Discontinuous Transcription Generates Heterogeneity at the Leader Fusion Sites of Coronavirus mRNAs

SHINJI MAKINO,* LISA H. SOE, CHIEN-KOU SHIEH, AND MICHAEL M. C. LAI

Department of Microbiology, School of Medicine, University of Southern California, Los Angeles, California 90033

Received 23 March 1988/Accepted 4 July 1988

Coronavirus mRNA is synthesized by a discontinuous transcription process, which involves a free leader RNA species. As a result, each virus-specific mRNA contains an identical leader RNA derived from the 5' end of the genomic RNA. In this study, we demonstrate by primer extension studies that the leader-fusion sites on a given species of coronavirus subgenomic mRNA are heterogeneous. The heterogeneity was due to variation in the number of pentanucleotide (UCUAA) repeats present at the leader fusion site. This pentanucleotide repeat region was complementary between the free leader RNA and the transcription start sites on the template RNA. This result suggests that the discontinuous transcription of coronavirus mRNAs occurs within the complementary sequences localized in two different RNA segments and that RNA joining occurs at variable sites.

Mouse hepatitis virus (MHV), a member of the *Coronaviridae*, contains a single-stranded RNA genome of positive sense (11, 25). It replicates exclusively in the cytoplasm (5, 26) and synthesizes six subgenomic mRNAs, each of which is composed of a 72-nucleotide leader RNA derived from the 5' end of genomic RNA that is fused to sequences of various lengths from the 3' end of the genome (9, 10, 23). Considerable evidence suggests that these mRNAs are generated by a mechanism of leader-primed transcription in which a leader RNA is synthesized from the 3' end of the negative-strand template RNA, dissociates from the template, and then rejoins the template at the distant downstream sites where mRNAs initiate (3). In support of this transcription model, several free leader RNA species have been detected in the cytoplasm of MHV-infected cells (4). In addition, the leader sequence on the MHV mRNAs could be freely exchanged between two viruses during mixed infections, suggesting that free leader RNAs can be incorporated in *trans* into coronavirus subgenomic mRNAs (12). These leader RNAs are not added to mRNAs posttranscriptionally, since they have been detected on the nascent RNA chains in replicative intermediates (3). Furthermore, consensus splice donor or acceptor sequences are not present in coronavirus RNA (9, 19). Thus, this transcription mechanism is distinct from *trans* splicing observed in several different eucaryotic systems (7, 8, 15, 24).

Sequence analysis of genomic and subgenomic mRNAs showed that there is considerable sequence homology between the 3' end of the leader RNA and the transcription start sites for every mRNA (6, 19). Presumably, these homologous sequences are involved in leader RNA binding and subsequent initiation of transcription, as suggested by the finding that there is a rough correlation between the degree of nucleotide homology at the initiation start sites and the amount of mRNAs transcribed (19). Within this homologous sequence, there are three UCUAA repeats at the 3' end of leader RNA (nucleotides 60 to 74), while there are one or two imperfect UCUAA sequences at the different intergenic sites (1, 2, 6, 20-22). The presence of multiple pentanucleotide repeats at the leader-mRNA junction sites sug-

gests a possibility that the binding of leader RNA to template RNA during the synthesis of subgenomic mRNAs may be imprecise, resulting in mRNAs with heterogeneous numbers of UCUAA repeats. Indeed, a cDNA clone of mRNA 6 of the A59 strain of MHV has been found to contain an additional copy of the UCUAA repeat (1). However, it was not known whether this clone was the result of cloning artifacts and how frequently this RNA heterogeneity occurred.

In the studies reported here, we examined whether MHV mRNAs were indeed heterogeneous in terms of the UCUAA repeats. We used primer extension studies to examine the sequences of leader-mRNA junction sites. We found that coronavirus mRNAs demonstrate sequence heterogeneity at these junction points. For a given mRNA, distinct subspecies were detected which differed in the number of the pentanucleotide repeat, whereas the leader RNA of the genomic RNA and intergenic regions were homogeneous. This result has interesting implications for the mechanism of coronavirus mRNA transcription.

To determine whether the leader fusion sites of MHV mRNAs are heterogeneous, we performed primer extension studies using as primers two synthetic oligonucleotides, oligo 6 (5'-GGCCTGAGTGGTACTACT-3') and oligo 7 (5'-TTGCCAGGAACAAAAGA-3'), which are complementary to nucleotides 83 to 100 from the 5' end of mRNA 6 (17) and 87 to 104 from the 5' end of mRNA 7 (21), respectively. The 5'-end-labeled oligo 7 was hybridized to the purified mRNA 7, whereas oligo 6 was hybridized to mRNA 6. The hybridized primers were extended by reverse transcriptase. Primer extension products were then analyzed by electrophoresis on 6% polyacrylamide gels containing 8 M urea (14). Three different MHV strains were studied to assess the generality of the findings. Among these viruses, MHV-A59 and MHV-JHM are prototypes of MHVs, whereas MHV-JHM-2c is an MHV-JHM-derived small-plaque mutant isolated from a persistently infected cell culture (13). The results are shown in Fig. 1A and B. It is evident that several mRNAs were heterogeneous at their 5' end regions. The mRNA 6 of both MHV-A59 and MHV-JHM-2c yielded two distinct primer-extended products, which differed in size by 5 nucleotides. The mRNA 7 of both MHV-JHM and MHV-

* Corresponding author.

FIG. 1. Primer extension analysis of mRNA 6 (A) and mRNA 7 (B) from MHV-A59 (lanes A59), MHV-JHM (lanes JHM), and MHV-JHM-2c (lanes 2c). Marker DNAs are also shown; numbers indicate nucleotides. The synthetic oligodeoxyribonucleotides oligo 6 and oligo 7 were 5' end labeled with [γ - 32 P]ATP by polynucleotide kinase (16). Gel-purified mRNA 6 and mRNA 7 were incubated in 8 μ l of distilled water containing 10 mM methyl mercury. After 10 min of incubation at room temperature, mRNA 6 and mRNA 7 were further incubated with 5'-end-labeled oligo 6 and oligo 7, respectively, in 50 μ l of buffer containing 60 U of RNasin (Promega Biotec), 10 mM MgCl₂, 100 mM KCl, 50 mM Tris hydrochloride (pH 8.3), 10 mM dithiothreitol, 1.25 mM each dATP, dCTP, dGTP, and TTP, 28 mM β -mercaptoethanol, and 20 U of avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc.) at 42°C for 1 h. Reaction products were then analyzed by electrophoresis on 6% polyacrylamide gels containing 8.3 M urea.

JHM-2c yielded three distinct products, which also differed in size by 5 nucleotides. Among the three primer-extended products of MHV-JHM-2c, the amounts of the two larger products were almost equal and that of the smallest product was roughly half of the larger cDNA products. Only mRNA 6 of MHV-JHM and mRNA 7 of MHV-A59 yielded a single cDNA product. The relative amounts of these distinct products were reproducible from experiment to experiment; thus, they were not artifacts of reverse transcription. Since primer extension studies detect only major RNA species, it cannot be ruled out that additional RNA subspecies may also exist.

FIG. 2. Sequences of the primer extension products. Primer-extended products (Fig. 1) were eluted from gels and sequenced by the method of Maxam and Gilbert. Only the sequences at the leader-junction regions are shown. The 5-nucleotide repeats are underlined and numbered. The fastest-, medium-, and slowest-migrating primer-extended products (Fig. 1) are represented as L, M, and U, respectively. Asterisks represent the major cDNA products among the cDNA bands in each sample.

Gilbert (14). Figure 2 shows that the heterogeneity at the 5' ends of mRNA 6 and mRNA 7 was due to the difference in the number of UC_{UAA}, UA_{UAA}, or UC_{CA $\bar{A}$} repeats at the leader-body junction sites. In some cases, these pentanucleotides were not perfectly repeated. We conclude that the mRNAs of MHV are heterogeneous at the leader-body fusion sites as a result of variation of a pentanucleotide repeat sequence.

The heterogeneity of the pentanucleotide repeats in these mRNAs could have been due to heterogeneity of the 5' end of the genomic RNA, from which the leader RNA is derived. We therefore performed primer extension studies using a specific primer (5'-AATGTCAGCACTATGACA-3') (oligo 1) complementary to nucleotides 123 to 140 from the 5' end of the genome (19), which represents sequences downstream of the leader RNA. The genomic RNAs from the three plaque-cloned MHV strains were examined. A single cDNA product was obtained from all three viral RNAs, indicating the size homogeneity of the 5' end sequence of each viral genome (Fig. 3). Sequence analysis of the extended primers showed that the 5' ends of genomic RNA of MHV-A59, MHV-JHM, and MHV-JHM-2c contained 2, 3, and 4 UCUAA repeats, respectively (data not shown). In addition, MHV-JHM-2c has a 9-nucleotide deletion downstream of the leader sequence (data not shown). We have also shown that the intergenic regions of all three virus strains are homogeneous in their sequences (data not shown). Thus, the heterogeneity of the pentanucleotide repeats in the subgeno-

FIG. 3. Primer extension analysis of the genomic RNA of MHV. The synthetic oligodeoxyribonucleotide (oligo 1) complementary to nucleotides 123 to 140 from the 5' end was hybridized to the genomic RNAs of MHV-A59 (lane A59), MHV-JHM (lane JHM), and MHV-JHM-2c (lane 2c) and primer extended with reverse transcriptase. The products were electrophoresed on 6% polyacrylamide gels containing 8 M urea. Marker DNAs were run in parallel.

mic mRNAs is likely generated during leader RNA binding to the template RNA and subsequent RNA transcription.

Since the pentanucleotides in the leader sequences at the 5' ends of mRNAs are not perfectly repeated, we compared the sequences of the pentanucleotide repeats present in genomic RNA, intergenic regions, and mRNAs. The sequences of the first two regions were also determined by Maxam-Gilbert sequencing of the primer-extended products (data not shown). It was found that the pentanucleotides in these three regions were not perfectly matched. Notably, most of the mismatching nucleotides in the pentanucleotides of mRNAs were identical to those of the intergenic regions; however, a few were identical to the corresponding region at the 5' end of genomic RNA. An example of sequence comparison of the three regions is presented in Fig. 4. The mechanism for this diversity is not clear.

This study demonstrates a unique kind of heterogeneity in subgenomic mRNAs of MHV. Since the 5' end of the genomic RNA, from which the leader RNA is derived, and the intergenic regions, which serve as the template for the initiation of mRNA synthesis, were homogeneous in their

FIG. 4. Sequence comparison of leader RNA, intergenic region between genes E and F, and mRNA 6 of MHV-JHM-2c. Sequences were obtained by Maxam-Gilbert sequencing of the primer-extended products, using different synthetic primers. The sequences presented are complementary to the DNA sequences. The triangles and arrows identify the base mismatches between the leader RNA and intergenic regions. U*, Sequences of the major, slower-migrating cDNA band; L, minor, faster-migrating cDNA band.

sequences, the heterogeneity of mRNAs was most likely the result of imprecise binding of the leader RNA to the template RNA at the different pentanucleotide repeats. This result further suggests that the coronavirus mRNA synthesis involves the leader RNA binding to the complementary regions on the template RNA. Thus, coronavirus transcription is different from influenza virus "cap-snatching" transcription (18) or trypanosome *trans* splicing (15, 24). It should be noted that a leader RNA with a fixed number of UC₄AA repeats can generate not only mRNA subspecies containing a smaller number of repeats but also mRNA subspecies containing a larger number of repeats (Fig. 2). Also, the mRNA heterogeneity was detected not only with viruses (MHV-JHM and MHV-JHM-2c) which contain different numbers of the pentanucleotide repeats between the leader RNA and intergenic regions but also with viruses (e.g., MHV-A59) which have identical numbers of repeats at both places. This result further suggests that the leader RNA binding is imprecise.

For a given virus strain and mRNA species, the relative ratio of the mRNA subspecies containing different numbers of pentanucleotide repeats is reproducible. The factors controlling the amounts of different RNA subspecies transcribed are unknown at the present time. The variation of different RNA subspecies could have resulted from different strengths of binding between the leader RNA and the template RNA. Alternatively, since several free leader RNA species of different lengths have been detected in MHV-infected cells (4), different free leader RNAs may be utilized in generating mRNAs with different numbers of pentanucleotide repeats. It should be pointed out that the primer extension studies used here reflected accurately the relative ratio of different RNA subspecies. However, very minor RNA subspecies

could have been missed. Thus, it cannot be ruled out that additional RNA subspecies with even more or fewer pentanucleotide repeats may exist. Finally, the finding that the leader sequences of mRNAs, in most cases in which sequence mismatches between the 5' end of genomic RNA and intergenic sequences exist, reflect those of intergenic sites on the RNA template rather than the sequences of the leader RNA at the 5' end of genomic RNA suggests that the MHV RNA polymerases may have a proofreading activity, which is commonly associated with DNA polymerases but not with RNA polymerases. Thus the RNA polymerase of MHV may have very interesting biochemical properties, the characterization of which is currently in progress.

We thank Susan C. Baker for helpful discussions and Carol Flores for preparation of the manuscript.

This work was supported by Public Health Service research grants AI19244 and NS18146 from the National Institutes of Health, National Multiple Sclerosis Society research grant RG1449, and National Science Foundation grant DMB-9332. L.H.S. was a post-doctoral fellow of the Bank of America-Giannini Foundation.

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. 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.
4. Baric, R. S., S. A. Stohlman, M. K. Razavi, and M. M. C. Lai. 1985. Characterization of leader-related small RNAs in coronavirus-infected cells: further evidence for leader-primed mechanism of transcription. *Virus Res.* **3**:19-33.
5. Brayton, P. R., R. G. Ganges, and S. A. Stohlman. 1981. Host cell nuclear function and murine hepatitis virus replication. *J. Gen. Virol.* **50**:457-460.
6. Budzylowicz, C. J., S. P. Wilczynski, and S. R. Weiss. 1985. Three intergenic regions of coronavirus mouse hepatitis virus strain A59 genome RNA contain a common nucleotide sequence that is homologous to the 3' end of the viral mRNA leader sequence. *J. Virol.* **53**:834-840.
7. Koller, B., H. Fromm, E. Galun, and M. Edelman. 1987. Evidence for in vivo *trans* splicing of pre-mRNAs in tobacco chloroplasts. *Cell* **48**:111-119.
8. Krause, M., and D. Hirsh. 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* **49**:753-761.
9. Lai, M. M. C., R. S. Baric, P. R. Brayton, and S. A. Stohlman. 1984. Characterization of leader RNA sequence on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic virus. *Proc. Natl. Acad. Sci. USA* **81**:3626-3630.
10. 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.
11. Lai, M. M. C., and S. A. Stohlman. 1978. RNA of mouse hepatitis virus. *J. Virol.* **26**:236-242.
12. Makino, S., S. A. Stohlman, and M. M. C. Lai. 1986. Leader sequences of murine coronavirus mRNAs can be freely reassorted: evidence for the role of free leader RNA in transcription. *Proc. Natl. Acad. Sci. USA* **83**:4204-4208.
13. Makino, S., F. Taguchi, N. Hirano, and K. Fujiwara. 1984. Analysis of genomic and intracellular viral RNAs of small plaque mutants of mouse hepatitis virus, JHM strain. *Virology* **139**:138-151.
14. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
15. Murphy, W. J., K. P. Watkins, and N. Agabian. 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. *Cell* **47**:517-525.
16. Pedersen, F. S., and W. A. Haseltine. 1980. A micromethod for detailed characterization of high molecular weight RNA. *Methods Enzymol.* **65**:680-687.
17. Pfeleiderer, M., M. A. Skinner, and S. G. Siddell. 1986. Coronavirus MHV-JHM: nucleotide sequence of the mRNA that encodes the membrane protein. *Nucleic Acids Res.* **14**:6338.
18. Plotch, S. H., M. Bouloy, I. Ulmanen, and R. M. Krug. 1981. A unique cap (m⁷GpppXm)-dependent influenza virion endonuclease cleaves capped mRNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**:847-858.
19. Shieh, C.-K., L. H. Soe, S. Makino, M.-F. Chang, S. A. Stohlman, and M. M. C. Lai. 1987. The 5'-end sequence of the murine coronavirus genome: implications for multiple fusion sites in leader-primed transcription. *Virology* **156**:321-330.
20. Skinner, M. A., D. Ebner, and S. G. Siddell. 1985. Coronavirus MHV-JHM mRNA 5' has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. *J. Gen. Virol.* **66**:581-592.
21. Skinner, M. A., and S. G. Siddell. 1983. Coronavirus JHM: nucleotide sequence of the mRNA that encodes nucleocapsid protein. *Nucleic Acids Res.* **15**:5045-5054.
22. Skinner, M. A., and S. G. Siddell. 1985. Coding sequence of coronavirus MHV-JHM mRNA 4. *J. Gen. Virol.* **66**:593-596.
23. Spaan, W., H. Delius, M. Skinner, J. Armstrong, P. Rottier, S. Smeekens, B. A. M. van der Zeijst, and S. G. Siddell. 1983. Coronavirus mRNA synthesis involves fusion of noncontiguous sequences. *EMBO J.* **2**:1939-1944.
24. Sutton, R. E., and J. C. Boothroyd. 1986. Evidence for *trans* splicing in trypanosomes. *Cell* **47**:527-535.
25. Wege, H., A. Müller, and V. ter Meulen. 1978. Genomic RNA of the murine coronavirus JHM. *J. Gen. Virol.* **41**:217-227.
26. Wilhelmssen, K. C., J. L. Leibowitz, C. W. Bond, and J. A. Robb. 1981. The replication of murine coronaviruses in enucleated cells. *Virology* **110**:225-230.