Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 8665–8668, November 1989
Biochemistry

Physical map of the genome of sonchus yellow net virus, a plant rhabdovirus with six genes and conserved gene junction sequences

(genome map of negative-strand virus/rhabdovirus evolution/rhabdovirus intergenic sequence)

LOUIS A. HEATON*, BRADLEY I. HILLMAN†, BRENDA G. HUNTER, DOUWE ZUIDEMA‡, AND ANDREW O. JACKSON§

Department of Plant Pathology, University of California, Berkeley, CA 94720

Communicated by Myron K. Brakke, August 14, 1989 (received for review April 27, 1989)

ABSTRACT **We provide evidence that a plant rhabdovirus, sonchus yellow net virus (SYNV), is similar to most animal rhabdoviruses in the order of structural genes and in the nucleotide sequences at the gene junctions but that it differs in the presence and location of a putative nonstructural gene. From the patterns of hybridization of a library of recombinant DNA clones, we have shown that the SYNV genome is transcribed into a short 3′-terminal "leader RNA" and six mRNAs. The proteins encoded by the SYNV mRNAs, in order of the appearance of their genes in the SYNV genome, are designated 3′-N-M2-sc4-M1-G-L-5′ (N, nucleoprotein; M, matrix protein; sc, protein encoded by SYNV complementary RNA; G, glycoprotein; L, large protein). The intergenic and flanking gene sequences are conserved and consist of a central core of 14 nucleotides (3′-UUCUUUUUGGUUGU/A-5′) whose sequence is similar to the sequence at the gene junctions of vesicular stomatitis and rabies viruses. The SYNV core consists of an 8-nucleotide (3′-UUCUUUUU-5′) transcription termination signal at the 5′ terminus of each gene, a dinucleotide (GG) spacer whose complement does not appear in mRNA, and a tetranucleotide (3′-UUGU/A-5′) that is complementary to the first four nucleotides at the 5′ terminus of the SYNV mRNAs. These results, when compared with structural information available on animal rhabdoviruses, suggest that organization of structural genes and maintenance of signals thought to play important roles in regulation of transcription have been conserved during evolution in plant, insect, and vertebrate hosts. However, differences in number and location of putative nonstructural genes reveal some flexibility in genome organization that may be important in deducing taxonomic and evolutionary relationships among viruses causing diseases in phylogenetically diverse hosts.**

The Rhabdovirideae represent one of only two virus families with members that infect and cause serious diseases in both plants and animals (1). The members of the family infect hosts that are widely distributed throughout both kingdoms and have been isolated from many species of plants, mammals, fish, insects, and other invertebrates (2, 3). The prototype rhabdovirus, vesicular stomatitis virus (VSV), has been the subject of intensive investigation over the past 20 years and, as such, it has become a useful model that is now a standard against which the physicochemical properties of other rhabdoviruses are compared (4). Virions of VSV and other rhabdoviruses are of a complex bullet-shaped or bacilliform morphology and contain lipid (20%), carbohydrate (8%), RNA (3%), and five structural proteins. The RNAs of the rhabdoviruses consist of an 11- to 13-kilobase single-stranded RNA of negative-sense polarity. The five structural proteins of VSV include a large protein (L), a glycoprotein (G), a nucleoprotein (N), a matrix protein (M), and a phosphoprotein (P), which for historical reasons is often designated NS (4). The structural proteins of other rhabdoviruses have been given similar designations, but they vary slightly in size and relative abundance. Comparative analyses of the plant and animal rhabdoviruses therefore provide rare opportunities to study the replication and evolution of viruses within a single family that infect hosts in both the plant and animal kingdoms.

Sonchus yellow net virus (SYNV) has become the most highly characterized of the plant rhabdoviruses (2, 5–8). We have previously identified five discrete SYNV-complementary (sc) RNAs, or viral-encoded mRNAs, that were designated scRNAs 1–5 and have postulated that these mRNAs encode the five structural proteins L, G, N, M1, and M2, respectively (5). The putative locations of these proteins within the virion and their electrophoretic patterns (9) are consistent with the nomenclature proposed for rhabdoviruses (10), but the genome organization and sequences at the junctions of all of the genes have not previously been determined.

In contrast to SYNV and other plant rhabdoviruses, the genomes of three animal rhabdoviruses have been mapped, and VSV and rabies RNAs have been subjected to detailed nucleotide sequence analyses. More than a decade ago, Abraham and Banerjee (11) and Ball and White (12) determined that the gene order of VSV is 3′-N-P(NS)-M-G-L-5′. Subsequently, elucidation of the complete primary structure of the VSV (IND) genome (13–16) verified the gene order and confirmed the presence of five genes, although a previously unidentified protein product has recently been shown to be expressed from the phosphoprotein mRNA by a unique mechanism (17). Variation in the number of genes encoded by rhabdovirus genomes was first demonstrated by the finding that six discrete mRNAs are transcribed by infectious hematopoietic necrosis virus (IHNV), a fish rhabdovirus (18, 19). Five of the proteins are structural and the sixth is nonvirion (NV), since it is induced in infected cells but has not been detected in mature virions. Further differences among rhabdoviruses were revealed by analysis of the rabies virus genome, which is composed of five structural genes but also contains a long intergenic sequence between the G and L genes (20). Thus, from the analyses of a relatively few

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: VSV, vesicular stomatitis virus; SYNV, sonchus yellow net virus; scRNA, SYNV complementary RNA; IHNV, infectious hematopoietic necrosis virus; NV, nonvirion; gcDNA, genomic cDNA.

*Present address: Department of Plant Pathology, Throckmorton Hall, Kansas State University, Manhattan, KS 66056.
†Present address: Department of Plant Pathology, Cook College, Rutgers University, New Brunswick, NJ 08903.
‡Present address: Department of Virology, Agricultural University, Binnenhaven 11, 6709 PD Wageningen, The Netherlands.
§To whom reprint requests should be addressed.

rhabdoviruses, it has become evident that members of the family vary in the number of genes and in their organization.

Despite the variation in genome organization, the nucleotide sequences at the intergenic and flanking gene regions of two animal rhabdoviruses, VSV and rabies virus, are strikingly similar. VSV (IND) contains a highly conserved 23-nucleotide gene junction consensus sequence (21), and a nearly identical, although slightly more variable, consensus is present at the rabies virus gene junctions (20). Because these putative regulatory signals at the gene junctions of these viruses are so highly conserved and because the number of mRNAs transcribed from different animal rhabdovirus genomes such as VSV and IHNV differ, we have extended our analyses of the organization and expression of the SYNV genome. We have determined the gene order of SYNV, detected six mRNA species encoded by the SYNV genome, and sequenced the intergenic and flanking regions at each gene junction. The results now permit us to compare the organization of a plant rhabdovirus genome with those of the animal rhabdoviruses.

MATERIALS AND METHODS

General Procedures. The procedures used for SYNV (ATCC PV-263) propagation, maintenance, purification, and RNA extraction have been described (22, 23). Isolation of poly(A) RNA and hybridization analyses were as outlined by Rezaian *et al.* (5). Construction of cDNA libraries from SYNV RNA and from the poly(A) fraction of SYNV-infected tobacco plants and methods used for nucleic acid sequence analysis were as described by Zuidema *et al.* (7).

Identification of cDNA Clones. Approximately 4800 ampicillin-resistant colonies with cDNA inserts derived from the poly(A) RNA of infected tobacco leaves were hybridized with randomly labeled SYNV genomic RNA. Plasmids purified from 50 positive colonies were arbitrarily designated pAS 1–50. The inserts were found to range from 300 to 1500 base pairs and were grouped according to scRNA specificity by screening with recombinant plasmids that hybridized to specific scRNAs (5). The largest recombinant plasmid from each scRNA specificity group was chosen to screen the genomic cDNA (gcDNA) library described below.

A gcDNA library was constructed essentially as described for the pAS library except that purified virion RNA was used as a template and reverse transcription was accomplished via random self-priming (13). Approximately 600 ampicillin-resistant colonies were probed with randomly labeled SYNV genomic RNA, and the plasmids purified from the 260 positive colonies were arbitrarily designated pGL 1–260. The inserts in these plasmids varied from 500 to 1800 base pairs. One hundred of the newly constructed gcDNA clones were grouped according to scRNA specificity by screening with appropriate inserts from the pAS library.

Mapping Genes Encoding the scRNAs. To differentiate the scRNAs on RNA blots and to determine the order of genes on the SYNV genome, selected gcDNA plasmids were nick-translated and used as hybridization probes of blots of poly(A) RNA from SYNV-infected tobacco leaves. If a single gcDNA plasmid hybridized to two scRNAs, then the plasmid contained intergenic and flanking gene sequences, and the genes encoding the two scRNAs were concluded to be adjacent. Such plasmids were called "gene junction" clones. Other plasmids that hybridized to a single scRNA were assigned to the corresponding scRNA specificity group. An overlapping group of the cDNA and gcDNA inserts was obtained by DNA blot hybridization by using appropriate plasmid probes to identify inserts derived from adjacent regions of the genome. With this strategy, we aligned a set of overlapping inserts that represented SYNV RNA from the 3′ terminus into the L gene.

RESULTS AND DISCUSSION

The order of genes that encode the scRNAs, as determined from the hybridization patterns of SYNV-specific probes, is shown in Fig. 1. The cumulative sizes of the cDNA inserts and their hybridization to individual scRNAs constitute evidence that six distinct mRNAs are transcribed from the SYNV genome instead of the previously estimated five (5). The additional scRNA, which we now designate scRNA 6, moves only slightly in advance of scRNAs 4 and 5 in agarose gels (Fig. 1), and the difficulties in resolving this RNA contributed to our failure to identify it previously. Our present results with five gcDNA plasmids, each of which hybridize to two scRNAs, show that scRNAs 3 and 5, 5 and 4, 4 and 6, 6 and 2, and 2 and 1, respectively, are encoded by adjacent genes (Fig. 1). These conclusions have been verified by sequence analyses of individual genes (data to be presented elsewhere). The order of the RNAs relative to the SYNV genome is 3′-scRNA3-scRNA5-scRNA4-scRNA6-scRNA2-scRNA1-5′.

Assignment of Proteins Encoded by the scRNAs. We (6) have previously shown that the 144-nucleotide "leader RNA" gene ("l") is located at the 3′ terminus of the SYNV genomic RNA and precedes the nucleocapsid protein gene from which scRNA 3 is transcribed (7). The mRNA, scRNA 5, which is transcribed from the next downstream gene, encodes the M2 protein, which has limited amino acid sequence relatedness to phosphoproteins of other rhabdoviruses (8). Identification of both the scRNA 3 and scRNA 5 gene products was obtained by protein blot analyses with antisera elicited by fusion proteins constructed from sequenced cDNA that was derived from the respective scRNA (7, 8). Based on our unpublished data from similar serological experiments, scRNA 6 is the mRNA for the M1 protein, which we believe is the matrix protein. We have been unable to detect a structural polypeptide encoded by scRNA 4 and therefore provisionally designate its putative translation product "sc4." We have not yet determined whether sc4 is a nonstructural polypeptide that is synthesized at a specific stage of replication. From comparisons of their coding capacities with the molecular weights of their putative protein products and from relatedness of amino acid sequence deduced from nucleotide sequence data, scRNAs 1 and 2 encode the L and G proteins, respectively (unpublished data). Therefore, the

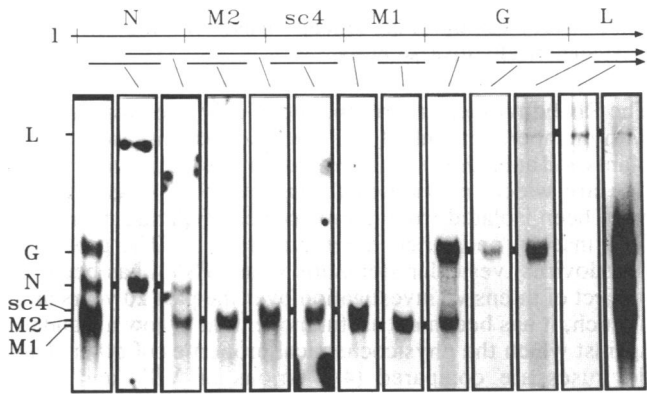

FIG. 1. Construction of a genome map for SYNV. Selected SYNV cDNA clones derived from genomic RNA or poly(A) RNA were nick-translated and used to probe blots of poly(A) RNA from SYNV-infected tobacco leaves. The gene order was deduced from the hybridization patterns of overlapping cDNA clones. The lanes to the far left and far right were probed with randomly labeled SYNV genomic RNA and serve as markers for SYNV mRNAs. The far right lane was overexposed to reveal the L protein mRNA, which is present at ≈5-fold lower abundance than scRNAs 2–6. L, scRNA 1; G, scRNA 2; N, scRNA 3; sc 4, scRNA 4; M2, scRNA 5; M1, scRNA 6.

gene order of SYNV is 3′-"l"-N-M2(P?)-sc4-M1(M?)-G-L-5′.

Comparison of our results with those previously published for animal rhabdoviruses (11–20) reveals that the organization of the SYNV genome is similar to other rhabdoviruses in some respects, but some interesting variations in gene organization are also evident (Fig. 2). For instance, in the prototype VSV, the 3′-most gene (excluding the leader RNA gene) is the N gene, which is followed by genes for the P (or NS), M, G, and L proteins, respectively, which are all structural proteins (4). A distinct gene that encodes a nonstructural protein is not present. However, Herman (17) recently identified a sixth VSV protein that is synthesized by an in-frame internal initiation of translation on the P mRNA. Variation in the complexity of rhabdovirus genes was first detected with the fish rhabdovirus, IHNV, which is similar to SYNV both in the order of its structural genes and in the presence of a sixth gene that encodes a NV polypeptide (18). The IHNV NV protein is present in infected cells but not in mature virions (19). Although the number of genes in SYNV is similar to that in IHNV, their location differs because the IHNV NV gene lies between the G and L genes, while the SYNV sc4 gene is positioned between the M2 and M1 genes. Another variation is exemplified by the genome of rabies virus, which has the same order of structural genes as the other rhabdoviruses but differs from SYNV, IHNV, and VSV by having a 417-nucleotide intergenic region between the G and L genes (20). Sequence comparisons have prompted speculation that this intergenic region is a vestigial, or remnant, NV gene ("rg") that was derived from an ancestor with six genes. Therefore, the available structural data show that the rhabdoviruses are all similar in having a "leader" RNA and the N and P protein genes near the 3′ terminus with the G and L protein genes toward the 5′ terminus. But it is clear from the results with SYNV and IHNV that rhabdoviruses vary both in the number of encoded genes and in the presence and locations of nonstructural genes. Thus, the presence of a sixth gene capable of encoding a functional protein in the SYNV and IHNV genomes represents major differences between the VSV and rabies virus genomes. Moreover, the different relative positions of the sc4 gene in SYNV and the NV gene of IHNV show that rhabdoviruses have utilized different strategies of genome organization during evolution. Whether these different patterns of genome organization are significant in controlling gene expression and replication is not yet clear.

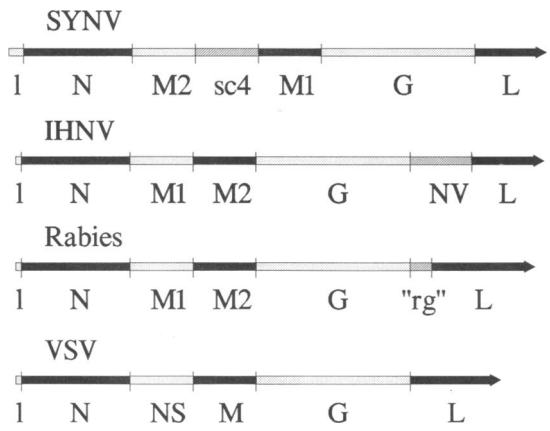

FIG. 2. Comparison of the genome maps of SYNV, IHNV, rabies virus, and VSV. The viruses have similar orders of their putative structural proteins. Similar patterns in the maps correspond to genes that are thought to encode structural proteins of similar function. However, there is no evidence for similarity of the functions of sc4, NV, or rg.

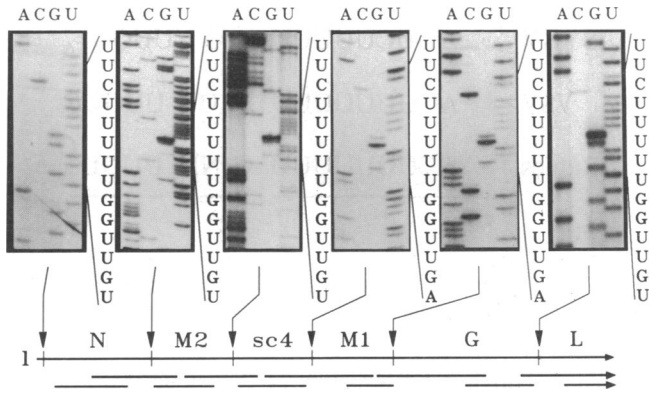

FIG. 3. Analyses of sequences at each of the SYNV gene junctions. Oligonucleotide primers were used to initiate dideoxynucleotide sequencing of single-stranded DNA templates that contained gene junction sequences. The nucleotide sequences at the gene junctions are illustrated and their positions on the genome are indicated on the map below.

Nevertheless, these variations may form a useful framework for defining taxonomic and evolutionary relationships of plant and animal rhabdoviruses as more comprehensive analyses of the genomes of other members of the group become available.

Nucleotide Sequence Analysis of the Gene Junction Regions. The 14-base sequence (3′-UUCUUUUUGGUUGU-5′) that separates the leader RNA from the 3′ terminus of the N protein gene (6) was previously found to be repeated between the N and M2 and the M2 and sc4 genes (8). Because this 14-base motif is identical to a portion of the VSV gene junction consensus sequence, we sequenced the inserts of the gene junction clones separating the remaining genes (Fig. 3). Our results show that the nucleotide sequences at each of the SYNV gene junctions are nearly identical to one another (Fig. 4). It is of particular interest from a comparative virology perspective that these sequences are also similar to those of VSV and rabies virus (Fig. 5). In all three viruses, the gene junction sequences consist of three distinct components. These include a putative signal for termination of transcription (or processing) and polyadenylylation, a short untranslated gene spacer, and a signal for the start of transcription of the downstream gene. The region that we believe signals termination of transcription of each of the SYNV genes consists of a uridylate-rich octanucleotide (3′-UUCUUUUU-5′) that is probably a reiterative template for the transcription of poly(A) tails at the 3′ ends of the mRNAs. The SYNV genes are separated by dinucleotide (GG) spacers that are not transcribed into N (7), M2 (8), M1 (unpublished data), sc4, or G mRNAs (unpublished data). All of the SYNV mRNAs begin with the tetranucleotide (3′-AACU/A-5′), which is similar to the first four nucleotides of all known

leader	UUUCUUUUU GG UUGUA	N
N	AUUCUUUUU GG UUGUC	M2
M2	AUUCUUUUU GG UUGUC	sc4
sc4	AUUCUUUUU GG UUGAA	M1
M1	AUUCUUUUU GG UUGAA	G
G	AUUCUUUUU GG UUGUA	L
	1 2 3	

FIG. 4. Comparison of the sequences at the SYNV gene junctions. The underlined nucleotides of the SYNV gene junctions are as follows: 1, the putative polyadenylylation signal at the 5′ terminus of each gene; 2, the untranslated intergenic sequence; 3, the transcription initiation site at the 3′ terminus of each gene.

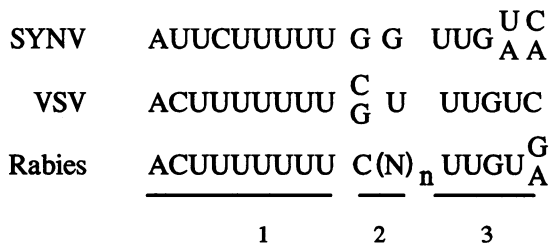

FIG. 5. Comparison of gene junction sequences of SYNV, VSV, and rabies virus. The underlined sequences are as described in Fig. 4.

animal rhabdovirus mRNAs (20, 24–26). Four of the SYNV mRNAs begin with AACA, while the remaining two begin with AACU (Fig. 4).

Consensus sequences similar to rhabdovirus gene junction sequences have been reported to exist in the human parainfluenza virus type 3 and other paramyxoviruses (27, 28), but, like the rhabdoviruses, the number of genes and nucleotide sequences that separate the genes vary somewhat (29). The present data showing nearly identical gene junctions of plant and animal rhabdoviruses (Fig. 5) thus provide a persuasive argument that plant and animal rhabdoviruses and the nonsegmented negative-strand viruses utilize very similar mechanisms to regulate transcription. These similarities and differences in the presence and locations of nonstructural genes also imply that the plant rhabdoviruses (and other nonsegmented negative-strand viruses) are closely related and that members of the group may not have been subjected to prolonged periods of separation during evolution. One plausible means whereby rhabdoviruses may have been able to cross boundaries separating the taxa relates to the ability of some members of the group to multiply in and be transmitted by insects that feed on plant and vertebrate hosts.

We thank Ms. Joan Quay and Julie Sun for technical assistance and photography and Mr. Bruce Malcolm (University of California Biochemistry Department) for synthesis of oligonucleotides. Drs. Becky Boston, Peter Goldsbrough, David Marks, and T. J. Morris provided especially helpful discussions during the course of the work. We also appreciate thoughtful reviews of the manuscript provided by Drs. T. J. Morris, Brian Staskawicz, Rick Jones, Dave Knorr, and Maureen Whalen. This research was supported by Grant DMB-87 16467 from the National Science Foundation.

1. Matthews, R. E. F. (1982) *Intervirology* **17**, 1–199.
2. Jackson, A. O., Francki, R. I. B. & Zuidema, D. (1987) in *The Rhabdoviruses*, ed. Wagner, R. R. (Plenum, New York), pp. 427–508.
3. Shope, R. E. & Tesh, R. B. (1987) in *The Rhabdoviruses*, ed. Wagner, R. R. (Plenum, New York), pp. 509–534.
4. Wagner, R. R. (1987) in *The Rhabdoviruses*, ed. Wagner, R. R. (Plenum, New York), pp. 427–508.
5. Rezaian, M. A., Heaton, L. A., Pederson, K., Milner, J. J. & Jackson, A. O. (1983) *Virology* **131**, 221–229.
6. Zuidema, D., Heaton, L. A., Hanau, R. & Jackson, A. O. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5019–5023.
7. Zuidema, D., Heaton, L. A. & Jackson, A. O. (1987) *Virology* **159**, 373–380.
8. Heaton, L. A., Zuidema, D. & Jackson, A. O. (1987) *Virology* **161**, 234–241.
9. Jackson, A. O. (1978) *Virology* **87**, 172–181.
10. Wagner, R. R., Prevec, L., Brown, F., Summers, D. F., Sokol, F. & Mcleoud, R. (1972) *J. Virol.* **10**, 1228–1230.
11. Abraham, G. & Banerjee, A. K. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1054–1508.
12. Ball, L. A. & White, C. N. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 442–446.
13. Schubert, M., Harmison, G. G. & Meier, E. (1984) *J. Virol.* **51**, 505–514.
14. Gallione, C. J., Greene, J. R., Iverson, L. E. & Rose, J. K. (1981) *J. Virol.* **39**, 529–535.
15. Rose, J. K. & Gallione, C. J. (1981) *J. Virol.* **39**, 519–528.
16. McGeoch, D. J. & Dolan, A. (1979) *Nucleic Acids Res.* **6**, 3199–3211.
17. Herman, R. C. (1986) *J. Virol.* **58**, 797–804.
18. Kurath, G. & Leong, J. C. (1985) *J. Virol.* **53**, 462–468.
19. Kurath, G., Ahern, K. G., Pearson, G. D. & Leong, J. C. (1985) *J. Virol.* **53**, 469–476.
20. Tordo, N., Poch, O., Ermine, A., Kieth, G. & Rougeon, F. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3914–3918.
21. Rose, J. K. (1980) *Cell* **19**, 415–421.
22. Jackson, A. O. & Christie, S. R. (1977) *Virology* **77**, 344–355.
23. Milner, J. J. & Jackson, A. O. (1979) *Virology* **97**, 90–99.
24. Kurilla, M. G., Cabradilla, C. D., Holloway, B. P. & Keene, J. D. (1984) *J. Virol.* **50**, 773–778.
25. Roy, P., Gupta, D. C. & Kiuchi, A. (1984) *Virus Res.* **1**, 189–202.
26. Rose, J. K. (1978) *Cell* **14**, 345–353.
27. Spriggs, M. K. & Collins, P. L. (1986) *J. Virol.* **59**, 646–654.
28. Elango, N., Varsanyi, T. M., Kovamees, J. & Norrby, E. (1988) *J. Gen. Virol.* **69**, 2893–2900.
29. Johnson, P. R. & Collins, P. L. (1988) *J. Gen. Virol.* **69**, 2901–2906.