JOURNAL OF VIROLOGY, Mar. 1985, p. 834–840
0022-538X/85/030834-07$02.00/0

Vol. 53, No. 3

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

CAROL J. BUDZILOWICZ, SHARON P. WILCZYNSKI, AND SUSAN R. WEISS*

Department of Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received 9 October 1984/Accepted 30 November 1984

cDNA clones that represent various portions of the coronavirus mouse hepatitis virus strain A59 genome RNA have been constructed. cDNAs were synthesized by transcription of genome RNA by using either oligo(dT)$_{12-18}$ or random oligomers of calf thymus DNA as primers. These cDNAs were converted into double-stranded DNA and cloned into pBR322 by standard techniques. The resulting cloned viral DNA fragments were mapped to viral genes by hybridization with Northern blots of intracellular RNA from mouse hepatitis virus strain A59-infected cells. These cDNA clones map in six of the seven viral genes. Clone g344, 1.8 kilobases, is the largest and encompasses gene 5 (which encodes a nonstructural protein) and gene 6 (which encodes the E1 viral glycoprotein) as well as the intergenic regions preceding genes 5, 6, and 7. Sequencing of parts of this cloned DNA show that these three intergenic regions contain a common 11-nucleotide sequence. This sequence shares homology with the 3′ end of the viral mRNA leader sequence. Thus, this common intergenic sequence may contain a binding site for a leader RNA that hybridizes to negative-strand viral RNA at the beginning of each gene to prime mRNA synthesis. The different degrees of homology between the leader and its putative binding site may influence the differential rates of transcription of the various viral mRNAs.

Coronaviruses form a group of enveloped positive-stranded RNA viruses that cause a variety of diseases in many species of animals (20). One member of this group is the murine virus, mouse hepatitis virus (MHV). The genome of MHV strain A59 is a polyadenylated RNA of approximately 18,000 nucleotides (15, 18). During infection of cultured cells MHV A59 generates a nested set of seven subgenomic mRNAs that all overlap in sequence with the 3′ end of genome RNA (8, 12, 18, 29, 33) (Fig. 1).

UV transcriptional mapping studies have shown that the synthesis of each MHV A59 subgenomic RNA is initiated independently (10), suggesting that these RNAs are not spliced from larger precursors. However, it has recently been shown that the subgenomic RNAs contain a leader sequence of approximately 70 nucleotides that is probably derived from the 5′ end of genome RNA (1–3, 11, 13, 28, 29). These data, taken together with the finding of one genome-sized negative-strand RNA (6), suggest that transcription of subgenomic RNAs is initiated from a full-genome-length negative-strand template and that during the genesis of these subgenomic RNAs there is a joining of nonadjacent sequences.

We describe here the cloning, mapping, and partial sequencing of MHV A59 virus-specific cDNA sequences. The cloned fragments described below map in the sequences unique to each of the subgenomic RNAs (except RNA 4) and thus in six of the seven MHV A59 genes. These are the first cloned DNAs derived from MHV genome RNA (as opposed to mRNAs) and the first to the 5′ side of the two genes most proximal to the 3′ end of the genome, N and E1. Two of these cloned DNA fragments were used to sequence the intergenic regions preceding genes 5, 6, and 7. Comparison of these intergenic regions shows a common sequence of 11 nucleotides. Furthermore, this common sequence shares homology with the 3′ end of the leader sequence found on viral mRNAs (1, 2, 11, 22, 28). This conserved intergenic sequence may be involved in the control of transcription of the mRNAs.

MATERIALS AND METHODS

Virus and cells. MHV A59 (19), a tissue-culture-adapted strain of MHV, was obtained from Julian Leibowitz and grown in monolayer cultures of 17CL-1 mouse fibroblasts. For preparation of virus-specific RNA, cells were infected at a multiplicity of infection of 1 and harvested approximately 16 h postinfection (33).

Preparation of RNA. Virus to be used as a source of genome RNA was isolated from the medium above infected cells that had been labeled with [^{3}H]uridine. The virus was purified by a combination of sucrose velocity and density gradients as previously described for coronaviruses (15). RNA was extracted by sodium dodecyl sulfate-proteinase K treatment followed by phenol extraction and ethanol precipitation, sedimented through a sucrose velocity gradient, and isolated as a uniform peak containing 57S RNA (33). RNA was extracted from the cytoplasm of MHV A59-infected cells that had been lysed with 0.1% sodium dodecyl sulfate and 0.5% Nonidet P-40 and purified by sodium dodecyl sulfate-proteinase K treatment, phenol extraction, and ethanol precipitation (33).

Agarose gel electrophoresis, blotting, and hybridization. DNAs were electrophoresed in 1.0 to 2.0% agarose gels in 0.04 M Tris acetate–0.001 M EDTA, and the DNA bands were visualized by UV light after the gel was stained with ethidium bromide. For preparative electrophoresis, low-melting-point agarose was used. After electrophoresis and staining, slices containing the DNA fragments were melted at 65°C for 20 min and the DNA was purified by phenol-chloroform extraction followed by ethanol precipitation. Such

* Corresponding author.

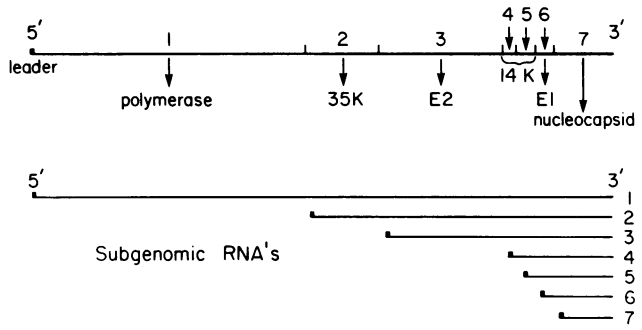

FIG. 1. MHV A59 genome RNA and subgenomic intracellular mRNAs. The gene order was deduced from cell-free translation experiments, assuming that each RNA is translated from the 5' unique sequences (17, 25, 26), and from sequencing of mRNAs 6 and 7 (1–3). The boundaries and approximate sizes of the genes were estimated from the sizes of the unique portions of the subgenomic RNAs (12, 18, 29) and from the sequencing data for genes 6 and 7 (1–3). The structural proteins are the nucleocapsid, E1 (small glycoprotein), and E2 (large peplomer glycoprotein). The nonstructural proteins include the 35K, 14K, and polymerase proteins (26). The location of the polymerase gene is tentative and is deduced from the facts that genome RNA is infectious and must therefore encode polymerase(s) and that eukaryotic translation usually proceeds from the 5' end of RNA. It has not been determined whether the 14K protein is translated from RNA 4 or 5 (17). The leader sequence is shown by the bar at the 5' end of each mRNA.

fragments were used for nick translations or restriction digests.

RNAs were electrophoresed in 1.0% agarose gels under denaturing conditions in the presence of formaldehyde (16) and blotted onto nitrocellulose for Northern blots. Bacterial supernatants containing recombinant M13 phage were applied directly onto nitrocellulose for dot blot analyses. Northern blots and dots blots were carried out as described by Thomas (32).

Probes for viral sequences were [^{32}P]dCTP-labeled cDNA transcribed from viral genome primed by oligomers of calf thymus DNA (31, 33), nick-translated (24) pBR322 recombinant plasmid DNAs containing viral DNA, or nick-translated purified viral DNA fragments. All probes had specific activities of approximately 10^8 cpm/μg.

Cloning into pBR322. Synthesis and cloning of double-stranded virus-specific DNA was carried out by standard techniques (9). Purified MHV A59 genome RNA was used as a template for the synthesis of cDNA with oligo(dT)$_{12-18}$ as primer and reverse transcriptase. Genome RNA was also transcribed in similar reactions by using oligomers of calf thymus DNA as primers at a primer/template ratio of approximately 6:1 (wt/wt) (23). Second-strand synthesis was carried out with the Klenow fragment of *Escherichia coli* polymerase I. The resulting double-stranded DNA was digested with S1 nuclease, and the larger DNA (500 to 4,000 base pairs) was selected by chromatography on an agarose A150 column. Homopolymeric tails were added to the DNA by using terminal transferase and dCTP. The tailed DNA was then hybridized to bacterial plasmid pBR322 DNA that had previously been linearized with endonuclease *Pst*I and tailed with dGTP. These hybrid DNAs were used to transform *E. coli* HB101. Bacterial colonies containing plasmids were selected and screened for viral inserts by two rounds of hybridization with ^{32}P-labeled cDNA (33) and colony filter hybridization techniques. Approximately 95% of the colonies contained viral sequences. Plasmid DNAs from some colonies were nick translated (24) and shown to hybridize with RNA from MHV A59-infected 17CL-1 cells but not with RNA from uninfected cells; this further confirmed that the plasmids contained viral sequences. Clones derived from oligo(dT)$_{12-18}$-primed DNA (a total of 300) were designated g1, g2, etc.; those derived from calf-thymus-primed DNA (a total of 120) were designated ct1, ct2, etc.

Subcloning into M13 and DNA sequencing. Viral inserts were removed from pBR322 by endonuclease *Pst*I digestion and ligated into the *Pst*I site of the replicative form (RF) of bacteriophage M13, strain mp8 (21). Alternatively, the viral inserts were digested with endonucleases *Hpa*II or *Hae*III and the resulting fragments were treated with T4 polymerase and blunt-end ligated into the *Sma*I site of the mp8 RF (21). The recombinant RFs were used to transform *E. coli* JM 103. The resulting phage were plaque purified and screened for insert size by agarose gel electrophoresis and for viral sequences by dot blot hybridization with nick-translated probes representing the viral insert or restriction fragment desired. Phage clones containing viral positive-strand DNA were differentiated form those containing negative-strand DNA by hybridization with virus-specific cDNA which detects positive-strand viral sequences only (33). Sequencing was accomplished by the Sanger dideoxynucleotide-chain-terminating method as modified by Biggins et al. (5).

RESULTS

Cloning and mapping of sequences from MHV A59 genome RNA. To sequence MHV intergenic regions, cDNA clones representing various portions of genome RNA were constructed. These cloned fragments were then mapped on the viral genome, and clones potentially containing intergenic sequences were selected for further study. Thus, virus-specific DNA sequences were transcribed from polyadenylated MHV A59 genome RNA by using oligo(dT) to prime specifically at the 3' end or random oligomers of calf thymus DNA to prime at random sites along the genome. These cDNAs were converted into double-stranded DNA and inserted into the bacterial plasmid pBR322 as described above. The resulting recombinant plasmid DNAs were digested with endonuclease *Pst*I, and the viral fragments were analyzed on agarose gels. The average size of the viral inserts from the clones derived from oligo(dT)-primed DNA (designated g1, g2, etc.) was 650 base pairs, and the average size of the viral inserts from the clones derived from randomly primed DNA (designated ct1, ct2, etc.) was 350 base pairs. Clone g344 contained a viral insert of 1,800 base pairs, much larger than any of the other cloned DNAs.

To map these cloned DNA fragments along the viral genome, recombinant plasmid DNAs (or in some cases purified viral inserts) were nick translated, and the resulting probes were hybridized to Northern blots of intracellular RNAs from MHV A59-infected cells. MHV A59 intracellular RNA consists of a nested set of subgenomic RNAs overlapping at the 3' ends (8, 12, 18, 29) (Fig. 1). These subgenomic RNAs therefore represent progressively longer segments of genome RNA with common 3' ends. Cloned fragments were assigned to the various genes on the basis of the smallest RNA with which they hybridized. A representative experiment is shown in Fig. 2. The map positions of the cloned fragments derived from randomly primed DNA were distributed among all the genes with the exception of the smallest, gene 4. Of the 35 clones so far mapped, 21 (60%) mapped in gene 1, the largest and putative polymerase gene. Three to

four clones mapped in each of the remaining five genes. (None of these clones contained the 70-nucleotide leader sequence derived from the 5′ end of genome RNA [1–3, 11, 13, 28, 29]. The viral insert from a clone containing the leader sequence would hybridize to all seven RNAs and thus appear to map in gene 7. However, such a clone would not contain sequences from the coding region of mRNA 7. All of the clones obtained that mapped in mRNA 7 contained sequence homology with clone c8 which is derived from the 3′ portion of gene 7 [data not shown]. These clones were therefore derived from the coding region of gene 7 and not from the leader sequence.) All of the clones derived from oligo(dT)-primed DNAs (with the exception of clone g344, which will be discussed in detail below) shared sequence homology as determined by dot blot hybridization (data not shown). All cloned DNAs derived from oligo(dT) priming mapped to gene 7 by Northern blot analysis (as in Fig. 2f). However, sequencing of a few of these clones demonstrated that they were not derived from the extreme 3′ end of the genome, as they did not contain polyadenylate tails. Clone g344, although derived from oligo(dT)-primed DNA, was not homologous to the smaller oligo(dT)-primed clones and extended from gene 7 at least into gene 5 as demonstrated by its homology to DNAs from clones ct54, ct55, and ct71 that all mapped in gene 5. This clone was selected for further study of intergenic sequences.

Intergenic regions. To more precisely map the 1,800-base-pair viral insert from clone g344 and to isolate intergenic regions in smaller DNA fragments appropriate for sequencing, restriction digests were carried out. Thus the viral insert was excised from g344 plasmid DNA, separated from pBR322 DNA by preparative agarose gel electrophoresis, and then digested with endonuclease *Hpa*II. *Hpa*II digestion followed by agarose gel electrophoresis resulted in three major bands of approximately 1,400 (*Hpa* A), 250 (*Hpa* B), and 150 base pairs (*Hpa* C). To map these DNAs on the viral genome, the fragments were eluted from agarose gels, nick translated, and hybridized to Northern blots of intracellular RNA from MHV A59-infected cells as described above. The *Hpa* A fragment hybridized to RNAs 1 through 6, while the *Hpa* B and *Hpa* C fragments hybridized to RNAs 1 through 7 (Fig. 3). This indicates that, whereas the *Hpa* B and *Hpa* C fragments contain sequences from gene 7, the 3′ end of the *Hpa* A fragment must be in gene 6. These fragments were aligned on the viral genome by using these Northern blot data as well as the known *Hpa*II restriction sites in genes 6 and 7 (1, 2) (Fig. 4). The position of the *Hpa* A and *Hpa* C bands as end fragments was later confirmed by partial sequencing of these fragments, which demonstrated oligodeoxycytidylic acid tracts as well as the sequences from the ends of the original g344 insert (data not shown). This cloned DNA therefore mapped from approximately 200 base pairs in gene 4 to 200 base pairs into gene 7 and must include the intergenic regions preceding genes 5, 6, and 7. The *Hpa* B fragment had to contain the intergenic sequence preceding gene 7.

To obtain DNA fragments containing the other intergenic regions, the viral insert from g344 DNA was digested with endonuclease *Hae*III. This digestion resulted in eight fragments which were mapped to viral genes by hybridization to Northern blots of intracellular RNA from MHV A59-infected cells as described above. The blots of the fragments likely to contain intergenic regions, *Hae* B and *Hae* C, are

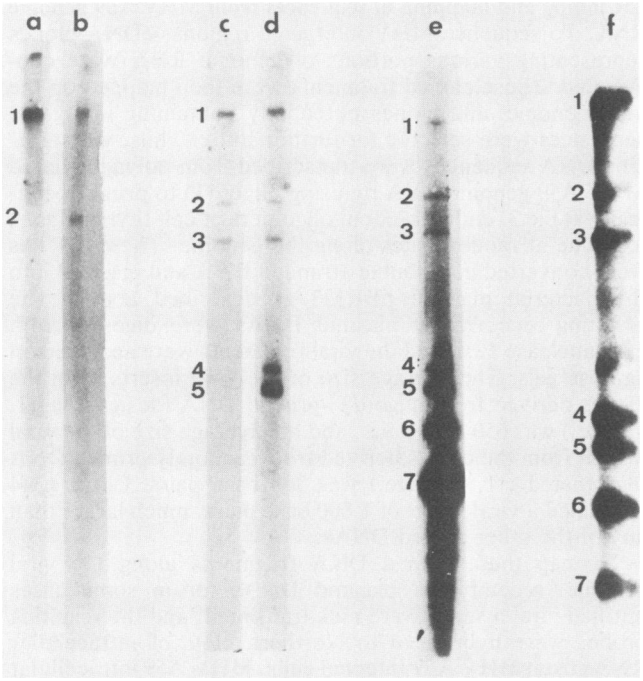

FIG. 2. Mapping of MHV A59 cloned DNAs by Northern blot hybridization. Viral inserts were excised from recombinant plasmids, purified by electrophoresis and elution from agarose gels, and nick translated with ^{32}P[dCTP]. Each probe (10^6 cpm) was hybridized to a strip from a Northern blot of RNA from MHV A59-infected 17CL-1 cells (32). Hybridization and exposure to X-ray film were each done for 24 h (32). Lanes: (a) Clone ct42 (gene 1); (b) clone ct50 (gene 2); (c) clone ct106 (gene 3); (d) clone ct54 (gene 5); (e) clone ct30 (gene 7); and (f) clone g344 (gene 7). Lanes a-b, c-d, e, and f were from different Northern blots. In each case another strip from the same blot was hybridized with MHV A59 cDNA to locate the positions of all seven virus-specific intracellular RNAs.

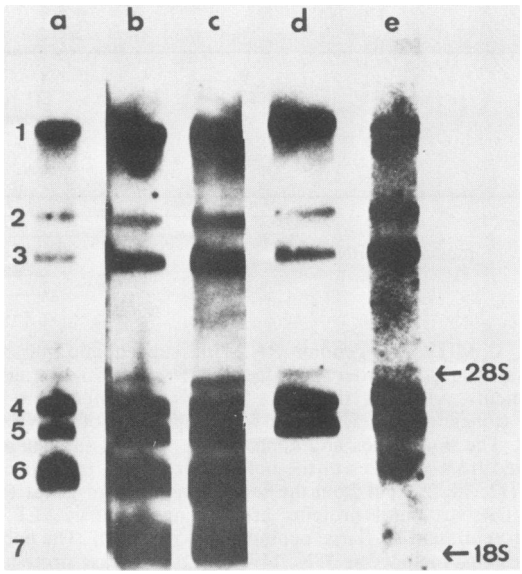

FIG. 3. Mapping of clone g344 restriction fragments by Northern blot hybridization. *Hae*III- or *Hpa*II-digested g344 restriction fragments were eluted from agarose gels as described in the text and nick translated, and the resulting probes were hybridized with strips from Northern blots of RNA from MHV A59-infected cells as described in Fig. 2. Lanes: (a) *Hpa* A; (b) *Hpa* B; (c) *Hpa* C; (d) *Hae* B; (e) *Hae* C. A parallel Northern blot strip was hybridized with virus-specific cDNA to locate the seven virus-specific RNAs. The arrows indicate the migration of 18S and 28S marker rRNAs.

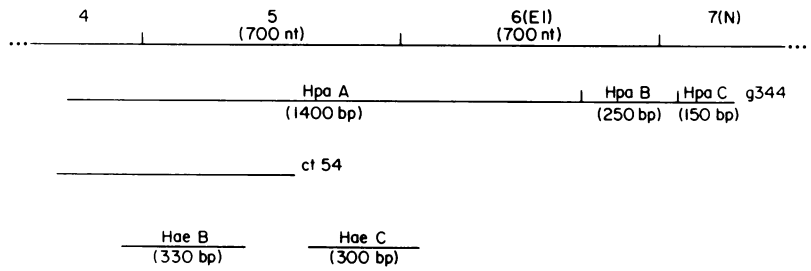

FIG. 4. Map positions of g344 and ct54 viral DNAs along the viral genome. The *Hpa*II digest restriction fragments of g344 DNA were aligned on the viral genome by using the results of Northern blot hybridization (Fig. 3) and the known *Hpa*II restriction sites in genes 6 and 7 (1, 2). The *Hae* B and *Hae* C fragments were aligned by using Northern blot data (Fig. 3) and known restriction sites in genes 6 and 7 (1, 2) and were more precisely aligned by sequence data. The position of ct54 DNA was determined by its hybridization with mRNA 5 (Fig. 2), homology with g344, and its *Hae*III restriction digest pattern (not shown). Abbreviations: nt, nucleotides; bp, base pairs.

shown in Fig. 3. The *Hae* B fragment hybridized to RNAs 1 through 5, indicating that its 3′ end must be in gene 5, whereas the *Hae* C fragment hybridized with RNAs 1 through 6, indicating that its 3′ end is in gene 6. These fragments were aligned on the genome again by using the Northern blot data and the known *Hae*III restriction sites in genes 6 and 7 (1, 2) and were more precisely aligned after sequencing as described below (Fig. 4).

For sequencing, the *Hpa*II- and *Hae*III-digested g344 DNAs were subcloned into the RF of bacteriophage M13, strain mp8. The resulting phage clones containing the individual restriction fragments were selected by hybridization with nick-translated probes derived from each of these fragments. Subclones containing the *Hae* B, *Hae* C, and *Hpa* B fragments were sequenced by the dideoxynucleotide-chain-terminating method (5). For each fragment, at least two subclones were sequenced and each subclone was sequenced at least four times. The gene 4-5 intergenic sequence was confirmed by sequencing the *Hae* B restriction fragments derived from an independent pBR322 clone, clone ct54, that overlapped with g344 and mapped in gene 5 (Fig. 2 and 4). The intergenic sequences (converted into RNA sequences) are shown in Fig. 5. A common 11-nucleotide sequence, CUAAUCU(C)AAAC, is present and for genes 6 and 7 closely precedes an AUG sequence. It is clear from the published sequencing data of genes 6 and 7 (1–3) that these are the initiation codons for the N and E1 proteins, respectively. Furthermore, this common sequence is also present at the end of the noncoding regions just 5′ to the E1 and N genes in mRNAs 6 and 7, respectively (1–3, 22, 28) (Fig. 5). Nucleotide 7 is a U in the common intergenic sequences preceding genes 5 and 7 and a C preceding gene 6. This is consistent with the sequences obtained for the 5′ noncoding regions mRNAs 6 and 7 (1–3, 22, 28). This conserved sequence is discussed below in relation to the viral leader sequences and possible function in transcriptional control.

DISCUSSION

A leader sequence of approximately 70 nucleotides, probably derived from the 5′ end of genome RNA, has been found on the 5′ ends of all the MHV A59 subgenomic RNAs (11) and has been sequenced for mRNAs 6 and 7 (1–3, 22,

A		1 5 ↓ 10		
Gene 4/5 (Genome)	AUUAUGUUA	CUAAUCUAAAC	CUCAUCU.....AUGAGACCACAG	
B				E1 Coding Sequence
Gene 5/6 (Genome)	GGAUGAUAU	CUAAUCCAAAC	AUU	AUGAGUAGUACU
Gene 6 (mRNA)	AGUUUAAAU	CUAAUCCAAAC	AUU	AUGAGUAGUACU
C				N Coding Sequence
Gene 6/7 (Genome)	GUUGAGAAC	CUAAUCUAAAC	UUUAAGG	AUGUCUUUUGUU
Gene 7 (mRNA)	NGUUUAAAU	CUAAUCUNAAC	UUUAAGG	AUGUCUUUUGUU

FIG. 5. MHV A59 intergenic sequences. The sequences obtained for the MHV A59 intergenic regions are shown in comparison with the 5′ noncoding regions of the mRNAs of the corresponding genes. The beginnings of the coding regions for genes 6 and 7 are also shown. The common nucleotides are numbered 1 through 11. Homology between the genomic sequences and the corresponding mRNA sequences is shown by underlining. The arrow over nucleotide 7 indicates the only variable nucleotide within the 11-nucleotide common region. (A) The gene 4-5 boundary obtained from sequencing of the *Hae* B fragment; the first AUG is 53 nucleotides downstream from nucleotide 1. (B) The gene 5-6 boundary obtained from sequencing of the *Hae* C fragment is compared with the corresponding sequence from mRNA 6 taken from Armstrong et al. (1) and Niemann et al. (22). (C) The gene 6-7 boundary obtained from sequencing of the *Hpa* B fragment is compared with the corresponding region of mRNA 7 taken from Spaan et al. (28).

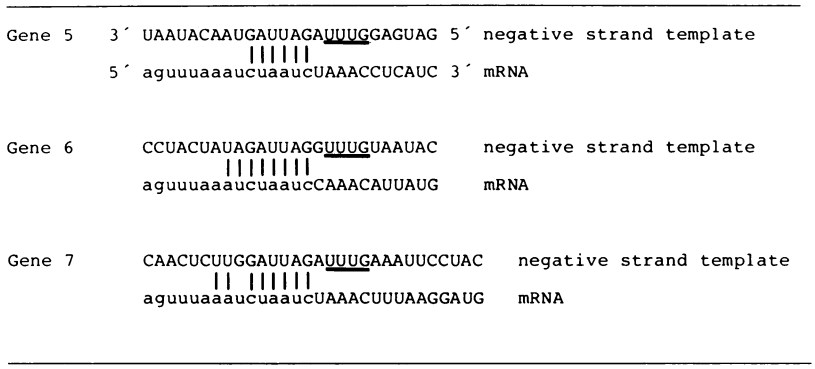

FIG. 6. Hypothetical binding of leader RNA to negative-strand template. The 5′ noncoding regions of the mRNAs are aligned with the negative-strand templates (complement of the intergenic sequences shown in Fig. 5). The sequences for mRNAs 6 and 7 are taken from Armstrong et al. (1–3), while the sequence of mRNA 5 is inferred from the fact that all the mRNAs have the same leader sequence (11) and from the sequencing of g344 DNA. THe lower-case letters represent the 3′ end of the putative leader RNA. Base pairing between the leader and template is indicated by vertical lines. The probable site at which the leader ends and the body of the mRNA begins is suggested by the divergence of the common nucleotide sequence at nucleotide 7. The body of the mRNA is shown in capital letters. The underlined nucleotides UUUG (AAAC, in the positive strand genome) are those that are conserved among the intergenic regions but are probably not derived from the leader RNA.

28). The JHM strain of MHV, which has a genome that is about 75% homologous in nucleotide sequence to MHV A59 (33), generates a very similar leader sequence (27). Two models have been proposed for the genesis of the MHV subgenomic RNAs (4, 28). In the first, a positive-stranded leader RNA is transcribed from the 3′ end of genome-sized negative-strand RNA and subsequently released (perhaps in association with an RNA polymerase); this RNA then is used to prime RNA synthesis at the beginning of each of the viral genes. In the second model, after transcription the leader remains associated with the negative-strand template and there is a looping out of template as the polymerase jumps and continues transcription at the beginning of each gene. The first model is currently favored because Baric et al. (4) have shown that only one size of RF RNA can be derived by ribonuclease treatment of MHV A59 replicative intermediates.

Either model suggests that common sequences might precede each of the seven viral genes that serve as transcriptional control signals, perhaps recognition sites for the binding of a polymerase, an RNA primer, or a primer RNA-polymerase complex. Several other classes of RNA viruses have conserved regions of 10 to 20 nucleotides that are thought to serve as control signals for transcription or translation (30). To explore this possibility for coronaviruses, we have cloned and sequenced three intergenic regions derived from viral genome RNA. The intergenic regions preceding genes 5, 6, and 7 contain a common 11-nucleotide sequence (with one variable nucleotide) (Fig. 5). Furthermore, this sequence has been found in the 5′ noncoding region adjacent to the coding regions of mRNAs 6 and 7 (1–3, 22, 28) (Fig. 5), the probable site of fusion of the leader sequence and the body of the mRNAs. (We have sequenced at least 200 nucleotides in from the initiator AUG codons of genes 6 and 7 as well as almost all of gene 5 [data not shown]. The common 11-nucleotide sequence was not observed within these coding regions or within the published coding regions of mRNA 6 and 7 [1–3].) The fact that nucleotide 7 differs in the 5′ noncoding regions of mRNAs 6 and 7 as well as in the corresponding intergenic regions strongly suggests that this is the site at which the leader sequence ends and the body of the mRNA begins. (It is unlikely that this difference at nucleotide 7 is a cloning artifact, as the common sequence preceding gene 6 contained a C at nucleotide 7 when obtained from at least two independent cDNAs clones synthesized from mRNA [1, 22] and from the genomic cDNA clone g344, while the common sequence preceding gene 7 contained a U at nucleotide 7 when obtained from a cDNA synthesized from mRNA [2, 3], from primer-directed sequencing on mRNA 7 directly [28], and from the genomic cDNA clone g344.) The homology between the 5′ portion of the common intergenic sequence and the 3′ end of the leader sequence suggests that the intergenic region may contain a primer binding site for leader RNA (Fig. 6).

Figure 6 shows a model for priming of transcription of the viral mRNAs by using the data discussed above. The putative leader RNA is aligned with the corresponding regions on the negative-stranded templates (the complement of the intergenic regions) for each mRNA, and the complementary nucleotides are indicated. Two of the three nucleotides just 5′ to the common regions preceding genes 6 and 7 are complementary to the mRNA leader sequence so that there are eight complementary base pairs in these intergenic regions. The total number of complementary base pairs in the intergenic region preceding gene 5 is six. In this model, the leader RNA would hybridize to each of the intergenic regions and then be used as a primer for mRNA transcription. Since only 6 to 8 base pairs would be involved in base pairing, it seems unlikely that this alone accounts for the ability of the primer RNA to find and base pair with the intergenic region. A primer-polymerase complex is more likely to be involved in binding. Nucleotides 8 through 11, which are probably not part of the leader, are conserved in the intergenic regions preceding genes 5, 6, and 7 and thus may be involved in transcriptional control. Perhaps part or all of the 11-nucleotide conserved region serves as a recognition site for binding of a polymerase-RNA complex which brings the primer into proximity with the intergenic regions so that base pairing and subsequent priming occurs.

The mRNAs of MHV are present in infected cells at widely varying concentrations, and these amounts remain constant at various times during infection (18). It is likely that these varying molar amounts are due to synthesis and not to degradation of the mRNAs, as the same ratios are obtained after 1 h and steady-state pulses of label (18). This

variation in transcription rates among the viral mRNAs could be at least partially explained by the differences in the homology between leader RNA and its seven putative intergenic binding sites which might affect the rate of transcription. The molar ratios of mRNAs 7:6:5 in the infected cell are 100:32:11. The fact that the intergenic region preceding gene 5 has less homology with the leader sequence than with the sequences preceding genes 6 and 7 (Fig. 5 and 6) may be one of the factors, resulting in the less frequent transcription of mRNA 5. Other factors, such as the sequences found between the common 11 nucleotides and the beginning of the coding region, presumably play a role in the control of transcription. For example, the distance between the conserved sequence and the beginning of the coding region is much greater for gene 5 than for genes 6 and 7. More intergenic regions as well as the extreme 5′ end of genome RNA must be sequenced to more completely understand these factors.

For genes 6 and 7, the common intergenic sequence is closely followed by an AUG sequence, and it is clear from the published sequences of the E1 and N genes that these are the initiation codon for the E1 and N proteins (1–3). Furthermore, in the case of the gene 6-7 intergenic boundary, it is clear from the published sequence of the E1 mRNA that the UAA in positions 2 to 4 in the sequence is the termination codon for the E1 protein. The sequence that we obtained for the gene 6-7 boundary differs by one nucleotide from that obtained by primer extension (28). However, our sequence is consistent with the composition of oligonucleotide 17 (14) that is present in all the mRNAs except mRNA 7 and is thought to represent the sequence just upstream from the coding region of mRNA 7. In the case of the gene 4-5 intergenic region, the first AUG is 53 nucleotides downstream for the conserved 11 nucleotides. Since neither of these genes have been completely sequenced, we cannot yet determine the protein start and stop regions within this intergenic sequence.

The two intergenic boundaries preceding the two most 3′ proximal genes of avian infectious bronchitis virus, another coronavirus, have been sequenced. These regions have a 9-base-pair homology which appears twice for the intergenic region preceding gene B and only once for gene A (7). There is no significant homology between the MHV A59 and infectious bronchitis virus intergenic sequences.

The cDNA clones that are described here have many further applications. Because the MHV A59 mRNAs form a nested set, it has been difficult to construct probes for individual genes. This is the first report of gene-specific cloned fragments not proximal to the 3′ end of genome RNA and representing nonstructural genes. The clones described here will be useful for the sequence analysis of more intergenic regions as well as for the sequencing and expression of nonstructural genes. This includes genes encoding the 35,000-molecular-weight (35K), 14K (26), and polymerase proteins.

ACKNOWLEDGMENTS

This research was supported by Public Health Service grant A1 17418 from the National Institutes of Health and grant RG 1421 from the National Multiple Sclerosis Society.

We thank Julian Leibowitz for helpful comments and Laura D. Pastore for typing the manuscript.

LITERATURE CITED

1. **Armstrong, J., H. Niemann, S. Smeekens, P. Rottier, and G. Warren.** 1984. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. Nature (London) **308**:751–752.
2. **Armstrong, J., S. Smeekens, and P. Rottier.** 1983. Sequence of the nucleocapsid gene from murine coronavirus MHV-A59. Nucleic Acids Res. **11**:883–891.
3. **Armstrong, J., S. Smeekens, W. Spaan, P. Rottier, and B. A. M. van der Zeijst.** 1984. Cloning and sequencing the nucleocapsid and E1 genes of coronavirus MHV-A59, p. 155–162. *In* P. J. M. Rottier, B. A. M. van der Zeijst, W. J. M. Spaan, and M. C. Horzinek (ed.), Molecular biology and pathogenesis of coronaviruses. Plenum Publishing Corp., New York.
4. **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.
5. **Biggins, M. D., 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. U.S.A. **80**:3963–3965.
6. **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.
7. **Brown, T. D. K., and M. E. G. Boursnell.** 1984. Avian infectious bronchitis virus genomic RNA contains sequence homologies at the intergenic boundaries. Virus Res. **1**:15–24.
8. **Cheley, S., R. Anderson, M. J. Cupples, E. C. M. Lee Chan, and V. L. Morris.** 1981. Intracellular murine hepatitis virus-specific RNAs contain common sequences. Virology **112**:596–604.
9. **Gough, N. M., E. A. Webb, S. Corey, and J. M. Adams.** 1980. Molecular cloning of seven immunoglobulin K chain messenger ribonucleic acids. Biochemistry **19**:2702–2710.
10. **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.
11. **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. U.S.A. **81**:3626–3630.
12. **Lai, M. M. C., P. R. Brayton, R. 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.
13. **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.
14. **Lai, M. M. C., C. D. Patton, and S. A. Stohlman.** 1982. Further characterization of mRNA's of mouse hepatitis virus: presence of common 5′-end nucleotides. J. Virol. **41**:557–565.
15. **Lai, M. M. C., and S. A. Stohlman.** 1978. RNA of mouse hepatitis virus. J. Virol. **26**:236–242.
16. **Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker.** 1977. RNA molecular weight determinations of gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry **96**:4743–4751.
17. **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.
18. **Leibowitz, J. L., K. C. Wilhelmson, and C. W. Bond.** 1981. The virus-specific intracellular RNA species of two murine coronaviruses: MHV-A59 and MHV-JHM. Virology **114**:39–51.
19. **Manaker, R. A., C. V. Piczak, A. A. Miller, and M. F. Stanton.** 1961. A hepatitis virus complicating studies with murine leukemia. J. Natl. Cancer Inst. **27**:29–51.
20. **McIntosh, K.** 1974. Coronaviruses: a comparative review. Curr. Top. Microbiol. Immunol. **63**:80–129.
21. **Messing, J., R. Crea, and P. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
22. **Niemann, H., G. Heisterberg-Moutsis, R. Geyer, H.-D. Klenk, and M. Wirth.** 1984. Glycoprotein E1 of MHV-A59; structure of the O-linked carbohydrates and construction of full-length recombinant cDNA clones, p. 201–213. *In* P. J. M. Rottier, B. A. M. van der Zeijst, W. J. M. Spaan, and M. C. Horzinek (ed.), Molecular biology and pathogenesis of coronaviruses. Plenum Publishing Corp., New York.

23. **Ou, J.-H., C. M. Rice, L. Dalgarno, E. G. Strauss, and J. H. Strauss.** 1982. Sequence studies of several alphavirus genome RNAs containing the start of the subgenomic RNA. Proc. Natl. Acad. Sci. U.S.A. **79**:5235–5239.
24. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
25. **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.
26. **Siddell, S., H. Wege, and V. ter Meulen.** 1982. The structure and function of Coronaviruses. Curr. Top. Microbiol. Immunol. **99**:131–163.
27. **Skinner, M., and S. Siddell.** 1984. Nucleotide sequencing of mouse hepatitis virus strain JHM messenger RNA, p. 163–171. *In* P. J. M. Rottier, B. A. M. van der Zeijst, W. J. M. Spaan, and M. C. Horzinek (ed.), Molecular biology and pathogenesis of coronaviruses. Plenum Publishing Corp., New York.
28. **Spaan, W. J. M., H. Delius, M. A. Skinner, J. Armstrong, P. J. W. Rottier, S. Smeekens, S. G. Siddell, and B. A. M. van der Zeijst.** 1984. Transcription strategy of coronaviruses: fusion of non-contiguous sequences during mRNA synthesis, p. 173–186. *In* P. J. M. Rottier, B. A. M. van der Zeijst, W. J. M. Spaan, and M. C. Horzinek (ed.), Molecular biology and pathogenesis of coronaviruses. Plenum Publishing Corp., New York.
29. **Spaan, W. J. M., P. J. M. Rottier, M. C. Horzinek, and B. A. M. van der Zeijst.** 1982. Sequence relationships between the genome and the intracellular RNA species 1, 3, 6, and 7 of mouse hepatitis virus strain A59. J. Virol. **42**:432–439.
30. **Strauss, E. G., and J. H. Strauss.** 1983. Replication strategies of the single-stranded RNA viruses. Curr. Top. Microbiol. Immunol. **105**:1–98.
31. **Taylor, J. M., R. Illmensee, and S. Sommers.** 1976. Efficient transcription of RNA into DNA by avian sarcoma virus polymerase. Biochim. Biophys. Acta **442**:324–330.
32. **Thomas, P.** 1980. Hybridization of RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A. **77**:5201–5205.
33. **Weiss, S. R., and J. L. Leibowitz.** 1983. Characterization of murine coronavirus RNA by hybridization with virus-specific cDNA probes. J. Gen. Virol. **64**:127–133.