Model for Polymerase Access to the Overlapped L Gene of Respiratory Syncytial Virus

RACHEL FEARNS AND PETER L. COLLINS*

Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892-0720

Received 21 April 1998/Accepted 23 September 1998

The last two genes of respiratory syncytial virus (RSV), M2 and L, overlap by 68 nucleotides, an arrangement which has counterparts in a number of nonsegmented negative-strand RNA viruses. Thus, the gene-end (GE) signal of M2 lies downstream of the L gene-start (GS) signal, separated by 45 nucleotides. Since RSV transcription ostensibly is sequential and unidirectional from a single promoter within the 3’ leader region, it was unclear how the polymerase accesses the L GS signal. Furthermore, it was previously shown that 90% of transcripts which are initiated at the L GS signal are polyadenylated and terminated at the M2 GE signal, yielding a short, truncated L mRNA as the major transcription product of the L gene. Despite these apparent down-regulatory features, we show that the accumulation of full-length L mRNA during RSV infection is only sixfold less than that of its upstream neighbor, M2. We used cDNA-encoded genome analogs in an intracellular transcription assay to investigate the mechanism of transcription of the overlapped genes. Expression of L was found to be dependent on sequential transcription from the 3’ end of the genome. Apart from the L GS signal, the only other strict requirement for initiation at L was the M2 GE signal. This implies that the polymerase accesses the L GS signal only following arrival at the M2 GE signal. Thus, polymerase which terminates at the M2 GE signal presumably scans upstream to initiate at the L GS signal. This also would provide a mechanism whereby polymerase which terminates prematurely during transcription of L could recycle from the M2 GE signal to the L GS signal, thereby accounting for the unexpectedly high level of synthesis of full-length L mRNA. The sequence and spacing between the two signals were not critical. Furthermore, the polymerase also was capable of efficiently transcribing from an L GS signal placed downstream of the M2 GE signal, implying that the overlapping arrangement is not obligatory. When copies of the L GS signal were placed concurrently upstream and downstream of the M2 GE signal, both were utilized. This finding indicates that a polymerase situated at a GE signal is capable of scanning for a GS signal in either the upstream or downstream direction and thereafter initiating transcription.

Human respiratory syncytial virus (RSV) is the leading viral agent of pediatric respiratory tract disease worldwide. It is a pneumovirus of the paramyxovirus family of the order Mononegavirales, the nonsegmented negative-strand RNA viruses, and has a genome of 15,222 nucleotides (nt) (9, 25).

The genome of the prototype nonsegmented negative-strand RNA virus, exemplified by the rhabdovirus vesicular stomatitis virus (VSV) and the paramyxovirus Sendai virus, is tightly encapsidated with the nucleocapsid (N) protein and associated with the phosphoprotein P and the polymerase protein L, which together comprise the minimum RNA-dependent RNA polymerase (for reviews, see references 22 and 33). The encapsidated genome is the template for both transcription and RNA replication. During RNA replication, the polymerase synthesizes a complete encapsidated positive-sense copy of the genome, the antigenome, which in turn serves as the template for synthesis of progeny genomes. During transcription, the polymerase initiates at a single promoter at the 3’-NS1-NS2-N-P-M-SH-G-F-M2-L-5’ end of the genome. Apart from the L GS signal, the gene-end (GE) signal directs polyadenylation and termination at the mRNA 3’ end (3, 17, 28). Between the genes lie short intergenic regions which are not transcribed but which can, as in the case of VSV, play a role in transcription (4, 28). There is a gradient of mRNA abundance, with mRNAs derived from the 3’ end of the genome being more abundant that those copied from the 5’ end. This is due to attenuation of transcription, which for the first four genes of VSV was shown to occur primarily at the gene junctions (18) and might involve polymerase dissociation during polyadenylation or failure to reinitiate. In the case of RSV, studies using a dicistronic genome analog indicated that the degree of attenuation across a single gene junction was such that transcription of the downstream gene was approximately one-third of that of the upstream one (19).

Although RSV is similar in pattern of replication and transcription to the prototypic nonsegmented negative-strand RNA viruses, it differs in several details (for a review, see reference 9). First, it encodes more mRNAs and proteins (10 and 11, respectively), with the RSV proteins as follows: the above-mentioned N, P, and L proteins; the fusion (F), attachment (G), and small hydrophobic (SH) transmembrane surface proteins; the internal matrix (M) protein; the nonstructural NS1 and NS2 proteins; and two proteins encoded by separate overlapping open reading frames (ORFs) in the M2 mRNA, namely, the M2-1 and M2-2 proteins. The viral gene order is 3’-NS1-NS2-N-P-M-SH-G-F-M2-L-5’. Another difference is that the viral transcriptase requires the M2-1 protein in addition to N, P, and L for fully processive transcription (10).
third difference is that the RSV intergenic regions are noncon-
served in length and sequence and do not play a demonstrable
role in transcription (7, 19). Finally, the last two RSV genes,
M2 and L, overlap by 68 nt rather than being separated by an
intergenic region (8) (Fig. 1).

UV mapping studies showed that transcription of the first
nine RSV genes, at least, is sequential (11). Structure-function
studies with cDNA-encoded minigenomes provided indepen-
dent confirmation that transcription of a downstream RSV
gene is dependent on transcription and termination at the GE
signal of its upstream neighbor (20). Since the L GS signal lies
upstream rather than downstream of the GE signal of its up-
stream neighbor, M2, it is unclear how the RSV polymerase
accesses the L gene. In addition, polymerases which do initiate
at the L GS signal will transcribe only a short region of L
before encountering the M2 GE signal. We previously showed
that this results in polyadenylation and termination approxi-
ately 90% of the time, such that only 10% of transcription
starts have the potential to yield full-length L mRNA (8).

There are several possibilities which could explain how the
polymerase accesses the L GS signal. First, an alternative ver-
ion of the sequential transcription model which is consistent
with the previously available data postulates that polymerase
molecules bind to GS signal sequences along the length of the
genome but do not initiate transcription until activated by a
polymerase molecule completing transcription of the preced-
ing gene (30). If this “cascade” model is correct, the gene
overlap would pose no bar at all to polymerase entry at L, and
a polymerase transcribing the M2 gene could trigger a poly-
merase at the L GS signal to initiate transcription. Alterna-
tively, there could be an internal promoter located close to the
L GS signal which enables L to be transcribed independently of
the rest of the genome. A third possibility is that a fraction of
polymerase either fails to initiate at the M2 GS signal or
terminates prematurely within M2, and slides through M2
scanning for a GS signal. A similar “sliding” model has been
suggested to explain transcription of a gene lying downstream
of a nonfunctional GE signal in the paramyxovirus simian virus
41 (31). Finally, it is possible that although the polymerase
transcribes in a unidirectional manner, it has sufficient flexibil-
ity to terminate at the M2 GE signal and then to reach back
and reinitiate synthesis at the L GS signal.

Other viruses of the order Mononegavirales have subse-
quently been shown to have overlapping genes. Marburg virus
has one and Ebola virus has three sets of overlapping GS and
GE signals (5, 13, 26), a number of rhabdoviruses have been
shown to have gene overlaps which range from 21 to 33 nt (23,
29, 34), and Borna disease virus has an 18-nt overlap between
its p40 (N) and p23 (P) genes (27). As gene overlaps are found
in a number of nonsegmented negative-strand viruses, it is
possible that there is a common feature of negative-strand
virus transcription which enables the polymerase to accommo-
date these junctions.

In this study, we used a reconstituted intracellular transcrip-
tion assay to investigate the mechanism by which the RSV
polymerase accesses the L GS signal. We determined that L
is transcribed sequentially following transcription of the up-
stream genes. The data that we obtained are consistent only
with the last model proposed above, in which the polymerase
reaches backward to access the L GS signal following termina-
tion at the M2 GE signal.

MATERIALS AND METHODS

Protein expression plasmids. pTM1 plasmids containing the ORFs of the N,
P, M2-1, and L proteins under the transcriptional control of the promoter for T7
RNA polymerase and the translational control of the internal ribosome entry site
of encephalomyocarditis virus were constructed in previous work (10, 15).

Minigenome cDNAs. Plasmids encoding the minigenome RNAs used in this
study were constructed by using an intermediate plasmid, C65R. C65R is derived
from Rl4C (20) and encodes a minigenome with authentic 3’ and 5’ RSV ter-
mini which flank a negative-sense copy of chloramphenicoltransferase (CAT)
ORF which has a GS signal at its upstream end. This is followed in turn by a
negative-sense copy of the luciferase (LUC) ORF. This ORF was
modified such that 661 nt were removed and replaced by a second copy of the
CAT ORF. This chimeric CAT/LUC sequence has a GE signal at its downstream
end. The encoded minigenome is bordered at its 5’ end by the T7 RNA poly-
merase promoter and at its 3’ end with a self-cleaving hammerhead ribozyme.

The structure of each minigenome is diagrammed in the relevant figure and
further described in Results and the figure legends. To construct plasmids A, F,
G, H, I, and Q, PCR-generated DNA fragments spanning various lengths of
the M2/L gene junction region were inserted into the SalI and BglII sites of C65R.
Mutagenic PCR primers were used to alter the positioning of the L GS signal in
minigenomes O and P or to generate an additional L GS signal, minigenome S.
To delete the L GS signal to generate minigenomes B and R, plasmids A and S,
respectively, were digested with SspI and BamHI (there is a naturally occurring
BamHI site immediately following the L GS signal), and the recessed 3’ termini
were filled in with Klenow DNA polymerase and then religated. To increase the
length of the overlap region for the construction of minigenome N, compen-
satory oligonucleotides encoding the L GS signal and 22 nt of M2-specific sequence
were inserted into the SalI and BamHI sites of C65R. DNA minigenome cDNAs.

Plasmids encoding minigenomes C and D were generated by using a PCR-
based mutagenesis method (6), using plasmid A as a template. To delete M2
sequence downstream of the L GS signal to generate minigenome E, the 500-nt M gene
PCR fragment described above was inserted into the BamHI/BssHII sites of
plasmid A which had not yet received M sequence.

Another intermediate plasmid, J (-M), was constructed to facilitate mutagen-
esis of the overlap. To generate J (-M), a PCR-generated fragment encoding the
68-nt M2/L overlap was inserted into the SalI/BglII sites of C65R. Then a PCR
fragment encoding the first 100 nt of L-specific sequence was inserted into the
SalI and BglII sites of C65R. The final stage in the construction of each of these
cDNAs was to increase the size of the CAT/LUC gene by the purpose of improving
the electropherographic properties of the T7/LUC mRNA (see Results for
explanation). A 500-nt fragment of the M gene of RSV was amplified by PCR using
M-specific oligonucleotides flanked with BamHI and BsrWI restriction sites and
inserted into the BglII/BsrWI restriction sites of each minigenome cDNA.

Plasmids encoding minigenomes C and D were generated by using a PCR-
based mutagenesis method (6), using plasmid A as a template. To delete M2
sequence downstream of the L GS signal to generate minigenome E, the 500-nt M gene
PCR fragment described above was inserted into the BamHI/BssHII sites of
plasmid A which had not yet received M sequence.

Another intermediate plasmid, J (-M), was constructed to facilitate mutagen-
esis of the overlap. To generate J (-M), a PCR-generated fragment encoding the
68-nt M2/L overlap was inserted into the SalI/BglII sites of C65R. Then a PCR
fragment encoding the first 100 nt of L-specific sequence was inserted into the
BglII and BssHII window of J (-M). These plasmids were modified by insertion of a 500-nt M gene PCR fragment similar to that
described above but flanked at each end with a BsrWI restriction site. This
fragment was inserted into the BsrWI site of each plasmid, and its orientation was
verified to that it was consistent for all minigenomes.

In the case of the dicistronic minigenomes, the complete M2/L specific se-
quence of each plasmid was confirmed by sequencing. For plasmids encoding
dicistronic minigenomes, the F/M2 gene junction and 60 nt flanking either side and
the M2/L gene junction and 100 nt flanking either side were sequenced.

Transfections. Monolayers of Hep-2 cells in six-well dishes were transfect-
d with the following mixture of plasmids per single well of a six-well dish: 0.2 μg of
minigenome DNA, 0.4 μg of pTM1 N, 0.2 μg of pTM1 P, 0.1 μg of pTM1 L, and
0.1 μg of pTM1 M2-1. The cells were simultaneously transfected with plasmids
and infected at 10 PFU per cell with vaccinia virus VTF7-3 (provided by Thomas
Fuerst and Bernard Moss), which expresses the T7 RNA polymerase (14), as

FIG. 1. Structure of the M2/L overlap. (A) Diagram (not to scale) of the
portion of the RSV genome containing the M2/L overlap. GS or GE signals are
represented by open or filled boxes, respectively, and the F, M2, and L genes are
indicated with brackets. (B) Sequence of the M2/L overlap shown as negative-
sense RNA. The L GS and M2 GE signals are within boxes. The UAC which
specifies the initiating methionine of the L ORF is indicated in bold text.
fetal bovine serum was added, and the cells were incubated for a further 24 h. After 22 to 24 h, the transfection-infection mixture was replaced with OptiMem containing 2% fetal bovine serum and actinomycin D at 2 μg/ml. The actinomycin-D-containing medium was removed after 2 h, fresh OptiMem containing 2% fetal bovine serum was added, and the cells were incubated for a further 24 h.

RSV infection time course. Monolayers of HEp-2 cells in 25-cm² flasks were infected with RSV (A2) at a multiplicity of infection of 4 PFU/cell. Following 1 h of adsorption, the virus inoculum was removed, the cell monolayer was washed, and OptiMem containing 2% fetal bovine serum was added. Cells for the 0-h time point were harvested at this time. The remaining flasks were incubated at 37°C, and cells were harvested at 3-h intervals.

RNA isolation, oligo(dT) chromatography, and Northern blot hybridization. Total intracellular RNA was extracted from cell pellets by using Trizol reagent (Life Technologies) according to the supplier’s protocol except that the RNAs were extracted with phenol-chloroform following the isopropanol precipitation. Poly(A)+ RNAs were purified from approximately 50 μg of total cellular RNA by oligo(dT) chromatography using a minibatch method (15). Approximately 4 μg of total RNA sample (Fig. 8) or one-third of oligo(dT)-purified mRNA sample (Fig. 2 to 7) was analyzed by electrophoresis in a 1.5% agarose gel containing 0.44 M formaldehyde, transferred to nitrocellulose (Schleicher & Schuell), and fixed by UV cross-linking (Stratagene). The blots shown in Fig. 2 to 7 were probed with negative-sense CAT- or LUC-specific riboprobes as indicated. The blot shown in Fig. 8 was probed with a negative-sense 5‘-labeled oligonucleotide which binds to the region between the L GS and M2 GE signals of the M2/L overlap. Quantitation was carried out with a PhosphorImager 445SI (Molecular Dynamics).

RESULTS

Reconstituting transcription of an overlapped gene in a minigenome. The overlap between the M2 and L genes is shown in Fig. 1. Several models that might account for expression of the overlapped L gene were described in the introduction. We sought to distinguish between these models by recreating the overlap in a minigenome and investigating its expression by mutational analysis. To examine transcription from the re-created overlapping gene junction, we used a reconstituted transcription-replication system which has been described previously (10, 15). In this system, a plasmid-encoded analog of negative-sense genomic RNA, consisting of a reporter gene (or genes) flanked by the downstream two-thirds of the LUC ORF, is synthesized intracellularly in the presence of RSV polymerase proteins which are encoded by cotransfected support plasmids. Plasmid expression is mediated by T7 RNA polymerase supplied by the vaccinia virus recombinant vTF7-3. Minigenome A was constructed to simulate the overlapped genes. It contains three genes, in 5‘-to-5‘ order: (i) the CAT ORF fused to the last 55 nt of the F gene, followed by the authentic F-M2 intergenic region, (ii) the complete M2 gene (including the M2/L overlapping region), and (iii) 100 nt of L-specific sequence fused to 500 nt of RSV M-gene sequence fused in turn to a second copy of CAT fused to the downstream two-thirds of the LUC ORF. This last gene is the surrogate for the L gene and will be referred to as M/CAT/LUC. Each of these genes is flanked by an independent set of GS and GE signals (Fig. 2A). The purpose of this construct was to place the 68-nt M2/L gene overlap in its authentic context preceded by sequence including the F/M2 gene junction and the complete M2 gene. Because the first and third genes both have the downstream CAT-specific sequence, the products of both these genes can be detected with a negative-sense CAT-specific riboprobe and compared to each other. As a negative control for L transcription, we constructed minigenome B, which is similar to minigenome A but lacks F and M2 sequence up to and including the L GS signal (Fig. 2A). This construct would be expected to generate only the upstream CAT mRNA. cDNA encoding either minigenome A or minigenome B was transfected into cells with pTM1 plasmids which express the L, P, N, and M2-1 proteins. At 48 h posttransfection, cells were harvested, and polyadenylated RNAs were isolated by oligo(dT) chromatography and analyzed by Northern blotting.

Figure 2B shows RNAs detected with a CAT-specific riboprobe. Minigenomes A and B each produced the expected CAT mRNA representing the first gene (Fig. 2B, lanes 1 and 2). Minigenome A also produced the M2 mRNA representing its second gene, which hybridized with an M2-specific probe (not shown) but not with the CAT probe. It also produced the M/CAT/LUC mRNA (lane 2), which is transcribed from the third gene and represents L mRNA. As described later, the relative level of the M/CAT/LUC mRNA compared to its upstream neighbors was comparable to that of L mRNA in RSV-infected cells. Unexpectedly, both minigenomes also produced another mRNA, designated X, which hybridized with the CAT probe and is described below. To confirm the identities of the CAT and M/CAT/LUC mRNAs, a duplicate of the blot shown in Fig. 2B was probed with a LUC-specific probe (Fig. 2C). As expected, the LUC probe hybridized with the M/CAT/LUC mRNA generated from minigenome A but not with the CAT mRNA. The LUC probe also detected mRNA X. These results demonstrate that minigenome A contains all of the sequences necessary for the RSV polymerase to synthesize L mRNA from the gene overlap.

The synthesis of mRNA X was unexpected. Several obser-
positive control, showing the mRNAs generated by minigenome A; lane 2 contains a negative control in which L plasmid was omitted; and lane 3 contains another control, showing the products from minigenome B which does not express M/CAT/LUC mRNA (Fig. 2).

In minigenome C, the 3'-proximal GS signal was ablated. Previously we have shown that this signal is important for the overall level of transcription; if it is removed, mRNA levels of the first gene and subsequent genes are reduced 10- to 20-fold (20). Residual mRNA expression represents transcription in which the polymerase reads from the leader into the first gene independently of the promoter-proximal GS signal and thereafter engages in normal stop-start transcription (20). In agreement with these findings, minigenome C yielded a barely detectable level of CAT mRNA (Fig. 3B, lane 4). The CAT mRNA that was synthesized migrated slightly more slowly than CAT from minigenome A, consistent with it being attached to leader RNA. The level of M/CAT/LUC mRNA that was synthesized from minigenome C was also significantly reduced (although not to the same extent as CAT). This showed that M/CAT/LUC transcription is highly dependent on transcription from the 3'-proximal promoter. That expression of the downstream gene (M/CAT/LUC) was not diminished to the same extent as expression of the leader-proximal gene (CAT) is consistent with our previous observations for a minigenome with a conventional gene junction (20) and might reflect differential stabilities of the mRNAs, as the leader-CAT readthrough mRNA would not be capped or methylated.

Next, we investigated whether M/CAT/LUC transcription is dependent on, or affected by, transcription of its upstream neighbor, the M2 gene. In minigenome D, the M2 GS signal was deleted so as to preclude the synthesis of monocistronic M2 mRNA. As expected, this minigenome yielded CAT mRNA (Fig. 3B, lane 5), and hybridization with an M2-specific riboprobe confirmed that this minigenome also yielded CAT-M2 readthrough mRNA (this species is obscured by RNA X on the blot probed with CAT) but not monocistronic M2 mRNA (data not shown). However, minigenome D yielded only a low level of M/CAT/LUC mRNA. This result shows that L transcription is largely dependent on transcription initiation at the M2 GS signal. The residual synthesis of M/CAT/LUC mRNA can be attributed to polymerase which transcribed M2 as the CAT-M2 dicistronic mRNA. Thus, a prerequisite for M/CAT/LUC transcription appeared to be that polymerase must approach the overlap in a transcribing mode.

We then asked whether M2 gene sequence lying downstream of the L GS signal is required for L transcription. In minigenome E, the RSV-specific sequence downstream of the L GS signal was deleted, leaving the L GS signal intact. This minigenome yielded monocistronic CAT mRNA and M2/M/CAT/LUC readthrough mRNA; however, no monocistronic M/CAT/LUC mRNA was expressed (Fig. 3B, lane 6). This result clearly shows that sequence located downstream of the L GS signal is required for L transcription.

Deletion analysis of sequences flanking the M2/L overlap. Minigenome A, which was used in the experiments described above, contains the F/M2 gene junction region and the complete M2 gene. To determine if the F and M2 sequence upstream of the L GS signal contains a cis-acting element which is required for M/CAT/LUC gene expression, we constructed a minigenome in which the F sequence, the F/M2 gene junction, and most of the M2 gene were deleted (Fig. 4A, minigenome F). Deletion of the M2 gene also served another purpose: when the complete M2 gene is present in a minigenome, it can be transcribed into functional mRNA encoding the M2-1 and M2-2 proteins, both of which are known to affect RNA syn-

FIG. 3. Expression of the overlapped gene is dependent on sequential transcription. (A) Structures (not to scale) of minigenome A and its derivatives C, D, and E; (B) Northern blot of their encoded mRNAs. (A) Minigenome C is similar to minigenome A except that it lacks the 3'-proximal GS signal; minigenome D lacks the M2 GS signal; in minigenome E, the overlap sequence downstream of the L GS signal, including the M2 GE signal, is deleted (dotted lines). GS or GE signals are indicated with open or filled boxes, respectively. The sequence derived from the F, M2, and M2/L overlap region of RSV is hatched. (B) Northern blot was probed with a negative-sense CAT-specific riboprobe.
thesis. Experiments were carried out which showed that variation of M2-1 protein does not affect expression at the M2-1 overlap (data not shown) but deletion of the M2 gene avoids this consideration altogether. Minigenome F is a dicistronic minigenome which contains the CAT gene fused to the last 100 nt of L-specific sequence, the M2/L gene overlap, followed by the first 100 nt of L-specific sequence fused to M/CAT/LUC. Thus, the M2 GE signal is the termination signal for the CAT gene, and as in the previous minigenomes, the M/CAT/LUC gene initiates with the L GS signal.

Northern blot analysis using a CAT-specific riboprobe showed that minigenome F yielded CAT mRNA and a significant level of M/CAT/LUC mRNA (Fig. 4B, lane 4) compared to minigenome A (Fig. 4B, lane 1). The somewhat higher level of M/CAT/LUC mRNA expressed by minigenome F than by minigenome A is expected, since it now is encoded by the second rather than the third gene. This demonstrates that the presence of the M2 gene is not specifically required for M/CAT/LUC transcription. Next the sequences lying immediately upstream of the L GS signal and/or downstream of the M2 GE signal were deleted to investigate if they play a role in M/CAT/LUC mRNA expression (Fig. 4A, minigenomes G, H, and I). Minigenome G yielded M/CAT/LUC mRNA at a level equivalent to that produced by minigenome F, showing that the M2 sequence lying immediately upstream of the L GS signal plays no role in L transcription. Further analysis of the L-specific sequence lying downstream of the M2 GE signal demonstrated that the enhancing effect could be conferred by 20 nt of L-specific sequence, which was the smallest sequence tested (data not shown).

**Mutation analysis of the 68-nt overlap reveals a critical role for the M2 GE signal.** The M2/L overlap can be considered as three components: the 11-nt L GS signal, which is necessary for L transcription; the 12-nt M2 GE signal; and the 45 nt lying between these two motifs. The idea that the L GS signal consists of 11 nt, compared to the 10 nt of the alternative form of the GS signal that is conserved among the other nine genes (21), is based on the finding that position 11 is sensitive to the substitution of G for the native C, whereas position 12 is insensitive (data not shown). To determine which of these elements are required for L transcription, mutagenesis of the overlap region was carried out. To facilitate manipulation, minigenome G, which contains all sequences necessary for optimal expression of M/CAT/LUC, was modified to create a BglII restriction site almost immediately downstream of the M2 GE signal (Fig. 5A, minigenome J). Minigenome J yielded a similar level of M/CAT/LUC mRNA expression as minigenome G (Fig. 5B; compare lanes 1 and 4), indicating that it was a suitable construct in which mutations of the overlap region could be studied. This also showed that positions 3, 5, and 7 downstream of the M2 GE signal are not important for the enhancement of M/CAT/LUC mRNA synthesis associated with the 20 nt of L-gene sequence at this location.

Minigenome K, in which the M2 GE signal was deleted, yielded dicistronic CAT-M/CAT/LUC readthrough mRNA but no monocistronic M/CAT/LUC mRNA (Fig. 5B, lane 5). Thus, in the absence of the M2 GE signal, the polymerase continues transcription through the M2/L gene junction into the M/CAT/LUC gene and does not respond to the L GS signal. This finding is in agreement with the result for minigenome E (Fig. 3, lane 6) and suggests that the polymerase must terminate transcription at the M2 GE signal to be able to transcribe M/CAT/LUC.

Next the role of the 45 nt which lie between the L GS and the M2 GE signals was investigated. We constructed two minigenomes in which this sequence was replaced with alternative negative-sense RSV sequences of equal length; minigenome L contained the 44-nt SH-G intergenic region followed by an A residue (which normally lies upstream of the M2 GE signal), and minigenome M contained 45 nt of SH gene sequence (Fig. 5A). Minigenome L yielded M/CAT/LUC mRNA at the same level as minigenome J (Fig. 5B; compare lane 6 with lane 4), indicating that the sequence lying between the L GS and M2 GE signals can be substituted by heterologous sequence without a deleterious effect on M/CAT/LUC transcription. Minigenome M also yielded M/CAT/LUC mRNA but at a lower level than minigenome J, indicating that although there is not a requirement for a specific sequence between the L GS and M2 GE signals, some heterologous replacement sequences...
might be tolerated better than others (Fig. 5B; compare lane 7 with lane 4).

As the region between the L GS and M2 GE signals does not require a specific sequence, it was of interest to know if the intervening sequence length is important. Minigenome G was modified to position the L GS signal either 22 nt further from or 22 or 44 nt closer to the M2 GE signal (Fig. 6A, minigenomes N, O, and P). Minigenomes N and O both expressed M/CAT/LUC mRNA at similar levels as minigenome G (Fig. 6B; compare lane 1 to lanes 4 and 5), showing that the relative spacing of the L GS and M2 GE signals is not critical for M/CAT/LUC expression. Surprisingly, minigenome P, in which the L GS and M2 GE signals lie almost adjacent to each other, expressed the M/CAT/LUC mRNA at a greater level than minigenome G (Fig. 6B; compare lane 1 to lane 6). Thus, the relative positions of the L GS and M2 GE signals can affect the efficiency of synthesis of M/CAT/LUC mRNA, but only when placed in very close proximity.

The polymerase can initiate transcription at a GS signal(s) lying on either side, or both sides, of the M2 GE signal. The data presented above suggest that the polymerase transcribes the M2 gene, terminates at the M2 GE signal, and then initiates synthesis at the L GS signal. If this is the case, events at the M2/L gene junction might be similar to events at a conventional gene junction, except that following termination the polymerase apparently must reach backward on the template rather than forward. To investigate if it is possible for the polymerase to initiate at a GS signal downstream of the M2 GE signal, we constructed minigenome R, in which the L GS signal was removed from its authentic location and placed 45 nt downstream of the M2 GE signal. Another minigenome (S) was constructed so as to contain the L GS signal in its normal location as well as a second copy inserted downstream. Minigenomes R and S contain 45 nt of L-specific sequence downstream of the M2 GE signal, compared to 100 nt for minigenome G, used in previous experiments as a marker and control. Therefore, minigenome G was modified to contain the same 45-nt segment. The resulting minigenome Q would encode an M/CAT/LUC mRNA that is identical in size to that initiated at the upstream signal of minigenome S (Fig. 7A).

M/CAT/LUC mRNA was confirmed to be generated from minigenome Q with equal efficiency as from minigenome G (data not shown). A Northern blot showing the mRNAs expressed from these minigenomes is shown in Fig. 7B. The M/CAT/LUC mRNA expressed from minigenome R was slightly smaller than that expressed from minigenome Q, which would be expected as its GS signal has been moved 114 nt downstream. The M/CAT/LUC mRNA was expressed from minigenome R as efficiently as from minigenome Q, indicating that the polymerase is capable of reaching or moving forward from the M2 GE signal (Fig. 7B; compare lanes 3 and 4).

When minigenome S was used as a template, two M/CAT/LUC transcripts were visible as a doublet, representing transcription from both GS signals (Fig. 7B, lane 5). The relative amount of mRNA in each band of the doublet was quantitated with a PhosphorImager. It was found that the mRNA initiating from the authentic GS signal was 1.5-fold in excess of the mRNA originating at the GS signal inserted downstream of the M2 GE signal.

Quantitation of M2 and L mRNA levels in RSV-infected cells. As demonstrated previously (8), the presence of the M2 GE signal within the L gene causes premature termination and polyadenylation of 90% of L-gene transcripts, resulting in the synthesis of a short mRNA of 68 nt [exclusive of poly(A)]. This would be expected to greatly reduce the accumulation of complete L mRNA compared to the other RSV mRNAs. However, the data presented above show that in the reconstituted transcription system, the M/CAT/LUC gene, which represents L, is transcribed relatively efficiently (In Fig. 4, minigenome G expressed M/CAT/LUC mRNA at one-eighth of the level of CAT mRNA.) Because quantitation of the relative molar amounts of RSV mRNAs had not been reported previously, we investigated the relative accumulation of full-length L mRNA by Northern blot analysis. Figure 8 shows a Northern blot of total cellular RNA probed with a negative-sense 5'-labeled oligonucleotide which hybridizes to the region lying between the L GS and M2 GE signals. This probe detects the monocistronic M2 and L mRNAs and the polycistronic F-M2, M2-L, and G-F-M2 mRNAs. The above-mentioned 68-nt transcript is not detectable under these conditions of Northern blotting, but its detection was unnecessary since the desired comparison was of full-length mRNAs. Probing with this oligonucleotide permits direct quantitation of L mRNA relative to M2 mRNA. PhosphorImager analysis of the Northern blot showed that L mRNA is present at approximately one-sixth of the level of M2 mRNA at all times during the course of infection, which is similar to the ratio between M/CAT/LUC and CAT mRNA from minigenome G. This shows that the minigenome system provides a faithful reconstruction of mRNA expression from the gene overlap. Thus, L is expressed at an unexpectedly high...
level compared to M2, given that a typical intergenic region yields a threefold drop in transcription (19) and that most polymerases which initiate at the L GS signal terminate prematurely.

**DISCUSSION**

In this study, a minigenome system was used to investigate the cis-acting requirements for L-gene transcription, with the aim of determining the mechanism by which the polymerase expresses L mRNA. Three models have been proposed to explain how the polymerase accesses the L GS signal (introduction): a cascade model involving separate polymerase entry at each GS signal, a second-promoter model involving independent polymerase entry at the L gene, and a sliding model involving access by nontranscribing polymerases. The results described here are incompatible with these models and support instead a model in which the polymerase must arrive at the M2 GE signal and then apparently scan backward (and apparently also forward) to locate the L GS signal.

Several lines of data are incompatible with the previously proposed models. For example, expression of the L gene was found to depend on sequential transcription from the 3' end of the genome (Fig. 3, minigenomes C and D), which is inconsistent with the second-promoter model. Also, the observation that all of the overlap sequence could be deleted except for the L GS and M2 GE signals (Fig. 5, minigenomes B and K to M) seems inconsistent with a second-promoter model because these are transcription signals which bear no resemblance to the 3' genomic leader region and the promoter site contained within. While the requirement for sequential transcription could be consistent with a sliding model, such a model predicts that inactivation of the M2 GS signal would increase the quantity of polymerase available for initiation at the L GS signal. However, the observed result was that M/CAT/LUC gene transcription decreased (Fig. 3, minigenome D). This indicated that polymerase must arrive at the overlap in a transcribing mode rather than a nontranscribing or sliding mode. Also, the finding that almost all of the M2 gene could be deleted (Fig. 4, minigenome G) indicates that L-gene transcription does not depend on possible unknown cis-acting elements within the M2 gene, which might be involved in switching polymerase into a sliding mode. The most striking observation was the absolute requirement for the M2 GE signal for transcription from the L GS signal (Fig. 3 and 5, minigenomes E and K). If either the cascade, second-promoter, or sliding model is correct, deleting the M2 GE signal should result in an approximately 10-fold increase in L-gene expression due to alleviation of premature termination. However, the observed result was that if the M2 GE signal was deleted, monocistronic M/CAT/LUC mRNA transcription was ablated in favor of readthrough to make a dicistronic mRNA. Thus, a model for transcription of the L gene must include an absolute requirement for the M2 GE signal. These results suggest the following model: the polymerase initiates at a single promoter site at the 3' end of the genome and then transcribes each gene sequentially; once the polymerase has terminated transcription at the M2 GE signal, it recognizes the L GS signal and initiates L mRNA synthesis.

An interesting adjunct to the model described above is that the polymerase molecules which terminate L transcription prematurely at the M2 GE signal should be able to reinitiate L transcription just as readily as those which terminate following transcription of the complete M2 gene. Thus, a polymerase molecule presumably might have multiple attempts at transcribing complete L mRNA. The idea that the polymerase can...
proteins (lanes 2 to 5) or N, P, and M2-1 but not L (lane 1). The blot was probed (lane 4), and S (lane 5) complemented with plasmids encoding N, P, L, and M2-1 poly(A) signal was placed 45 nt downstream of the M2 GE signal. (B) Northern blot of S, the L GS signal was retained in its authentic location and an additional L GS usual location and placed 45 nt downstream of the M2 GE signal. In minigenome genomes R and S. In minigenome R, the L GS signal has been deleted from its location of the L GS signal has been altered; (B) Northern blot of their movement is required. To determine if there was a require-
ment for specific sequence within the M2/L overlap sequence
including the L GS signal was increased because this construction involved a change of C to U at position 11 of the L GS signal. However,
we note that previous saturation mutagenesis of the canonical GS signal showed that either C or U is functional in the final position and also that there is no prior instance of a mutation in a GS signal which enhances transcription (21).

To determine if the polymerase is constrained to retrograde movement at the M2/L overlap, we constructed minigenomes in which the L GS signal was placed downstream of the M2 GE signal. It was found that M/CAT/LUC mRNA was expressed with similar efficiencies whether the L GS signal was located upstream or downstream of the M2 GE signal (Fig. 7, minigenomes Q and R), indicating that the polymerase is capable of reaching forward at this site. If L GS signals were placed both upstream and downstream of the M2 GE signal (Fig. 7, minigenome S), M/CAT/LUC mRNA was generated from both GS signals. In this latter situation, transcripts generated from the authentic L GS signal position were 1.5-fold in excess of transcripts generated from the downstream GS signal. Examination of the level of M/CAT/L mRNA would not provide a reliable comparison of the efficiency of initiation at the upstream GS signal versus the downstream one. This is because the efficiency of recycling from the M2 GE signal to the upstream L GS signal is unknown. For example, if 100 polymerase molecules arrive at the M2 GE signal, then by analogy with other GE signals, 90 to 95 molecules would polyadenylate and release the M2 mRNA and remain template bound, and the remaining 5 to 10 would read through to make a polycistronic transcript and pass from consideration. Of the 90 to 95 molecules, an unknown number might dissociate from the template and pass from consideration. Of the remainder, an unknown number would recycle to the upstream L GS signal, and an unknown number would scan downstream with the potential of initiation at the newly introduced downstream L GS signal. An additional complicating factor is that the presence of this introduced downstream GS signal might capture polymerases which would otherwise have either recycled or released from the template. Then, the fraction of polymerases which recycled would arrive at the M2 GE signal and be reintroduced into the pathways described above. While reliable comparison of initiation efficiency seems precluded, the upstream L GS signal clearly is comparable to the introduced downstream GS signal in the efficiency of producing a full-length mRNA.

Given that no specific sequence which induces the polymerase to scan or move backward at the M2 GE signal could be identified, bidirectional scanning might not be limited to this particular junction. For conventional RSV gene junctions, we assume that the polymerase terminates at the GE signal, releases the completed mRNA, and is available to initiate at the next GS signal. In RSV strain A2, this signal is located from 1 to 5 nt further downstream. This variability in length (discounting possible effects from higher-order nucleocapsid structure), and the dispensability of the intergenic region altogether for transcription (19), suggests that the GE signal and adjacent intergenic sequence do not automatically place the polymerase in a unique, optimal position for initiation. Rather, we assume that the polymerase must scan the template for the GS signal. We imagine that this involves movement of the polymerase along the intergenic region without synthesis, analogous to bacterial RNA polymerases, which are thought to contact the DNA template and then scan until a promoter is encountered, although it is possible that localized dissociation and reassociation occur. During transcription, polymerase movement along the template presumably is dictated by extension of the nascent transcript. Once the polymerase has terminated transcription, there may be nothing to define the direction in which it can move on the template. In this respect, the finding that it can initiate at an L GS signal whether it is placed upstream or downstream of the M2 GE signal indicates that the polymerase scans the local sequence in both directions for a GS signal.

As described in the introduction, there are also other members of the order Mononegavirales which have overlapping gene junctions. In the situations in which overlapping gene junctions occur, both genes are often expressed efficiently. Similarly, we found that L is expressed relatively efficiently compared to M2, suggesting that the overlapping arrangement does not significantly attenuate the overall level of L expression (Fig. 8). Indeed, the levels of expression of complete M/CAT/LUC mRNA were similar whether the L GS signal was located upstream or downstream of the M2 GE signal (Fig. 7, minigenomes Q and R). As described above, we suggest that the overlap in RSV is not severely attenuating for the synthesis of full-length L mRNA because the mechanism of initiation at the L GS signal also provides a mechanism for recycling polymerases which terminate prematurely during L-gene transcription. It is reasonable to expect that this same mechanism can operate in other members of the order Mononegavirales containing overlapped genes. Pneumonia virus of mice, a close relative of RSV, has the same gene order as RSV but does not have an overlap between the M2 and L genes. This indicates that the overlap is not an obligatory feature of pneumovirus replication. Thus, it is possible that the gene overlap does not have a significant effect on the biology of the virus. This can now be directly analyzed by preparing recombinant RSV in which the gene overlap has been modified or removed altogether.

ACKNOWLEDGMENTS

We thank Myron Hill and Ena Camargo for technical assistance and Robert Chanock, Brian Murphy, Michael Teng, Stephen Whitehead, Alex Bukreyev, Katalin Juhasz, and Alison Berringham for critical comments on the manuscript.

REFERENCES

order, and regulatory elements of the replication cycle. Virus Res. 24:1–19.
15. Grosfeld, H., M. G. Hill, and P. L. Collins. 1995. RNA replication by respiratory syncytial virus (RSV) is directed by the N, P, and L proteins; transcription also occurs under these conditions but requires RSV superinfection for efficient synthesis of full-length mRNA. J. Virol. 69:5677–5686.