

Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/W^v mice. *Cell* **42**, 71–79.

19 KELLER, G., PAIGE, C., GILBOA, E. & WAGNER, E. F. (1985). Expression of a foreign gene in myeloid and lymphoid cells derived from multipotent haematopoietic

precursors. *Nature* **318**, 149–154.

20 LEMISCHKA, I. R., RAULET, D. H. & MULLIGAN, R. C. (1986). Developmental potential and dynamic behavior of hematopoietic stem cells. *Cell* **45**, 917–927.

21 ANDERSON, W. F. (1984). Prospects for human gene therapy. *Science* **226**, 401–409.

ELI GILBOA is in the Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10021, USA.

Coronavirus Leader-RNA-Primed Transcription: an Alternative Mechanism to RNA Splicing

Michael M. C. Lai

Summary

Many viral and cellular mRNA species contain a leader sequence derived from a distant upstream site on the same gene by a process of RNA splicing. This process usually involves either nuclear functions or self-splicing of RNA molecules. Coronavirus, a cytoplasmic RNA virus, unfolds yet another mechanism of joining RNA, which involves the use of a free leader RNA molecule. This molecule is synthesized and dissociates from the template RNA, and subsequently reassociates with the template RNA at downstream initiation sites of subgenomic mRNAs to serve as the primer for transcription. This leader-primed transcriptional process thus generates viral mRNAs with a fused leader sequence. A similar mechanism might also operate in the mRNA transcription of African trypanosomes.

Introduction

Coronaviruses are a newly recognized group of viruses which infect many animal species, causing a variety of respiratory and gastrointestinal illnesses. Among these viruses are porcine transmissible gastroenteritis virus (TGEV), bovine coronavirus (BCV), avian infectious bronchitis virus (IBV), feline infectious peritonitis virus (FIPV), etc., which cause severe epidemiological problems in livestock and other domestic animals. Human coronaviruses are responsible for a significant number of common colds and diarrhoea, and have also been implicated in multiple sclerosis. Another member of the virus group, mouse hepatitis virus (MHV), is a frequent contaminant of laboratory mouse colonies.

Coronaviruses share several common

morphological and structural features.¹ They consist of an enveloped virus particle with characteristic petal-shaped spikes, which give an appearance of a crown, or corona, thus giving the name for the virus. The viral envelope contains two glycoproteins, one of which forms the spikes and the other the matrix proteins. The spikes interact with virus receptors on the cell surface, thus determining the target-cell specificity of the virus. Inside the viral envelope is a helical nucleocapsid which is composed of a nucleocapsid protein N and the RNA genome. The genome is a single piece of RNA of more than 18,000 nucleotides. The RNA is of positive sense, meaning that the viral genome can be directly translated into proteins.

Replication Pathway of Coronaviruses

Coronaviruses generally infect cells of the same animal species of origin in tissue culture. Immediately after the virus enters infected cells, the viral RNA is released from the virus particles by a process termed 'uncoating'. The viral RNA is used as messenger for an RNA-dependent RNA polymerase, which transcribes the genomic RNA into a full-length negative-stranded RNA.² This negative-stranded RNA species then serves as the template for the synthesis of seven virus-specific mRNAs. Each mRNA is used for the synthesis of either a viral structural or nonstructural protein. The nonstructural proteins are probably used to regulate the replication and transcription of virus. One of these nonstructural proteins is the RNA-dependent RNA polymerase which is needed to syn-

thesize more mRNAs as well as virion genomic RNA. The structural proteins, together with the genomic RNA, are used to assemble the virus particles. These particles, unlike most other enveloped viruses, bud into the endoplasmic reticulum, instead of directly into the plasma membrane.³ The mature virus particles travel through the Golgi complex and are eventually released into the extracellular space. The virus acquires the virus-specific glycoproteins, E1 and E2, during the budding and transport process.

Coronavirus mRNAs Have a Unique Nested-set Structure with a Common Leader RNA Sequence

Coronavirus mRNA is comprised of a genomic-sized RNA and six subgenomic mRNA species. All of these RNAs are associated with polysomes, thus representing functional mRNA species

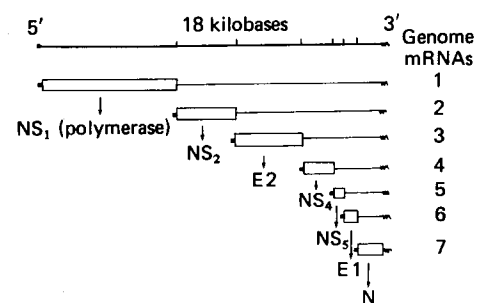


Fig. 1. mRNA structure of coronaviruses. The structure is based on that of mouse hepatitis virus. The open boxes represent the functional parts of mRNAs. The solid squares are the leader sequences and the wavy lines represent poly A tails. NS proteins represent nonstructural proteins which are not packaged into virus particles. E1, Matrix protein; E2, spike protein; N, nucleocapsid protein.

REVIEW ARTICLES

in infected cells. The structure of these sequences is such that they each start from the 3'-end of the genomic RNA and extend for various distances toward the 5'-end, depending on the size of the mRNA species (Fig. 1).⁴ Thus the sequence of each subgenomic mRNA is completely contained within the 3'-portion of the next larger mRNA, i.e. a nested-set structure. The 5'-portion of each mRNA therefore contains a unique sequence which does not overlap with the smaller mRNAs. This unique portion is the functional part of each mRNA. For instance, when mRNA 3 was added to an *in vitro* translation system, such as reticulocyte lysate,⁵ or was injected into frog oocytes,⁶ only the E2 protein was synthesized, although this mRNA contains several downstream genes. Thus each mRNA is functionally monocistronic. This phenomenon is probably a result of the inability of mammalian cells to utilize internal translation initiation codons of mRNAs.

The abundance of different mRNAs varies. The small mRNAs are relatively more abundant than the larger mRNA species,⁷ but this inverse relationship between the mRNA size and abundance is not universal. Furthermore, the genetic map of different coronaviruses is not co-linear. For instance, IBV contains an additional gene inserted between the genes encoding E1 and N. Thus genetic recombination, insertion or deletion must have occurred during the evolution of the viruses.

Another unique feature of coronavirus mRNAs is that the 5'-ends of all the mRNAs, including genomic RNA, contain an identical stretch of 50–70 nucleotides,^{8,9} termed the leader sequence. These leader nucleotides are not repeated in the internal region of the genomic RNA. Therefore the leader of the subgenomic mRNAs must be derived from the 5'-end of the genomic RNA, analogous to the leader sequences of many eukaryotic mRNAs. The presence of the leader sequences on the coronavirus mRNAs was first suggested by an unusual oligonucleotide,¹⁰ which turns out to be the fusion site between the leader RNA and mRNAs. The identity of the leader sequences has now been confirmed by cDNA cloning and sequencing of genomic RNA and mRNAs. In the case of MHV, the leader sequence is approximately 70 nucleotides long.

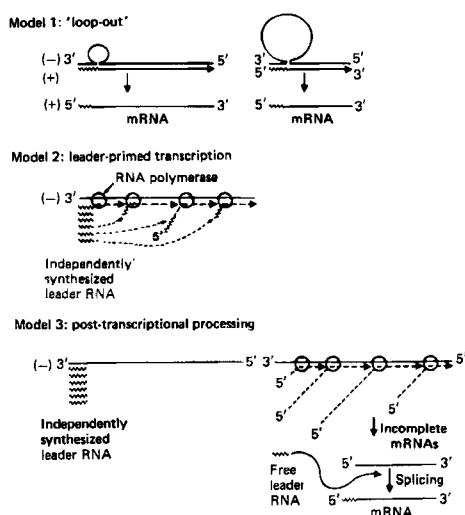


Fig. 2. Three possible models of coronavirus transcription.

The Coronavirus leader RNA is not Derived by the Conventional Pre-mRNA Splicing Mechanism

Although the presence of a leader sequence is a common feature of eukaryotic mRNAs, the synthesis of coronavirus mRNAs does not appear to utilize the conventional RNA splicing mechanism, which involves cleavage of intervening sequences from a precursor RNA. Two pieces of data suggest that the leader sequences present at the 5'-end of coronavirus mRNAs are probably derived by a novel mechanism: (1) coronaviruses replicate exclusively in the cytoplasm of infected cells,^{11,12} while conventional RNA splicing takes place in the nucleus; and (2) ultraviolet (UV) irradiation of coronavirus-infected cells resulted in the inhibition of individual mRNA synthesis at a rate proportional to the size of the mRNA.¹³ The UV target size of each coronavirus mRNA is approximately the same as its physical size, suggesting that each subgenomic mRNA is transcribed independently. If these mRNAs were derived from the cleavage of a precursor RNA, as in the case of conventional RNA splicing, the UV target size for inhibition of the synthesis of all the mRNAs would be expected to be equal to that of the genomic RNA. Thus a novel mechanism for the generation of these mRNAs is most likely utilized by coronaviruses.

Several transcriptional models for coronavirus mRNAs have been proposed (Fig. 2).¹⁴ (1) The 'loop-out' model in which the RNA template forms a loop in the 'intron' region, thus bringing the leader RNA in close proximity to the initiation sites of subgenomic mRNAs. The RNA poly-

merase can thus jump from the leader sequence to the mRNAs in continuous transcription. For smaller mRNAs the loops will be larger. (2) The 'leader-primed transcription' model in which the leader RNA is transcribed and becomes dissociated from the RNA template. This 'free' leader RNA then binds to the template at the initiation sites of various mRNAs and serves as the primer for transcription. (3) The 'post-transcriptional processing' model in which the leader RNA and the body sequences are transcribed independently and then joined together by a *trans-splicing* mechanism post-transcriptionally. A considerable body of data¹⁴ suggests that the first and third models are not compatible with the replication mechanism of coronavirus RNA. Thus, the 'leader-primed transcription' model is considered to be the most likely mechanism utilized for the transcription of coronavirus mRNAs.

Evidence for the 'Leader-primed Transcription' Model

The direct evidence in support of the 'leader-primed transcription' model is multi-fold. (1) Several free leader RNA species of 50–90 nucleotides have been detected in the cytoplasm of MHV-infected cells.¹⁵ They are discrete RNA species, some of which are dissociated from the RNA template. (2) A temperature-sensitive mutant has been isolated which synthesizes only the small leader RNA but not mRNAs at the non-permissive temperature,¹⁵ suggesting that the synthesis of the leader RNA and the synthesis of the mRNAs are discontinuous and require different viral proteins. (3) During a mixed infection with two different MHVs, the mRNAs of each virus contain a population with the leader RNA sequences derived from the co-infecting virus.¹⁶ This result suggests that the leader RNA is a separate transcriptional unit and can be freely exchanged between the mRNAs of two co-infecting viruses. Thus the free leader-containing RNA species detected in the cytoplasm of MHV-infected cells represent *bona fide* transcriptional intermediates, rather than abortive transcription products. (4) An RNA species complementary to the leader RNA ('anti-sense' leader RNA) could be expressed in L-2 cells by a mammalian expression vector. When the cells expressing this anti-sense leader RNA were infected with MHV, the RNA transcription of the superinfecting virus was inhibited, though not completely (L. Soe, unpublished observation). This

result suggests that the availability of the leader RNA sequences is required for MHV RNA synthesis. These data are most compatible with the interpretation that the leader RNA serves as a primer for transcription.

The Mechanism of Leader-primed Transcription

The precise mechanism of leader RNA-primed transcription has been suggested from the sequence analysis of both the 5'-end and the mRNA initiation sites on the RNA genomes. The sequence of the 5'-end of MHV genomic RNA reveals a possible leader RNA termination signal, i.e. a hairpin loop followed by an AU-rich sequence, approximately 84 nucleotides from the 5'-end (Fig. 3).¹⁷ This is consistent with the size of the free leader-containing RNA species detected in the infected cells.¹⁵ Close to the 3'-end of the putative leader RNA, there is a stretch of 9–18 nucleotides, which is homologous to the genomic sequences at the initiation points of various mRNAs.^{17, 18} Thus the leader RNA would be able to bind to the RNA template at these sites. The homologous sequences are not located at the exact 3'-end of the free leader RNA, and there is also a mismatched nucleotide within some of the homologous regions. These mismatched nucleotides in the leader RNA have to be removed before transcription takes place. Thus the mechanism of MHV RNA transcription involves the synthesis of a free leader RNA species from the 3'-end of the negative-sense RNA template, followed by its binding to the intergenic regions via the homologous sequences. The leader RNA is then cleaved at the mismatched point, generating the primer for the initiation of mRNA transcription (Fig. 3). Therefore, the leader-body fusion sites vary with the mRNA species, depending on the location of the mismatched nucleotide.

There is a close parallel between the number of homologous nucleotides at the intergenic sites preceding each mRNA and the amount of RNA synthesized in the infected cells.¹⁷ For instance, there are 18, 14 and 9 homologous nucleotides upstream of mRNAs 7, 6 and 4, respectively. Correspondingly, the molar ratio of these three RNAs in infected cells is roughly 100:31:1.7.⁷ In addition, other regulatory sequences may also be involved. The free leader RNA may be associated with RNA polymerase, which could assist in leader RNA binding.

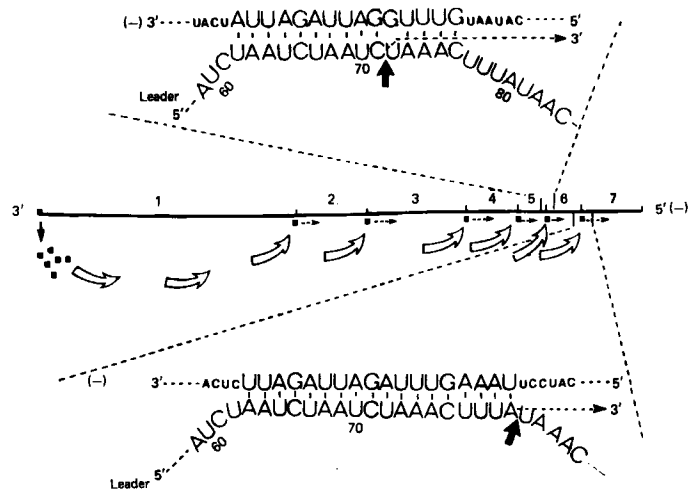


Fig. 3. The postulated model of leader-primed transcription. The solid squares represent free leader RNA, which binds to the initiation points of mRNAs on the template RNA. The expanded regions show the nucleotide sequences of the intergenic regions of mRNAs 6 and 7, and the mechanism of leader RNA binding. The solid arrows represent the postulated cleavage points of the leader RNA. Transcription starts from the 3'-end of the cleaved primer. The resulting mRNAs consist of the leader sequence fused to the gene.

RNA Replication May be Discontinuous and 'Stop-and-go'

After sufficient mRNAs are synthesized in the infected cells, RNA synthesis shifts to the replication of genomic RNAs to be packaged into virions. The transcription of mRNAs and replication of genomic RNA appear to be catalyzed by separate RNA polymerases. Although the replication of the genomic RNA represents faithful copying of the RNA template, it probably follows a discontinuous and 'stop-and-go' mode. This mechanism was suggested by the discovery of many discrete species of small leader-containing RNAs in the MHV-infected cells.¹⁵ The sizes of these RNA species correspond to the lengths between the 5'-end and the probable hairpin loops on the genomic RNA. Thus they may represent transcriptional 'pausing' intermediates which stop at the regions of secondary structure on the RNA template. These intermediates would subsequently participate in continuing transcription. This mode of discontinuous RNA transcription has also been demonstrated in several other systems, such as QB, MS2 and T7 phages.^{19, 20}

The evidence that these pausing RNA intermediates are *bona fide* intermediates of RNA replication came from the demonstration that coronaviruses can undergo RNA recombination at a very high frequency,²¹ which almost matches the frequency of reassortment of RNA segments in reoviruses or influenza virus. This result suggests that, although coronavirus contains a

non-segmented RNA genome, its replication may involve the generation of freely dissociated RNA intermediates, which result from transcriptional 'pausing' and that these intermediates may switch RNA templates. Thus this 'stop-and-go' mode of RNA replication may result in RNA recombination, which has now been demonstrated in several animal and plant virus systems.

The Discontinuous, Leader-primed RNA Transcription May be a Common Phenomenon for Gene Regulation in Eukaryotic Cells

The unique feature of leader RNA-primed transcription may not be limited to coronaviruses. A similar phenomenon has been noted in African trypanosomes, in which a 35-nucleotide leader RNA sequence is associated with Variant Surface Glycoprotein (VSG) mRNAs and many other cellular mRNA species.^{22, 23} This leader RNA sequence is located on a separate chromosome, and is expressed as a 137-nucleotide leader-containing RNA species. The mechanism of mRNA synthesis in this case is not clear, but may involve either a leader-primed transcription or a *trans*-splicing event.

An RNA-primed transcription has also been noted in influenza virus, which requires capped (m⁷ GpppXm) cellular mRNAs as primers for its mRNA synthesis.²⁴ One notable difference between coronavirus and influenza virus is the lack of a specific sequence requirement for the primer RNA in the case of influenza virus. A similar

REVIEW ARTICLES

primer-dependent RNA transcription has also been detected in yet another RNA virus, bunyavirus.²⁵ In addition, retroviruses use specific tRNAs to prime cDNA reverse transcription.

The leader RNA in coronavirus transcription may play a role in regulating the rate of RNA synthesis. This can be achieved by the extent of homology at the intergenic regions and also by the specific sequences in the leader RNA. The leader sequences in different MHVs appear to vary significantly in their ability to promote RNA transcription and recombination. Further efforts in delineating the leader sequences and also the properties of the virus-encoded RNA polymerases should give further insights into the novel mechanisms of leader-primed transcription and discontinuous RNA replication.

REFERENCES

- 1 STURMAN, L. S. & HOLMES, K. V. (1983). The molecular biology of coronaviruses. *Adv. in Virus Res.* **28**, 35–112.
- 2 LAI, M. M. C., PATTON, C. D. & STOHLMAN, S. A. (1982). Replication of mouse hepatitis virus: negative-stranded RNA and replicative form RNA are of genome length. *J. Virol.* **44**, 487–492.
- 3 MASSALSKI, A., COULTER-MACKIE, M. & DALES, S. (1981). Assembly of mouse hepatitis virus strain JHM. In *Biochemistry and Biology of Coronaviruses* (ed. V. ter Meulen, S. Siddell and H. Wege), p. 111–118.
- 4 LAI, M. M. C., BRAYTON, P. R., ARMEN, R. C., PATTON, C. D., PUGH, C. & STOHLMAN, S. A. (1981). Mouse hepatitis virus A59: messenger RNA structure and genetic localization of the sequence divergence from the hepatotropic strain MHV 3. *J. Virol.* **39**, 823–834.
- 5 LEIBOWITZ, J. L., WEISS, S. R., PAAVOLA, E. & BOND, C. W. (1982). Cell-free translation of murine coronavirus RNA. *J. Virol.* **43**, 905–913.
- 6 ROTTIER, P. J. M., SPAAN, W. J. M., HORZINEK, M. & VAN DER ZEIJST, B. A. M. (1981). Translation of three mouse hepatitis virus (MHV-A59) subgenomic RNAs in *Xenopus laevis* oocytes. *J. Virol.* **38**, 20–26.
- 7 LEIBOWITZ, J. L., WILHELMSSEN, K. C. & BOND, C. W. (1981). The virus-specific intracellular RNA species of two murine coronaviruses: MHV-A59 and MHV-JHM. *Virology* **114**, 29–51.
- 8 SPAAN, W., DELIUS, H., SKINNER, M., ARMSTRONG, J., ROTTIER, P., SMEEKENS, S., VAN DER ZEIJST, B. A. M. & SIDDELL, S. G. (1983). Coronavirus mRNA synthesis involves fusion of noncontiguous sequences. *EMBO J.* **2**, 1939–1944.
- 9 LAI, M. M. C., BARIC, R. S., BRAYTON, P. R. & STOHLMAN, S. A. (1984). Characterization of leader RNA sequences 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., PATTON, C. D. & STOHLMAN, S. A. (1982). Further characterization of mouse hepatitis virus: presence of common 5'-end nucleotides. *J. Virol.* **41**, 557–565.
- 11 BRAYTON, P. R., GANGES, R. G. & STOHLMAN, S. A. (1981). Host cell nuclear function and murine hepatitis virus replication. *J. Gen. Virol.* **56**, 457–460.
- 12 WILHELMSSEN, K. C., LEIBOWITZ, J. L., BOND, C. W. & ROBB, J. A. (1981). The replication of murine coronaviruses in nucleated cells. *Virology* **110**, 225–230.
- 13 JACOBS, L., SPAAN, W. J. M., HORZINEK, M. C. & 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**, 401–406.
- 14 BARIC, R. S., STOHLMAN, S. A. & 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**, 633–640.
- 15 BARIC, R. S., STOHLMAN, S. A., RAZAVI, M. K. & 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**, 19–33.
- 16 MAKINO, S., STOHLMAN, S. A. & 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**, 4204–4208.
- 17 SHIEH, C. K., SOE, L., MAKINO, S., STOHLMAN, S. A. & LAI, M. M. C. (1986). The 5'-end sequence of murine coronavirus genome: implication for multiple fusion sites in leader-primed transcription. *Virology* (In press).
- 18 BUDZILOWICZ, C. J., WILCZYNSKI, S. P. & WEISS, S. R. (1985). Three intergenic regions of coronavirus mouse hepatitis virus strain A59 genomic RNA contain a common nucleotide sequence that is homologous to the 3'-end of the viral mRNA leader sequence. *J. Virol.* **53**, 834–840.
- 19 MILLS, D. R., DABKIN, C. & KRAMER, F. R. (1978). Template-determined, variable rate of RNA chain elongation. *Cell* **15**, 541–550.
- 20 KASSAVETIS, G. A. & CHAMBERLIN, M. J. (1981). Pausing and termination of transcription within the early region of bacteriophage T7 DNA *in vitro*. *J. Biol. Chem.* **256**, 2777–2786.
- 21 MAKINO, S., KECK, J. G., STOHLMAN, S. A. & LAI, M. M. C. (1986). High-frequency RNA recombination of murine coronaviruses. *J. Virol.* **56**, 729–737.
- 22 CAMPBELL, D. A., THORNTON, D. A. & BOOTHROYD, J. C. (1984). Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. *Nature* **311**, 350–355.
- 23 SATHER, S. & AGABIAN, N. (1985). A 5'-spliced leader is added *in trans* to both α - and β -tubulin transcripts in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. USA* **82**, 5695–5699.
- 24 KRUG, R. M. (1981). Priming of influenza viral RNA transcription by capped heterologous RNAs. *Curr. Topics in Microbiol. and Immunol.* **93**, 125–150.
- 25 PATTERSON, J. L., HOLLOWAY, B. & KOLAKOSKY, D. (1984). La Crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependent endonuclease. *J. Virol.* **52**, 215–222.

MICHAEL M. C. LAI is at the Departments of Microbiology and Neurology, University of Southern California, School of Medicine, Los Angeles, CA 90033, USA.

REVIEWS FOR BIOESSAYS

Although review articles for *BioEssays* are normally solicited by members of the Editorial Board, unsolicited reviews will be considered. If you are interested in writing a review article, please send a short résumé outlining the subject and scope of the proposed piece, with a brief description of your scientific background, to the Staff Editor, *BioEssays*, Cambridge University Press, The Edinburgh Bldg., Shaftesbury Rd., Cambridge CB2 2RU, United Kingdom.