

TRANSCRIPTION, REPLICATION, RECOMBINATION, AND ENGINEERING OF CORONAVIRUS GENES

Michael M. C. Lai

Howard Hughes Medical Institute and
Department of Molecular Microbiology
University of Southern California School of Medicine
Los Angeles, California 90033

INTRODUCTION

When one discusses coronavirus RNA synthesis, several unique features of coronavirus RNAs immediately come to mind¹: the extraordinary length (27-31 kb) of the RNA genome, the subgenomic mRNAs with a leader RNA derived from the 5'-end of the genome, and the presence of common sequences between the 3'-end of the leader RNA and the sequence (intergenic sequence, IG) preceding each gene (Figure 1). These characteristics indicate that coronavirus RNA synthesis has to be discontinuous, so that the leader RNA can be joined with mRNAs. This feature, coupled with several unique phenomena associated with coronavirus replication, e.g., high frequency of RNA recombination² and rapid generation and evolution of defective interfering (DI) RNA³, suggests that coronavirus RNA polymerase is probably nonprocessive, being able to jump between different RNA regions and RNA molecules during synthesis. This discontinuous nature of coronavirus RNA synthesis appears to contrast with another requisite function of RNA polymerase, i.e., faithful replication of the long RNA genome. These seemingly conflicting demands on coronavirus RNA polymerase suggest that the coronavirus polymerase is very versatile and the mechanism of coronavirus RNA synthesis is unique.

The first macromolecular event after virus entry is the synthesis of a viral RNA polymerase from the incoming viral genome, which is then copied into a (-)-strand RNA. Subsequently, the (-) strand RNA serves as a template for the synthesis of mRNAs and genomic RNA. Thus, the discussion of coronavirus RNA synthesis should cover four separate but intertwined issues:

1. Synthesis, processing and function of RNA polymerase
2. Synthesis of (-)-strand RNA
3. Synthesis of subgenomic mRNAs (transcription)
4. Synthesis of genomic RNA (replication).

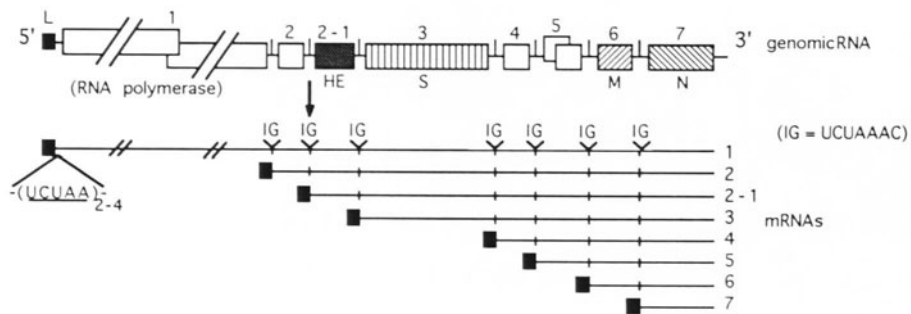


Figure 1. The structure of MHV genomic RNA and subgenomic mRNAs.

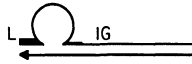
In this review, I will discuss only the last three issues. The studies of the synthesis and processing of polymerases have just begun to come of age, and will be extensively reviewed elsewhere in the book.

MODELS OF CORONAVIRUS SUBGENOMIC mRNA SYNTHESIS

Several potential models for explaining coronavirus subgenomic mRNA synthesis have been proposed, which can be classified into continuous and discontinuous mechanisms (Figure 2). One continuous mechanism proposes that the joining of mRNAs with the leader sequence occurs by a continuous transcription using a looped-out template. This may occur during either (+)- or (-)-strand RNA synthesis. Alternatively, the mRNAs are proposed to be amplified from the virion-packaged subgenomic mRNAs (termed “subgenomic replicons”)⁴ by a faithful continuous transcription. The discontinuous transcription mechanisms include three models: The first is leader-primed transcription, in which the leader RNA is synthesized from the 3'-end of the full-length, (-)-strand RNA template, dissociates from the template, and then rejoins the template at the IG sites to serve as a primer for mRNA transcription⁵. In this model, IG sequences serve as promoter and transcription initiation sites. The second model proposes that discontinuous transcription occurs during (-)-strand RNA synthesis⁶; thus, IG sites serve as transcription terminators. By an unknown mechanism, the (-)-strand, subgenomic transcript then joins the leader RNA and continues to transcribe an antigenomic complement of the leader sequence. The antigenomic, subgenomic RNAs serve as the templates for faithful transcription of the subgenomic mRNAs. The third model is the classical *trans*-splicing mechanism⁷, in which subgenomic mRNAs and leader RNAs are transcribed separately and then joined together by a post-transcriptional *trans*-splicing mechanism. This may occur during either (+)- or (-)-strand RNA synthesis. Although the continuous transcription models cannot be rigorously ruled out, the existing evidence favors the discontinuous models because of recent evidence showing that the subgenomic mRNAs and leader RNAs are derived from two separate RNA molecules^{8,9}. The different discontinuous transcription models, however, have not been clearly distinguished. The two favored models, the leader-primed transcription and discontinuous (-)-strand RNA synthesis models, are both consistent with some of the experimental data. It has thus been proposed that both models might operate during two different stages of transcription, so-called primary transcription and secondary transcription¹⁰. Although these two models differ in the stage of transcription at which discontinuous synthesis occurs, i.e., either during (+)- or (-)-strand RNA synthesis, both models have a common feature, i.e., both require that the IG sequence

CONTINUOUS

- 1) Continuous RNA synthesis on the "looped-out" template

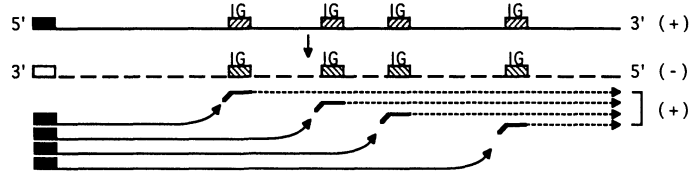


- 2) Amplification of virion-packaged subgenomic mRNAs (subgenomic "replicon")

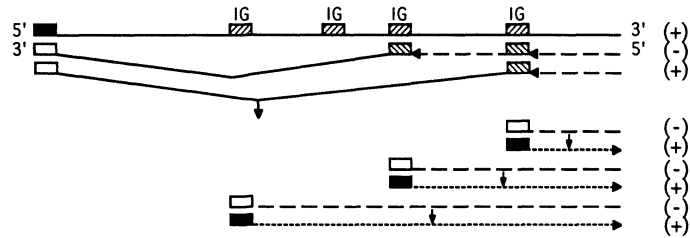


DISCONTINUOUS

- 1) Leader-primed transcription during positive-strand synthesis



- 2) Discontinuous transcription during negative-strand synthesis



- 3) *Trans*-splicing

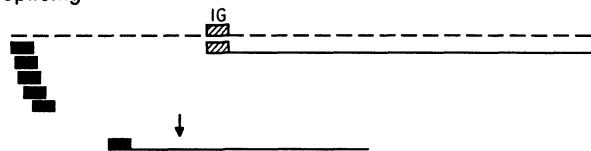


Figure 2. Models of coronavirus subgenomic mRNA transcription.

interacts with the 3'-end of the leader sequence, so that the leader RNA can join the mRNAs at precise sites. These two models cannot yet be unequivocally distinguished.

The following discussion will focus on the regulation of RNA synthesis, including the *trans*- and *cis*-acting regulatory sequences for the synthesis of various RNA species, and the interaction between the leader and IG sequences. These features are critical to coronavirus RNA synthesis regardless of the actual mechanism of RNA transcription.

THE REGULATORY RNA SEQUENCES FOR CORONAVIRUS mRNA TRANSCRIPTION AND GENOMIC RNA REPLICATION

The discontinuous transcription models predict that mRNA transcription is regulated by the leader and IG sequences. The importance of these regulatory elements has been

examined extensively, using a DI RNA as an experimental model, in the past several years^{11,12}. The DI RNA can be considered a mini-genome, because it contains both the 5'- and 3'-ends of the genomic RNA and can replicate autonomously in the presence of a helper virus, which provides RNA polymerases. The replication signals for coronavirus RNA have been determined using these DI RNAs, and localized to approximately 400 nucleotides at both ends^{13,14}. For some DI RNA species, a short stretch of internal sequences is also needed for RNA replication, probably to maintain proper RNA conformation^{13,14}. By inserting an intergenic sequence into these DI RNA vectors, it has been unequivocally established that IG sequences are the necessary regulatory elements for initiating RNA synthesis^{11,12}. As few as seven nucleotides (UCUAAAC) are sufficient¹⁵. More recently published reports further established that the upstream sequences at the 5'-end of the genome are also required for mRNA transcription initiating from the IG sequence¹⁶. These upstream sequences include the leader sequence and some ill-defined sequences located at the 5'-end of the genomic RNAs. The most surprising finding is that the leader sequence is not sufficient. If the 5'-end sequences immediately downstream of the leader are replaced with exogenous sequences or even the sequences of subgenomic mRNAs, no mRNA transcription occurs despite the presence of both the leader and IG sequences¹⁶. These results indicate that certain sequences from the genomic RNAs are required for mRNA transcription, and that only the genomic RNA can be used for mRNA synthesis. These findings thus suggest that mRNAs are not likely the results of amplification of the virion-packaged subgenomic mRNAs⁴. The leader RNA also provides an enhancer-like function for mRNA transcription¹⁶. Thus, the leader RNA has dual functions: providing the leader RNA for mRNAs and enhancing transcription. Most interestingly, recent data indicate unequivocally that the leader RNA of mRNAs is derived from the helper virus RNA, thus representing a *bona fide trans*-acting leader RNA^{8,9,16}. Furthermore, the efficiency of transcription is affected by the nature of the leader sequence not only of the DI RNA, but also of the helper viral RNA⁹. Therefore, the leader RNA can regulate transcription, not only *in cis* but also *in trans*, and the extent of the influence of the leader RNA on transcription varies with the nature of the IG sequence^{9,16}. These studies thus establish that mRNA transcription is regulated by three components: IG, the *cis*-acting leader and a *trans*-acting leader.

THE POSSIBLE MECHANISM OF THE LEADER-IG INTERACTIONS

Because of the presence of homologous sequences between the IG sites and 3'-end of the leader RNA, it was previously proposed that the leader RNA can bind to the intergenic sequence via this stretch of sequences, which are complementary between the (+)-strand leader and the (-)-strand IG⁵. Site-specific mutagenesis of IG sequences indeed caused alteration of the transcription efficiencies from the IG site^{12,17}. However, there is no reasonable correlation between the transcription efficiency and the extent of sequence homology between the IG site and the leader¹⁷. In fact, in most coronaviruses (one exception being MHV), the amounts of subgenomic mRNAs do not correlate with the extents of homology. Therefore, direct RNA-RNA interaction cannot explain satisfactorily the transcriptional regulation of mRNAs. The possibility that the coronaviral transcriptional regulation is mediated by factors other than direct RNA-RNA interaction recently has been suggested by a couple of studies, which demonstrated that the leader-mRNA fusion occurs outside of these complementary sequences^{17,18}. The most dramatic example is the transcription by a murine coronavirus, MHV-2C. This virus transcribes mRNAs which contain unusually aberrant leader-fusion sites; there is no sequence homology between the leader

and these sites¹⁸. These data strongly suggest that direct RNA-RNA interaction may not be the driving force behind coronavirus mRNA transcription.

An alternative possibility is that the transcriptional regulation is mediated by RNA-protein and protein-protein interactions. It is likely that certain cellular or viral proteins bind to the RNA regulatory regions; these proteins then interact with each other to bring the essential RNA components together to form a transcription or replication complex¹⁸. Thus, the selection of the transcriptional initiation sites may be mediated by the specificity of these RNA-binding proteins, which can be considered transcription factors. This model is similar to DNA-dependent RNA transcription, in which transcription is regulated mainly by the interactions of transcription factors. Some of the cellular proteins interacting with the leader RNA and intergenic sequences of both strands have been detected¹⁹. These cellular proteins conceivably are the transcription factors which regulate coronavirus mRNA transcription.

THE REGULATION OF (-)-STRAND RNA SYNTHESIS

The (-)-strand RNA is the template used for the synthesis of the subgenomic mRNAs and genomic RNA. Thus, it controls the synthesis of all the major viral RNA species. Several features of coronavirus (-)-strand RNAs are already known: (1) Both subgenomic- and genomic-sized (-)-strand RNAs are present in the infected cells^{4,6}. (2) (-)-strand RNAs contain both an antigenomic-sense leader RNA at the 3'-end and poly(U) sequences at the 5'-end, thus representing complementary copies of the (+)-strand RNAs^{20,21}. (3) (-)-strand RNA synthesis peaks slightly ahead of (+)-strand RNA synthesis and gradually declines during the viral replication cycle²²; however, the ability to synthesize (-)-strand RNA persists throughout the infection¹⁰. (4) (-)-strand RNAs in the cells are present in the double-strand (ds) RNA forms^{22,23}; thus, the ds RNA may be the actual template for RNA transcription and replication.

This possibility has not been rigorously tested. The regulation of (-)-strand RNA synthesis has not been as well studied as (+)-strand RNA because the amount of (-)-strand RNA is very small in the infected cells. Recently, our laboratory has optimized a ribonuclease protection method and utilized a transfected DI RNA to study the mechanism of the regulation of (-)-strand RNA synthesis²⁴. The results demonstrated that the *cis*-acting signal for (-)-strand RNA synthesis resides entirely within the 55 nucleotides at the 3'-end, plus poly(A) sequences. No specific 5' upstream sequence is required²⁴. This sequence requirement is considerably less stringent than those required for RNA transcription and replication. Most notably, coronavirus RNA replication requires more than 400 nucleotides of virus-specific sequences at the 3'-end^{13,14}. Why RNA replication requires a longer stretch of specific sequences at the 3'-end of the genomic RNA than that required for (-)-strand RNA synthesis is of considerable interest. This finding, therefore, suggests that the 3'-end sequences are required not only for (-)-strand RNA synthesis but also for (+)-strand RNA synthesis. (+)-strand RNA synthesis is initiated from the 3'-end of the (-)-strand template (corresponding to the 5'-end of the genomic RNA), and the signal for RNA replication would be expected to be at the 5'-end of genomic RNA. Thus, the presence of the replication signal at the 3'-end of the genome raises an interesting possibility that (+)-strand RNA synthesis requires the interaction between the 5'- and 3'-end sequences of the template RNA. Furthermore, we have also shown that transcription from an internal IG site of the DI RNA inhibits (-)-strand RNA synthesis of the very RNA, again suggesting that transcription requires certain RNA interactions, which interfere with (-)-strand RNA synthesis²⁴. The interaction between the 5'- and 3'-end sequences has previously been shown to be required for RNA synthesis of some other RNA viruses²⁵. Thus, this interaction may be a common feature of RNA-dependent RNA synthesis.

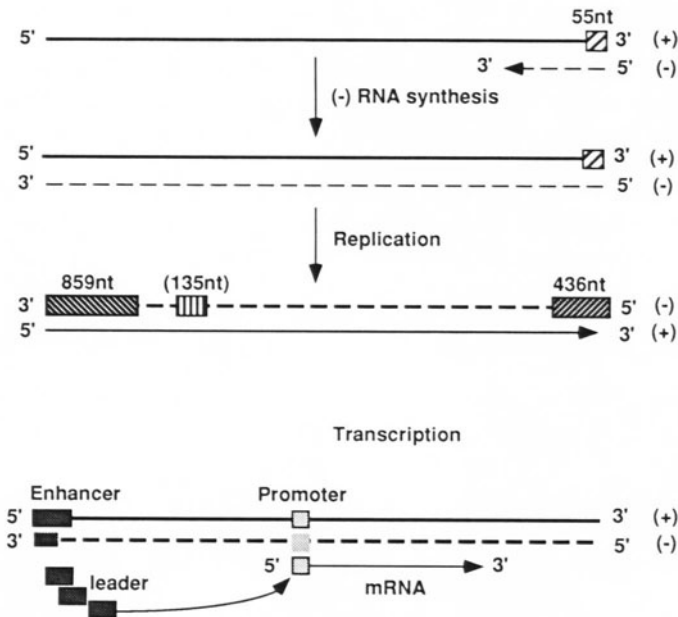


Figure 3. *Cis*-acting signals for (-)-strand RNA synthesis, RNA replication and subgenomic mRNA transcription. The boxed regions are the required sequences for the indicated steps of RNA synthesis. The figure is reproduced from Reference 29, with permission.

Another surprising feature of (-)-strand RNA synthesis is that RNAs which transcribe a subgenomic mRNA make smaller amounts of (-)-strand RNA than nontranscribing RNAs²⁴. Therefore, the amounts of mRNA transcription do not parallel the quantities of (-)-strand RNA templates available, and that the quantitative regulation of RNA synthesis likely occurs mainly at the step of (+)-strand RNA synthesis rather than (-)-strand RNA synthesis. When the *cis*-acting signals for (-)-strand RNA, mRNA and genomic RNA synthesis are compared (Figure 3), it appears that the signal for (-)-strand RNA synthesis is strikingly less complex than that for the other two RNA species. This is conceptually logical, since (-)-strand RNA has to be synthesized directly from the incoming genomic RNA at the time when very few viral proteins have been synthesized. Since the 5'-end sequences of the genomic RNA are not required for (-)-strand RNA synthesis, any RNAs which contain the 55 nucleotides from the 3'-end of the genome and poly(A) will be able to initiate (-)-strand RNA synthesis. This is consistent with the detection of the subgenomic (-)-strand RNA species which correspond to all of the subgenomic mRNAs in the virus-infected cells^{4,6}. In contrast, both RNA replication and transcription require longer stretches of sequences at both the 5'- and 3'-ends (Lin and Lai, unpublished observation), suggesting a complex interaction between various RNA regions.

SUMMARY OF THE MECHANISM OF CORONAVIRUS mRNA SYNTHESIS

Although the precise models of coronavirus mRNA transcription cannot yet be unequivocally deciphered, several features of mRNA synthesis have already been established: (1) mRNA transcription is discontinuous, since the leader sequences on most mRNAs are derived *in trans* from a different RNA molecule^{8,9,16}. (2) mRNA synthesis can be regulated *in trans* by RNA elements located on different RNA molecules⁹. (3) leader-mRNA fusion may not be precise, suggesting that complementary RNA sequences between the

leader and IG sequences are not necessarily directly involved in leader fusion^{17,18}. (4) the RNA regulatory regions are complexed with cellular and, probably, viral proteins¹⁹. These proteins might mediate RNA- and protein-protein interactions to form transcription or replication complexes. (5) The amounts of (-)-strand RNA do not correspond to the amounts of mRNAs or genomic RNAs synthesized, suggesting that the quantitative regulation of RNAs is not at (-)-strand RNA synthesis, but rather at (+)-strand RNA synthesis²⁴. (6) Subgenomic (-)-strand RNAs have been shown to be derived from subgenomic replicative form RNA, indicating that subgenomic (-)-strand RNA can function as a template⁶. However, it is not clear what the predominant mechanism for mRNA transcription is. The most crucial question is at what step, either (+)-strand or (-)-strand synthesis, discontinuous transcription occurs. This issue will require future studies. It is possible that these different mechanisms proposed are not mutually exclusive and may operate simultaneously or at different stages of the viral replication cycle.

RNA RECOMBINATION AND GENETIC ENGINEERING

The studies of RNA viruses in general have been aided tremendously by the availability of infectious viral RNA or cDNA. This approach, unfortunately, has not been available to coronaviruses. The task of achieving such a technical feat is also formidable, because of the extremely large size of the viral RNA. However, progress has been made in the genetic engineering of coronavirus RNA by taking advantage of the fact that coronavirus RNA undergoes a very high frequency of recombination. RNA recombination now has been shown to occur not only in MHV, but also in several other coronaviruses, although at a much lower frequency (see this volume). It can occur between two viral RNAs, and also between viral RNA and a transfected RNA fragment^{26,27} or defective-interfering RNA²⁸. This capacity to undergo RNA recombination has been harnessed to replace the 5'- and 3'-ends of viral genomic RNA with desired sequence^{27,28}. However, the success of this approach so far has been limited to both ends of the genomic RNA; it has not yet been possible to replace the internal sequences or insert foreign genes, probably because multiple cross-overs are required. We have recently developed a DI RNA vector for expressing foreign genes. This DI vector is made to contain either an internal ribosomal entry site (IRES)¹³ or IG sequence¹⁶. When a viral gene is inserted behind either IG or IRES sequences, it can be expressed, and the protein product can be incorporated into the virions, thus generating a pseudotyped virus (Liao and Lai, unpublished). Furthermore, DI RNA can be incorporated into virions, thus generating a pseudo-recombinant virus, which contains both the wild-type viral genome and a defective-interfering RNA containing a foreign gene. This represents a first step toward the genetic engineering of coronavirus. This system can potentially be used to express a foreign gene of interest; for instance, it can be used to express cytokines or cytokine antagonists to study the effects of these molecules on viral infections. This is particularly powerful, since these molecules will be expressed only in virus-infected cells. It can be used to express spike proteins of different viruses and, thereby, alter the host range of the virus. Finally, it can be used to express a desired protein to induce antibodies, thus serving as a virus vaccine. Therefore, even though an infectious full-length cDNA clone is not yet available, the era of genetic engineering of coronaviruses may be at hand.

CONCLUSIONS

The large size of coronavirus RNA genome and the discontinuous nature of its mRNA synthesis are the unique features of coronavirus. Although the detailed mechanism of RNA

synthesis has not yet been unequivocally established, the presently available evidence suggests that it has many interesting features, enabling coronavirus RNA to serve as a paradigm for the studies of other viral RNAs. Progress in the studies of RNA synthesis also has made possible the genetic engineering of coronavirus. However, many unanswered questions remain, e.g., what is the nature of coronavirus RNA polymerase? What are the transcription factors? How are the various steps of RNA synthesis regulated? What is the nature of the templates for RNA synthesis? These and other questions will demand further studies in the future.

REFERENCES

- 1 Lai M M C Coronavirus Organization, Replication and Expression of Genome *Ann Rev Microb* 1990,44 303-333
- 2 Makino S , Keck J G , Stohlman S A , Lai M M C High-frequency RNA recombination of murine coronaviruses *J Virol* 1986,57 729-737
- 3 Makino S , Fujioka N , Fujiwara K Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus *J Virol* 1985,54 329-36
- 4 Sethna P B , Hung S L , Brian D A Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons *Proc Natl Acad Sci USA* 1989,86 5626-5630
- 5 Lai M M C Coronavirus leader RNA-primed transcription An alternative mechanism to RNA splicing *BioEssays* 1989,5 257-260
- 6 Sawicki S G , Sawicki D L Coronavirus transcription Subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis *Proc Natl Acad Sci USA* 1990,64 1050-1056
- 7 Sutton R E , Boothroyd J C Evidence for trans splicing in trypanosome *Cell* 1986,47 527-535
- 8 Jeong Y S , Makino S Evidence for coronavirus discontinuous transcription *J Virol* 1994,68 2615-2623
- 9 Zhang X , Lai M M C Coronavirus leader RNA regulates and initiates subgenomic mRNA transcription, both in trans and in cis *J Virol* 1994,68 4738-4746
- 10 Jeong Y S , Makino S Mechanism of coronavirus transcription Duration of primary transcription initiation activity and effects of subgenomic RNA transcription on RNA replication *J Virol* 1992,66 3339-3346
- 11 Makino S , Joo M , Makino J K A system for study of coronavirus mRNA synthesis A regulated, expressed subgenomic defective-interfering RNA results from intergenic site insertion *J Virol* 1991,65 6031-6041
- 12 Joo M , Makino S Mutagenic analysis of the coronavirus intergenic consensus sequence *J Virol* 1992,66 6330-6337
- 13 Lin Y-J , Lai M M C Deletion mapping of a mouse hepatitis virus defective-interfering RNA reveals the requirement of an internal and discontinuous sequence for replication *J Virol* 1993,67 6110-6118
- 14 Kim Y-N , Jeong Y S , Makino S Analysis of cis-acting sequences essential for coronavirus defective interfering RNA replication *Virology* 1993,197 53-63
- 15 Makino S , Joo M Effect of intergenic consensus sequence flanking sequences on coronavirus transcription *J Virol* 1993,67 3304-3311
- 16 Liao C -L , Lai M M C Requirement of the 5'-end genomic sequence as an upstream cis-acting element for coronavirus subgenomic mRNA transcription *J Virol* 1994,68 4727-4737
- 17 Van der Most R G , de Groot R J , Spaan W J M Subgenomic RNA synthesis directed by a synthetic defective interfering RNA of mouse hepatitis virus A study of coronavirus transcription initiation *J Virol* 1994,68 3656-3666
- 18 Zhang X , Lai M M C Unusual heterogeneity of leader-mRNA fusion in a murine coronavirus Implications for the mechanism of RNA transcription and recombination *J Virol* 1994,68 6626-6633
- 19 Furuya T , Lai M M C Three different cellular proteins bind to the complementary sites on the 5'-end positive- and 3'-end negative-strands of mouse hepatitis virus RNA *J Virol* 1993,67 7215-7222
- 20 Hofmann M A , Brian D A The 5'-end of coronavirus minus-strand RNAs contains a short poly(U) tract *J Virol* 1991,65 6331-6333
- 21 Sethna P B , Hofmann M A , Brian D A Minus-strand copies of replicating coronavirus mRNAs contain antileaders *J Virol* 1991,65 320-325
- 22 Sawicki S G , Sawicki D L Coronavirus minus-strand RNA synthesis and effect of cycloheximide on coronavirus RNA synthesis *J Virol* 1986,57 328-334

- 23 Perlman S , Ries D , Bolger E , Chang L J , Stoltzfus C M MHV nucleocapsid synthesis in the presence of cyclohexamide and accumulation of negative-strand MHV RNA *Virus Res* 1986,6 261-272
- 24 Lin Y-J , Liao C L , Lai M M C Identification of the cis-acting signal for minus-strand RNA synthesis of a murine coronavirus Implications for the role of minus-strand RNA in RNA replication and transcription *J Virol* 1994,(In press)
- 25 Luo G , Luytjes W , Enami M , Palese P The polyadenylation signal of influenza virus RNA involves a stretch of uridines followed by the RNA duplex of the panhandle structure *J Virol* 1991,65 2861-2867
- 26 Liao C -L , Lai M M C RNA recombination in a coronavirus Recombination between viral genomic RNA and transfected RNA fragments *J Virol* 1992,66 6117-6124
- 27 Koetzner C A , Parker M M , Ricard C S , Sturman L S , Masters P S Repair and mutagenesis of the genome of a deletion mutant of the coronavirus mouse hepatitis virus by targeted RNA recombination *J Virol* 1992,66 1841-1848
- 28 Van der Most R G , Heijnen L , Spaan W J M , de Groot R J Homologous RNA recombination allows efficient introduction of site-specific mutations into the genome of coronavirus MHV-A59 via synthetic co-replicating RNA *Nucl Acids Res* 1992,20 3375-3381
- 29 Lai M M C , Liao C -L , Lin Y-J , Zhang X Coronavirus How a large RNA viral genome is replicated and transcribed *Infect Agents and Dis* 1994,3 98-105