SEQUENCES INVOLVED IN THE REPLICATION OF CORONAVIRUSES

P.J. Bredenbeek, J. Charite, J.F.A. Noten, W. Luytjes,
M.C. Horzinek, B.A.M. van der Zeijst[1], and W.J.M. Spaan

Institute of Virology, P.O. Box 80.150 and
[1]Section Bacteriology, P.O. Box 80.171, Veterinary Faculty
State Univervity Utrecht, The Netherlands

INTRODUCTION

Coronaviruses are enveloped viruses with a positive stranded RNA genome of about 20 kilobases. Upon infection 6 subgenomic mRNAs are produced from a negative stranded template of genome length (Stern and Kennedy, 1980; Spaan et al., 1981; Lai et al., 1982). RNase T1 fingerprinting and hybridization studies have shown that these subgenomic RNAs form a 3' coterminal nested set (Leibowitz et al.,1981; Lai et al., 1981; Spaan et al., 1982). Only the gene located at the 5' end of each subgenomic RNA is translated.

U.V. transcription mapping studies have shown that the subgenomic RNAs are synthesized independently (Jacobs et al., 1981; Stern and Sefton, 1982). However, in apparent contradiction with the idea of an independent initiation of coronaviral RNA synthesis, it was subsequently found that all subgenomic RNAs share a 5' leader sequence of 65 nucleotides encoded by the 3' end of the negative stranded template (Spaan et al., 1983; Lai et al., 1984).

To explain these data we and others (Spaan et al., 1983; Baric et al., 1983) have proposed that the viral mRNAs are synthesized by a discontinuous transcription process. Several models have been suggested to explain the fusion of the non-contiguous transcripts; jumping of the polymerase by looping out of the negative stranded template, co-transcriptional or post-transcriptional ligation of the leader and the bodies of the mRNAs and, priming of the mRNA body transcription by the leader.

Recently Makino et al. (1986) have shown that fusion of the leader and the mRNA body is a trans-acting process. This excludes jumping of the polymerase by looping out of the negative stranded template. The presence of leader sequences on the replicative intermediates makes post-transcriptional ligation very unlikely (Baric et al., 1983).

The remaining two models to explain the unusual transcription process of coronaviruses are shown in fig. 1. In the first model (A) independent initiation of transcription of leader and body sequences is proposed. The leader transcript is ligated to the body RNA while the latter is still being trancribed. In the second model (B) the leader serves as a primer for

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

the initation of the transcription at the conserved intergenic regions on the minus strand RNA. Reinitiation of transcription is completely leader dependent. At this moment we cannot discriminate between the two mechanisms.

In order to identify the sequences involved in the joining of the MHV-A59 leader and mRNA body sequences we have determined the sequence of the so called fusion regions located upstream of each gene. To investigate the possibility for the MHV leader to act as a primer during viral transcription in more detail we have in addition determined the nucleotide sequence of the extreme 5' end of the genome. Our data show that the fusion region is also located at the 5' end of the genome of MHV. Therefore a putative free leader would have a structure that is complementary to the internal reinitiation sites on the negative stranded template.

METHODS

cDNA Synthesis and Cloning

cDNA was prepared from genomic RNA and poly (A) selected RNA, using pentanucleotides or oligo-dT as a primer. First and second strand cDNA synthesis was carried out essentially as described by Gubler and Hoffman (1983). Double stranded cDNA was dC-tailed, annealed into Pst 1 cleaved dG tailed pUC9 and subsequently transformed (Hanahan,1983) into *E. coli* strain JM109.

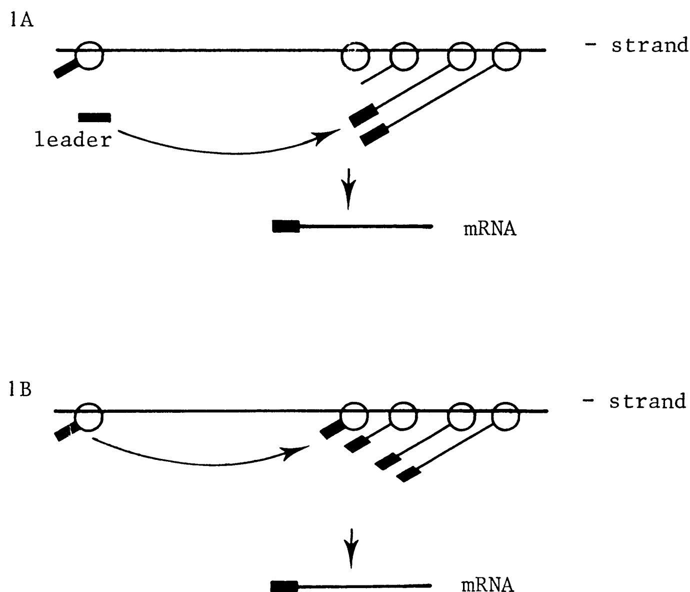

Fig. 1. Two possible models for the discontinuous transcription of coronaviruses.

Screening and Analysis of the Recombinants

In situ hybridization of bacterial colonies with kinase-labeled oligonucleotides was performed according to Maniatis et al. (1982) and Meinkoth and Wahl (1984). Before hybridization the filters were treated with 50 μg/ml proteinase K to reduce background hybridization. Plasmid DNA was prepared from positive colonies according to Birnboim and Doly (1979). Restriction enzyme digestions were performed using standard methods.

DNA Sequence Analysis

Virus specific DNA fragments were separated by agarose gel electrophoresis and purified by binding to NA 45 paper (Schleicher and Schuell) and subsequently recloned into bacteriophage M13 mp9. Sequencing was carried out using the chain termination procedure of Sanger et al. (1980). The data were analyzed on a DEC 20/60 computer using the programs of Staden (1986).

RESULTS

cDNA cloning and mapping of recombinant plasmids

Initially, molecular clones of cDNA prepared from poly (A) selected RNA were obtained using oligo-dT as a primer. The largest clone isolated from this library contained a cDNA insert corresponding to a full-length mRNA 5. To clone the remaining part of the MHV genome a cDNA library from random primed viral genomic RNA was prepared which library contained recombinant plasmids with virus-specific inserts up to 8.5 kilobases (kb). Several of the obtained recombinant plasmids hybridized with a leader-specific probe.

From these two libraries several large clones were selected. They were used to construct a continuous map of approximately 17 kb starting at the 3' end of the genome (Fig 2). Hybridization studies did not reveal any overlap between cDNA clones located at the 5' end part of this alignment and the recombinant plasmids containing leader sequences (data not shown). Considering the length of the largest MHV leader positive insert which is approximately 3 kb and the continuous map of 17 kb it is obvious that the MHV-A59 genome is larger than 20 kb.

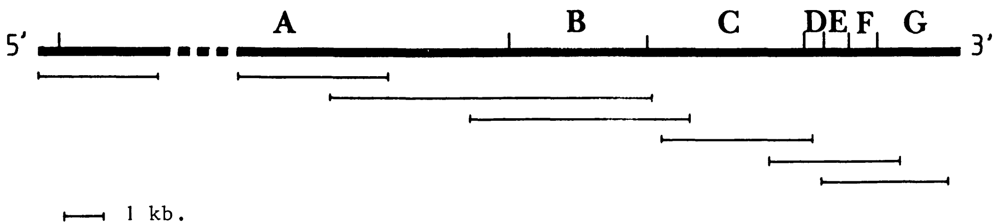

Fig. 2. Map of the MHV-A59 genome showing the localization of the most important clones obtained. Letters A to G indicate the viral genes.

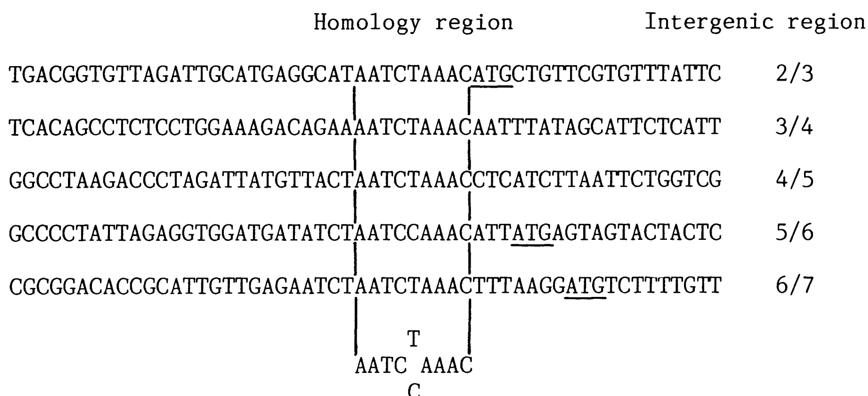

Fig. 3. Comparison of the sequences of intergenic regions on the MHV-A59 genome. The vertical lines enclose the homology region. The start codons of the E2, E1 and nucleocapsid gene are underlined. The sequence of the intergenic region 4/5 has been published by Budzilowicz et al., (1985), the sequences of the intergenic regions 5/6 and 6/7 are from Spaan et al., (1983).

Sequences involved in the replication of MHV-A59

From the random-primed cDNA library several clones were isolated to determine the sequence of the MHV-A59 E2 gene (W. Luytjes et al., manuscript in preparation). Immediately upstream and downstream of the open reading frame encoding the peplomer protein (E2), two hitherto unknown intergenic regions are located. They show a remarkable homology with the three intergenic regions already known (Fig. 3) (Spaan et al., 1983; Budzilowicz et al., 1985).

Restriction enzyme digestions of the leader-positive clones obtained from the genomic cDNA library revealed that these clones have an identical map. The inserts of two clones were recloned in bacteriophage M13 and sequenced (Fig. 4). The 5' end of this sequence is identical to the leader sequence of mRNAs 5,6, and 7. The presented sequence is incomplete at the 5' end. The first 21 nucleotides (as determined by Lai et., 1984) are not present in the clones. The underlined sequence AATCTAAAC is very similar to the nucleotide sequence at the leader/body junctions of these mRNAs. Starting at position 99 there is a very small open reading frame with the potential to encode a protein of only 8 amino acids.

5'
22 TACGTACCCTCTCAACTCTAAAACTCTTGTAGTTTAAATCTAATCTAAACTTTATAAACG 81

GCACTTCCTGCGTGTCCATGCCCGTGGGCCTGCTCTTGTGATAGTGTCTGACATTTGTAG 141

Fig. 4. Sequence of the 5' end of the MHV-A59 genome. The numbering of the nucleotides is taken from Lai et al. (1984).

DISCUSSION

In this paper the isolation and sequence analysis of cDNA clones containing important sequences for the replication of MHV-A59 is described. Although we have cloned approximately 20 kb of the MHV-A59 genome there is still a gap in gene A.

Except for the intergenic region between gene A and B the nucleotide sequences of all intergenic regions have now been determined. From these data we can conclude that the sequence 5' AATC$^{T}/_{C}$AAAC 3' is repeated at the intergenic regions on the viral genome. These regions of homology are thought to play an important role in the (re)initation of the MHV RNA polymerase (Spaan et al., 1983; Budzilowicz et al., 1985) and can be considered as a promoter or recognition signal for the viral polymerase.

At this moment the termination site on the negative RNA strand for the transcription of the leader is unknown. Previously we have postulated a donor and acceptor site sequence homology between the 3' end of the leader and the intergenic regions (Spaan et al., 1983). The sequence of the 5' end of the genome confirms that such a model is possible. Comparison between the sequences of the 5' non-coding regions of mRNAs 5,6 and 7 shows that for each mRNA potentially a different number of nucleotides can be derived from the 5' end of the genome (Fig. 5). S1 mapping studies indicate that the length of the leader transcript varies from 72 to 75 nucleotides (Baric et al., 1985). In addition, leader transcripts from 130 - 250 nucleotides have been detected in MHV-A59 infected cells (Baric et al., 1985). These multiple leader-related transcripts suggest that different leader RNAs are fused to different mRNAs.

It has been suggested previously that base pairing between the 3' end of the leader and the intergenic regions on the negative stranded template plays an important role during the reinitiation of the polymerase in the leader-primed transcription model of coronaviruses (Spaan et al., 1983; Budzilowicz et al., 1985; Bredenbeek et al., 1986.). We have investigated the possibilities for base pairing, assuming leader transcripts longer than 72 nucleotides (Fig. 6). In this model fusion between the leader transcript and the body sequence should take place upstream of the nucleotide indicated by an arrowhead, which has to be transcribed from the intergenic region. The differences in the degree of base pairing could account for the different molar ratios of the viral mRNAs. If leader transcripts of at

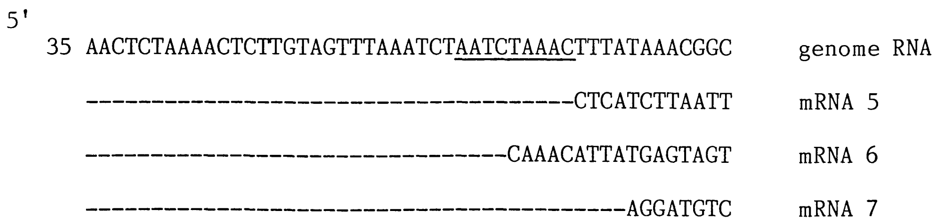

Fig. 5. Comparison of the leader sequences starting at nucleotide 35 of MHV-A59 mRNA 5, 6, 7 and genome RNA. The part of the leader sequence of the different mRNAs that can be derived from the leader is indicated with dashes.

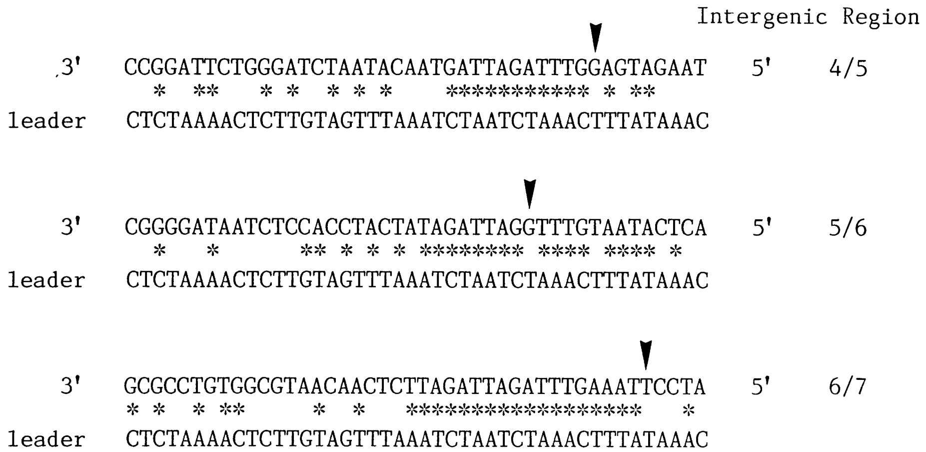

Fig. 6.Base pairing model for the synthesis of MHV-A59 mRNAs 5, 6 and 7 if the leader transcript is longer than 72 nucleotides. The upper strand represents the intergenic regions on the viral minus strand RNA. The lower strand reprsents the leader transcript. Base pairing is indicated with asteriks.

least 75 nucleotides are involved in the reinitiation of the polymerase, trimming of the leader to the correct length by a nuclease must be postulated . This model resembles that of the mRNA transcription of infuenza virus. However in this system is host-cell encoded in contrast to the viral encoded primer of coronaviruses (Plotch et al.,1981).

REFERENCES

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

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

Birnboim, H. C. and Doly, J., 1979, A rapid alkaline extraction procedure for screening recombinant plasmid DNA, Nucl. Acids Res., 7:1513.

Bredenbeek, P. J., Noten, J. F. A., Lenstra, J. A., Horzinek, M. C., van der Zeijst, B. A. M. and Spaan, W. J. M., 1986, The nucleotide sequence of the extreme 5' end of the avian coronavirus genome; implications for the discontinuous mRNA synthesis, Nucl. Acids Res., 14:7806.

Budzilowicz, C. J., Wilczynski C. 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., 53:840.

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

Hanahan, D., 1983, Studies on transformation of escherichia coli with plasmids, J. Mol. Biol., 166:557.

Jacobs, L., Spaan, W. J. M., Horzinek, M. C. and van der Zeijst, B. A. M. , 1981, Synthesis of subgenomic mRNAs of mouse hepatitis virus is initiated independently:evidence from U.V.transcription mapping, J. Virol., 39:401.

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

Lai, M. M. C., Patton, C. D. and Stohlman, S. A., 1983, Replication of mouse hepatitis virus: Negative-stranded RNA and replicative form RNA are of genome length, J. Virol., 44:487.

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.

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

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

Maniatis, T., Fritsch, E. F. and Sambrook, J., 1982, Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory.

Meinkoth, J. and Wahl, G., 1984, Hybridization of nucleic acids immobilized on solid supports, Anal. Biochem., 70:75.

Plotch, J. S., Bouloy, M., Ulmanen, I. and Krug, R. M., 1981, A unique cap (m^7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription, Cell, 23:847

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A., 1980, Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing, J. Mol. Biol., 143:161.

Spaan, W. J. M., Rottier, P. J. M., Horzinek, M. C. and van der Zeijst, B. A. M., 1981, Isolation and identification of virus-specific mRNAs in cells infected with mouse hepatitis virus (MHV-A59), Virology, 108:424.

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 the intracellular RNA species 1, 3, 6, and 7 of mouse hepatitis virus strain A59, J. Virol., 42:432.

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

Staden, R., 1986, The current status and portability of our sequence handling software, Nucl. Acids Res., 14:217.

Stern, D. F. and Kennedy, S. I. T., 1980, Coronavirus multiplication strategy. II. Mapping the avian infectious bronchitis virus intracellular RNA species to the genome, J. Virol., 36:440.

Stern, D. F. and Sefton, B. M., 1982, Synthesis of coronavirus mRNAs: Kinetics of inactivation of infectious bronchitis virus RNA synthesis by U.V. light,J. Virol., 42:755.