Coronavirus mRNA synthesis involves fusion of non-contiguous sequences

Willy Spaan*, Hajo Delius1, Michael Skinner2, John Armstrong3, Peter Rottier, Sjef Smeekens, Bernard A.M. van der Zeijst and Stuart G. Siddell2

Institute of Virology, State University, NL-3800 TD Utrecht, The Netherlands, 1European Molecular Biology Laboratory, Postfach 10.2209, 6900 Heidelberg, and 2Institute of Virology and Immunobiology, 8700 Würzburg, FRG

Communicated by W. Doerfler
Received on 12 July 1983

Positive-stranded genomic RNA of coronavirus MHV and its six subgenomic mRNAs are synthesized in the cytoplasm of the host cell. The mRNAs are composed of leader and body sequences which are non-contiguous on the genome and are fused together in the cytoplasm by a mechanism which appears to involve an unusual and specific 'polymerase jumping' event.

Key words: coronavirus MHV/RNA synthesis/polymerase jumping

Introduction

RNA viruses display a variety of replication strategies but share many biosynthetic pathways with the host cell. They are therefore excellent probes for the investigation of cell processes and their study has played an important role in the discovery of several aspects of gene expression in animal cells. To date, these studies have concentrated on a relatively small number of viruses which were particularly suited to biochemical analyses, but there are also a number of RNA virus families for which the replication strategy is almost, or completely, unknown. In spite of their pathogenetic properties, and the resulting economic losses, coronaviruses were, until recently, such a group. However, in the last few years many basic aspects of coronavirus replication have been elucidated (for reviews, see Siddell et al., 1983; Sturman and Holmes, 1983).

Coronaviruses are enveloped positive-stranded RNA viruses. Their genome is linear, unsegmented and 15 000 - 20 000 bases in length. The most studied member of the group is murine hepatitis virus (MHV). MHV replicates in the cell cytoplasm and viral genetic information is expressed in infected cells as one genomic sized and multiple subgenomic mRNAs. These mRNAs are synthesized in non-equimolar amounts, but in relatively constant proportions throughout infection (Leibowitz et al., 1981). The template for viral mRNA synthesis is a genomic length negative strand (Lai et al., 1982a). Analysis of virion and mRNA of MHV-A59 by two independent methods reveals that the viral mRNAs have a 'nested set' structure with 3'-co-terminal ends and sequences extending for different lengths in a 5'-direction (Leibowitz et al., 1981; Lai et al., 1981, 1982b; Spaan et al., 1982; Cheley et al., 1981). Each subgenomic mRNA is capped and polyadenylated, as is genome RNA (Lai et al., 1982b), and is translated independently to produce a single protein the size of which corresponds to the coding capacity of the 5'-sequences not found in the next smallest mRNA (Siddell, 1983; Rottier et al., 1981; Leibowitz et al., 1982). There is at present no indication that the replication of MHV involves a nuclear phase or nuclear factors. MHV is reported to grow in enucleated cells and its replication is not inhibited by actinomycin D or alpha-amanitin (Brayton et al., 1981; Wilhelmsen et al., 1981; Maby et al., 1983). Also, the synthesis of each mRNA is inactivated by u.v. irradiation in proportion to its own length (Jacobs et al., 1981). Thus the subgenomic mRNAs are not produced by the processing of larger RNAs.

Analysis of oligonucleotides generated by RNase T1 digestion of viral RNA reveals unique oligonucleotides that do not fit into the 'nested set' structure. T1 oligonucleotides 10 and 19 from mRNA7 of MHV-A59 (using the nomenclature of Lai et al., 1982b) are not present in the corresponding 3'-end of the genome (Spaan et al., 1982; Lai et al., 1983), suggesting that these oligonucleotides are derived from a leader sequence which all mRNAs might share. Oligonucleotide 10 is also found in the larger mRNAs and genome RNA (Lai et al., 1983) but mRNA-specific differences in electrophoretic mobility have been detected for oligonucleotide 19. It is present as oligonucleotide 19, 19a and 3a in mRNA7, mRNA6 and mRNA5, respectively. Oligonucleotides 19 and 19a have very similar base compositions. Oligonucleotide 17, which is found in mRNA6 and larger mRNAs, but not in mRNA7 (Lai et al. 1982b, 1983; Spaan et al., 1982), also has a base composition similar to oligonucleotides 19 and 19a (Armstrong et al., 1983). Finally, the mRNAs share at least five nucleotides at their 5'-end (Lai et al., 1982b).

These data can be interpreted in the model shown in Figure 1. In this model, sequences present at the 5'-end of genome RNA, or mRNA1, are also found at the 5'-end of each subgenomic mRNA (these sequences will be referred to as leaders). MHV-A59 oligonucleotide 10 would be encompassed within these sequences. Part of the MHV-A59 oligonucleotides 19 and 19a would also be contained within the leader, but differences would arise from fusion of the leader sequence with the various bodies of the mRNAs. Oligonucleotide 17 would represent sequences at the 5'-end of the mRNA7 body, part of which would be lost during the construction of mRNA7 but not for example mRNA6.

In the experiments described here we have investigated the structural organization of MHV mRNA by electron microscopic analysis of hybrids formed between single-stranded cDNA copies from mRNA7 with genomic RNA or mRNA6. In addition, we have determined the nucleotide sequence at the 5'-end of MHV-A59 mRNA7 and the corresponding region of the genome. The 5'-sequences of MHV-JHM mRNA7 have also been determined. The MHV-A59 and the MHV-JHM strains have unrelated RNase T1 fingerprints (Lai and Stohlmans, 1981), but the recognition sites for the fusion of the leader sequence and the mRNA body might be expected to be conserved between strains as well as in different mRNAs of the same strain.

*To whom reprint requests should be sent.
**Results**

*Electron microscopy of hybrids between single-stranded cDNA copied from mRNA 7 with genome RNA or mRNA6*

Single-stranded cDNA was prepared as described (Armstrong et al., 1983) and after hydrolysis of the RNA template it was treated with glyoxal and dimethylsulfoxide and analyzed by agarose gel electrophoresis. A main band approximately the size of mRNA7 was identified (data not shown). This cDNA was annealed to genome RNA and prepared for electron microscopy by cytochrome spreading. A sequence homology between the 5' end of the RNA copied into cDNA, and the 5' end of the genome RNA should lead to the formation of a looped hybrid and indeed such structures were observed. Figure 2A shows such a hybrid molecule accompanied by a tracing outlining the possible arrangement of the RNA and DNA strands. Only linear molecules were observed when preparations of RNA alone were analysed in the same way and circularization, therefore, has to be attributed to the hybrid formation. The length of the hybrid region in these molecules corresponds to 1995 ± 160 bp, using PM2 DNA as a standard and after correction for the shortened hybrid length (Priess et al., 1980). The size of the loop was determined as 19.4 ± 1.0 kb. This value is an approximation due to the lack of a suitable RNA standard of this size. No double-strand could be discerned at the point of the re-entry of the genome RNA into the hybrid near the 5' end. This excludes a double-stranded region much larger than 50 nucleotides. Figure 2B shows a hybrid between the same cDNA and mRNA6. Again, a double-stranded region (1890 ± 140 bp) caused by hybridization between the cDNA and the 3' end of mRNA6, and a single-stranded loop structure (in this case of 600 ± 80 nucleotides) were observed. This single-stranded loop probably represents the E1 gene. Its size would be sufficient for it to encode a polypeptide of 22 000 ± 2900 mol. wt., the approximate size of the non-glycosylated form of polypeptide E1 found in MHV-A59-infected cells (Rottier et al., 1981). The most likely explanation of the loop formation in these hybrids is once again the presence of common leader sequences in mRNA6 and mRNA7, although again no hybrid stretch on the 5' side of the loop could be detected. The data strongly support the model shown in Figure 1.

Sequence analysis of mRNA7 and the region of the genome between the E1 and N genes

A further prediction of the model shown in Figure 1 is that the nucleotide sequence of the region immediately upstream from the nucleocapsid gene of mRNA7 should diverge from the region upstream from the nucleocapsid gene in the genome. Therefore, we determined the nucleotide sequence in these regions by two independent procedures. Firstly, two recombinant DNA clones were analysed. One clone, MS38, was obtained from mRNA7 of MHV-JHM. An outline of the procedures to obtain and sequence this clone is given in Materials and methods and full details will be given elsewhere (Sklärner and Siddell, 1983). This clone corresponds to the 5' half of mRNA7. Two large T oligonucleotides are predicted from this sequence. One oligonucleotide (position −53/−29) would have a base composition very similar to that reported for MHV-A59 oligonucleotide 10, the other (position −25/−17) was similar to oligonucleotide 19 (Figure 2, the differences will be discussed below). At position 1 an AUG codon is found that generally conforms to a consensus for initiation sequences (Kozak, 1981), particularly in the occurrence of an A residue at position −3. It is followed by an extensive open reading frame (Sklärner and Siddell, 1983).

The second recombinant DNA cloned and sequenced was prepared from a mixture of MHV-A59 mRNAs. Clone S9 covered sequences upstream from the initiation codon preceding a sequence of 1664 nucleotides at the 3' end of the genome, including the nucleocapsid gene (Armstrong et al., 1983) (Figure 4). Clone S9 started to diverge (reading 3' to 5') from clone MS38 at position −22, exactly at the same point as oligonucleotides 17 and 19 diverge from each other, when ordered for maximal homology with the DNA sequences (Figure 3). These data therefore support the model of Figure 1, and imply that clone S9, which lacked sequences corresponding to oligonucleotide 10, represents sequences not found in mRNA7, but present in other mRNA species and genome RNA (see below).

The second approach we have used was to sequence directly the 5' end of mRNA7 and the corresponding region of the genome of MHV-A59 using DNA primers and reverse transcriptase for dideoxy sequencing on RNA (McReynolds et al., 1978; Zimmerm and Kaesberg, 1978; Zain and Roberts, 1979). The primers used were a synthetic primer complementary to positions 11−24 (New England Biolabs) and a restriction fragment isolated from the RF DNA of clone S38 (Figure 4) after digestion with restriction endonuclease HpaII. A single-stranded primer was obtained by treating the fragment with exonuclease III prior to hybridisation (Zain and Roberts,
Coronavirus MHV mRNA synthesis

Fig. 2. Electron micrographs of hybrids between coronavirus MHV-A59 RNAs and cDNA copied from mRNA7. (A) Hybrid between cDNA and the genome RNA. (B) Hybrid between cDNA and mRNA6. The tracings show an interpretation of the arrangement of the cDNA (D) and RNA (R). The structure of the joint between the 5' end of the RNA loops and the cDNA cannot be identified in the electron micrographs. Viral RNAs were isolated as described (Spaan et al., 1982).

1979). The resulting sequence for genomic RNA and mRNA7 are shown in Figure 3. The following conclusions can be drawn. Firstly, the sequences of mRNA7 of A59 and JHM are almost identical, the only major exception being that JHM appears to contain an extra copy of the sequence AUCUA or AAUCU. The extreme end of MHV-A59 mRNA7 could not be read, but a sequence (starting at position -53) similar to oligonucleotide 10 was found. Also oligonucleotide 19 was detected. Secondly, the sequence obtained from the MHV-A59 genome was identical to that of mRNA7 up to position -22, the same position at which recombinant DNA clones MS38 and S9 diverged, as well as oligonucleotides 17 and 19. Finally, it is clear the MHV-A59 clone S9 represents the sequence present in the genome. The
Fig. 3. Comparison of nucleotide sequences of the 5' end of MHV mRNA7 and the homologous part of genome RNA and the alignment with T1 oligonucleotides. The sequences were obtained as described in Materials and methods. The alignment of T1 oligonucleotides with the other sequences was made by maximizing homology, using the base compositions determined by Lai et al. (1982b). Homologies between adjacent sequences are indicated by asterisks.

| Clone A59 | -70 | -60 | -50 | -40 | -30 | -20 | -10 | 1 | 10 |
|-----------|-----|-----|-----|-----|-----|-----|-----|---|---|
| Genome A59 | CGGACCGCAUGUUGAGA | AUCUAUCUAACUUGUGGUAAGUGCUUUUGGUGGCUUGGG | | | | | | |
| T1 oligo- | CGGACCGCAUGUUGAGA | AUCUAUCUAACUUGUGGUAAGUGCUUUUGGUGGCUUGGG | | | | | | |
| Nucleotides A59 | | | | | | | | |
| RNA7 A59 | GUCGUACGUACCCUCUACUCUAAAACUCUUGUAGUUA | ACCUAUCAUCAUAAUUGAUGCUUUUGGUGGCUUGGG | | | | | | |
| RNA7 JHM | GUCGUACGUACCCUCUACUCUAAAACUCUUGUAGUUA | ACCUAUCAUCAUAAUUGAUGCUUUUGGUGGCUUGGG | | | | | | |

Fig. 4. Localization of DNA clones of MHV used to determine the sequences upstream from the gene encoding the nucleocapsid protein (N). The physical map of the sequence of the nucleocapsid gene is shown. Protein (N) is translated from position 0 to 1362. Sequence divergence between the 5' end of mRNA7 and the corresponding region of the genome (Figure 3) is indicated in MS38.

Origin of clone F10 which previously was suspected to represent the genome of MHV-A59 (Armstrong et al., 1983) remains to be determined.

Localization of the fusion site of leader and body sequences of MHV mRNAs

From the above data, summarized in Figure 5, it is clear that the 5' terminus of the MHV-A59 mRNA body cannot extend beyond position -21, the site of divergence of mRNA7 and the genome. As mentioned before, T1 oligonucleotide 19a (specific for RNA6) and T1 oligonucleotide 19 (specific for RNA7) have a very similar base composition. By predicting the sequence of 19a from the base composition (Lai et al., 1982b) and by comparison with the MHV-A59 sequence between -24 and -2 (Figure 5) it can be seen that the 3' terminus of the leader sequence fused to mRNA6 cannot extend beyond the first base difference (reading in a 5' to 3' direction) in oligonucleotide 19a and 19a. Thus the data suggest that for MHV-A59 the fusion of leader and body sequences producing oligonucleotides 19 and 19a occurred within the sequence 5' AAUCUAUACUAACCU 3', a sequence that does not contain the consensus (A/C)AG/G established for splice junctions in viral and cellular mRNAs (Flint, 1981). It was interesting that the only major difference between A59 and JHM virus found in the 5'-terminal region was within this fusion sequence. The repeated palindromic pentamer AUCUA (or the sequence AAUCU) might function as a recognition signal during the fusion process.

Discussion

How are the leader and body sequences of coronaviral mRNAs fused?

Previous studies have demonstrated that MHV subgenomic mRNAs are not produced by the processing of larger RNAs and that conventional splicing mechanisms are not involved in MHV mRNA synthesis (Brayton et al., 1981; Wilhelmson et al., 1981; Mahy et al., 1983; Jacobs et al., 1981). The data presented here could, however, be explained by a polymerase jumping mechanism. Translocations of RNA viral polymerases have been postulated before to explain the generation of mutant defective interfering (DI) particles (Perneur, 1981; Lazzarini et al., 1981; Fields and Winter, 1982). In the case of MHV, however, the mechanisms would be part of the normal pathway of viral replication, and thus would be the first example of specific fusion of long non-contiguous RNA sequences in the cytoplasm. The mechanism we propose for the generation of coronavirus mRNA would involve the synthesis of a short RNA transcript from the 3' end of the negative stranded template. The polymerase/leader complex would then be translocated to an internal position on the negative stranded template where transcription would resume. This translocation may or may not involve dissociation of the polymerase/leader complex from the template. Translocation would have to occur to specific positions and at specific frequencies. It seems likely that this specificity would be, at least in part, related to sequences in the 5' non-coding region of MHV RNAs. We have argued that the fusion of leader and body sequences during the synthesis of mRNA7 and mRNA6 occurs within the sequence 5' AAUCUAUACUAACCU 3'. This sequence or its 5' end might also be present within region X of the leader (Figure 1) creating a donor and acceptor site sequence homology.

This mechanism superficially resembles that proposed for the formation of DI influenza virus RNA. Fields and Winter (1982) suggested that after the termination of transcription in a U-rich region of influenza genomic RNA, sequences at the 3' terminus of the nascent chain can base pair to sites downstream on the template. A similar mechanism might also explain the high frequency of recombination between
Fig. 5. Localization of the fusion site between body and leader of MHV mRNA. For details see text.

Achtenkrafft et al., 1982. Therefore, the synthesis of MHV mRNAs could be considered as a refined form of a mechanism which occurs in the generation of mutant forms of other RNA viruses. The distinguishing features of the mechanism in coronavirus infection are that it is highly efficient and specific, being the normal route for viral replication. Clearly, further sequence analyses of the extreme 5' end of the MHV genome and of more intergenic regions are needed to elucidate the details of this unusual mechanism of RNA synthesis.

Materials and methods

**Viruses and cells**

The growth of coronavirus MHV in Sac- cells has been described previously (Siddell et al., 1980; Spaan et al., 1981).

**Electron microscopy**

The cDNA was incubated with RNA at a concentration of ~1 µg/ml in 50% formamide containing 10 mM Tris-HCl, 1 mM EDTA and 0.2 M CaCl₂ for 30 min at 40°C. A 10-fold diluted aliquot was spread from 30% formamide, 0.1 M Tris-HCl, 1 mM EDTA, 0.1 µg/ml PM2 DNA with CBB-treated cytochrome (Delius et al., 1972) in a hypophase of distilled water. Samples were picked up on Parlodion-coated grids, stained with uranyl acetate and rotary shadowed with platinum.

**cDNA cloning and sequencing**

The procedures used to clone and sequence the 5' region of MHV-JHM mRNA7 were briefly as follows. Polyadenylated RNA from MHV-JHM infected Sac- cells was used as a template for avian myeloblastosis virus reverse transcriptase in a reaction primed with oligo(dG). After alkaline hydrolysis of RNA, single-stranded cDNA was tagged with oligo(dC) using terminal transferase. Second strand synthesis was primed with oligo(dG). Double-stranded cDNA was tagged with oligo(dC) and annealed to the plasmid pAT153 which had been linearized with PstI and tagged with oligo(dG). After transformation with this DNA, Escherichia coli HB101 colonies with an A⁴/Th⁴ phenotype were screened for coronavirus sequence inserts by hybridization to single-stranded [³²P]cDNA copied from virion RNA. One clone isolated by this procedure, MS38, was further characterized by Northern blotting and restriction enzyme mapping. After subcloning into M13, the inserted DNA was sequenced by the chain termination method of Sanger et al. (1977). The procedures used to obtain and sequence MHV A59 clone S9 have been described in detail (Armstrong et al., 1983).

**Sequencing of RNA with DNA primers and reverse transcriptase**

Clone S83 (Armstrong et al., 1983; Figure 4) was used to isolate a DNA primer for sequencing directly on the viral RNA. Replicative-form DNA was isolated from a 21 culture of S83-infected bacteria, by Triton lysis and equilibrium centrifugation in CsCl. 50 µg was digested with 60 units of endonuclease HpaII (Bethesda Research Laboratories) for 1 h at 37°C in a volume of 100 µl. After the digestion, the DNA was treated with bacterial alkaline phosphatase (P-L Biochemicals Inc.), labelled using [³²P]dATP (3000 Ci/ mmol, Radiochemical Centre, Amersham) and T4 polynucleotide kinase (Bethesda Research Labs.) and purified by gel electrophoresis using standard procedures. The fragment was eluted and purified on a DEAE column (Smith, 1980) and dissolved in 50 µl 10 mM Tris-HCl pH 8.0, 1 mM EDTA.
Lazzarini, R.D., Keene, J.D. and Schubert, M. (1981) Cell, 26, 145-154.
Leibowitz, J.I., Wilhelmsen, K.C. and Bond, C.W. (1981) Virology, 114, 39-51.
Leibowitz, J.I., Weiss, S.R., Paavola, E. and Bond, C.W. (1982) J. Virol., 43, 905-913.
Mahy, B.W.J., Siddell, S., Wege, H. and ter Meulen, V. (1983) J. Gen. Virol., 64, 103-111.
McReynolds, L., O'Malley, B.W., Nisbet, A.D., Forthegill, J.E., Givol, D., Fields, S., Robertson, M. and Brownloe, G.G. (1978) Nature, 273, 723-728.
Perrault, J. (1981) Curr. Top. Microbiol. Immunol., 93, 151-207.
Priess, H., Koller, B., Hess, B. and Delius, H. (1980) Mol. Gen. Genet., 178, 27-34.
Rottier, P.J.M., Spaan, W.J.M., Horzinek, M. and van der Zeijst, B.A.M. (1981) J. Virol., 38, 20-26.
Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
Siddell, S.G. (1983) J. Gen. Virol., 64, 113-125.
Siddell, S.G., Wege, H., Barthel, A. and ter Meulen, V. (1980) J. Gen. Virol., 33, 10-17.
Siddell, S.G., Wege, H. and ter Meulen, V. (1983) J. Gen. Virol., 64, 761-776.
Skinner, M.A. and Siddell, S.G. (1983) Nucleic Acids Res., 11, 5045-5054.
Smith, A. (1980) Methods Enzymol., 65, 560-580.
Spaan, W.J.M., Rottier, P.J.M., Horzinek, M.C. and van der Zeijst, B.A.M. (1981) Virology, 108, 424-434.
Spaan, W.J.M., Rottier, P.J.M., Horzinek, M.C. and van der Zeijst, B.A.M. (1982) J. Virol., 42, 432-439.
Sturman, I.S. and Holmes, K.V. (1983) Adv. Virus Res., 28, in press.
Wilhelmsen, K.C., Leibowitz, J.I., Bond, C.W. and Robb, J.A. (1981) Virology, 110, 225-230.
Zain, B.S. and Roberts, R. (1979) J. Mol. Biol., 131, 341-352.
Zimmern, D. and Kaesberg, P. (1978) Proc. Natl. Acad. Sci. USA, 75, 4257-4261.