 1987a; Shieh *et al.*, 1987). However, as predicted for MHV-A59 or for JHM, when the nucleotide 1–150 (includes the complete 5' terminus) are analyzed, nucleotides 75 and 76 are contained within the second stem and are not left free.

The 3' ends of positive- and negative-strand genome-size RNA presumably encode important regulatory sequences involved in the replication of genome. It was of interest therefore to compare the 3' terminal sequence of plus- and minus-strand RNAs. Conservation of a primary nucleic acid sequence or of a secondary structure at the 3' ends of both plus- and minus-strand RNAs would suggest that the same or a similar replication complex recognizes both RNA species. Indeed, sequence homology between the 3' ends of both plus- and minus-strand RNAs has been identified in the related coronavirus, IBV. This sequence is about 60 nucleotides in length and is located approximately 50 nucleotides from the 3' ends of both RNA species

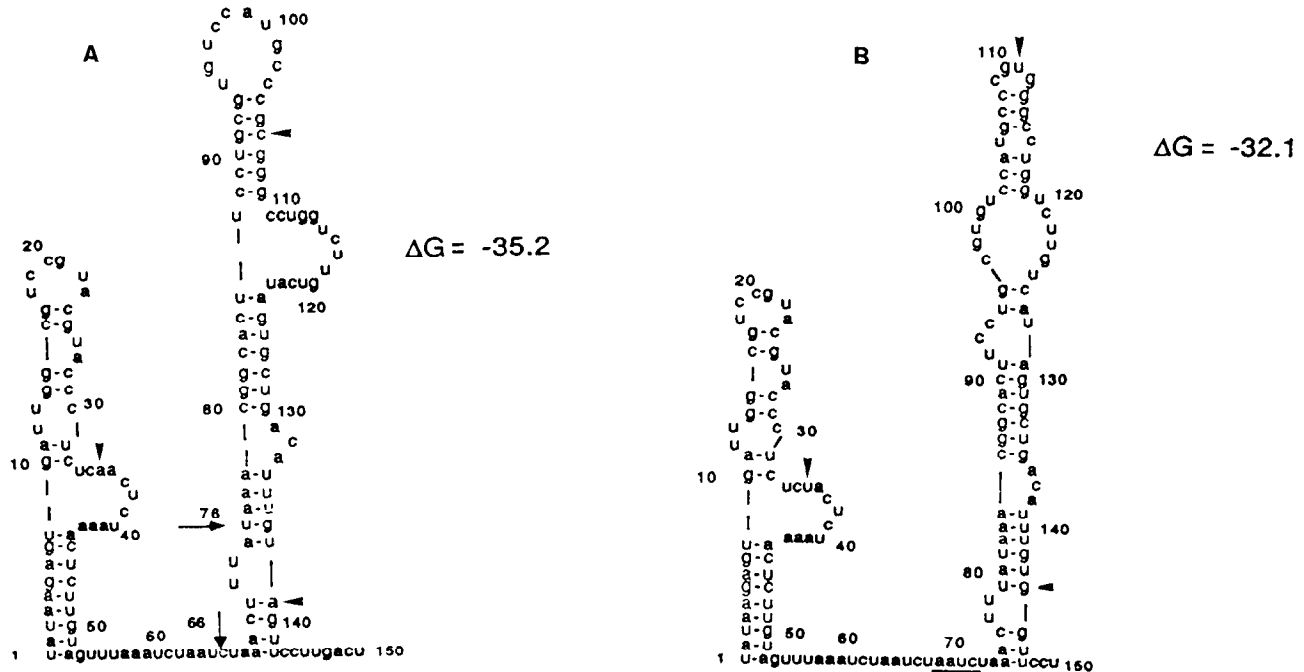


Fig. 5. Computer prediction of the secondary structure of the 5' end of genome RNA. The first 150 nucleotides of the MHV A59 and JHM genome sequence were analyzed by the program of Zucker and Steigler (1981). Arrowheads are used to mark the differences between the sequences of the two strains. (A) MHV-A59 sequence. The loop encompassing nucleotides 1–52 has a $\Delta G = -12.4$ kcal while the loop encompassing nucleotides 69–142 has a $\Delta G = -22.8$ kcal. The total $\Delta G = -35.2$ kcal. Nucleotide 66 marks the last nucleotide that is conserved in the leader sequences of all mRNAs and is thus a possible site of termination of leader RNA. Nucleotide 76 is the site proposed by Shieh *et al.* (1987) as a possible termination site for transcription of larger leader transcripts. (B) JHM sequence taken from Soe *et al.* (1987). The loop encompassing nucleotides 1–52 has a $\Delta G = -12.4$ kcal while the loop encompassing nucleotides 74–147 has a $\Delta G = -19.7$ kcal. The total $\Delta G = -32.1$ kcal. The extra AAUCU found in the JHM sequence is underlined.

(Bournsnel *et al.*, 1987). However, this homology was not observed in the case of MHV-A59. This is not surprising since in the cases of some other classes of RNA viruses, for example, flaviviruses (Brinton and Disputo, 1988) and alphaviruses (Levis *et al.*, 1986), the regulatory sequences or secondary structures are different on plus and minus strands.

ACKNOWLEDGMENTS

We thank Dr. Masayori Inouye for construction of the leader sequence clone. We acknowledge the excellent technical assistance of Carol Budzilowicz, Edward Murray, and Ans Noren. This work was supported by NIH Grants AI-17418 and NS-21954 to S.R.W. C.P. was supported by NIH Grant AI-17418 and Training Grant T32 AI-07325. P.B. was sponsored by the Netherlands Organization for the Advancement of Pure Research (N.W.O.). P.Z. was supported by NIH Grant NS-21954 and Training Grant T32 NS-07180.

REFERENCES

- ARMSTRONG, J., SMEEKENS, S., and ROTTIER, P. J. M. (1983). Sequence of the nucleocapsid gene from murine coronavirus MHV-A59. *Nucleic Acids Res.* **11**, 883–891.
- BARIC, R. S., SHIEH, C. K., STOHLMAN, S. A., and LAI, M. M. C. (1987a). Studies into the mechanism of MHV transcription. In "Coronaviruses. Advances in Experimental Medicine and Biology" (M. M. C. Lai and S. A. Stohlman, Eds.), Vol. 218, pp. 137–149. Plenum, New York.
- BARIC, R. S., SHIEH, C. K., STOHLMAN, S. A., and LAI, M. M. C. (1987b). 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). Characteristic 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.
- BIGGINS, M. D., GIBSON, T. J., and HONG, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
- BOURNSELN, M. E. G., BROWN, T. D. K., FOULDS, I. J., GREEN, P. H., TOMLEY, F. M., and BINNS, M. M. (1987). The complete sequence of the genome of avian infectious bronchitis virus. *J. Gen. Virol.* **68**, 57–77.
- 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**, 197–204.

- BRIERLY, I., BOURSNEILL, M. E. G., BINNS, M. M., BILIMORIA, B., BLOK, V. C., BROWN, T. D. K., and INGLIS, S. C. (1987). An efficient ribosomal frame-shifting signal in the polymerase encoding region of the coronavirus IBV. *EMBO J.* **6**, 3779–3785.
- BRINTON, M. A., and DISPOTO, J. H. (1988). Sequence and secondary structure analysis of the 5' terminal region of the flavivirus genome RNA. *Virology* **162**, 290–299.
- BUDZIŁOWICZ, C. B., WILCZYŃSKI, S. P., and WEISS, S. R. (1985). Three intergenic regions of coronavirus mouse hepatitis virus strain A59 genome RNA contain a common nucleotide sequence that is homologous to the 3' end of the viral mRNA leader sequence. *J. Virol.* **53**, 834–840.
- 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.
- DENISON, M. R., and PERLMAN, S. (1987). Identification of a putative polymerase gene product in cells infected with murine coronavirus A59. *Virology* **157**, 565–568.
- GUBLER, U., and HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.
- HENIKOFF, S., KELLY, J. D., and COHEN, E. H. (1983). Transcription terminates in yeast distal to a control sequences. *Cell* **33**, 607–614.
- JACKS, T., and VARMUS, H. E. (1985). Expression of the Rous sarcoma virus *pol* gene by ribosomal frame shifting. *Science* **230**, 1237–1242.
- 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**, 3626–3639.
- LAI, M. M. C., BARIC, R. S., MAKINO, S., KECK, J. G., EGBERT, J., LEIBOWITZ, J. L., and STOHLMAN, S. A. (1985). Recombination between nonsegmented RNA genomes of murine coronaviruses. *J. Virol.* **56**, 449–456.
- LAI, M. M. C., BRAYTON, P. R., ARMEN, P. C., PATTON, C. D., PUGH, C., and STOHLMAN, S. A. (1981). Mouse hepatitis virus A59: mRNA structure and localization of 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). Further characterization of mRNA's of mouse hepatitis virus: Presence of common 5'-end nucleotides. *J. Virol.* **46**, 1027–1033.
- LAI, M. M. C., and STOHLMAN, S. A. (1978). RNA of mouse hepatitis virus. *J. Virol.* **26**, 236–242.
- LEHRACH, H., DIAMOND, D., WOZNEY, J. M., and BOEDTKER, H. (1977). RNA molecular weight determinations of gel electrophoresis under denaturing conditions: A critical reexamination. *Biochemistry* **96**, 4743–4751.
- LEIBOWITZ, J. L., WEISS, S. R., PAAVOLA, E., and BOND, C. W. (1982). Cell-free translation of murine coronavirus RNA. *J. Virol.* **43**, 905–913.
- LEIBOWITZ, J. L., WILHELMSON, D. 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.
- LEVIS, R., WEISS, B. G., TSIANG, M., HUANG, H., and SCHLESINGER, S. (1986). Deletion mapping of Sindbis virus DI RNAs derived from cDNAs defines the sequences essential for replication and packaging. *Cell* **44**, 137–145.
- LOMNICZI, B. J. (1977). Biological properties of avian coronavirus RNA. *Gen. Virol.* **36**, 531–533.
- LUYTJES, W., BREDENBEEK, P. J., NOTEN, A. F. S., HORZINEK, M. C., and SPAAN, W. J. M. (1988). Sequence of mouse hepatitis virus A59 mRNA: Indications for RNA recombination between coronaviruses and influenza C virus. *Virology* **166**, 415–422.
- LUYTJES, W., STURMAN, L., BREDENBEEK, P. J., CHORITE, J., VAN DER ZEIJST, B. A. M., HORZINEK, M. C., and SPAAN, W. J. M. (1987). Primary structure of the glycoprotein E2 of mouse hepatitis virus strain A59 and identification of the trypsin cleavage site. *Virology* **161**, 479–487.
- MAKINO, S., STOHLMAN, S. A., and LAI, M. M. C. (1986). Leader sequences of murine coronavirus RNAs can be freely reassorted: Evidence for the role of free leader RNA in transcription. *Proc. Natl. Acad. Sci. USA* **83**, 4204–4208.
- MANAKER, R. A., PICZAK, C. V., MILLER, A. A., and STANTON, M. F. (1961). A hepatitis virus complicating studies with murine leukemia virus. *J. Natl. Cancer Inst.* **27**, 29–51.
- MANIATIS, T., FRITSCH, E. F., and SAMBROOK, J. (1982). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- MESSING, J., CREA, R., and SEEBURG, P. H. (1981). A system for shotgun DNA sequencing. *Nucleic Acids Res.* **9**, 309–321.
- MILLS, D. R., DABKIN, C., and KRAMER, F. R. (1980). Template determined, variable rate of chain elongation. *Cell* **15**, 541–550.
- RIGBY, P. W. J., DIECKMANN, M., RHODES, C., and BERG, P. (1977). Labeling of deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**, 237–251.
- 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., 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.
- SOE, L., SHIEH, C. K., BAKER, S., 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.
- SOUTHERN, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517.
- SPAAN, W. J. M., CAVANAGH, D., and HORZINEK, M. C. (1988). Coronaviruses: Structure and genome expression. *J. Gen. Virol.* **69**, 2939–2952.
- SPAAN, W. J. M., DELIUS, J., SKINNER, M. A., ARMSTRONG, J., ROTTIER, P. J. W., SMEEKENS, S., VAN DER ZEIJST, B. A. M., and SIDDELL, S. G. (1983). Coronavirus mRNA synthesis involves the fusion of non-contiguous sequences. *EMBO J.* **2**, 1839.
- SPAAN, W. J. M., ROTTIER, P. J. M., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M. (1982). Sequence relationship between the genome and intracellular RNA species 1, 3, 6 and 7 of mouse hepatitis virus strain A59. *J. Virol.* **42**, 432–439.
- THOMAS, P. (1980). Hybridization of RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
- WEISS, S. R., and LEIBOWITZ, J. L. (1983). Characterization of murine coronavirus RNA by hybridization with virus-specific cDNA probes. *J. Gen. Virol.* **64**, 127–133.
- ZARET, K. S., and SHERMAN, F. (1982). DNA sequences required for efficient transcription termination in yeast. *Cell* **28**, 563–373.
- ZUCKER, M., and STEIGLER, D. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**, 133–148.