JOURNAL OF VIROLOGY, Dec. 1990, p. 6196–6203
0022-538X/90/126196-08$02.00/0

Vol. 64, No. 12

A Novel Mechanism for the Initiation of Tacaribe Arenavirus Genome Replication

DOMINIQUE GARCIN AND DANIEL KOLAKOFSKY*

Department of Microbiology, University of Geneva Medical School, CMU, 9 Avenue de Champel, CH1211 Geneva 4, Switzerland

Received 28 June 1990/Accepted 5 September 1990

The ends of arenavirus genome and antigenome RNAs are highly conserved and where determined directly, always contain a 3′ G (referred to as position +1). However, primers extended to the 5′ ends of Tacaribe virus genomes and antigenomes extend to position −1. When genomes and antigenomes are annealed either inter or intramolecularly and treated with RNase A or T_1, there appears to be a single unpaired G at the 5′ ends of the hybrids. A single extra G is also found by cloning the 5′ ends of S antigenomes, and studies with capping enzyme detect (p)ppG at the 5′ ends of genome and antigenome chains. A model is proposed in which genome replication initiates with pppGpC to create the nontemplated extra G. In contrast, the nontemplated bases at the 5′ ends of the N mRNAs, which extend to positions −1 to −5, were found to be capped and also heterogeneous in sequence.

Tacaribe virus (TAC) is a member of the family *Arenaviridae*, which contains two RNA genome segments, S (small, ca. 3.5 kb) and L (large, ca. 7 kb). The genomes replicate via full-length complementary strand intermediates, which we refer to as antigenomes, and both genomes and antigenomes are found assembled with N protein as helical nucleocapsids (NCs) (9, 26).

Arenaviruses are similar in many respects to simple negative-strand RNA viruses but differ from the latter in that both segments of arenaviruses employ an ambisense coding strategy (5). This means that mRNAs are transcribed from both the genome and antigenome of each segment, although only from the 3′ side of each template (Fig. 1A). The S and L genomes are transcribed into the nucleocapsid (N) protein and the polymerase (L) protein mRNAs, respectively, while the S and L antigenomes are transcribed into the glycoprotein precursor (GPC) and the Z (or P11) mRNAs, respectively (2, 16, 27). The 3′ ends of the ambisense mRNAs from each segment are thought not to overlap but to be separated by a short intragenic region on each segment. The intragenic regions have the potential to form considerable secondary structure, which may be involved in the 3′ end formation of the mRNAs.

When cellular extracts of arenavirus-infected cells are centrifuged on CsCl density gradients, the viral mRNAs are found in the pellet fraction along with the vast majority of the cellular RNAs. However, the genomes and antigenomes band as NCs at a density of 1.31 g/ml, a region of the gradient which is relatively devoid of cellular RNAs. We have used this separation protocol to examine the 5′ ends of TAC S segment RNAs by primer extension (26). We found that primers extended on the genome or antigenomes detected a unique 5′ end, whereas those extended on N or GPC mRNAs yielded a restricted ladder of bands, which were from zero to four nt longer than the same primers extended on antigenome or genome RNAs. This suggested that the majority of the mRNAs contain a small number of nontemplated bases at their 5′ ends. These extra bases are consistent with a "cap-snatching" mechanism being used for mRNA initiation, as found previously for influenza virus and bunyavirus (15, 17, 18, 24). The same studies suggested that genomes might also contain a single nontemplated G residue at their 5′ end, which would be unique for this group of RNA viruses. The work presented here confirms this suggestion, extends it to all the NC RNAs, and discusses how the unusual initiation of genome and antigenome synthesis might take place.

MATERIALS AND METHODS

Virus, cells, and plasmids. Tacaribe virus, obtained from David Auperin, Centers for Disease Control, Atlanta, Ga., was cloned in BHK-21 cells. A single plaque was passaged twice in BHK cells, and the medium was harvested at 48 h p.i. for virus stock.

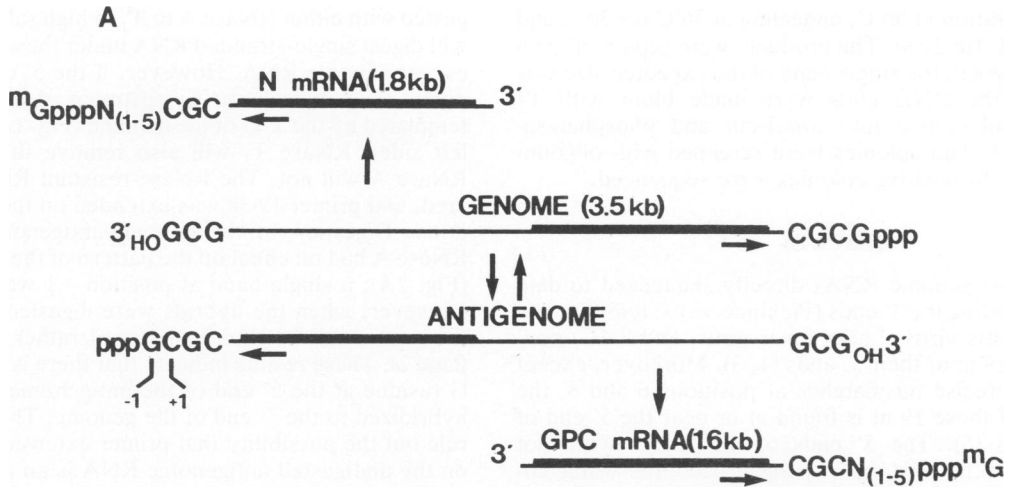
A *Cla*I-*Kpn*I fragment containing part of the polylinker of the pAT vector and nt 1 to 439 of the TAC S genome (12) was obtained from Maria Franze-Fernandez. The fragment was cloned between the same sites of pBluescript KS to generate pTAC S 1-439.

Preparation of intracellular RNAs. Confluent cultures of BHK cells were infected with a 1:50 dilution of the virus stock and maintained in minimal essential medium plus 2% fetal bovine serum at 33°C for 72 h. This culture medium was then used undiluted to "rush" infect other cultures. At 48 h, when the cytopathic effect was clearly visible, the cells were harvested by scraping into phosphate-buffered saline and recovered by centrifugation. RNAs were separated into band and pellet fractions on CsCl density gradients according to Raju et al. (26).

TAC oligonucleotide probes and primers. Six oligonucleotides complementary to the 5′ ends of the S antigenome and N mRNAs, were used: 3′-TGTCACCTAGGATCCGTT TA-5′ (positions 4 to 23), 3′-ACCTAGGATCCGTTTAA CAG-5′ (positions 8 to 27), 3′-GTTTAACAGATTGAG AAAGT-5′ (positions 19 to 38), 3′-CAGATTGAGAAAGT GACTCG-5′ (positions 25 to 44), 3′-GAGAAAGTGACTC GAG-5′ (positions 31 to 46), and 3′-GGAACGAAAC TAGCGGTATTACCGAG-5′ (positions 58 to 83). The positions refer to the 3′ end of the S genome (12).

Primer extensions. Ten micrograms of CsCl pellet RNA (ca. 10% of a 10-cm [diameter] dish) or a one-fourth cell

* Corresponding author.



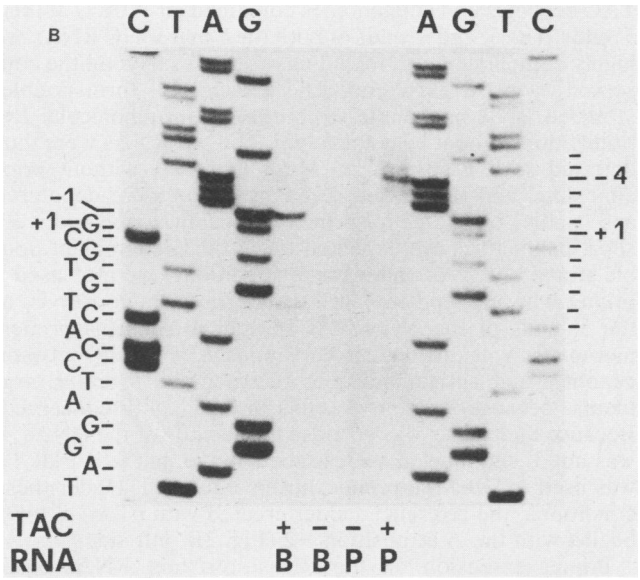


FIG. 1. (A) Schematic representation of the S segment RNAs. RNAs are shown as horizontal lines, the thicker portions representing protein-coding sequences. The 5′ and 3′ ends are shown as triphosphate and hydroxyl groups, respectively. Vertical arrows show the direction of RNA synthesis. Horizontal arrows show the primers used to analyze the 5′ ends of the RNAs. Positions −1 and +1 are indicated. (B) Mapping the 5′ ends of the S antigenome and N mRNA by primer extension. Primer 31-46 was extended on CsCl band (lanes B) and pellet (lanes P) RNA from both infected (lanes TAC +) and uninfected (lanes TAC −) cells, as well as on pTAC S 1-439 DNA (which contains the 3′ end of the S genome) in the presence of ddNTPs, as indicated above. The sequence of the 3′ end of the genome, from positions +1 to +16 is marked on the left. Position −1 marks the 5′ end of the antigenome, and position −4 marks the major band of the N mRNA.

equivalent amount of CsCl band RNA (plus 1 μg of glycogen), was combined with 1 pmol of 5′-end-labeled primer in 2.5 M ammonium acetate, and ethanol precipitated. Primer extension reactions were performed as previously described (26), except that a final volume of 50 μl was used.

RNase A and T_1 digestion. For each reaction, CsCl band RNA from one-fourth of a 10-cm dish was either used directly (see Fig. 2B) or first annealed for 20 min at 65°C in 40 μl of 2.5× buffer A (0.375 M NaCl, 25 mM Tris hydrochloride [pH 7.4], 2.5 mM EDTA) (Fig. 2A). The samples were diluted to 300 μl in the same buffer and digested with either RNase A or T_1, as indicated in Fig. 2, for 30 min at 25°C. The RNase was then digested with 150 μg of proteinase K per ml in the presence of 0.5% sodium dodecyl sulfate for 15 min at 37°C. After phenol extraction and ethanol precipitation, the remaining RNAs were analyzed by primer extension.

Immunoselection with anti-cap antibodies. Anti-cap group antiserum (provided by Ted Munns, Washington University, St. Louis, Mo.) or control antiserum was adsorbed to protein A-Sepharose (14). For each sample to be analyzed, 50 μl of a 50% suspension of protein A-Sepharose was incubated with 15 μl of antiserum, with the total volume being 150 μl after the addition of NEN (150 mM NaCl, 1 mM EDTA, and 0.5% Nonidet P-40). Incubation was at 4°C for 2 h, followed by three washes of the antibody-coupled beads with NEN. Prior to immunoselection, primer extensions on CsCl pellet or band RNA were carried out as above, except that 2 mM sodium pyrophosphate was included in the reaction buffer. Following the reaction, the products were diluted fivefold with NEN and incubated with 50 μl of a 50% suspension of antibody-coupled protein A-Sepharose for 1 h at 4°C. Immunoselected mRNA-cDNA duplexes were washed three times in NEN, phenol extracted, ethanol precipitated, and electrophoresed on 8% sequencing gels.

Cloning the 5′ ends of the N mRNA and S antigenome. The 5′ ends of the RNAs were amplified by an anchored polymerase chain reaction protocol (21). First-strand synthesis was carried out on either CsCl band RNA from three 10-cm dishes or on CsCl pellet RNA from one dish (100 μg) with primer nt 58 to 83 (primer 58-83) in 150-μl reactions. The products were chromatographed on Sephadex G-50 to remove dNTPs, and the excluded volume was precipitated with ethanol in 2.5 M ammonium acetate. The cDNA-RNA hybrids were G tailed in 0.1 M K cacodylate–2 mM $CoCl_2$–0.1 mM dithiotheitol, with 40 μM dGTP and 300 U of terminal deoxynucleotidal transferase (Boehringer GmbH) per ml in a total volume of 100 μl (1 h at 37°C). The products were extracted with phenol–chloroform and precipitated with ethanol. Amplification was carried out with 20 pmol of 5′-ACGGTA $(C)_{14}$ and primer 31-46 (which was 5′ end labeled) in 67 mM Tris hydrochloride (pH 8.8)–16.6 mM ammonium sulfate–10 mM 2-mercaptoethanol–0.17 mg of bovine serum albumin per ml–3 mM $MgCl_2$ and 1.25 U of Taq polymerase (Genofit) in a total volume of 100 μl for 30

cycles (denaturation at 90°C, annealing at 56°C for 30 s, and reaction at 70°C for 20 s). The products were separated on a 10% acrylamide gel, the single band of the expected size was isolated, and the DNA ends were made blunt with T4 polymerase and cloned into *Sma*I-cut and phosphatase-treated pGEM3. The colonies were screened with oligonucleotide 4-23, and positive colonies were sequenced.

RESULTS

All arenavirus genome RNAs directly sequenced to date by pCp end labeling the 3′ ends (Pichinde virus, lymphocytic choriomeninghitis virus, Lassa fever virus, and TAC) contain the same 19 nt at their 3′ ends (1, 3). Moreover, except for the same precise mismatches at positions 6 and 8, the complement of these 19 nt is found at or near the 5′ end of these genomes (5). The 5′ ends of the RNAs were not determined directly, but by cloning procedures using homopolymer tailing to protect the 5′ end. For Pichinde virus and Lassa fever virus, what appeared to be an extra or nontemplated G was found at the 5′ end (2, 3), and the same extra G was found by directly sequencing the cDNA from primer extended on the TAC S genome (26). At this stage, however, we note that none of the arenavirus antigenome 3′ ends have been directly determined for technical reasons. There is thus no way to be sure that the presumptive extra 5′ G of the genome is not templated by a 3′ C residue present on the antigenome (Fig. 1A, right side). The reason for thinking that this would not be so was that all genomes contain a 3′ G and that the precise 3′ ends of genomes and antigenomes would be conserved, so that RNA synthesis would begin in a similar manner on both of these templates.

One way to settle this problem is to determine the 5′ end of the antigenome RNA, as the 3′ end of its template is not in doubt. In a previous study (26), we were unable to chemically sequence the cDNA of the primer extended on antigenome RNA to determine its precise 5′ end due to insufficient product. In the interval, plasmid DNA containing the precise 3′ end of the TAC S genome was obtained (pTAC S 1-439, as described in Materials and Methods). The same experiment could now be done more easily, by simply comparing the position the primer extended on the S antigenome against a sequence ladder generated by the same primer on pTAC S 1-439 DNA.

Figure 1B shows the results of such an experiment, in which a primer representing positions 31 to 46 from the 3′ end of the S genome was extended on both CsCl pellet and band RNAs. In our numbering system, position +1 is defined as the 3′ end of the genome or antigenome. As found previously (26), the primer extended predominantly to a single band on antigenome RNA and to a short ladder of bands slightly longer on N mRNA. It is clear that the 5′ end of the antigenome by this test is one base upstream of the 3′ end of the genome, or position −1. The 5′ ends of the N mRNAs are slightly longer, with bands at positions −2 to −5 being visible here. The mRNA band at position −4 is the most intense. The same result was obtained with primers 4-23, 8-27, 19-38, and 25-44 (data not shown).

As mentioned previously (26), our reaction conditions can accurately determine the 5′ ends of La Crosse bunyavirus genomes and antigenomes, so that if our results on the TAC RNAs are somehow artifactual, the artifact must be peculiar to the particular sequences at the ends of these RNAs. To rule out this possibility, the following experiments were carried out. The CsCl band RNA, which contains an excess of genomes (data not shown), was annealed and then digested with either RNase A or T_1 in high salt. Both nucleases will digest single-stranded RNA under these conditions, e.g., excess genome RNA. However, if the 5′ end of the antigenome contains an extra G upstream of the first C residue templated by the 3′ G of the genome (Fig. 1A, position +1 on left side), RNase T_1 will also remove this base, whereas RNase A will not. The RNase-resistant RNAs were recovered, and primer 19-38 was extended on them after denaturation. Digestion of the genome-antigenome hybrids with RNase A had no effect on the pattern of the primer extension (Fig. 2A); a single band at position −1 was found (lane 3). However, when the hybrids were digested with RNase T_1, the primer extended to position +1 rather than position −1 (lane 2). These results indicate that there is a single unpaired G residue at the 5′ end of the antigenome RNA when it is hybridized to the 3′ end of the genome. They also appear to rule out the possibility that primer extension to position −1 on the undigested antigenome RNA is an artifact.

We next examined by a similar approach whether all the TAC genomes and antigenomes contained an extra G at their 5′ ends. The 5′ and 3′ ends of both the S and L NC RNAs are highly complementary, including sequences beyond the conserved 19 nt at each end. These ends will form double-stranded RNA panhandle structures by intramolecular hybridization without heat annealing. The NC RNAs were thus digested with RNase A or T_1 in high salt without prior annealing, and the RNase-resistant RNAs were denatured and used as templates for primer extension. Since relatively short panhandles are expected to be the predominant double-stranded RNAs under these conditions, we also used a primer whose 3′ end was well within the 19 nt conserved at the 5′ ends of the NC RNAs (primer 4-23). This primer, moreover, will distinguish very poorly, if at all, between genomes and antigenomes, or between the S and L segments, because the four 5′ ends are so highly conserved. Because the primer was so close to the ends of the RNAs, it was not 5′ end labeled as described above, but [α-^{32}P]dCTP was used to label the primer during extension. Under these conditions, the reference ladder created with pTAC S 1-439 begins with the A at position −2 (Fig. 2B, left side).

Primer extension on the RNase-resistant RNAs again yielded a major band at position −1 (as judged by the spacing [Fig. 2B, lane 1]). However, several other bands were also present by using this labeling procedure (most notably positions −5 and −6) which were also present, although weaker, with mock-infected samples (lane 11). These other bands presumably appear because there are other oligonucleotides present which act as primers. The major band at position −1 is the one expected from primer 4-23, and its intensity is unchanged as expected when ddATP and ddTTP are included in the extension reaction (there is no A or T in positions 1 to 3), whereas the other bands disappear (lanes 2 and 12). Otherwise, the same results are obtained with this general primer as when an S antigenome-specific primer was used (Fig. 2B). In this experiment, we also used increasing concentrations of each RNase. At all concentrations, RNase A had no effect on the position of the major primer extension product, whereas RNase T_1 digestion of the panhandle resulted in the primer extending to position +1 rather than position −1. The lower intensity of the band at position +1 is also expected, as this product would incorporate one dCMP instead of two.

It therefore appears that all genomes and antigenomes contain an extra G residue at their 5′ ends. The panhandle structures formed by intramolecular annealing of the ends of these RNAs, and which presumably maintain the NCs as

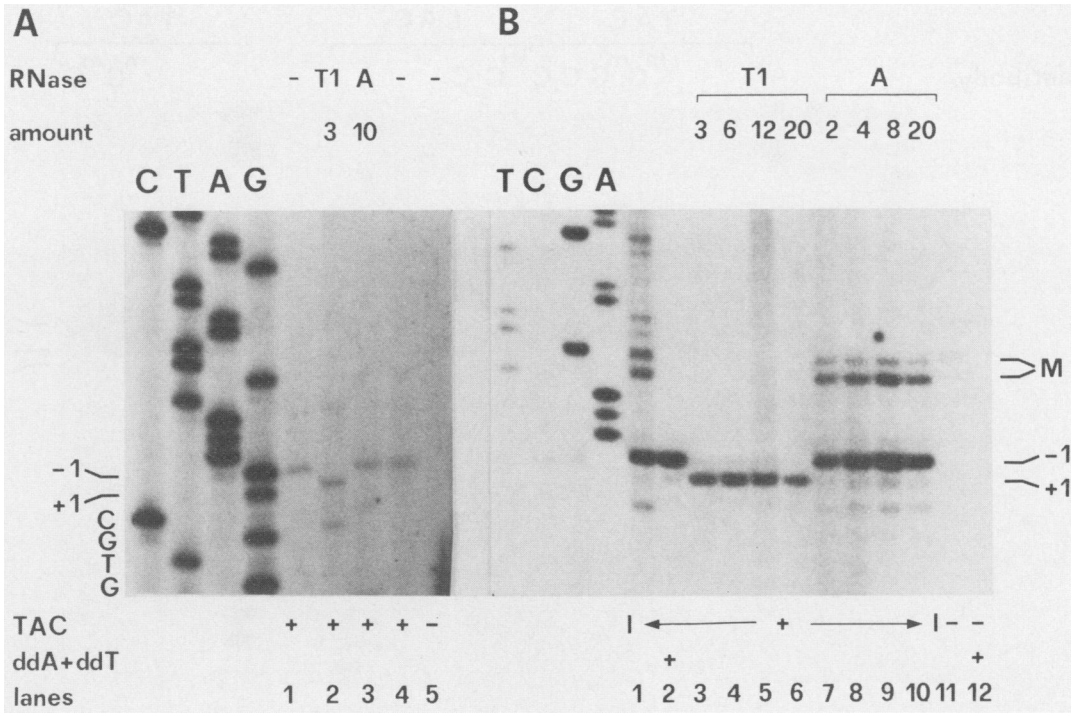


FIG. 2. An apparently nontemplated G at the 5′ ends of genomes and antigenomes. CsCl band RNA was either heat annealed (panel A) or not (panel B), and digested with either RNase T_1 or A in a high-salt solution (see Materials and Methods). The amounts of RNase refer to micrograms of RNase A or units of RNase T_1 per 300 μl. The remaining double-stranded RNAs were recovered and used for primer extension after heat denaturation. 5′-end-labeled primer 19-38 was used in panel A, and nonradioactive primer 4-23 in the presence of [^{32}P]dCTP was used in panel B. Lanes: +, infected; −, uninfected. Reactions run in the absence of dATP and dTTP and presence of ddATP and ddTTP are shown (lanes 2 and 12). The sequence ladders were generated with the same primers on pTAC S 1-439. Positions −1 and +1 are indicated on the sides. M, Bands at positions −5 and −6 that are also present in the uninfected controls.

circular structures (23), therefore do not have flush ends. This is unique for these viruses with segmented RNA genomes.

The chemical nature of the 5′ ends of the TAC RNAs. We expect that the TAC mRNAs will be capped to function efficiently in translation, as there are few exceptions to this rule. However, this need not be so for genomes and antigenomes. It is unlikely that genomes and antigenomes are translated in vivo even though they contain protein-coding sequences near their 5′ ends, as they are found almost exclusively in NC structures which are sufficiently stable to band in CsCl density gradients (26). Moreover, Leung et al. (20) were unable either to translate PIC virion RNA in vitro or to find cap groups on these RNAs.

We examined whether the TAC RNAs were capped with the aid of anti-cap antibodies (Materials and Methods). Primer 58-83 was extended on the CsCl pellet and band RNAs as described above, except that the reverse transcriptase reaction also contained 2 mM pyrophosphate to limit the RNase H activity of the reverse transcriptase. Before gel analysis, the cDNA-RNA hybrids were then selected with the anti-cap antibodies bound to staph A beads. The ability to recovery the radiolabeled cDNA by this selection would mean that the template on which the primer was extended was capped (15, 28). We were able to select ca. 30% of the primers extended on the pellet RNA (Fig. 3, lane 3), and with the same relative distribution of bands at positions −1 to −5 as the unselected sample (lane 1). In the same experiment, on the other hand, we were unable to select any of the cDNA which extended to position −1 on antigenome RNA (lane 4), even though this band comigrated with the bottom band of that from mRNAs, which was selectable. Overexposure of this area of the gel shows that the minor amount of mRNAs present in the CsCl band fraction (positions −3 and −4) are also selected by these antibodies (Fig. 3, right side). It was also possible to select the N mRNA with anti-cap antibodies from the CsCl pellet RNA before primer extension (data not shown). In a further control, we were unable to select any of the cDNAs which were made from a GTP-initiated RNA synthesized with T7 polymerase via a T7 promoter in vitro (data not shown).

The inability to select 100% of the cDNAs from mRNAs is probably not because only a fraction of the mRNAs are capped, as we could only select a similar fraction of the cDNAs made from La Crosse virus mRNAs (lanes 7 and 9), all of whose mRNAs are thought to be capped. It is more likely that we were not able to limit all the RNase H activity of the reverse transcriptase during the primer extension reaction (15).

If the genome RNA is not capped, and if the extra G at the 5′ end is the result of chain initiation at this point, the 5′ end of the genome RNAs should carry a free triphosphate. It would also be useful to determine the 5′ base of the NC RNA more directly. We first attempted to metabolically label this RNA with $^{32}PO_4$ to examine possible 5′ phosphorylated structures from hydrolysates, but we were unable to obtain the several million cpm of clean RNA required for this. We next attempted to specifically kinase label the 5′ end of phosphatased RNA but found the RNA to be too degraded for subsequent analysis. We therefore examined whether we could cap the TAC NC band RNA in vitro with [α-^{32}P]GTP in the presence of *S*-adenosyl-methionine, by using the

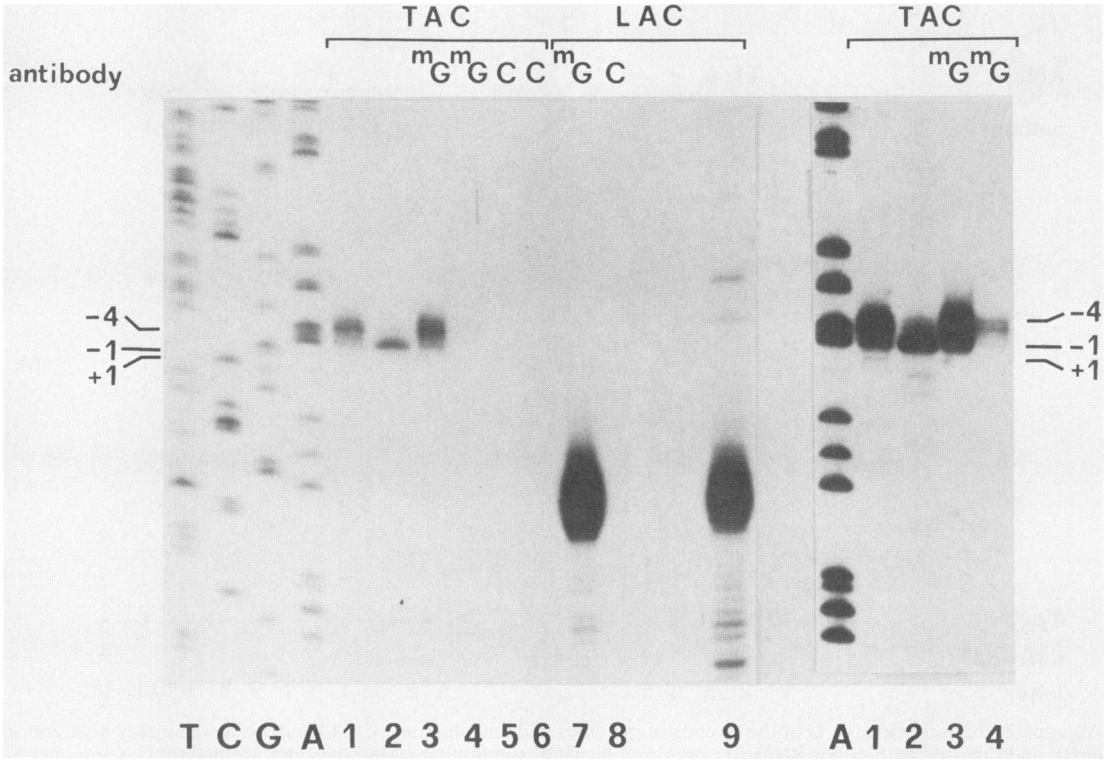


FIG. 3. Immunoselection of cDNA-RNA hybrids with anti-cap antibodies. Primer 58-83 was extended on CsCl band RNA from one dish (lanes 2, 4, and 6) and on 50 μg of CsCl pellet RNA (lanes 1, 3, and 5) of infected cells (TAC). As a control, a primer specific for the N mRNA of La Crosse bunyavirus (LAC) (14) was extended on 20 μg of CsCl pellet RNA from Lac-infected cells (lanes 7 to 9). Of each of the reaction products, 45% was selected with either anti-cap antibodies (lanes ^{m}G) or a control antiserum to the Sendai virus C protein (lanes C) (Materials and Methods). The selected cDNAs were recovered and electrophoresed along with the remaining 10% of the unselected cDNAs (lanes 1, 2, and 9), and a sequence ladder was generated with the TAC primer on pTAC S 1-439. Positions relative to the 3′ end of the genome are marked on the sides. An overexposure of lanes A and 1 to 4 is shown on the right.

vaccinia virus guanylyltransferase or capping enzyme. This enzyme will only modify the 5′ ends of RNAs that contain di- or triphosphate ends (30), and unlike the polynucleotide kinase method, the labeling is unaffected by the 5′ OH groups generated upon RNA degradation. As controls, we capped the same amount, in cellular equivalents, of RNA from a parallel CsCl gradient region of uninfected cells, as well as an in vitro transcript made from a T7 promoter initiated with GTP. The radiolabeled RNAs were recovered and digested with nuclease P1, a non-base-specific nuclease which degrades RNA to 5′ monophosphates but which will not digest the 5′-5′ linkage of caps. To ensure that only cap groups were being examined, a duplicate sample was also digested with alkaline phosphatase. The products were then separated by ionophoresis. The NC band RNAs from infected cells yielded three spots, which migrated similar to marker GpppG, GpppA, and ^{m}GpppG, all of which were resistant to phosphatase (Fig. 4). When quantitated, >85% of the NC RNAs appeared to have (p)ppG ends. The RNA from the uninfected control showed no products, indicating that this region of the CsCl gradient is essentially free of host RNAs that can be capped in vitro. In addition, the GpppG and ^{m}GpppG spots from the infected RNAs behaved similarly in all respects to those obtained from the control pppG-initiated T7 transcript (the nature of the spot migrating just ahead of ^{m}GpppG here is unclear). We also controlled the efficiency of our phosphatase treatment, by showing that these conditions would quantitatively convert pppG to P_i (data not shown).

A limitation of this experiment is that it is impossible to determine what fraction of the RNAs has free triphosphate ends. Nevertheless, these results show that at least some of the TAC NC band RNAs contain (p)ppG at their 5′ ends, and a much smaller fraction may have (p)ppA ends. We have not attempted to determine whether the TAC mRNAs can also be capped in vitro, as these RNAs represent only a very small fraction of the CsCl pellet RNA.

Cloning the 5′ ends of the S antigenome-sense RNAs. The two other families of single-stranded RNA viruses with segmented genomes, influenza virus and bunyavirus, initiate their mRNAs with capped primers derived from host mRNAs. These mRNAs have 10 to 14 (18) and 10 to 18 (25) nontemplated bases at their 5′ ends, respectively, which are also heterogeneous in sequence (6, 7, 29). The length heterogeneity of the primers here presumably reflects the preference for cleavage of the viral endonuclease on different host mRNAs. The extra sequences on the TAC mRNAs, although much shorter than the primers used by influenza virus and bunyavirus, also appear to be capped. We therefore next determined whether these extra bases were also heterogeneous in sequence.

The 5′ ends of the N mRNA were cloned by G tailing the cDNA-RNA hybrid generated by extension of primer 58-83 on CsCl pellet RNA as described above. Part of this sequence was then amplified by polymerase chain reaction with a primer containing a *Kpn*I site followed by oligo(dC) and a nested second primer (see Materials and Methods). A DNA of the expected size was the only band seen when the

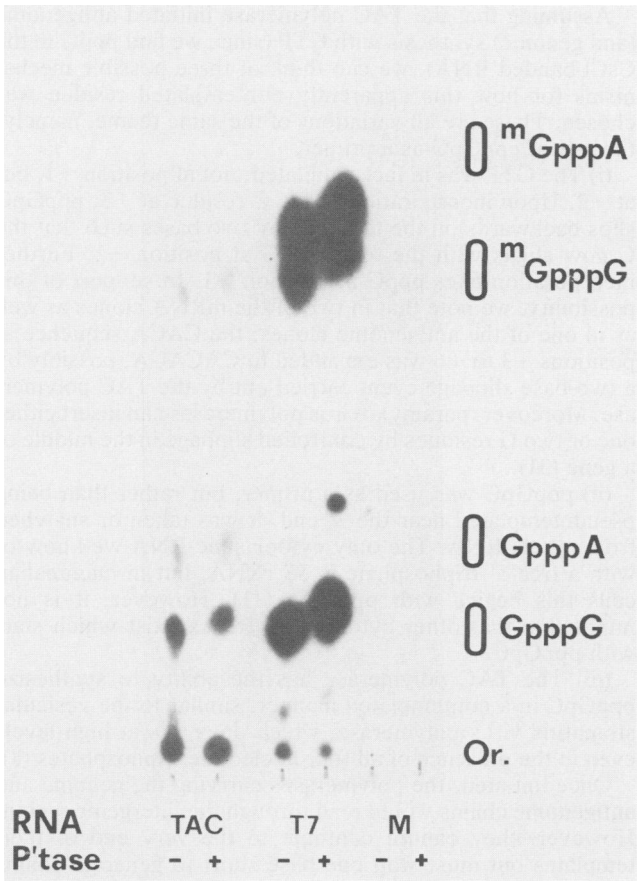


FIG. 4. Capping the 5′ ends of the CsCl band RNAs. CsCl band RNAs from 10 dishes of TAC-infected (TAC) or uninfected (M) cells, and ca. 1 μg of a pppG-initiated transcript made in vitro with T7 polymerase (T7), were capped with the vaccinia virus guanylyl- and methyltransferases according to Moss (22) with 100 μCi of [α-^{32}P]GTP in the presence of 100 μM *S*-adenosyl-methionine in 100-μl reactions. After phenol extraction and Sephadex G-50 chromatography, the RNAs were recovered and samples were digested with 10 μg of nuclease P1 in 15 μl of 1 mM sodium acetate, pH 5.3, for 2 h at 37°C. Duplicate samples were also treated with 22 U of calf intestinal alkaline phosphatase (P'tase; Boehringer). The digests were electrophoresed on Whatmann DE81 paper in pyridine acetate, pH 3.5, along with marker cap groups (indicated on the right). The origin (Or) is also marked, and migration is towards the anode at the top.

TABLE 1. TAC N mRNA and S antigenome 5′ ends

CsCl pellet RNA	No. of clones[a]
CGCACAGUGG	7
⌊CA⌋	
CGCACAGUGG	
or	1*
CG CACACAGUGG	
A CGCACAGUGG	2
U CGCACAGUGG	4
⌊CA⌋	
U CGCACAGUGG	
or	1*
UCG CACACAGUGG	
GA CACACAGUGG	1
UA CGCACAGUGG	1
GU CGCACAGUGG	1
AA CGCACAGUGG	2
UU CGCACAGUGG	1
GUG CGCACAGUGG	1
UGU CGCACAGUGG	1
GACA CGCACAGUGG	1
UCUG CGCACAGUGG	1
AGUG CGCACAGUGG	2
UGACG CGCACAGUGG	1
: : : :	
−5 −1 +1 +10	
CsCl band RNA	
CGCACAGUGG	1
⌊CA⌋	
G CGCACAGUGG	1*
G CGCACAGUGG	10
U GGCACAGUGG	1

[a] Clones marked with an asterisk have an ambiguous sequence alignment. These mRNAs have either inserted CA within the CACA sequence at positions +3 to +6 and have only one or no extra bases at their 5′ ends or have three or two extra bases but have an A rather than the expected G at position +2 (underlined and in boldface).

products were separated on a gel, and this was isolated and cloned. We found that the majority (90%) of the clones had rearranged the inserted sequence when the host bacterium was either *Escherichia coli* HB101 or DH5. However, we found 11 which contained the entire expected sequence to position +1, shortly followed by a C stretch, the *Kpn*I site, and the expected downstream polylinker region. To determine whether the sequences obtained were skewed by the propensity to rearrange in these hosts, the same DNA was cloned in the SURE strain of *E. coli* (Stratagene). The majority (17 of 19) of these clones which were sequenced were not rearranged, and their range of sequences was very similar to the previous 11 clones.

The relevant sequences of all of these clones are shown in Table 1. The 28 mRNAs appear to contain from zero to five extra bases at their 5′ ends. There is, however, an ambiguity in this cloning method of determining the precise 5′ end. mRNAs which contain an extra C residue(s) at their ends cannot be distinguished from those that have no extra sequences, because this end in the clone contains a C tail. This presumably accounts for the seven (and possibly eight) mRNAs listed as having no extra bases. These mRNAs probably have one or more Cs at their 5′ ends, as we did not find any mRNAs which extended only to position +1 in the initial primer extension reaction (data not shown), which was similar to that shown in Fig. 1B. Moreover, animal cell mRNAs which start with C are not uncommon (4).

In the case of two of the clones (Table 1, asterisk), there is also an ambiguity in aligning the sequence. These mRNAs have either inserted CA within the CACA sequence at positions +3 to +6 and have only one or no extra bases at their 5′ ends or have three or two extra bases but have an A rather than the expected G at position +2 (underlined and in boldface in Table 1). Moreover, one other mRNA has an A at position +2, and this was the only other base change we detected in all these clones. The extra bases at the 5′ ends of the mRNAs are largely, although not entirely, heteroge-

neous in sequence, and this too is consistent with a cap-snatching mechanism. If this mechanism applies, a subset of the host mRNAs appears to be used as substrates for the endonuclease, as within the 28 examples there are 4 (and possibly 5) which have a single extra U and 2 with an extra AGUG. A similar conclusion was reached for influenza virus mRNA synthesis (29).

In contrast to the heterogeneity at the ends of the mRNAs as determined by cDNA cloning, 13 clones made in an identical fashion in the SURE strain (except that CsCl band RNA was used as the RNA template) showed little heterogeneity. Of the 13, 10 contained the expected sequence of the S antigenome to position +1, then a single G residue, followed by the oligo(C) tail (Table 1). One clone marked with an asterisk contained a single extra G but also appeared to have an inserted CA within the CACA repeat, similar to two of the mRNA clones. Two others, one with no apparent extra bases, and one with an extra U as well as a G rather than a C at position +1, may have been derived from the small amount of mRNA found within the CsCl band. We note that the G-for-C change at position +1 has only occurred once among the 41 clones and might be an artifact. We also note that there is no example of a clone made from CsCl pellet RNA with a single extra G residue listed in Table 1.

DISCUSSION

There are five families of single-stranded RNA viruses whose genomes (and antigenomes) are expressed from helical NCs: rhabdovirus and paramyxovirus with nonsegmented genomes and influenza virus, bunyavirus, and arenavirus with segmented genomes. In all cases, the ends of the genomes contain inverted complementary sequences, such that the ends of these RNAs can anneal to form a panhandle structure. The inverted sequences are thought to have multiple functions. To the extent that these ends are complementary, the sequences at both ends of the genome and antigenome RNAs are identical. RNA synthesis then initiates at the same sequences at the 3′ ends of genome and antigenome NCs, and NC assembly presumably initiates at the same sequences at the 5′ ends of the nascent antigenome and genome chains. For those viruses with segmented genomes, these sequences are also responsible for circularizing the NCs, which may play a role in their expression and in the packaging of the different NC segments into virus particles.

The present study shows that arenavirus segments also conform to this rule, the only difference being that they contain a single extra G residue at their 5′ ends. The panhandles formed by the intramolecular annealing then do not contain flush ends, as is so for the other virus families, but a single G overhang on the 5′ side. The significance of this extra G is unclear, but we suspect that it is related to the fact the precise 3′ ends of these segments are purines (G residues). All viral polymerases described to date which initiate with a triphosphate do so with either ATP or GTP (4). Except for the *Nairovirus* and *Hantavirus* genera of the *Bunyaviridae* (10), which also contain a 3′ purine residue and whose 5′ ends have not been directly determined, all the others in this group contain 3′ U residues and are known to initiate genome replication at position +1 with ATP. The 3′ G of the arenavirus segments would mean that if this polymerase were to initiate at position +1, it would have to do so with CTP, which would be unique for a viral (but not a cellular) polymerase, which apparently is not the case for the TAC polymerase.

Assuming that the TAC polymerase initiated antigenome (and genome) synthesis with GTP (since we find pppG in the CsCl-banded RNA), we can think of three possible mechanisms for how this apparently nontemplated residue was chosen. These are all variations of the same theme, namely, the use of pppGpC as a primer.

(i) The GTP was in fact templated, not at position +1, but at +2. Upon incorporation of the C residue at +3, pppGpC slips backwards on the template by two bases such that the C now aligns with the template G at position +1. Further incorporation fixes pppG at position −1. In support of this possibility, we note that in two of the mRNA clones as well as in one of the antigenome clones, the CACA sequence at positions +3 to +6 was expanded to CACACA, possibly by a two-base slippage event carried out by the TAC polymerase. Moreover, paramyxovirus polymerases can insert either one or two G residues by controlled slippage in the middle of a gene (31).

(ii) pppGpC was used as a primer, but rather than being pseudotemplated near the 3′ end, it was taken or snatched from a host RNA. The only cytoplasmic RNA we know of with a free 5′ triphosphate is 5S rRNA, but in mammalian cells this begins with pppGpU (11). However, it is not impossible that other cytoplasmic RNAs exist which start with pppGpC.

(iii) The TAC polymerase has the ability to synthesize pppGpC in a nontemplated manner, similar to the vesicular stomatitis virus polymerase, which does this at high levels even in the presence of all four nucleotide triphosphates (8).

Once initiated, the polymerases carrying the genome and antigenome chains would read through the intergenic region. However, they cannot continue to the very end of their templates but must stop one base short to generate chains with the 3′ G at position +1. The apparent inability of the polymerase to incorporate the last templated base is also curious. However, we note that such a mechanism is required to maintain the 3′ ends of the genomes and antigenomes if the 5′ ends are formed by initiation with a primer. This situation is somewhat analogous to that of bacteriophage Qβ (32). Here the polymerase adds a single nontemplated A to the 3′ ends of the chains during replication but maintains the 5′ ends by initiating with GTP at the penultimate C residue at the 3′ end of the templates.

The TAC mRNAs, which are found predominantly in the CsCl pellet fraction, contain from one to five additional bases at their 5′ ends. We now also know that these extra bases are heterogeneous in sequence and that all the 5′ ends, regardless of length, appear to contain cap groups, as they can be specifically selected by anti-cap antisera. These findings are consistent with a cap-snatching mechanism of mRNA initiation. However, whereas the mechanisms described for influenza virus and bunyavirus are remarkably similar in detail, that of arenavirus is different from the two other families in two respects. Not only are the extra sequences on the TAC mRNAs quite different in length, but there appears to be little or no heterogeneity at position +1. Heterogeneity at this position is found for both influenza virus (29) and bunyavirus (7). For influenza virus, in vitro studies have also shown that base pairing between the 3′ end of the capped primer and position +1 of the template is not a requirement for transcription stimulation (19). It is unclear whether these are simply differences in detail, e.g., that the primers used by the TAC polymerase extend to position +2 rather than +1, where there appears to be some heterogeneity, and that the overall mechanism for arenaviruses remains one of cap snatching. On the other hand, these

differences may be of a more fundamental nature. Moreover, the interpretation of this data is complicated by the uncertainty about how genomes are initiated. For example, if pppGpC initiates genomes and this dinucleotide is generated by the viral polymerase, the polymerase may also be capable of making mixed oligonucleotides to prime mRNA synthesis which could be capped subsequently.

Further progress will require testing some of these ideas with an in vitro system which carries out RNA synthesis, as recently described for lymphocytic choriomeningitis virus (13).

ACKNOWLEDGMENTS

We thank Ted Munns, Washington University, for the anti-cap antibodies and Maria Franze-Fernandez, Buenos Aires, Argentina, for the DNA containing the 3′ end of the TAC S segment.

This work was supported by the Swiss National Science Fund.

LITERATURE CITED

1. **Auperin, D. D., R. W. Compans, and D. H. L. Bishop.** 1982. Nucleotide sequence conservation at the 3′ termini of the virion RNA species of New World and Old World arenaviruses. Virology **121**:200–203.
2. **Auperin, D. D., V. Romanovski, M. Galinski, and D. H. L. Bishop.** 1984. Sequencing studies of Pichinde arenavirus S RNA indicate a novel coding strategy, an ambisense viral S RNA. J. Virol. **52**:897–904.
3. **Auperin, D. D., D. R. Sasso, and J. B. McCormick.** 1986. Nucleotide sequence of the glycoprotein gene and intergenic region of the Lassa virus S genome RNA. Virology **154**:155–167.
4. **Banerjee, A. K.** 1980. 5′ terminal cap structure in eucaryotic messenger ribonucleic acids. Microbiol. Rev. **44**:175–205.
5. **Bishop, D. H. L., and D. D. Auperin.** 1987. Arenavirus gene structure and organization. Curr. Top. Microbiol. Immunol. **133**:5–17.
6. **Bishop, D. H. L., M. E. Gay, and Y. Matsuoko.** 1983. Nonviral heterogeneous sequences are present at the 5′ ends of one species of snowshoe hare bunyavirus S complementary RNA. Nucleic Acids Res. **11**:6409–6418.
7. **Bouloy, M., N. Pardigon, P. Vialat, S. Gerbaud, and M. Girard.** 1990. Characterization of the 5′ and 3′ ends of viral messenger RNAs isolated from BHK21 cells infected with germiston virus (Bunyavirus). Virology **175**:50–58.
8. **Chanda, P. K., and A. K. Banerjee.** 1981. Identification of promoter-proximal oligonucleotides and a unique dinucleotide, pppGpC, from in vitro transcription products of vesicular stomatitis virus. J. Virol. **39**:93–103.
9. **Compans, R. W., and D. H. L. Bishop.** 1985. Biochemistry of Arenaviruses. Curr. Top. Microbiol. Immunol. **114**:153–175.
10. **Elliott, R. M.** 1990. Molecular biology of the Bunyaviridae. J. Gen. Virol. **71**:501–522.
11. **Erdmann, V. A., E. Huysmans, A. Vandenberghe, and R. de Wachter.** 1983. Collection of published 5S and 5.8S ribosomal RNA sequences. Nucleic Acids Res. **11**:r105–r129.
12. **Franze-Fernandez, M. T., C. Zetina, S. Iapalucci, M. A. Lucero, C. Bouissou, R. Lopez, O. Rey, M. Daheli, G. N. Cohen, and M. M. Zakin.** 1987. Molecular structure and early events in the replication of Tacaribe arenavirus S RNA. Virus Res. **7**:309–324.
13. **Fuller, F. V., and P. J. Southern.** 1989. Detection of virus-specific RNA-dependent RNA polymerase activity in extracts from cells infected with lymphocytic choriomeningitis virus: in vitro synthesis in full-length viral RNA species. J. Virol. **63**: 1938–1944.
14. **Hacker, D., R. Raju, and D. Kolakofsky.** 1989. La Crosse virus nucleocapsid protein controls its own synthesis in mosquito cells by encapsidating its mRNA. J. Virol. **63**:5166–5174.
15. **Hacker, D., S. Rochat, and D. Kolakofsky.** 1990. Anti-mRNAs in La Crosse bunyavirus-infected cells. J. Virol. **64**:5051–5057.
16. **Iapalucci, S., N. Lopez, O. Rey, M. M. Zakin, G. N. Cohen, and M. T. Franze-Fernandez.** The 5′ region of Tacaribe virus L RNA encodes a protein with a potential metal binding domain. Virology **173**:357–361.
17. **Krug, R. M.** 1981. Priming of influenza viral RNA transcription by capped heterologous RNAs. Curr. Top. Microbiol. Immunol. **93**:125–149.
18. **Krug, R. M.** 1983. Transcription and replication of influenza viruses, p. 70–98. *In* P. Palese and D. W. Kingsbury (ed.), Genetics of influenza viruses. Springer-Verlag, Vienna.
19. **Krug, R. M., B. A. Broni, A. J. LaFiandra, M. A. Morgan, and A. J. Shatkin.** 1980. Priming and inhibitory activities of RNAs for the influenza viral RNA template. Proc. Natl. Acad. Sci. USA **77**:5874–5878.
20. **Leung, W. C., H. P. Gosh, and W. E. Rawls.** 1977. Strandedness of Pichinde virus RNA. J. Virol. **22**:235–237.
21. **Loh, E. Y., J. F. Elliott, S. Cwirla, L. L. Lanier, and M. M. Davis.** 1989. Polymerase chain reaction with single-sided specificity: analysis of T cell receptor d chain. Science **243**:217–220.
22. **Moss, B.** 1977. Utilization of the guanylyltransferase and methyltransferases of vaccinia virus to modify and identify the 5′-terminals of heterologous RNA species. Biochem. Biophys. Res. Commun. **74**:374–383.
23. **Palmer, E. L., J. Obijeski, P. A. Webb, and K. M. Johnson.** 1977. The circular segmented nucleocapsid of an Arenavirus-Tacaribe virus. J. Gen. Virol. **36**:541–545.
24. **Patterson, J. L., B. Holloway, and D. Kolakofsky.** 1984. La Crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependent endonuclease. J. Virol. **52**:215–222.
25. **Patterson, J. L., and D. Kolakofsky.** 1984. Characterization of La Crosse virus small-genome transcripts. J. Virol. **49**:680–685.
26. **Raju, R., L. Raju, D. Hacker, D. Garcin, R. Compans, and D. Kolakofsky.** 1990. Nontemplated bases at the 5′ ends of Tacaribe virus mRNAs. Virology **174**:53–59.
27. **Salvato, M. S., and E. M. Shimomaye.** 1989. The completed sequence of lymphocytic choriomeningitis virus reveals a unique RNA structure and a gene for a zinc finger protein. Virology **173**:1–10.
28. **Schwer, B., P. Visca, J. C. Vos, and H. G. Stunnenberg.** 1987. Discontinuous transcription or RNA processing of vaccinia virus late messengers results in a 5′ poly(A) leader. Cell **50**:163–169.
29. **Shaw, M. W., and R. A. Lamb.** 1984. A specific sub-set of host-cell mRNAs prime influenza virus mRNA synthesis. Virus Res. **1**:455–467.
30. **Venkatesan, S., A. Gershowitz, and B. Moss.** 1980. Modification of the 5′ end of mRNA: association of RNA triphosphatase with the RNA guanylyltransferase. RNA (guanine-7-)methyltransferase complex from vaccinia virus. J. Biol. Chem. **255**:903–908.
31. **Vidal, S., J. Curran, and D. Kolakofsky.** 1990. A stuttering model for paramyxovirus P mRNA editing. EMBO J. **9**:2017–2022.
32. **Weber, H., and C. Weissmann.** 1970. The 3′-termini of bacteriophage Qbeta plus and minus strands. J. Mol. Biol. **51**:215–224.