

## Sequence Determination of the (+) Leader RNA Regions of the Vesicular Stomatitis Virus Chandipura, Cocal, and Piry Serotype Genomes

COLOMBA GIORGI,† BENJAMIN BLUMBERG, AND DANIEL KOLAKOFSKY\*

*Department of Microbiology, University of Geneva Medical School, 1205 Geneva, Switzerland*

Received 30 November 1982/Accepted 12 January 1983

Using 3'-end-labeled genome probes, cells infected with vesicular stomatitis virus Chandipura, Cocal, and Piry serotypes were shown to contain (+) leader RNAs of approximately 50 nucleotides in length. The nucleotide sequence of the leader RNA regions of these genomes was determined and compared with the previously reported sequences of both the (+) and (-) leader RNA regions of other vesicular stomatitis virus serotypes. Regions of strong conservation of nucleotide sequence among the various vesicular stomatitis virus serotypes suggest those nucleotides thought to be involved in control functions during vesicular stomatitis virus replication.

Vesicular stomatitis virus (VSV), a rhabdovirus, contains all of its genetic information (11 to 12 kilobases) in a single RNA chain of negative polarity (-). The VSV (-) genome codes for five viral proteins which are expressed via five monocistronic mRNAs. These mRNAs are transcribed in the order 3' N, NS, M, G, and L 5' both *in vivo* and *in vitro*, and the relative molar abundance of each mRNA is determined by its proximity to the 3' end of the genome, which is thought to be the only site at which the viral polymerase can enter the genome template (1, 21). The first RNA transcribed from the (-) genome template is not the first mRNA but a short, 47-base (+) leader RNA, which is complementary to the precise 3' end of the (-) genome (4). This leader RNA has recently been shown to contain the nucleation site for the initiation of nucleocapsid (NC) assembly (1a-3). One possible function of the (+) leader RNA is thus to detach this nucleation site from the body of the first mRNA so that this mRNA would not also be encapsidated.

A model for VSV replication has recently been advanced in which the binding of N protein to the nascent leader RNA chain modulates mRNA transcription and genome replication (3, 11). In this model, the termination site for the leader RNA is analogous to the attenuator of the tryptophan operon of bacteria and the binding of N protein to the nascent leader chain would both promote readthrough of the leader termination signal and simultaneously begin the encapsidation of the nascent antigenome strand. Thus,

many of the features which control VSV replication are thought to be encoded in the relatively short leader RNA region. In an effort to better understand how these control processes might operate, we have sequenced the leader regions of the Piry (Pi), Chandipura (Ch), and Cocal (Co) serotypes of VSV.

### MATERIALS AND METHODS

**Viruses and their growth.** VSV Indiana (In; Mudd-Summers strain) was obtained from Donald F. Summers, University of Utah School of Medicine, Salt Lake City. VSV Ch and Co were obtained from Craig Pringle, Institute of Virology, Glasgow, Scotland, and VSV Pi was obtained from Danielle Teninges, Laboratoire Genetique des Virus, Gif-sur-Yvette, France. Each serotype was plaque purified twice in succession, and a single plaque was then used to generate virus stocks on five 10-cm dishes of BHK cells. Virus was grown on either BHK or LLCMK2 cells (10 PFU/cell) and harvested at 18 h postinfection for all serotypes except Co (36 h postinfection). Radioactive virus was grown in the presence of 20  $\mu$ Ci of [ $^3$ H]uridine per ml and 0.5  $\mu$ g of actinomycin D per ml. The virus was then pelleted through 30% glycerol, suspended in TNE buffer, banded on a glycerol-tartrate gradient (16), and then recovered by pelleting. A sodium dodecyl sulfate-polyacrylamide gel electrophoresis pattern of the viral proteins of each serotype is shown in Fig. 1.

**Preparation of 3'-end-labeled virion RNA (genome probe).** Purified virus (1 to 3 mg/ml) was incubated for 20 min at 37°C with 0.5% sodium dodecyl sulfate and 500  $\mu$ g of proteinase K per ml and then directly centrifuged on 5 to 23% sucrose gradients (15). The 42S RNA was localized by counting samples of the gradient for [ $^3$ H]RNA and recovered by two ethanol precipitations. Virion 42S RNA (5 to 10  $\mu$ g) was then 3' end labeled with  $^{32}$ P-labeled pCp and RNA ligase in a 20- $\mu$ l reaction as previously described (15) and

† On leave from Laboratorio di Virologia, Istituto Superiore di Sanità, Rome, Italy.

chromatographed on Sephadex G-50, and the excluded RNA was ethanol precipitated. The precipitate was dissolved in 25  $\mu$ l of 10 mM  $\text{Na}_2\text{PO}_4$ , denatured with 40 mM  $\text{CH}_3\text{HgOH}$ , and then diluted to 100  $\mu$ l with TNE containing 100 mM BME and recentrifuged on a sodium dodecyl sulfate sucrose gradient. RNA which sedimented at 42S was again selected and recovered by ethanol precipitation.

**Sequence determination.** (i) **Purine reactions.** Partial digestion with ribonucleases  $T_1$  and  $U_2$  was carried out in buffer containing 20 mM sodium citrate (pH 5.0 for  $T_1$ , pH 3.5 for  $U_2$ ), 7 M urea, 1 mM EDTA, 0.25 mg of tRNA per ml, 0.25% xylene cyanol, and bromphenol blue (5, 12). The end-labeled RNA was added to this buffer, and the reaction mixture was distributed into three 20- $\mu$ l portions in siliconized Eppendorf tubes. The tubes were heated at 50°C for 5 min and quick chilled in ice. A 1- $\mu$ l portion of ribonuclease  $T_1$  ( $10^{-1}$  U/ml) or  $U_2$  ( $10^{-2}$  U/ml) was added to the first of the three tubes, and a single 10-fold serial dilution was carried out into the second tube. The third tube was incubated without enzyme. The tubes were incubated for 15 min at 50°C, chilled, and loaded onto 12 or 20% sequencing gels.

For the ladder, 1.5  $\mu$ l of end-labeled RNA plus 2  $\mu$ l of tRNA (2  $\mu$ g/ml) was mixed with 15  $\mu$ l of formamide. The reaction was incubated at 100°C for 10 to 25 min.

(ii) **Pyrimidine reactions.** For the U reaction, the 3' labeled RNA was lyophilized in the presence of tRNA carrier, dissolved in 50% hydrazine-50% water (vol/vol), and incubated on ice for 10 to 15 min (17).

The C > U reaction was carried out in 3 M NaCl-hydrazine incubated on ice for 20 to 30 min (17). The chemically modified RNA was precipitated twice with ethanol in the presence of 0.3 M sodium acetate. It was then dissolved in 20  $\mu$ l of 1.0 M aniline-acetate buffer, pH 4.5, and incubated in the dark at 60°C for 20 min. The reaction was terminated by freezing at -70°C. The samples were lyophilized, dissolved in water, chilled at -70°C, and lyophilized again. Finally, the samples were dissolved in 3 to 5  $\mu$ l of 8 M urea-20 mM Tris-hydrochloride (pH 7.4)-1 mM EDTA-0.05% xylene cyanol and bromphenol blue, heated 2 min at 90°C, and loaded onto the sequencing gel.

## RESULTS

**VSV Ch, Co, and Pi also contain (+) leader RNAs.** To date, only the In and New Jersey serotypes of VSV have been shown to initiate transcription with a (+) leader RNA. To investigate whether the Ch, Co, and Pi serotypes similarly contained leader RNAs, we prepared 42S (-) RNA from each of the various purified virions (Fig. 1), and these (-) genomes were radiolabeled at their 3' OH ends with  $^{32}\text{P}$ -labeled pCp and RNA ligase (see above). The presence of (+) leader RNA in virus-infected cells was examined by annealing CsCl pellet RNA (total cytoplasmic RNA from which viral NCs have been removed by CsCl density gradient centrifugation [15]) against the end-labeled genome probes, followed by ribonuclease digestion to degrade single-stranded regions of the RNA. As a control, a similar optical density amount of

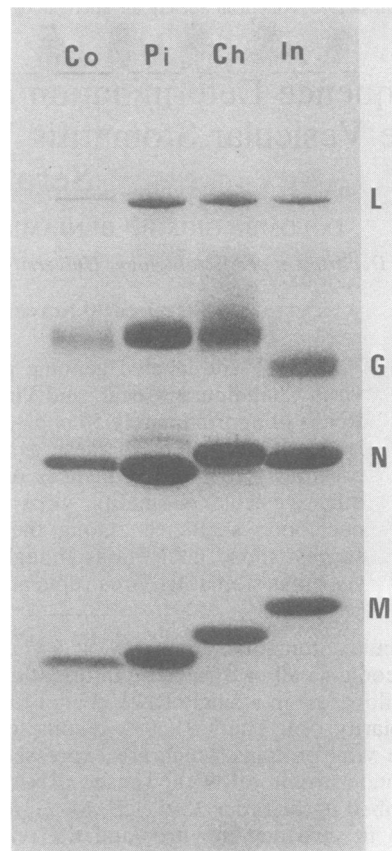


FIG. 1. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of the various VSV serotypes. Between 10 to 20  $\mu$ g of the various VSV serotypes were electrophoresed on a 10% acrylamide-0.8% bis(acrylamide) protein gel (13) which was then stained with Coomassie brilliant blue. The serotypes are noted at the top of the gel. The positions of the L, G, N, and M proteins of VSV In are indicated on the right.

CsCl pellet RNA from uninfected cells was annealed in parallel. The polyacrylamide gel of the remaining RNase-resistant double-stranded RNAs from this experiment is shown in Fig. 2. Note that when uninfected CsCl pellet RNA is annealed against the various 3' genome probes, only in the case of the Co genome is any of the labeled genome RNA resistant to RNase digestion. This short double-stranded RNA probably represents the complementary ends of the Co genome itself, similar to that seen in VSV New Jersey (7). However, when CsCl pellet RNAs from the various virus-infected cells were annealed against their homologous genome probes, several double-stranded RNAs resulted in each case. The results with VSV Ch and In appeared identical; three bands of decreasing intensity which migrated just in advance of the 75-base

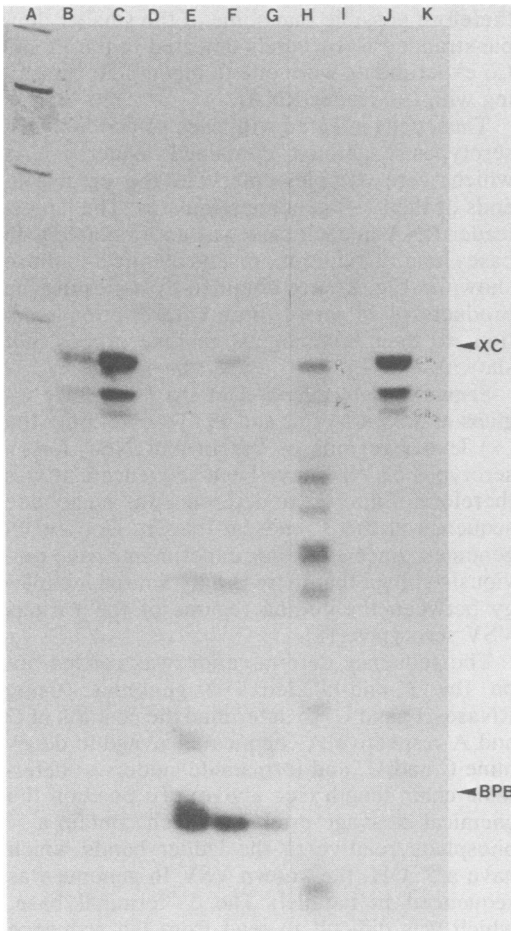


FIG. 2. Demonstration of (+) leader RNAs in the various VSV serotype-infected cells. CsCl pellet RNAs from the various VSV-infected BHK cells or from uninfected cells were annealed against their homologous 3'-end-labeled genome probes. The reactions were then digested with 20  $\mu$ g of RNase A per ml in 0.375 M NaCl as previously described (15), and the remaining RNA was recovered by ethanol precipitation and electrophoresed on a 10% acrylamide-0.5% bis(acrylamide) gel (0.4 mm by 20 cm by 40 cm). Lane A, *Hinf* digest of pBr322; lanes B to D, 0.5 and 1.5 optical density units (ODU) of VSV In-infected and 0.5 ODU of uninfected CsCl pellet RNA annealed against the VSV In genome probe; lanes E to G, 0.5 and 1.5 ODU of VSV Co-infected and 0.5 ODU of uninfected CsCl pellet RNA annealed against the VSV Co genome probe; lanes H to I, 1.25 ODU of VSV Pi-infected and 1.5 ODU of uninfected CsCl pellet RNA annealed against the VSV Pi genome probe; lanes J to K, 1.5 ODU of VSV Ch-infected and 1.5 ODU of uninfected CsCl pellet RNA annealed against the VSV Ch genome probe. XC, Xylene cyanol; BPB, bromophenol blue.

pair DNA marker resulted from the annealing. These bands were in the expected position of leader double-stranded RNA on these gels (10). However, the presence of three double-stranded RNA bands instead of a single band is probably due to the higher resolution of the long and thin gels used in this study. To more accurately determine the chain length of the 3' genome RNA protected against RNase digestion by these leader RNAs, the RNA was recovered from each of the three In and Ch bands, and a sample of these RNAs was partially digested with RNase T<sub>1</sub>. The partial digests were then electrophoresed along with untreated band material and a formamide ladder of the homologous end-labeled genome probes on a 7 M urea sequencing gel (data not shown). From the position of the guanine nucleotides in the sequence relative to the ladder, we estimate that the lengths of the 3'-end-labeled genome RNAs were 42 + 43 bases for the bottom band, 44 + 45 bases for the middle band, and 47 + 51 bases for the top band (bases 48 to 51 in each sequence are purines and cannot be cut by RNase A; see below). The remarkably similar pattern between VSV In and Ch shown in Fig. 2 parallels a very similar sequence at the 3' end of their genomes; only 2 out of the first 51 nucleotides are different between these two serotypes (see below). A similar length heterogeneity of the In (+) leader RNA has recently been demonstrated directly by Schubert et al. (19) by directly labeling *in vitro* transcripts with [ $\beta$ -<sup>32</sup>P]ATP. Thus, VSV Ch-infected cells, like VSV In-infected cells, contain (+) leader RNAs which are 42 to 47 bases in length.

The annealing of both Pi and Co CsCl pellet RNA against their homologous genome probes also resulted in multiple double-stranded RNA bands. In the case of VSV Co, two major double-stranded RNA bands are seen in addition to the putative complementary ends of the Co genome itself; the top band migrates just slightly slower than the top band of the In and Ch series, and the faster band in a separate experiment was estimated to be approximately 26 base pairs long. The reason why the shortest band increases in intensity on annealing with infected cell RNA is not clear, but it may be due to a third leader RNA of this size. In the case of VSV Pi, at least nine bands could routinely be detected, of which the top band migrated similarly to the top band of the In and Ch series. VSV Co- and VSV Pi-infected cells therefore also contain leader RNAs, but these leader RNAs are apparently heterogeneous over a wider length than their In and Ch counterparts. It should be noted, however, that although the infecting stocks used in these experiments were prepared from plaque-purified virus, the presence of small

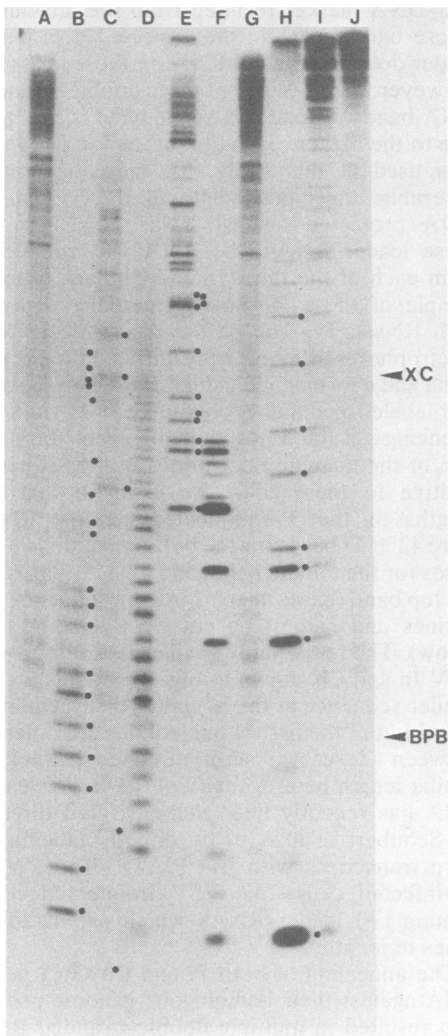


FIG. 3. Sequence determination of 3'-end-labeled VSV Pi genome. The conditions for partial digestion at specific bases are described in the text. Lane A, formamide ladder (10 min at 100°C); lane B, chemical reaction for U; lane C, chemical reaction for C > U; lane D, formamide ladder (25 min at 100°C); lane E,  $10^{-3}$  U of RNase U<sub>2</sub>; lane F,  $10^{-2}$  U of RNase U<sub>2</sub>; lane G, formamide ladder (10 min at 100°C); lane H,  $10^{-1}$  U of RNase T<sub>1</sub>; lane I,  $10^{-2}$  U of RNase T<sub>1</sub>; lane J, no enzyme added. BPB and XC on the right-hand margin show the positions of the bromphenol blue and xylene cyanol dye markers, respectively. Dots on the right-hand side of given bands indicate those bands which determine the sequence. The positions of the chemical cleavage products containing 5' phosphates relative to the ladder bands containing 5' OH groups were determined by sequencing VSV In on a separate gel. The dots show the sequence from the G at position 2 (lowest band, lanes H and I) to the AAAA sequence at positions 46 to 49, except for the U residues at positions 39, 40, 42, 43, and 44 which are detectable on the film but not on the photograph of this 20%

amounts of defective interfering genomes generated in our infections cannot be excluded. It is therefore possible that some of the smaller double-stranded RNA bands detected in the Pi and Co experiments were due to incomplete annealing with (-) leader RNA.

Thus, cells infected with each of various VSV serotypes examined contained leader RNAs which were complementary to the precise 3' ends of their (-) genome templates. The largest leader RNA in each case was approximately 50 bases long. In addition, results identical to those shown in Fig. 2 were obtained by annealing the products of *in vitro* virion transcription reactions to their homologous genome probes (not shown).

**Sequence determination of the (+) leader regions of VSV Ch, Co, and Pi.** To date, only the (+) leader regions of the In and New Jersey serotypes of VSV have been sequenced. It was therefore of interest to determine the nucleotide sequence at the 3' ends of the Ch, Co, and Pi genomes, since annealing experiments have previously shown that there is only limited homology between the coding regions of the various VSV serotypes (18).

The sequence determination was carried out on the 3'-end-labeled (-) genomes, using RNases T<sub>1</sub> and U<sub>2</sub> to determine the position of G and A respectively, chemical cleavage to determine C and U, and formamide ladders to determine chain length (see above). To position the chemical cleavage products which contain a 5' phosphate relative to the ladder bands which have a 5' OH, the known VSV In genome was sequenced in parallel. The 3' terminal base, which was difficult to read from the sequence gel, was determined by nearest neighbor transfer after alkaline hydrolysis and found to be U in all cases. An example of the sequencing gels, that for VSV Pi, is shown in Fig. 3. The sequence of the 3' ends of VSV Co, Ch, and Pi are shown in Fig. 4 along with previously reported sequences of VSV In and VSV New Jersey, and also the (-) leader regions of three In variants [which represent the 3' region of the (+) antigenome strand].

## DISCUSSION

The nucleotide sequences at the 3' end of the (-) genome of five VSV serotypes and of the 3' end of the (+) antigenome of three variants of the In serotype have several interesting features. Of the eight sequences shown, all of which differ from each other, there is remarkable homology

acrylamide sequencing gel. The various VSV genome sequences were also determined on 12% acrylamide gels to better resolve the bands above 40 nucleotides in length.



terminating the (+) leader RNA chains, we similarly examined these regions of the (-) genomes for nucleotide conservation. From the data presented in Fig. 2, we have established that the largest (+) leader RNAs of VSV Ch and Pi are 47 base pairs long, and the Co (+) leader RNA is one or two nucleotides longer. The nucleotide sequences which appear to have been conserved near the termination sites are also underlined in Fig. 4 and lead to the consensus sequence 3'-Y<sub>3-6</sub>NU<sup>↓</sup>GAAA, where Y represents pyrimidines and N is either a U or a G. The arrow shows the presumed point of chain termination. Keene et al. (8) have previously pointed out that (-) leader RNAs appear to be terminated by the sequences 3'-CCC<sup>↓</sup>AGAA. Thus (+) and (-) leader RNA termination sequences are related but somewhat different, and this difference may reflect the relative ease with which the viral polymerase reads through these termination signals during genome replication.

In conclusion, extensive conservation of the nucleotide sequence in the leader RNA (noncoding) regions of five different VSV serotypes provides suggestive evidence for the involvement of these sequences in control functions. In addition, the lengths of the (+) leader RNAs appear to have been highly conserved, suggesting that the distances between the concerned regions may also play a role in the control of VSV transcription and replication.

#### ACKNOWLEDGMENTS

We thank Rosette Bandelier and Colette Pasquier for expert technical assistance.

This work was supported by a grant from the Swiss National Science Foundation.

#### LITERATURE CITED

- Banerjee, A. K., G. Abraham, and R. J. Colonno. 1977. Vesicular stomatitis virus; mode of transcription (a review). *J. Gen. Virol.* 34:1-8.
- Blumberg, B. M., C. Giorgi, and D. Kolakofsky. 1983. VSV N protein selectively encapsidates VSV leader RNA *in vitro*. *Cell* 32:554-567.
- Blumberg, B. M., and D. Kolakofsky. 1981. Intracellular vesicular stomatitis virus leader RNAs are found in nucleocapsid structures. *J. Virol.* 40:568-576.
- Blumberg, B. M., M. Leppert, and D. Kolakofsky. 1981. Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication. *Cell* 23:837-845.
- Colonno, R. J., and A. K. Banerjee. 1977. Mapping and initiation studies on the leader RNA of vesicular stomatitis virus. *Virology* 77:260-268.
- Donis-Keller, H., A. M. Maxam, and W. Gilbert. 1977. Mapping adenines, guanines and pyrimidines in RNA. *Nucleic Acids Res.* 4:2527-2538.
- Isaac, C. L., and J. D. Keene. 1982. RNA polymerase-associated interactions near template promoter sequences of defective interfering particles of vesicular stomatitis virus. *J. Virol.* 43:241-249.
- Keene, J. D., M. Schubert, and R. A. Lazzarini. 1979. Terminal sequences of vesicular stomatitis virus RNA are both complementary and conserved. *J. Virol.* 32:167-174.
- Keene, J. D., M. Schubert, and R. A. Lazzarini. 1980. Intervening sequence between the leader region and the nucleocapsid gene of vesicular stomatitis virus RNA. *J. Virol.* 33:789-794.
- Keene, J. D., B. J. Thornton, and S. U. Emerson. 1981. Sequence-specific contacts between the RNA polymerase of vesicular stomatitis virus and the leader RNA gene. *Proc. Natl. Acad. Sci. U.S.A.* 78:6191-6195.
- Kolakofsky, D. 1982. Isolation of vesicular stomatitis virus defective interfering genomes with different amounts of 5'-terminal complementarity. *J. Virol.* 41:566-574.
- Kolakofsky, D., and B. M. Blumberg. 1982. A model for the control of non-segmented negative strand virus genome replication, p. 203-213. *In* B. W. J. Mahy, A. C. Minson, and G. K. Darby (ed.), *Virus persistence*, symposium 33. Society for General Microbiology Ltd. Cambridge University Press.
- Krupp, G., and H. J. Gross. 1979. Rapid RNA sequencing: nucleases from *Staphylococcus aureus* and *Neurospora crassa* discriminate between uridine and cytidine. *Nucleic Acids Res.* 6:3481-3490.
- Laemmli, U. K. 1979. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227:680-685.
- Leppert, M., and D. Kolakofsky. 1980. Effect of defective interfering particles on plus- and minus-strand leader RNAs in vesicular stomatitis virus-infected cells. *J. Virol.* 35:704-709.
- Leppert, M., L. Rittenhouse, J. Perrault, D. F. Summers, and D. Kolakofsky. 1979. Plus and minus strand leader RNAs in negative strand virus-infected cells. *Cell* 18:235-245.
- Objeski, J. F., D. H. L. Bishop, F. A. Murphy, and E. L. Palmer. 1976. Structural proteins of La Crosse virus. *J. Virol.* 19:985-997.
- Peattie, D. A. 1979. Direct chemical method for sequencing RNA. *Proc. Natl. Acad. Sci. U.S.A.* 76:1760-1764.
- Repik, P., A. Flamand, H. F. Clark, J. F. Objeski, P. Roy, and D. H. L. Bishop. 1974. Detection of homologous RNA sequences among six rhabdovirus genomes. *J. Virol.* 13:250-252.
- Schubert, M., G. G. Harmison, J. Sprague, C. S. Condra, and R. A. Lazzarini. 1982. *In vitro* transcription of vesicular stomatitis virus: initiation with GTP at a specific site within the N cistron. *J. Virol.* 43:166-173.
- Semler, B. L., and J. J. Holland. 1979. Persistent vesicular stomatitis virus infection mediates base substitutions in viral RNA termini. *J. Virol.* 32:420-428.
- Wagner, R. R. 1975. Reproduction of rhabdoviruses, p. 1-94. *In* H. Fraenkel-Conrat and R. Wagner (ed.), *Comprehensive virology*, vol. 4. Plenum Publishing Corp., New York.