

A NEW MODEL FOR CORONAVIRUS TRANSCRIPTION

S. G. Sawicki and D. L. Sawicki

Department of Microbiology and Immunology
Medical College of Ohio
Toledo, Ohio 43699

1. ABSTRACT

Coronaviruses contain an unusually long (27–32,000 ribonucleotide) positive sense RNA genome that is polyadenylated at the 3' end and capped at the 5' end. In addition to the genome, infected cells contain subgenomic mRNAs that form a 3' co-terminal nested set with the genome. In addition to their common 3' ends, the genome and the subgenomic mRNAs contain an identical 5' leader sequence. The transcription mechanism that coronaviruses use to produce subgenomic mRNA is not known and has been the subject of speculation since sequencing of the subgenomic mRNAs showed they must arise by discontinuous transcription. The current model called leader-primed transcription has subgenomic mRNAs transcribed directly from genome-length negative strands. It was based on the failure to find in coronavirus infected cells subgenome-length negative strands or replication intermediates containing subgenome-length negative strands. Clearly, these structures exist in infected cells and are transcriptionally active. We proposed a new model for coronavirus transcription which we called 3' discontinuous extension of negative strands. This model predicts that subgenome-length negative strands would be derived directly by transcription using the genome RNA as a template. The subgenome-length templates would contain the common 5' leader sequence and serve as templates for the production of subgenomic mRNAs. Our findings include showing that: 1. Replication intermediates (RIs) containing subgenome-length RNA exist in infected cells and are separable from RIs with genome-length templates. The RIs with subgenome-length templates are not derived by RNase treatment of RIs with genome-length templates. 2. The subgenome-length negative strands are formed early in infection when RIs are accumulating and the rate of viral RNA synthesis is increasing exponentially. 3. Subgenome-length negative strands contain at their 3' ends a complementary copy of the 72 nucleotide leader RNA that is found in the genome only at their 5' end. 4. RIs with subgenomic templates

serve immediately as templates for transcription of subgenomic mRNAs. Because subgenomic mRNAs are not replicated, i.e., copied into negative strands that in turn are used as templates for subgenomic mRNA synthesis, we propose that the subgenome-length negative strands must arise directly by transcription of the genome and acquire their common 3' anti-leader sequence after polymerase jumping from the intergenic regions to the leader sequence at the 5' end of the genome. This would make negative strand synthesis discontinuous and subgenomic mRNA synthesis continuous, which is the opposite of what was proposed in the leader primed model.

2. RESULTS AND DISCUSSION

At the VIth International Symposium on Corona and Related Viruses in Quebec, Canada, we first presented a new model for coronavirus transcription, which we called discontinuous extension of negative strands, to explain how subgenome-length negative strands would be derived directly from the genome (Sawicki and Sawicki, 1995). A new model was needed because the current model called leader-primed transcription (see review by Lai, 1990) was based on data which claimed to demonstrate that only genome-length negative strands existed in coronavirus infected cells (Lai, et al. 1982; Baric et al., 1983). Therefore, subgenomic mRNAs had to be derived from genome-length negative strands. The seminal observation made in David Brian's laboratory (Sethna et al., 1989) that transmissible gastroenteritis virus (TGEV) infected cells contained subgenome-length negative strands invalidated the premise on which the leader-primed model was based, mainly that only genome length-negative strands existed in coronavirus infected cells. Shortly afterward we demonstrated (Sawicki and Sawicki, 1990) that RIs containing subgenome-length negative strands existed and were transcriptionally active in mouse hepatitis virus (MHV) infected cells.

Because subgenomic mRNAs are not themselves able to be copied directly into negative strands (Brian et al., 1994; Makino et al., 1991; Masters et al., 1994), we suggested (Sawicki and Sawicki, 1995) that subgenome-length negative strands would be generated directly from the genome. We are able to separate RIs with subgenome-length templates from RIs with genome length templates, either by velocity gradient centrifugation on sucrose gradients or by agarose gel electrophoresis. The RIs with subgenomic templates incorporate ³H-uridine into positive and negative strands with the same kinetics as RIs with genome-length templates. The RIs with subgenome-length templates were formed quickly after infection and were detectable by 1.5 -2 hours post infection; but we could not determine whether they were formed directly from transcription of the genome or if they were derived by copying subgenomic mRNAs. The latter possibility seems unlikely based on the experiments cited above that subgenomic mRNAs are not replicated when transfected into cells. Sequences downstream of the leader are required for replication. Therefore, we concluded that subgenome-length negative strand templates arose directly from copying the genome RNA and derived their 3' ends, which are complementary to the leader sequence at the 5' end of the genome, by discontinuous transcription.

As we proposed three years ago the transcription apparatus of coronaviruses shows similarities to the E.coli RNA polymerase which transcribes DNA in a discontinuous fashion (Komissarova and Kashlev, 1997; Nudler et al., 1997). Nudler et al. (1997) stated that the basic features of the ternary complex of RNA polymerase are conserved in all living organisms but its detailed structure is unknown. We suggest that the coronavirus RNA polymerase also has an inherent capacity to transcribe its RNA template monotonously but

the intergenic sequences (IS elements) in the genome are sites of discontinuous elongation. Pausing at these sites would allow for detachment of the polymerase with its nascent negative strand. The polymerase with its nascent negative strand would reattach in cis or in trans at the same site or translocate specifically to the 5' end of the genome. The movement of the following polymerase may cover the site of detachment and force the detached polymerase to reattach at the 5' end of the genome. We predict that the genome would have multiple polymerase molecules copying it into negative strands. One possibility is that the initial or leading polymerase would be committed to produce a genome-length negative strand template and, thereby, ensure replication of the genome and the survival of the virus. Following polymerases would be subject to pausing. Loading of the genome with polymerases would force pausing of the following polymerase by the leading polymerase and account for the relationship of the relative abundance of the smaller compared to the larger subgenomic mRNAs with their position relative to the 3' of the genome.

The generation of RIs with subgenome-length templates would result from polymerase reattachment at the 5' end of the genome, either the same genome (cis) or another genome (trans). The high rate of recombination observed in coronaviruses would result from reattachment in trans to another genome but at the same site on the genome from which the polymerase had detached. Reattachment at the 5' end of the genome would be facilitated, as we suggested, by protein-protein interaction between the polymerase and a protein bound between the leader sequence at the 5' end of the genome and first open reading frame, which would be a region necessary for the genome to serve as a template for negative strand synthesis. The report by Chang et al. (1996) demonstrated that leader conversion of defective interfering (DI) RNA of bovine coronavirus was not guided by the IS element at the 5' end of the DI RNA but by a recombination event during negative strand synthesis. In their case the polymerase detached during negative strand synthesis before copying the leader sequence at the 5' end of the DI RNA and reattached in trans at the 5' end of the helper virus genome but downstream of the leader sequence. The recombinant DI RNA possessed the leader sequence including the IS element of the helper virus genome. Since the DI RNA did not contain an IS element at its 5' end the polymerase paused and detached up to 25 nucleotides before the leader sequence in the DI RNA. In this case the sequence downstream of the leader sequence in the DI RNA and in the helper virus were nearly identical. This allowed the detached negative strand to reinitiate RNA synthesis downstream of the IS element at the 5' end of the helper virus genome. In the case of generating subgenome-length templates, only the IS elements in the intergenic regions of the genome would be homologous with the IS element at the 5' end of the genome. Therefore, if the polymerase detaches at a site other than the IS elements when copying the genome it cannot easily elongate and complete negative strand synthesis at the 5' end of the genome because of base pair mismatch. Only when there are matching base pairs at the 3' end of the the detached nascent negative strand and the genome template will elongation of the nascent negative strands occur. This would also explain the heterogeneity of leader-body fusion sites on subgenomic RNA (see van der Most et al., 1994).

The IS elements in the genome of coronaviruses can be viewed as analogous to the intergenic dinucleotide and transcription start sequence in the genome of vesicular stomatitis virus, which are required for the efficient termination and initiation of transcription (Stillman and Whitt, 1997). It appears that like DNA-dependent RNA polymerases, RNA-dependent RNA polymerases transcribe a template in monotonous or in discontinuous fashion. The structural features of a template that determine discontinuous transcription remain to be elucidated.

At the last Symposium on Corona and Related viruses we stated that to our knowledge there is no experimental data that disproves our model. This remains true today. Although the details of the mechanism of discontinuous synthesis of negative strands remain a mystery, we can summarize the state of our understanding of coronavirus transcription as follows:

1. During the early phase of coronavirus infection, negative strand synthesis occurs coincidentally with positive strand synthesis.
2. Negative strand synthesis requires continued protein synthesis and is immediately halted by the addition of inhibitors of translation.
3. The rate of positive strand synthesis is determined by the number of negative strand templates.
4. Positive strand synthesis, in contrast to negative strand synthesis, is more stable and continues unabated for about 30 min. after inhibition of translation and then declines slowly. This suggests that a protein factor is required for the activity of the coronavirus positive strand RNA polymerase. There appears to be a pool of this factor sufficient to sustain the positive strand polymerase activity for 30 min.; and, if not replenished by protein synthesis, positive strand synthesis declines.
5. The decline in negative strand synthesis at about 5 hours post infection at 37°C in 17Cl-1 cells infected with the A59 strain of mouse hepatitis virus is correlated with leveling of positive strand synthesis. Late in the infection cycle, the rate of positive strand synthesis gradually declines.
6. The template for subgenome-length negative strands is the genome. The subgenome-length negative strands are produced by discontinuous transcription and are found in partially double stranded RIs, not as single stranded RNAs.
7. At all times post infection RIs with genome- and subgenome-length templates are actively producing genomes and subgenomic mRNAs.
8. The ratio of genomes to subgenomic mRNAs produced throughout infection does not vary and is determined by the ratio of RIs with genome- and subgenome-length templates.

REFERENCES

- Baric, R. S., Stohman, 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.
- Brian, D.A., Chang, R.-Y., Hofmann, M.A., Sethna, P.B., 1994, Role of subgenomic minus-strand RNA in coronavirus replication, *Arch. Virol.* (Suppl.) **9**:173–180.
- Chang, R.Y., Krishnan, R., and Brian, D.A., 1996, The UCUAAC promoter motif is not required for high frequency leader recombination in bovine coronavirus defective interfering RNA, *J. Virol.* **70**:2720–2729.
- Komissarova, N., and Kashlev, M., 1997, Transcription arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded, *Proc. Natl. Acad. Sci. USA* **94**:1755–1760.
- Lai, M. M. C., Patton, C. D., and Stohman, S. A., 1982, Replication of mouse hepatitis virus: negative-stranded RNA and replicative form RNA are of genome length, *J. Virol.* **44**:487–492.
- Lai, M.M.C., 1990, Coronavirus-organization, replication and expression of genome, *Annu. Rev. Microbiol.* **44**:303–333.
- Makino, S., Joo, M., and Makino, J.K., 1991, A system for study of coronavirus mRNA synthesis: a regulated, expressed subgenomic defective interfering RNA results from intergenic site insertion, *J. Virol.* **65**:6031–6041.
- Masters, P.S., Koetzner, C.A., Kerr, C.A., and Heo, Y., 1994, Optimization of targeted RNA recombination and mapping of a novel nucleocapsid gene mutation in the coronavirus mouse hepatitis, *J. Virol.* **68**:328–337.

- Nudler, E., Mustaev, A., Lukhtanov, E., and Goldfarb, A., 1997, The RNA-DNA hybrid maintains the register of transcription by preventing backtracking on RNA polymerase, *Cell*. **89**:33–41.
- Sawicki, S.G, and Sawicki, D.L., 1990. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis, *J. Virol.* **64**:1050–1056.
- Sawicki, S.G. and Sawicki, D.L., 1995, Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands, in: *Corona and Related Viruses* (P.J. Talbot and G.A. Levy, eds) Plenum Press, New York, pp. 499–505.
- Sethna, P.B., Hung, S-L., and Brian, D.A., 1989, Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons, *Proc. Natl. Acad. Sci. USA* **86**:5626–5630.
- Stillman, E.A., and Whitt, M.A., 1997, Mutational analysis of the intergenic dinucleotide and the transcriptional start sequence of vesicular stomatitis virus (VSV) define sequences required for efficient termination and initiation of VSV transcripts, *J. Virol.* **71**:2127–2137.
- van der Most, R.G., de Groot, R.J., and Spaan, W.J.W., 1994, Subgenomic RNA synthesis directed by a synthetic defective interfering RNA of mouse hepatitis virus: a study of coronavirus transcription initiation, *J. Virol.* **68**:3656–3666.