Sendai virus RNA polymerase scanning for mRNA start sites at gene junctions

PLATTET, Philippe, et al.

Abstract

Mini-genomes expressing two reporter genes and a variable gene junction were used to study Sendai virus RNA polymerase (RdRp) scanning for the mRNA start signal of the downstream gene (gs2). We found that RdRp could scan the template efficiently as long as the initiating uridylate of gs2 (3’ UCCCnnUUC) was preceded by the conserved intergenic region (3’ GAA) and the last 3 uridylates of the upstream gene end signal (ge1; 3’ AUUCUUUUU). The end of the leader sequence (3’ CUAAAA, which precedes gs1) could also be used for gene2 expression, but this sequence was considerably less efficient. Increasing the distance between ge1 and gs2 (up to 200 nt) led to the progressive loss of gene2 expression, in which half of gene2 expression was lost for each 70 nucleotides of intervening sequence. Beyond 200 nt, gene2 expression was lost more slowly. Our results suggest that there may be two populations of RdRp that scan at gene junctions, which can be distinguished by the efficiency with which they can scan the genome template for gs.

Reference


DOI : 10.1016/j.virol.2006.12.033
PMID : 17300823
Sendai virus RNA polymerase scanning for mRNA start sites at gene junctions

Philippe Plattet, Laura Strahle, Philippe le Mercier, Stéphane Hausmann, Dominique Garcin, Daniel Kolakofsky

Department of Microbiology and Molecular Medicine, University of Geneva School of Medicine, C.M.U., 1211 Geneva, Switzerland

Received 8 November 2006; returned to author for revision 7 December 2006; accepted 21 December 2006
Available online 14 February 2007

Abstract

Mini-genomes expressing two reporter genes and a variable gene junction were used to study Sendai virus RNA polymerase (RdRp) scanning for the mRNA start signal of the downstream gene (gs₂). We found that RdRp could scan the template efficiently as long as the initiating uridylate of gs₂ (3’ UCCcnUUUC) was preceded by the conserved intergenic region (3’ GAA) and the last 3 uridylates of the upstream gene end signal (ge₁; 3’ AUUCUUUUU). The end of the leader sequence (3’ CUAAAA, which precedes gs₁) could also be used for gene2 expression, but this sequence was considerably less efficient. Increasing the distance between ge₁ and gs₂ (up to 200 nt) led to the progressive loss of gene2 expression, in which half of gene2 expression was lost for each 70 nucleotides of intervening sequence. Beyond 200 nt, gene2 expression was lost more slowly. Our results suggest that there may be two populations of RdRp that scan at gene junctions, which can be distinguished by the efficiency with which they can scan the genome template for gs.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Sendai virus; RNA polymerase scanning; Gene junctions

The Paramyxoviridae family of non-segmented negative-strand RNA viruses (NNV) contains two subfamilies, Paramyxovirinae and Pneumovirinae (Lamb and Kolakofsky, 2001). Like all NNV, members of this virus family contain a linear array of 5–10 genes separated by gene junctions in which the intergenic region (IGR, i.e., those template nt not copied into mRNA during transcription) is either highly conserved or highly variable. For respiro-, morbilli- and henipaviruses, the gene junctions consist of a gene end signal (ge), an IGR that is precisely 3 nt long, and a gene start signal (gs), e.g., 3’ AUUCUUUUU GAA UCCcnUUUC for SeV (spacing highlights the IGR). For the remaining genera of the Paramyxovirinae (rubula- and avulaviruses) and the Pneumovirinae, the gene junctions contain ge and gs that are more or less conserved, but the IGR is highly variable, both in sequence and length (1 to 190 nt). Variable IGR length does not greatly affect transcription of the downstream gene in these viruses, presumably because the viral RNA-dependent RNA polymerase (RdRp) efficiently scans the template for another gs after releasing mature mRNA from the transcription elongation complex at ge (Fears and Collins, 1999; Cowton et al., 2006; Rassa et al., 2000; Stillman and Whitt, 1998; Stillman and Whitt, 1999).

The absence of conserved sequences upstream of the mRNA start site in viruses with variable IGRs (other than a single purine immediately upstream) suggested that the gs consisted solely of the conserved octa- to decanucleotide mRNA start site in this subgroup, and by extension to all the viruses. However, when the 2 nt IGR of the rhabdovirus VSVInd conserved gene junction (3’ AUACUUUUUUU GA UUCGnnUAG) was expanded from 2 to 21 or 88 nt, this did not appreciably diminish expression from the downstream gs (Stillman and Whitt, 1998; Hinzman et al., 2002). Moreover, in the latter case, both the 3’ GA (IGR) and the U7 run of ge were also found to be required for optimum expression from a downstream gs. VSV gs was thus much more extensive than previously appreciated, namely, 3’ UUUUUUU GA UUCGnnUAG (Hinzman et al., 2002). It would be useful to know what the exact limits of gs at gene junctions (or gs₂) are, i.e., exactly what vRdRp is looking for while scanning the template. This information is of interest.
not only for RdRp action at gene junctions, as RdRp scanning has also been proposed to occur for mRNA initiation at \( gs_1 \) (which follows the leader sequence, Fig. 1A). For example, reverse genetic studies of SeV have found that the strength of the 3′-end replication promoter negatively affects mRNA synthesis from \( gs_1 \). Similarly, the presence of a functional \( gs_1 \) negatively affects RNA synthesis from either the genomic or antigenomic replication promoters (Le Mercier et al., 2003; Vulliemoz et al., 2005). For SeV, it appears that these two RNA start sites (55 nt apart) compete with each other for a common pool of RdRp. As this competition could be quite severe (15-fold effects were seen), competition was proposed to operate in cis, i.e., after release of the leader RNA chain from the elongation complex, during RdRp scanning of the template to locate a new start site. In this case, the new start site could be either the genome 3′-end or \( gs_1 \), thus linking these two events (Kolakofsky et al., 2004).

VSV RdRp scanning to \( gs_1 \) has also been proposed as an alternate explanation of the remarkable results of Whelan and Wertz (2002). These authors found that initiation at \( gs_1 \) required prior initiation of \( le \) RNA at the genome 3′-end in vitro, confirming earlier results (Emerson, 1982). However, in vivo initiation at \( gs_1 \) did not appear to require prior initiation of \( le \) RNA at the genome 3′-end, i.e., initiation at \( gs_1 \) was independent of UV x-links in the genome between the replication promoter and \( gs_1 \). As scanning RdRp would not be impeded by UV x-links, these results could be explained as well by poor scanning efficiency in vitro and efficient scanning in vivo. Whelan and Wertz proposed that transcription and replication were controlled by initiation at different positions on the viral genome, i.e., during transcription, RdRp binds to specific sequences and initiates synthesis at \( gs_1 \), whereas for replication, RdRp binds to and initiates at the genome 3′-end. The possibility that VSV RdRp could directly bind to \( gs_1 \), however, appears increasingly unlikely given the recent structures of VSV and rabies virus N:RNAs (see Discussion). VSV RdRp scanning of the template to access \( gs_1 \) is thus an increasingly attractive explanation of Whelan and Wertz’s results for VSV (Kolakofsky et al., 2004). Strong evidence for RdRp scanning of the template to access \( gs_1 \) has also been recently reported for respiratory syncytial virus transcription (Cowton and Fears, 2005).

This paper describes SeV dicistronic mini-genomes with variable IGRs that express two reporter genes. These mini-genomes were used to characterize the cis-acting sequences of \( gs_2 \) that are required for mRNA synthesis by scanning RdRp.

**Results**

We constructed a dicistronic internal-deletion-type DI genome that expresses a red (\( gene_1 \)) followed by a green fluorescent reporter protein (\( gene_2 \)) to investigate SeV RdRp scanning for \( gs_2 \) at gene junctions. The two genes of the reference mini-genome (DI-RG, Fig. 1A) are separated by the conserved tri-purine IGR (for SeV; 3′ GAA is the IGR in the 1st four junctions, and 3′ GGG in the 5th). Various sequences were then inserted here to modify the cis-acting sequences on either side (\( ge_1 \) and \( gs_2 \)), as well as IGR length. The DI genomes were first recovered from DNA and amplified via transfection into BSR-T7 cells of the various pDI-RG along with pTM1-N, -P/Cstop (in which the C ORF is closed by a stop codon) and -L support plasmids (Fig. 1B). 2 days post-transfection (when most of the cells showed high levels of red fluorescence), the cells were infected with a low moi (ca. 0.1) of SeV-AGP55 and the culture continued for a further 3 days. SeV-AGP55 is a...
non-defective virus in which the trailer region of its antigenomic promoter has been replaced with the leader region, and this allows internal-deletion DIs to compete more effectively with the non-defective helper genome (Le Mercier et al., 2003; Finke and Conzelmann, 1999). During this time, 1.2 μg/ml of acetylated trypsin was added to the medium to allow the infection to spread. When this culture medium (sup1, diluted 1:10) was used to infect BSR-T7 (Fig. 1C) or LLC-MK2 cells (not shown), all the cells were found to express both fluorescent proteins. We also always carried out parallel infections of DI-R (Fig. 1A), a derivative of DI-RG in which the entire GFP coding sequence is deleted, to measure the green fluorescence of our cells that do not express GFP, and to control for the spillover of red light into the green channel. This allowed us to accurately determine the baseline for GFP expression. As DI-RG carries a wt gene junction, the level of GFP expressed relative to that of dsRED, i.e., the relative efficiency of gene2 expression, was set to 100%.

When the IGR of DI-RG was expanded by the addition of 12 or 36 nt (chosen at random from the L gene, but which terminated so that the displaced gs2 is still preceded by 3′ GAA, Fig. 2A), RFP levels (presumably a measure of DI genome levels) were only slightly reduced (Fig. 2C), but GFP levels were strongly reduced, to 3.5% and 4.1% relative to DI-RG, respectively (Fig. 2B). However, when the last 6 nt of ge (3′ CUUUUU) are substituted for the 3′ GAUAUA present upstream of 3′ GAA-gs2 in DI-RG +36, GFP expression was largely restored (DI-RG +36/CU5; 65% relative to DI-RG) (Fig. 2B). As mentioned above, this is similar to what has been described for VSV transcription, where the U7 run of the upstream 3′ AUACU7 VSV ge was required for optimum transcription from gs2 (Hinzman et al., 2002). SeV RdRp thus

**Fig. 2.** Gene2 expression requires part of ge directly upstream of gs2, and mRNA termination at ge1. (A) Sequences at the gene junction of the various DI-RG are presented as minus-strands (template sequences) written 3′ to 5′, such that transcription would proceed from left to right, and in groups of 6 nucleotides to indicate the hexamer phase. IGRs are highlighted on a gray background. Sequence differences are highlighted by reverse display, and the gs2 uridylate initiation site is underlined and in bold. The dotted lines within the DI-RG +36 constructs indicate the first 24 nt of the +36 that are not shown. (B) A bar graph of GFP/RFP expression as determined by FACS analysis, normalized to that of DI-RG which is set to 100. The normal background fluorescence in the green channel and possible spillover of red light was measured using DI-R. The results of two independent experiments, each with duplicate transfections for each DI-RG construct are shown; the error bars indicate their range. (C) A bar graph of RFP expression levels after infection with the various sup1 stocks, relative to that of DI-RG arbitrarily set to 100. Auto-fluorescence in the red channel (determined with DI genomes that express only GFP) was minimal, and has not been specifically accounted for.
appears to efficiently initiate at gs2, as long as gs2 is preceded by the conserved tri-purine IGR and the last 6 nt of ge1 (3’ CU5).

To determine whether the transcriptases which initiated at gs2 of DI-RG +36/CU5 had first to terminate at ge1, we inactivated ge1 by mutating its U5 run to 3’UGUGG, in DI-RG +36 and DI-RG +36/CU5 (Fig. 2A). All transcription from these DIs should thus start at gs1 and terminate at ge2 (and GFP should not be translated from the dicistronic mRNA). The apparent absence of a functional ge1 reduced GFP expression to background levels (0.4%) in both DI-RG +36 [-ge1] and DI-RG +36/CU5 [-ge1] (Fig. 2B). Thus, virtually all RdRp that express gene2 have apparently first terminated mRNA synthesis at ge1. To ensure that mutation of the U5 run of ge1 to 3’ UGUGG had effectively inactivated ge1, and to ensure that ge1 was operating normally in DI-RG +36/CU5 (and did not somehow lead to a gene1 mRNA which extended to the displaced gs2), the DI-RG mRNAs were analyzed by real-time RT/PCR as follows (the 3 possible DI-RG +36 transcripts are shown in Fig. 3A). Intracellular RNA of the various DI-RG +36 infections was first separated into encapsidated genomes/antigenomes and all other (unencapsidated) RNAs by CsCl density centrifugation. The mRNA-containing CsCl pellet fraction was then used for 3 reverse transcription reactions with 3 (leftward) primers located, respectively: (i) near the end of gene1 (RT1, which copies all 3 transcripts); (ii) within the expanded IGR (RT2, which copies dicistronic mRNAs and gene1 mRNAs that have extended to gs2); and (iii) within gene2 (RT3, which copies only dicistronic mRNAs). The cDNAs were then amplified by PCR with a common set of primers that flank the probe to gene1 (Fig. 3A). To control for the fidelity of these determinations, RT/PCRs were also carried out on the genomes/antigenomes extracted from the CsCl-banded nucleocapsids, as a proxy for pure dicistronic mRNA. As shown in Fig. 3B, all three amplifications of the CsCl-banded RNA detected nearly equivalent amounts of gene1 sequences, and the negative effect of increasing length in the RTase reaction is relatively modest. Virtually identical results are obtained when the CsCl-pellet RNA of the DI-RG +36 [-ge1] infection is examined, indicating that mutation of the U5 run to 3’UGUGG had effectively inactivated ge1, leading essentially to dicistronic mRNA synthesis. In contrast, when the CsCl-pellet RNAs of the DI-RG +36 and DI-RG +36/CU5 infections were examined, there was little or no evidence of dicistronic mRNAs, or gene1 mRNAs that have extended to gs2 (lanes RT2, RT3, Fig. 3B). Thus, ge1 of DI-RG +36 and DI-RG +36/CU5 appear to function efficiently. Since ge1 of DI-RG +36/CU5 operates normally, and expression of gene2 requires termination at ge1, transcriptases which initiated at gs2 of DI-RG +36/CU5 have apparently traversed the additional 36 nt of the IGR without synthesizing RNA, i.e., they scanned the genome template for another mRNA start site.

Fig. 3. RT/PCR analysis of DI-RG transcripts. (A) A diagram of three possible transcripts from the DI-RG +36 genomes: (1) termination at ge1 and initiation at the displaced gs2, (2) termination of gene1 just before gs2, and (3) no termination, read-through transcription. The primers used for RT/PCR (short arrows) and the fluorescent probe used to detect the amplified DNAs (barbell) are indicated. (B) Encapsidated genome RNA (CsCl band) and un-encapsidated mRNAs (CsCl pellet) of the various DI-RG +36 infections was reverse transcribed using RT1, RT2 or RT3 as indicated, and as described in the text. The resulting cDNAs were then amplified by the unlabelled primers on either side of the probe. The results of two independent, parallel experiments, each with duplicate transfections for each DI-RG construct, are plotted as bar graphs in which the RT1 reaction (which measures all transcripts) was set at 100.
The above results indicate that the functional SeV gs includes not only the conserved decanucleotide 3′ UCCCU-nUUUC at the start of each mRNA, but presumably some of the sequences upstream as well. Unlike gs at gene junctions that are preceded by a tri-purine IGR (3′ GRR) and the upstream ge, gs1 at position 56 is preceded by the 55 nt leader region. The leader region ends with 3 purines (3′ AAA), but the hexamer 3′ GUCCUA immediately upstream bears little resemblance to the last 6 nt of ge (3′ CUUUUU) required for SeV RdRp mRNA synthesis from gs2. SeV RdRp that have initiated at gs1 are thought to have first initiated a leader transcript at the genome 3′-end, and then to have scanned the template for gs1 after releasing the leader transcript (Vidal and Kolakofsky, 1989). These RdRp may also have directly scanned to gs1 from the genome 3′-end without having initiated a leader chain. Although RdRp that initiate at gs1 may not be equivalent to those that initiate at a gene junction (they have not as yet modified a transcript 5′-end, and may not be committed to transcription), it was nevertheless of interest to determine whether the leader sequence which precedes gs1 could be used in place of the 3′ CU3GAA normally found upstream, but clearly not as efficiently.

**Upstream extension of SeV gs2**

The conserved SeV gene junction (3′ AUUCUUUUUGR-RUCCCU-nUUUC) contains a tri-purine (3′ GRR) that separates two pyrimidine-rich regions (underlined). The highly conserved IGR of respiro-, morbilli- and henipaviruses is precisely 3 nt long (there are no exceptions) and most are 3′ GAA. 3′ GAA is also conserved following the last ge of these viruses, and thus like the conserved IGR 3′ GA of vesiculoviruses, may also function in mRNA termination (Barr et al., 1997; Stillman and Whitt, 1998). To determine whether 3′ GAA was required for gene expression from gs2 (independent of its requirement for ge1 function), we first mutated the IGR to its complement (3′ CUU) in DI-RG +36/CU5 (Fig. 5). In this case (DI-RG +36/CU5/CUU) where there are no purines between the upstream 3′ CU5 and the downstream 3′ UCCC of gs2 (Fig. 5A), expression from gs2 was largely ablated (from 65% to 2.5%, Fig. 5B). The restoration of a single purine in this tri-nucleotide (3′ GUU, CAU and CUA) increased relative gene2 expression to 10–18% (Fig. 5B). The restoration of two contiguous purines in this tri-nucleotide further increased relative gene2 expression, but in an unequal fashion: 3′ GAU almost fully restored gene2 expression (to 52% vs. 65%), whereas 3′ UAA only restored expression to 23% (Fig. 5B). Thus, the tri-nucleotide directly upstream of gs2 must contain at least one purine for activity, two contiguous purines are better, and the strictly conserved G of the 3′ GRR appears to be more important than the A adjacent to the initiating uridylate.

![Fig. 4](image-url)
We next examined the extent of the sequences upstream of 3′ GAA required for gs2 expression by scanning transcription factors. The CU5 pyrimidine run of DI-RG +36/CU5 was progressively shortened to 3′ GU5, 3′ GGU4, 3′ GGGU3, and 3′ GGGGU2 (Fig. 6A). Shortening this pyrimidine run up to GGGU3 had only a modest effect on relative gene2 expression, whereas shortening this run to GGGGU2 reduced expression to near background levels (1.5%, Fig. 6B). Thus, a minimum of 3 uridylates plus the 3′ GAA are required. For SeV, all but the last junctional gs is located such that mRNA synthesis begins at N protein hexamer position 1 (i.e., opposite the nucleotide-binding position closest to the genome 3′-end; phasing indicated by spacing below). Thus, the last 6 nt of the extended 3′ A UUCUUUU UUGAA ge signal are associated with a single N subunit in all DI-RG constructs. This N subunit phasing may help to optimize mRNA synthesis at downstream gs, as the hexamer phase of gs is conserved. A summary of RdRp interaction with the cis-acting sequences involved in mRNA initiation, editing, and termination is shown in Fig. 8.
Length dependence of scanning

RdRp that has released gene1 mRNA at ge1 can either dissociate from the template or scan to and reinitiate at a nearby gs2. This presumably occurs even at wild-type gene junctions, where RdRp must traverse at least 3 template nucleotides to reinitiate mRNA synthesis. If this is a stochastic process, it is only a matter of time before RdRp dissociates from the N:RNA when it has not found another gs. Thus, if RdRp scans the template for the next gs, one would expect that the efficiency of this process to be inversely proportional to the distance between ge1 and gs2. The additional sequences inserted into the IGR of DI-RG +36 were derived from the SeV L gene, and larger segments of the L gene were inserted into the IGR of DI-RG to generate DI-RG +54, +102, +192, +300, and +564. As for DI-RG +36, pairs of DIs were made that contained either 3′ GAU AUAGAA or 3′ CUU UUUGAA upstream of gs2 (sequence spacing indicates hexamer phase); the former serving as a negative control for the latter. The addition of up to 564 nt to DI-RG (1782 nt) affected its ability to be recovered from DNA or amplified by SeV-AGP55 only slightly, as all these DI infections expressed RFP at least 70% of the level of DI-RG (Fig. 7B). In contrast, steady-state GFP levels were clearly reduced in a progressive fashion with increasing IGR length, and the presence of 3′ CU5 upstream of the 3′ GAA was required for efficient expression from the displaced gs2 (Fig. 7A).

The rates of decay in the ratio of GFP/RFP were determined with a curve fitting program (KaleidaGraph, Synergy Software) according to the equation \( \Delta \text{GFP/RFP} = a \cdot \exp(-r \cdot x) \), where \( r \) is the rate of decay, \( a \) is the \( Y \)-intercept, and \( x \) is the nucleotide distance. When the first 4 points of the CU5-dependent data set are considered (IGR increased by up to 192 nt), loss of gene2 expression with increasing IGR length follows a simple exponential decay extremely well (0.0097 ± 0.0007 units/nt of

Fig. 7. Effect of IGR length on relative gene2 expression. Scanning to gs2. SeV stocks with DI-RG genomes containing 36, 54, 102, 192, 300, and 564 nt of the L gene added to the IGR, with either 3′ GAU AUAGAA (−CU5, diamonds) or 3′ CUU UUUGAA upstream of gs2 (+CU5, squares), were used to infect BSR T7 cells. GFP/RFP expression levels were determined by FACS, and normalized to that of DI-RG set to 100. The normal background fluorescence in the green channel (baseline) was determined using DI-R. The results of two independent, parallel experiments, each with duplicate transfections for each DI-RG construct, are shown. The curves were fitted to the data points as described in the text. The broken line considers only the first four data points; the solid line considers the last two data points as well. Scanning to gs1. The results of Cordey and Roux (2006) in which M gene sequences were used as stuffer to separate gs1 from the genome 3′-end (also described in the text), and in which GFP expression (FACS analysis) relative to DI genome levels (Northern blotting) were determined, are plotted for comparison (circles, solid line).

(B) A bar graph of RFP expression levels after infection with the various sup1 stocks, relative to that of DI-RG set to 100. The gradual decline of RFP expression with DI genome length presumably reflects a slight replication disadvantage of the longer DI genomes.
IGR length), with a half-loss of ~70 nt. This rate of decay is much faster than that determined by Cordey and Roux for RdRp scanning to gs1 (Cordey and Roux, 2006). These workers used mono-cistronic DI-genomes with tandem promoters; an antigenomic promoter at the very end for replication, followed by a replication-defective genomic promoter for GFP expression. These mono-cistronic DI genomes were recovered from plasmids similarly to DI-RG and rescued with the same AGP55 helper virus; the only major methodological difference in these two studies was that GFP expression was related to DI-genome levels determined by Northern blotting for the monocistronic DIs. By progressively increasing the distance between the tandem promoters, Cordey and Roux found that GFP expression progressively decreased, but at a slower rate. These results are plotted in Fig. 7 using the same curve-fitting program, and they show that the loss of GFP expression also follows a simple exponential decay, but at 0.0023±0.0002 units/nt of stuffer fragment, with a half-loss of 300 nt (Fig. 7A, scanning to gs1). These latter RdRp, which presumably scan the template for gs1 either directly from the DI-genome 3′-end, or after terminating a trailer transcript (within the antigenomic promoter), can apparently scan 4 times the distance than RdRp after terminating a trailer transcript (within the antigenomic scanning to 3′ GA)

The exact template uridylate on which RdRp stutters to polyadenylate the mRNA at gs is not known, but it is likely to be at the end of the run, as shown in Fig. 8. In this case, the L catalytic subunit of RdRp would presumably interact with sequences extending up to 8 nt upstream of the polymerase active site (asterisk, Fig. 8) and 3 nt downstream during polyadenylation and mRNA release (i.e., 3′ AUUCUUUUU-GAA, stuffer site highlighted). For mRNA initiation, SeV L would presumably interact with sequences extending up to 6 nt upstream of the active site and 9 nt downstream (3′ UUU-GAAUCCUcnuUUC). For P gene mRNA editing, where RdRp stutters on a template C (to add guanylates to the mRNA), SeV L interacts with sequences extending up to 10 nt upstream of the active site and at least 1 nt downstream (3′ UUAUUUUUUCC) (Hausmann et al., 1999). The difference between these events, of course, is that RdRp upstream of the active site interacts with an RNA duplex during polyadenylation and mRNA editing (presumably within a cleft that extends from active site to the mRNA exit channel), while during RdRp scanning for gs, only the single-stranded genome RNA is present. During scanning, L presumably samples the single-stranded template sequences for the extended gs as it moves along the genome RNA. When a suitable gs is encountered, RdRp presumably pauses and uses the topography of the cis-acting sequence to position its active site so that RNA synthesis can initiate. During this process, much of the same surfaces RdRp uses to interact with gs as part of the RNA duplex presumably interacts with the upstream portion of gs as ssRNA (Fig. 8). As might be expected for RdRp movement along the template in the absence of RNA synthesis, whether the nucleotide bases of gs are purines or pyrimidines appear to be as important for gs recognition as the H-bonding properties of the nucleotide bases. For example, for efficient gene2 expression, the upstream 3 nt must contain at least one, and probably two contiguous purines between the pyrimidine-rich ge and gs sequences. These purines may represent a

**Discussion**

We have used dicistronic mini-genomes with variable IGRs to study SeV RdRp scanning of the genome template for gs2 at gene junctions. As expected, we found that SeV RdRp could scan the template in a relatively efficient manner. However, efficient gene2 synthesis occurred only when the (displaced) gs2 also contained the conserved IGR (3′ GAA) immediately upstream, as well as 3 of the 5 uridylates of the conserved AUUCU3 ge. This situation is similar to that described for VSV, where the dinucleotide IGR (3′ GA) and the U7 run of the conserved gene junction (3′ AUACUC, GA UUUCGmUAG) were found to play a more limited role in mRNA synthesis from gs2 (Hinzman et al., 2002). The SeV gs thus includes the conserved 3′ UCCCnUUUC mRNA start site plus the upstream hexamer 3′ UUUGAA normally present at gene junctions, just as VSV gs includes the conserved 3′ UUCUCGn- nUAG mRNA start site and the upstream nonomer 3′ U-GA (Fig. 8). During template scanning, these RdRp would presumably interact with sequences located upstream as well as downstream of the initiating uridylicate.

![Fig. 8. NNV RdRp interaction with cis-acting template signals. The SeV and VSV gene junctions (top and bottom) and the SeV P gene editing signal (middle) are shown as minus-strand RNA, written 3′ to 5′. The asterisks mark the RdRp active site during mRNA initiation (gs), editing (edit) and termination (ge). Thick lines indicate the extent of the conserved cis-acting sequences involved in each event.](image-url)
landmark for the RdRp to position the active site relative to the start site uridylate.

The crystal structures of rabies virus and VSV N:RNA rings have recently been determined (Albertini et al., 2006; Green et al., 2006). The path of the RNA is clearly defined within these mini-nucleocapsids, and is virtually identical for each protomer of each ring (which are associated with precisely 9 nt). These are artificial structures in that they are closed rings. However, these N:RNA rings should accurately represent regions of linear NCs that are removed from the ends, such as those containing gs and ge. Both rhabdoviral N proteins are also constructed similarly, with two lobes which enclose and almost totally surround the RNA towards the inside of the N:RNA ring. Several of the bases of each RNA nonomer are completely shielded by the N protein such that W−C base-pairing cannot occur, and their ribose−PO4 backbone is rigidly contained and cannot form an RNA duplex. For both rhabdoviral N:RNAs, the genome RNA must be removed from the N protomers to act as a template for RNA synthesis. For viruses of the Paramyxovirinae, only a medium-resolution (12 Å) structure of the measles virus N:RNA is available, in which the path of the RNA is not clearly defined but whose overall structure is similar to that of the rhabdoviral N:RNAs (Schoenh et al., 2004). Given that the RNA bases of SeV N:RNAs are more resistant to chemical attack than those of VSV N:RNAs (Iseni et al., 2000, 2002), it seems reasonable that SeV genome RNA is similarly hidden within the N protomers of these nucleocapsids and must also be removed from the N protomers to act as a template for RdRp. This transient separation of the template RNA from the N protomers must also be required for NNV RdRp to scan the genome to locate the cis-acting sequences that form gs.

This paper extends the work on RdRp scanning at gene junctions begun with VSV in two ways. We have examined the effectiveness of the end of the leader sequence in directing expression from gs2, and we have determined the effect of scanning distance on the level of gene2 expression. It is noteworthy that virtually all RdRp that have initiated at gs2, even the approximately 4% that do so in a CU3-independent fashion when the IGR is extended by 36 nt, have apparently terminated transcription at ge1 (Fig. 2); that is, virtually none of the RdRp that initiate at gs2 have scanned the template directly from the genome 3′-end. Although the end of the leader sequence (3′CUAAAA) is clearly more effective than a random sequence followed by 3′ GAA (e.g., 3′ AUAGGA; 18% vs. 4%), it is considerably less effective than the 3′UUUGAAA present at gene junctions (65%). Two possible reasons for the relative inactivity of the leader sequence at gs2 are: (i) that the linear sequence itself is different (e.g., there is an adenosine rather than a guanosine at position −3, which appears to be important (Fig. 5)); and (ii) that gs1 naturally begins at hexamer phase 2 whereas gs2 begins at hexamer phase 1. Since the end of the leader sequence is presumably highly effective in directing RdRp initiation at gs1, its relative ineffectiveness at gene junctions may reflect differences in RdRp which operate at these two sites, e.g., RdRp which initiates at gs2 has already synthesized a mature mRNA that is modified at both ends and is committed to transcription, whereas RdRp which initiates at gs1 has yet to modify even the mRNA 5′-end. RdRp which initiates at gs1 also appear to be able to scan longer distances of the template more effectively than RdRp at gene junctions, and this may be related to the fact that SeV leaders are variable in length and some are >200 nt long (Leppert et al., 1979), whereas natural IGRs are only 3 nt long. SeV RdRp that scan to gs1 may then need to scan longer distances than those that scan to gs2. If RdRp scan the template to initiate at gs1 as well as gs2, and if cis-acting sequences upstream of the initiating uridylate are important in each case, then the RdRp surfaces that interact with the upstream sequence, and/or somehow sense the gs hexamer phase, may also be different. It is thus possible that the differences in RdRp action at gs1 and gs2 we describe, namely which sequences upstream of gs (and/or their hexamer phase) are preferred for mRNA initiation, and the extent to which longer distances can be scanned effectively, are related to the commitment of RdRp to transcription that is thought to occur after productive initiation at gs1 (Kolakofsky et al., 2004).

Methods

Virus and cells

BSR-T7/5 cells, a baby hamster kidney cell line expressing a T7 RNA polymerase (Buchholz et al., 1999), were propagated in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 5% fetal calf serum (FCS) in a 5% CO2 atmosphere. The AGP55 recombinant Sendai virus (rSeV-AGP55) was constructed and rescued as described (Le Mercier et al., 2002). Virus stocks, prepared in 9-day-old embryonated chicken eggs from three times plaque-purified virus, reached titres ranging between 5×10⁸ and 10⁹ pfu ml⁻¹.

DI-RG plasmids

A dicistronic mini-genome expressing dsRED2 and green fluorescent protein (GFP), called pDI-RG, was constructed based on the pDI wt/wt plasmid previously described (Le Mercier et al., 2003). The gene end (ge1) and gene start (gs2) sequences were derived from the SeV N−P gene junction. The N protein hexamer phase context of ge1, gs2 and IGR (intergenic region) of N−P gene junction were retained in all constructs. The extended IGR sequences were derived from the SeV L gene. All dicistronic mini-genomes obeyed the rule of six (Calain and Roux, 1993).

DI particle rescue, cell passage and FACS

BSR-T7 cells (1×10⁵) were plated in 6-well plates 20 h before transfection with a mix containing 0.25 μg of pTM1-L, 0.5 μg of pTM1-N, 0.5 μg of pTM1-P/Cstop (which does not express C proteins), 1 μg of the various pDI-RG constructs and Fugene. About 6 h later, the transfection mix was discarded and replaced with 2 ml of Glasgow MEM supplemented with 5% FCS. 48 h post-transfection, the cells were infected with rSeV-AGP55 at a moi of 0.1. 1 and 2 days post-infection 1.2 μg/ml of acetylated trypsin was added to the serum-free medium to allow
viral spread to occur. Supernatant was harvested 3 days post-infection, cleared by centrifugation, and frozen at −80 °C. Fresh BSR-T7 cells seeded in 6-well plates were infected for 2 h with 1:10 dilution of each mixed viral stock. 2 days post-infection, cells were washed, centrifuged, and re-suspended in PBS + 1% FCS, and analyzed by FACS for red and green fluorescence intensity. The baseline for the green fluorescence normally present in cells not expressing GFP was determined by parallel infection with DI-R (Fig. 1).

Analysis of encapsidated and messenger RNAs

Confluent BSR-T7 cells in 9-cm Petri dishes (2 × 10^7 cells) were infected with mixed DI-R/SeV-AGP55 stocks. 2 days post-infection, the cells were collected and the post-nuclear supernatants were made 5 mM in EDTA and loaded onto linear 20–40% (w/w) CsCl gradients. After centrifugation (40000 rpm, 12 °C, overnight), the nucleocapsids banding in the CsCl gradient and the non-encapsidated cellular and viral RNAs sent to the pellet were collected as described previously (Kolakofsky, 1976). The resulting RNAs were characterized by TagMan analysis using specific oligonucleotides and TagMan probes located in the 3′-end of the dsRED2 gene (Fig. 3). RNA was first converted to cDNA with MMLV reverse transcriptase for 1 h at 37 °C, with 1:10 of this RNA mixed with 0.5 μg of specific primer (RT1, RT2, or RT3). One-tenth of the cDNA was then used for real-time PCR, using TagMan Universal Master mix and the ABI Prism 7700 sequence detector. TagMan probe and primers sequences are available upon request.

Acknowledgments

We are grateful to Kyle Tanner (Geneva) for help with the mathematical analysis of the decay curves of Fig. 7. This work was supported by a grant from the Swiss National Science Foundation.

References

Cowton, V.M., Fears, R., 2005. Evidence that the respiratory syncytial virus polymerase is recruited to nucleotides 1 to 11 at the 3′ end of the nucleocapsid and can scan to access internal signals. J. Virol. 79, 11311–11322.