BACKGROUND PAPER

TRANSCRIPTION AND REPLICATION OF CORONAVIRUS RNA: A 1989 UPDATE

Michael M.C. Lai

Department of Microbiology
University of Southern California School of Medicine
Los Angeles
CA 90033
USA

INTRODUCTION

The genomic RNA of coronaviruses has two unique features. Firstly, it is one of the largest stable RNAs known to exist in nature and is unquestionably the largest viral genomic RNA. The complete sequence of avian infectious bronchitis virus (IBV) RNA shows that its size is 27.6 kilobases (kb) (Boursnell *et al.*, 1987). Although the complete sequences of other coronavirus genomic RNAs are not yet available, preliminary data from several laboratories indicate that the genomic RNA of mouse hepatitis virus (MHV) is as long as 32 kb (Shieh and Lai, unpublished observations). The large size of these RNAs poses a theoretical quandary for the replication of coronavirus RNA, considering the high error frequency of RNA-dependent RNA synthesis observed in some systems (Holland *et al.*, 1982; Steinhauer and Holland, 1986). How does coronavirus RNA replicate faithfully despite the high error frequency of RNA synthesis? It is possible that a proof-reading mechanism operates to correct unavoidable mistakes which are expected to occur in almost every RNA molecule of this size. Secondly, coronavirus RNA contains a leader sequence (approximately 72 nucleotides) which is repeated at the 5'-end of every subgenomic mRNA species. This structural organization appears to be similar to that of most of eukaryotic mRNAs which contain leader sequences derived by RNA splicing. Yet the coronavirus RNA genome does not have consensus splicing signals and the virus replicates exclusively in the cytoplasm (Wilhelmson *et al.*, 1981; Brayton *et al.*, 1981), where there is no conventional RNA splicing machinery. Furthermore, UV transcriptional mapping studies suggest that coronavirus mRNA species are transcribed independently (Jacobs *et al.*, 1981), instead of being derived by cleavage of a precursor RNA. Thus, a new transcriptional mechanism must operate to transcribe coronavirus mRNAs.

LEADER-PRIMED TRANSCRIPTION OF SUBGENOMIC mRNAs

Several possible mechanisms could explain the generation of coronavirus subgenomic RNAs (Baric *et al.*, 1983). Over the last 4-5 years, a considerable body of evidence has been accumulated in support of a transcription mechanism unique to coronavirus, i.e. "leader-primed transcription" (Fig. 1.). This mechanism proposes that a leader RNA is transcribed first from the template RNA. It dissociates from the template as a free leader RNA species, which then binds to the template RNA again at various transcription initiation sites and serves as a primer for mRNA transcription.

The essence of this mechanism has been proven: (1) "Free" leader RNA species of 60-90 nucleotides have been detected in the cytoplasm of MHV-infected cells (Baric *et al.*, 1985, 1987). These RNAs are not associated with the template RNA. (2) A temperature-sensitive mutant has been isolated, which makes leader RNA, but no mRNAs at nonpermissive

Fig. 1 Model of leader primed transcription. Solid squares represent leader RNA and open circles represent RNA polymerases.

temperature, suggesting that the synthesis of leader RNA and the synthesis of mRNAs are discontinuous and require different viral proteins (Baric *et al.*, 1985). (3) During mixed infection with two different MHV strains, the leader RNA of mRNAs can be freely exchanged between the coinfecting viruses as if the leader RNA exists as a separate transcription unit, independent of the rest of mRNAs (Makino *et al.*, 1986). (4) When two MHV strains which differ only in the sequence at the 3'-end of leader RNA were compared, it was found that the number of mRNA species transcribed by these two viruses was different (Makino and Lai, 1989). For instance, a derivative of JHM strain which contains two UCUAA repeat sequences at the 3'-end of leader RNA transcribes an additional mRNA, 2b, while the same virus with three UCUAA repeats does not (Makino and Lai, 1989). This result gives the clearest evidence that the transcriptional initiation of coronavirus subgenomic mRNAs is controlled by leader RNA sequence. (5) Each mRNA species of MHV is heterogeneous, varying in the number of UCUAA repeats at the 3'-end of leader RNA (Makino *et al.*, 1988). Since this repeat sequence is present at both the 3'-end of leader RNA and the transcriptional initiation sites, this result suggests that these sequences are involved in leader RNA binding and that the leader RNA binding is imprecise. These studies indicate that coronavirus mRNA transcription is the result of interactions between the 3'-end of leader sequences and the transcription initiation sequences at the template RNA.

CONSENSUS SEQUENCES OF TRANSCRIPTIONAL INITIATION SITES AND LEADER RNA

A large number of various coronavirus genomic RNA sequences have already been obtained. Thus, it is now possible to derive a consensus sequence at the transcription initiation sites. Table 1 shows a compilation of the possible transcriptional initiation sites of various coronavirus strains. Since most of these sequences have not been confirmed by sequencing of mRNAs to be the actual transcriptional initiation sites, the designation of these sequences as transcriptional initiation sites was based on sequence homology and their localization at

Table 1. Consensus sequences at transcriptional initiation sites

<u>Virus</u>	<u>Consensus sequence</u>
MHV, BCV, HCV-229E	UCUAAAC
HCV-OC43	UCUAAAU
TGEV, FIPV	ACUAAAC
IBV	UUAACAA

intergenic regions. It is clear that these sequences are very well conserved, in contrast to the rest of the genomic sequences. The only exception is IBV, which has diverged extensively from other coronaviruses. When transcriptional initiation sites for different mRNAs on the same viral genome is compared, some minor divergence is noted in a few transcriptional initiation sites. These seven-nucleotide sequences appear to be the core of the leader-binding sites and is homologous between the 3'-end of the leader RNA and the intergenic sequences. It should be noted that leader-RNA binding probably involves sequences adjacent to these core sequences as well, as between the leader RNA and the transcriptional start sites (Shieh *et al.*, 1987). The degrees of this homology at different sites parallel, to a certain extent, the amounts of the different mRNA species transcribed, but this correlation is not absolute (Shieh *et al.*, 1987). It is also noted that the core sequence is devoid of any G residues. The minor heterogeneity observed at different transcriptional sites involves only A, C and U interchanges. The mutation of A to G within the core sequence in one of the mRNAs (mRNA 2b) of A59 strain of MHV completely eliminates the transcription within this region (Shieh *et al.*, 1989).

The remaining sequence of the leader RNA is less well conserved. However, the leader RNA appears to maintain the same basic conformation in all of the coronaviruses which have been sequenced so far (Fig. 2). The functional significance of these secondary structures is not clear at the present time.

THE STATE OF TEMPLATE RNA AND LEADER RNA IN INFECTED CELLS

The state of the template RNA for coronavirus transcription has been less well studied due to the small quantity of template (negative strand) RNA. The single report characterizing the size of the coronavirus template RNA indicated that it is a genomic-sized RNA (Lai *et al.*, 1982). There is currently a controversy concerning the kinetics of its synthesis: it is not known whether the negative-stranded RNA is synthesized only early in the infection (Brayton *et al.*, 1984) or throughout the infection (Sawicki *et al.*, 1986). Recently, it has been shown that TGEV may synthesize subgenomic template RNAs which have sizes equivalent to those of subgenomic mRNAs (D. Brian, this volume). This finding is hard to reconcile with the existing data. Clearly more studies on the negative-strand RNA are warranted.

Fig. 2 Predicted secondary structures of MHV and IBV leader RNAs.

The understanding of the mechanism of coronavirus RNA replication has been aided by the studies of defective-interfering (DI) RNA. DI RNA has been generated from the JHM strain of MHV during high multiplicity passages of the virus (Makino *et al.*, 1984). It replicates at a very high efficiency, thus providing an opportunity to study the mechanism of RNA replication. Furthermore, DI RNAs are much smaller than the virion genomic RNA; for instance, the smallest DI RNA detected in MHV-infected cells is only 2.2 kb (Makino *et al.*, 1988) as compared to 32 kb for the entire virion genomic RNA. A complete cDNA clone of this DI RNA has been obtained and the *in vitro* transcribed RNA is capable of extremely efficient replication in the presence of a helper virus. Surprisingly, when the leader RNA in the DI RNA transfected cells was examined, it was found to contain the leader RNA from the helper virus, instead of its own leader RNA (Makino and Lai, this volume). This result suggests that a leader RNA supplied in trans is utilized in coronavirus DI RNA replication. Thus, the same mechanism of leader-primed RNA synthesis is likely to operate in the RNA replication step.

HIGH-FREQUENCY RNA RECOMBINATION SUGGESTS A DISCONTINUOUS MECHANISM OF RNA REPLICATION

Another interesting phenomenon which is relevant to coronavirus RNA replication is the high frequency of RNA recombination (Makino *et al.*, 1986). Recombination occurs both in tissue culture and in animals (Keck *et al.*, 1988b). Recombination frequency can be more than 10% under some conditions (Makino *et al.*, 1986). Such a high frequency of recombination resembles reassortment of segmented RNAs in influenza viruses, suggesting that segmented RNA intermediates may be generated during MHV RNA replication. RNA intermediates may account for RNA recombination by a copy-choice mechanism. Indeed, RNAs of discrete, non-message sizes have been detected in the cytoplasm of MHV-infected cells; these may correspond to transcriptional intermediates pausing at secondary structures in the template RNA (Baric *et al.*, 1987). These results suggest that coronavirus RNA replication proceeds by a discontinuous and non-processive mechanism, with transcriptional pauses occurring at sites of secondary structure. The occurrence of high-frequency RNA recombination may explain why the coronavirus RNA genome of enormous size (32 kb) can be maintained as a stable and infectious RNA.

ENZYMOLOGY OF RNA TRANSCRIPTION AND REPLICATION

Coronavirus RNA synthesis is mediated by virus-specific RNA-dependent RNA polymerase. The enzyme activity has been demonstrated in TGEV and MHV-infected cells; however, the protein responsible for the polymerase activity has not been identified. Based on the size of the possible gene products and analogy to other RNA viruses, it is assumed that the 5'-most gene (gene 1) encodes the polymerase proteins. This gene has been completely sequenced in IBV (Bournsell *et al.*, 1987). It is approximately 18 kb in length and has two open reading frames (ORFs) which overlap each other slightly. It has been shown that these two ORFs can be translated into a polyprotein by a ribosomal frame-shifting mechanism (Brierley *et al.*, 1987, 1989), resulting, presumably, in a polyprotein of M_r greater than 700,000. This polyprotein is most likely processed into multiple proteins (Brierley *et al.* this volume).

In the case of MHV, the sequencing of the gene encoding the polymerase protein has not been completed. This gene is approximately 23 kb long (Baker *et al.*, this volume). It has the capacity to encode a protein of M_r 800,000. Again, this polyprotein is most likely processed into multiple proteins. Indeed, it has already been shown that the N-terminus of this polyprotein is cleaved into a p28 protein (Denison and Perlman, 1986; Soe *et al.*, 1987); the cleavage is catalyzed by an autolytic activity inherent in this polyprotein (Baker *et al.*, this volume). The proteins encoded by this gene very likely contain other activities in addition to the polymerase activity. For instance, the putative endonuclease required for processing the leader RNA, and other activities such as those required for the regulation of RNA transcription may also be present within this polyprotein.

The polymerases are likely to be associated with membranes (Dennis and Brian, 1982; Brayton *et al.*, 1982). Most likely, the transcription and replication complexes are in a membranous compartment, so that a high concentration of leader RNAs can be maintained

locally and RNA recombination between coinfecting viruses can therefore occur at a high frequency.

PERSPECTIVES

Studies conducted so far have clearly established that coronaviruses utilize unique mechanisms for the transcription and replication of their RNAs. Many details of these mechanisms are, however, yet to be unveiled. Several areas of research in the future may be particularly useful in revealing additional insights into the mechanism of coronavirus RNA synthesis:

1. Establishment of An *In Vitro* Transcription and Replication System

Several *in vitro* systems capable of synthesizing either genomic or subgenomic RNAs have been reported (Dennis and Brian, 1982; Brayton *et al.*, 1982; Compton *et al.*, 1987; Leibowitz and DeVries, 1988). However, none of these are capable of utilizing exogenous RNAs as either template or primers. The ability of any *in vitro* system to transcribe or extend exogenous RNAs is a key to understanding the mechanism of RNA synthesis.

2. Temperature-Sensitive Mutants Defective in RNA Synthesis

An earlier report has grouped RNA (-) temperature-sensitive mutants into six complementation groups (Leibowitz *et al.*, 1982). However, no follow-up work has been reported. The characterization of phenotype and genotype of these mutants and genetic mapping of the temperature-sensitive lesions by complementation and recombination should yield valuable information on the possible gene products involved in coronavirus RNA synthesis.

3. Characterization of the Protein Products of the Gene Encoding the Putative Polymerase

This gene, ranging from 18 to 23 kb, very possibly encodes a polypeptide of M_r 700-800,000, which may subsequently be processed into multiple proteins. Its size is two to three times the size of the genomes of other positive-strand RNA viruses, such as flavi or picornaviruses. Thus, the primary gene product of this gene likely to undergo complex processing pathways, and possess multiple functions. The identification of these proteins and their functions will be an important contribution to the understanding of coronavirus RNA transcription and replication.

REFERENCES

- Baric, R. S., Nelson, G. W., Fleming, J. O., Deans, R. J., Keck, J. G., Casteel, N., and Stohlman, S. A., 1988, Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription, *J. Virol.*, 62:4280.
- Baric, R. S., Shieh, C.-K., Stohlman, S. A., and Lai, M. M. C., 1987, Analysis of intracellular small RNAs of mouse hepatitis virus: evidence for discontinuous transcription, *Virology*, 156:342.
- 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. M. C., 1985, Characterization of leader related small RNAs in coronavirus-infected cells: further evidence for leader-primed mechanism of transcription, *Virus Res.*, 3:19.
- Bournsell, M. E. G., Brown, T. D. K., Foulds, I. J., Green, P. F., Tomley, F. M., and Binns, M. M., 1987, Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus, *J. Gen. Virol.*, 68:57.
- Brayton, P. R., Ganges, R. G., and Stohlman, S. A., 1981, Host cell nuclear function and murine hepatitis virus replication, *J. Gen. Virol.*, 56:457.
- Brayton, P. R., Lai, M. M. C., Patton, C. D., and Stohlman, S. A., 1982, Characterization of two RNA polymerase activities induced by mouse hepatitis virus, *J. Virol.*, 42:847.
- Brayton, P. R., Stohlman, S. A., and Lai, M. M. C., 1984, Further characterization of mouse hepatitis virus RNA-dependent RNA polymerase, *Virology*, 133:197.
- Brierley, I., Bournsell, M. E. G., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. K., and Inglis, S. C., 1987, An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV, *EMBO J.*, 6:3779.

- Brierley, I., Digard, P., and Inglis, S. C., 1989, Characterization of an efficient coronavirus ribosomal frame-shifting signal: Requirement for an RNA pseudoknot, Cell, 57:537.
- Compton, S. R., Rogers, D. B., Holmes, K. V., Fertsch, D., Remenick, J., and McGowan, J. J., 1987, In vitro replication of mouse hepatitis virus strain A59, J. Virol., 61:1814.
- Denison, M. R., and Perlman, S., 1986, Translation and processing of mouse hepatitis virus virion RNA in a cell-free system, J. Virol., 60:12.
- Dennis, D. E., and Brian, D. A., 1982, RNA-dependent RNA polymerase activity in coronavirus-infected cells, J. Virol., 42:153.
- Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and van de Pol, S., 1982, Rapid evolution of RNA genomes, Science, 215:1577.
- Jacobs, L., Spaan, W.J.M., Horzinek, M.C. and van der Zeijst, B.A.M., 1981, The synthesis of the subgenomic mRNAs of mouse hepatitis virus is initiated independently: evidence from UV transcriptional mapping, J. Virol., 39:401.
- Keck, J. G., Hogue, B. G., Brian, D. A., and Lai, M. M. C., 1988a, Temporal regulation of bovine coronavirus RNA synthesis, Virus Res., 9:343.
- Keck, J. G., Matsushima, G. K., Makino, S., Fleming, J. O., Vannier, D. M., Stohlman, S. A., and Lai, M. M. C., 1988b, In vivo RNA-RNA recombination of coronavirus in mouse brain, J. Virol., 62:1810.
- Lai, M. M. C., Patton, C. D., and Stohlman, S. A., 1982, Replication of mouse hepatitis virus: negative-stranded RNA and replicative form RNA are of genome length, J. Virol., 44:487.
- Lai, M. M. C., 1986, Coronavirus leader-RNA-primed transcription: an alternative mechanism to RNA splicing, BioEssays, 5:257.
- Leibowitz, J. L., DeVries, J. R., and Haspel, M. V., 1982, Genetic analysis of murine hepatitis virus strain JHM, J. Virol., 42:1080.
- Leibowitz, J. L., and DeVries, J. R., 1988, Synthesis of virus-specific RNA in permeabilized murine coronavirus-infected cells, Virology 166:66.
- Makino, S., Taguchi, F., and Fujiwara, K., 1984, Defective interfering particles of mouse hepatitis virus, Virology, 133:9.
- Makino, S., Keck, J. G., Stohlman, S. A., and Lai, M. M. C., 1986, High-frequency RNA recombination of murine coronaviruses, J. Virol., 57:729.
- 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.
- Makino, S., Shieh, C.-K., Soe, L.-H., Baker, S. C., and Lai, M. M. C., 1988, Primary structure and translation of a defective interfering RNA of murine coronavirus, Virology, 166:550.
- Makino, S., Soe, L. H., Shieh, C.-K., and Lai, M. M. C., 1988, Discontinuous transcription generates heterogeneity at the leader fusion sites of coronavirus mRNAs, J. Virol., 62:3870.
- Makino, S., and Lai, M. M. C., 1989, Evolution of the 5'-end of genomic RNA of murine coronaviruses during passages in vitro, Virology, 169:227.
- Sawicki, S. G., and Sawicki, D. L., 1986, Coronavirus minus-strained RNA synthesis and effect of cycloheximide on coronavirus RNA synthesis, J. Virol., 57:328.
- Shieh, C.-K., Soe, L. H., Makino, S., Chang, M.-F., Stohlman, S. A., and Lai, M. M. C., 1987, The 5'-end sequence of the murine coronavirus genome: implication for multiple fusion sites in leader-primed transcription, Virology, 156:321.
- Shieh, C.-K., Lee, H.-J., Yokomori, K., La Monica, N., Makino, S., and Lai, M. M. C., 1989, Identification of a new transcriptional initiation site and the corresponding functional gene 2b in the murine coronavirus RNA genome, J. Virol., 63: 3729-3736.
- Soe, L. H., Shieh, C.-K., Baker, S. C., Chang, M.-F., and Lai, M. M. C., 1987, Sequence and translation of the murine coronavirus 5'-end genomic RNA reveals the N-terminal structure of the putative RNA polymerase, J. Virol., 61: 3968-3976.
- Steinhauer, D. A., and Holland, J. J., 1986, Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA, J. Virol., 57:219.
- Wilhelmsen, K. C., Leibowitz, J. L., Bond, C. W., and Robb, J. A., 1981, The replication of murine coronavirus in enucleated cells, Virology, 110:225-230.