Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 2703–2707, May 1987
Biochemistry

Segment-specific and common nucleotide sequences in the noncoding regions of influenza B virus genome RNAs

(viral transcription/viral assembly/inverted complementarity)

MARK Y. STOECKLE*, MICHAEL W. SHAW†, AND PURNELL W. CHOPPIN‡

Laboratory of Virology, The Rockefeller University, New York, NY 10021

Contributed by Purnell W. Choppin, January 7, 1987

ABSTRACT **The nucleotide sequences of the 3′ noncoding regions of all eight segments of influenza B virus RNA and the sequences of the 5′ noncoding regions of segments 4–8 were determined in virus strains isolated over a period of 40 years. Nearly complete conservation of the noncoding sequences was found. Nine nucleotides at the 3′ termini and 11 nucleotides at the 5′ termini were common to all segments examined. In the region immediately adjacent to the common 3′ terminal region, the nucleotides were specific for each segment and these segment-specific sequences were conserved in all strains examined. In each of the five segments in which both termini were examined, the segment-specific 3′ sequences exhibited perfect inverted complementarity to a segment-specific sequence adjacent to the common 5′ terminus. In addition, in the 3′ noncoding region of RNA segments 1–3, which encode proteins involved in RNA synthesis, a single nucleotide substitution at position 10 was found that distinguishes these segments from segments 4–8. Comparison of these data with published reports has revealed that some of the features found in the noncoding regions of influenza B virus are also present in influenza A and C virus RNAs. In the RNAs of all three virus types, there is a segment-specific sequence of nucleotides near the 3′ terminus that shows inverted complementarity to a sequence near the 5′ terminus. This segment-specific sequence may play a role in the transcription of individual segments or in sorting of segments during virion assembly.**

The genome of influenza B virus consists of eight single-stranded negative-sense RNA segments (1). In each of the six segments for which sequences have been reported, the protein coding region is flanked by noncoding sequences (2–7). Each segment has 30–60 noncoding nucleotides at the 3′ terminus and 30–100 noncoding nucleotides at the 5′ terminus. Although the functions of the noncoding regions are not yet known, they are presumed to be essential for viral replication. Because of the location of noncoding sequences at the termini of each segment where the synthesis of new positive- and negative-sense strands is initiated (1), it has been suggested that they may be involved in transcription (8–10). Investigation of the organization of the noncoding regions of influenza B virus RNAs may provide insight into mechanisms of viral replication and assembly that may be applicable to negative-strand viruses in general.

Early sequencing studies of influenza B virus RNAs indicated that the extreme 3′ and 5′ termini were nearly identical in all segments (8–10). However, more heterogeneity was reported in the terminal sequences of influenza B virus RNAs than in influenza A virus RNAs when the published sequences of different isolates were compared (11). To investigate further the organization of influenza B virus noncoding regions, we determined the 3′ noncoding sequences of all eight RNA segments and the 5′ noncoding sequences of segments 4–8 in strains isolated over 40 years.

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.

MATERIALS AND METHODS

Viral RNA Purification. Influenza B virus isolates Lee/1940 (Lee/40), Great Lakes/1760/1954 (GL/54), Maryland/1959 (MD/59), and North Dakota/1983 (ND/83) were grown in 10-day-old embryonated chicken eggs. Virus was collected from allantoic fluid and viral RNA was purified by phenol/chloroform extraction as described (12). RNA was isolated from virus-infected HKCC cells by CsCl centrifugation (13). Poly(A)$^+$ RNA was selected by oligo(dT) chromatography (14).

Gel Electrophoresis of Viral RNA. Viral RNA segments were separated by electrophoresis through 2.8% polyacrylamide gels containing 6 M urea (12). RNA was visualized with ethidium bromide and segments 1–3 were excised individually and the RNA was eluted (14).

Dideoxy Sequencing. Nucleotide sequences of the terminal regions of segments 5–8 were determined by using synthetic 15-mer oligonucleotides based on published Lee/40 sequences (2–6) to prime dideoxy-chain-terminator reactions (15) with viral RNA as template. Sequences of the 5′ terminal regions were obtained by using viral RNA (10 μg per set of reactions) and the 3′ terminal sequences were obtained by using mRNA (50 μg per set of reactions), as 5′ mRNA is complementary to 3′ viral RNA. For segments 1–3, partial 3′ sequences were obtained by using gel-purified viral RNA and a 10-mer oligonucleotide (AGCAGAAGCG) complementary to the 3′ termini. Oligonucleotides based on these sequences were then used to prime the sequencing reactions with mRNA.

RESULTS

3′ Terminal Sequences. The nucleotide sequences of cDNAs corresponding to the 3′ noncoding regions of each of the eight segments are shown in message sense in Fig. 1. Heterogeneity was found in the cDNA sequences upstream of the G at position 2, reflecting heterogeneity in the host-cell mRNA primer (16) and thus the terminal A shown for each sequence is based on published reports (8–10). The assignments for segments 1 and 3, for which sequences have not been previously reported for any influenza B virus strain, are based on their order of migration on the gel system used to separate viral RNA. In segments 2 and 4, differences between the data and published sequences of the same isolate (Lee/40) were found and these are marked in Fig. 1. The

*Present address: Laboratory of Viral Oncology, The Rockefeller University, New York, NY 10021.
†Present address: Influenza Branch, Centers for Disease Control, Atlanta, GA 30333.
‡Present address: Howard Hughes Medical Institute, Bethesda, MD 20814.

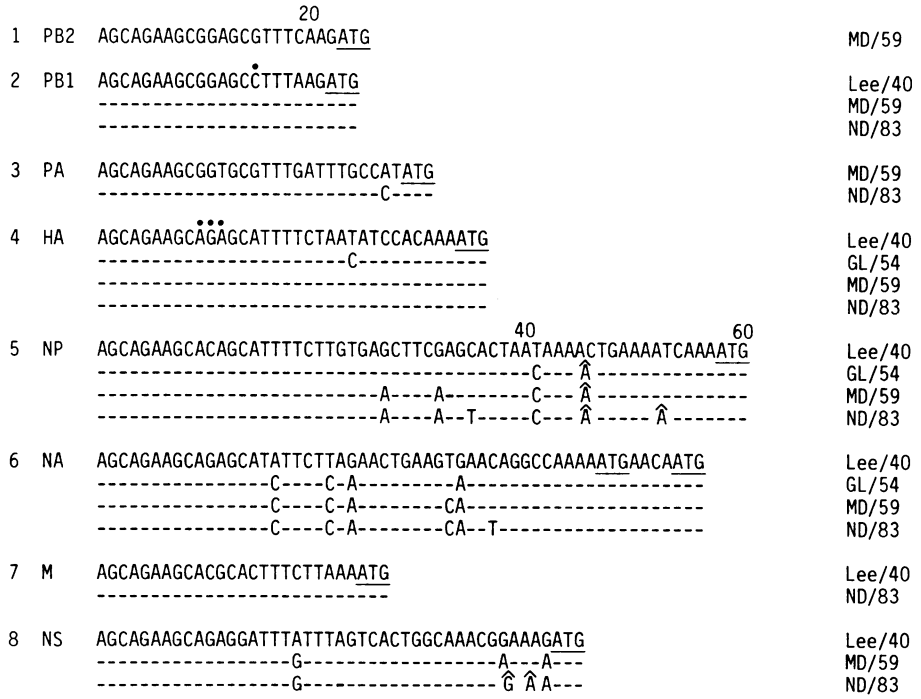

FIG. 1. Nucleotide sequences of cDNAs corresponding to the 3′ terminal regions of influenza B virus RNA segments as determined by primer-extension sequencing of viral mRNA. Sequences are shown in message sense 5′ to 3′. Initiation codons are underlined. Substitutions relative to the earliest isolate (top line) are indicated by the appropriate letters. -, Conservation; ∧, insertions; •, differences between the data and other reported sequences. The tentative assignments for segments 1 and 3 are based on the order of migration of these segments in gel electrophoresis. The terminal A shown for each sequence is based on published reports of RNA sequencing (8–10); heterogeneity is present in cDNA sequences at this position (16). PB1 and PB2, basic polymerase proteins 1 and 2; PA, acidic polymerase protein; HA, hemagglutinin; NP, nucleoprotein; NA, neuraminidase; M, membrane protein; NS, nonstructural proteins NS1 and NS2.

sequences of two or more isolates were determined for seven of the eight segments. For most segments, viral strains isolated over a period of 40 years were compared. Nearly complete conservation of the terminal sequences was found. Only two segments showed substitutions in the first 22 nucleotides (two in segment 6, one in segment 8). Two features of this conserved terminal region were common to all segments examined. The first 9 nucleotides at the 3′ terminus were identical and there was a T-rich region beginning with a T at position 16 and two to five more T residues at positions 17–22 in all segments (Fig. 2). The nucleotide sequence between these two common regions was different in each segment with the exception of segments 4 and 6, which showed the same sequence in this region. These segments encode the two surface glycoproteins hemagglutinin and neuraminidase. No substitutions in the 9 common nucleotides at the 3′ terminus or in the six segment-specific nucleotides adjacent to the common terminus were found in any of the isolates examined. Segments 1–3 were found to differ from segments 4–8 in having a G at position 10 where segments 4–8 have an A (Fig. 2). Thus, precise conservation of the 3′-terminal nucleotide sequences was found, including sequences common to all segments and sequences specific for each segment.

5′ Terminal Sequences. The nucleotide sequences of cDNAs corresponding to the 5′ noncoding regions of segments 4–8 of the ND/83 isolate were determined and are shown in message sense as compared to published sequences of the Lee/40 isolate (Fig. 3). Little divergence between the 1983 and the 1940 isolate was found. Two features were common to the 5′ noncoding region of all five segments examined (Fig. 4). The 11 terminal nucleotides were the same in each segment except at position 6. Segments 4, 6, and 8 have an A, and segments 5 and 7 have a T at this position. These differences between segments at position 6 in the 1983 isolate are also found in published sequences of the Lee/40 isolate. In addition to the common terminus, each segment examined had adenines at positions 17–22, which is known to

FIG. 2. Common features of the 3′ noncoding regions of influenza B virus RNAs. The common nucleotide sequences at the 3′ terminus and the T-rich regions beginning at position 16 are boxed. The sequences shown are the ND/83 isolate, except for segment 1, which is MD/59. Abbreviations are as in Fig. 1.

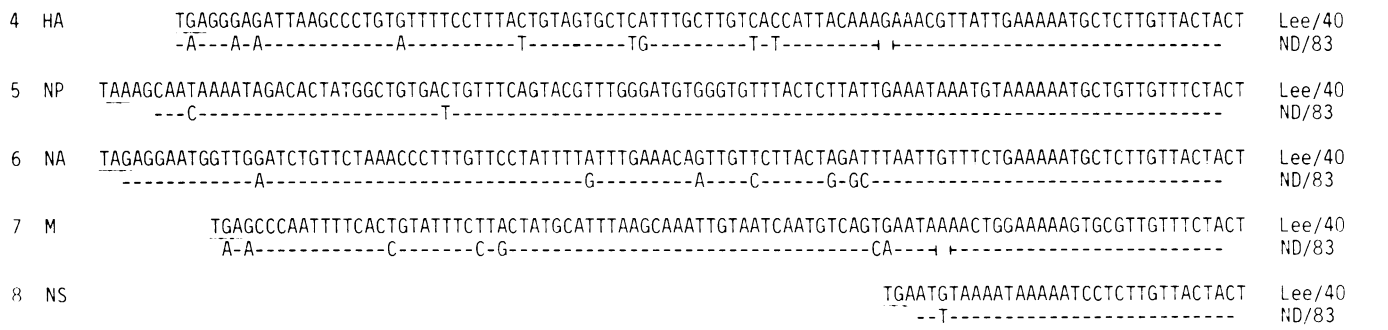

FIG. 3. Nucleotide sequences of cDNAs corresponding to the 5′ terminal regions of influenza B virus ND/83 RNAs as determined by primer-extension sequencing of viral RNA and compared to the published sequences of the Lee/40 strain (2–7). Sequences are shown in message sense 5′ to 3′. Termination codons are underlined. The substitutions in ND/83 relative to Lee/40 are shown. Deletions in ND/83 relative to Lee/40 are marked ⊣ ⊢. The terminal 1 or 2 nucleotides could not be determined by the methods used. Abbreviations are as in Fig. 1.

be the site for polyadenylylation during mRNA synthesis (17). The nucleotide sequences between these two common regions were different in each segment, except for segments 4 and 6, which were identical. No substitutions in either common region or the intervening segment-specific sequences were found when the data were compared to published sequences of the Lee/40 isolate. Further homologies between segments 4–7 were found in the noncoding regions upstream of the polyadenylylation site. Each of these segments was found to have a T-rich region near positions 50 and 70 (Fig. 4).

Inverted Complementarity of 3′ and 5′ Termini. The segment-specific sequences adjacent to the common 3′ and 5′ termini showed perfect inverted complementarity in all five segments in which both termini were examined. This complementarity is revealed when the 3′ and 5′ termini are aligned in a staggered fashion (Fig. 5). The complementary region begins at position 10 at the 3′ terminus and position 11 at the 5′ terminus and extends into the common 3′ T-rich region and the common 5′ polyadenylylation site. No substitutions that affect the complementarity of these regions were found in any isolates.

DISCUSSION

The noncoding regions of influenza virus RNAs are thought to contain sequences essential for the replication of the virus (1, 8–10). Analysis of the organization of the noncoding regions may provide insight into the mechanisms of viral replication. In this study, we determined the nucleotide sequences of the 3′ noncoding regions of all eight influenza B virus RNA segments and of the 5′ noncoding regions of segments 4–8 in virus strains isolated over a 40-year period. The data obtained include 3′ sequences of RNA segments 1 and 3 for an influenza B virus strain.

Precise conservation of the terminal noncoding sequences of the RNA segments examined was found. This suggests that the specific sequence is important. If it were not, one would expect nucleotide substitutions to accumulate over time. Divergence in the nucleotide sequences of the coding regions of influenza virus strains is well-documented (1). The frequency of nucleotide substitution in influenza A virus based on a comparison of nonstructural proteins NS1 and NS2 genes is estimated to be 2×10^{-3} per site per year (18). If this frequency applied to the terminal noncoding regions of influenza B virus, one would expect two substitutions (per 20 nucleotides) at each terminus of each segment in comparing the 1983 and 1940 isolates. Instead, no substitutions were found at the 5′ termini and only two segments showed substitutions at the 3′ termini. The sequence conservation found in the strains examined in this study suggests that some of the heterogeneity reported in previously published influenza B virus sequences (11) may represent sequencing errors and not true variation. In two cases, differences between the present results and published sequences of the same isolate were found. In one case, the reported sequence was determined by RNA sequencing (6). In the other case, the sequence was obtained by analysis of cloned cDNA (5). Primer extension sequencing of viral RNA, as was done in this study, may have avoided some errors that may occur with other methods.

Several features common to all RNA segments of the influenza B virus strains examined were found in the present study. Nine nucleotides at the 3′ terminus and 11 nucleotides at the 5′ terminus (except position 6) were common to all segments examined. This is a shorter common sequence than found in influenza A virus RNAs, which have been reported to have 12 common 3′ nucleotides (8–10).

In addition to the common 3′ terminal sequence, each RNA segment was found in this study to have a T-rich region beginning at position 16. This finding has not been previously reported for influenza B or A viruses. Each RNA segment examined was also found to have an adenine tract at positions 17–22 at the 5′ terminus that is present in all influenza virus RNAs sequenced to date and is the site for polyadenylylation during mRNA synthesis (17). In addition, further homologies were found in the 5′ noncoding regions of segments 4–7, upstream of the polyadenylylation site. These segments each

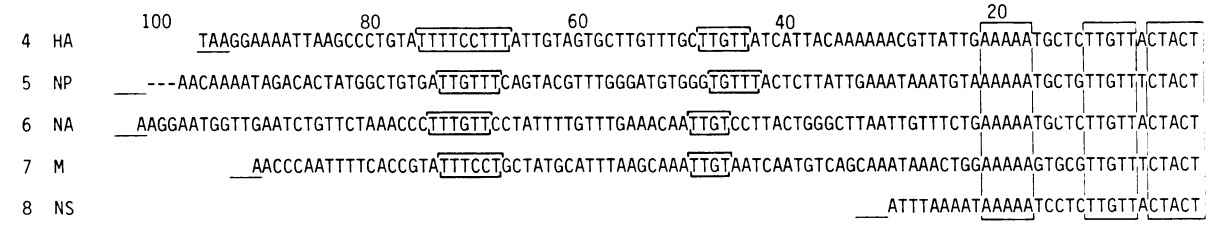

FIG. 4. Common features of the 5′ noncoding regions of influenza B virus RNAs. The sequences shown are the ND/83 isolate. Large boxes enclose the common terminal nucleotides and the polyadenylylation sites. Smaller boxes indicate the T-rich sequences found in segments 4–7. Abbreviations are as in Fig. 1.

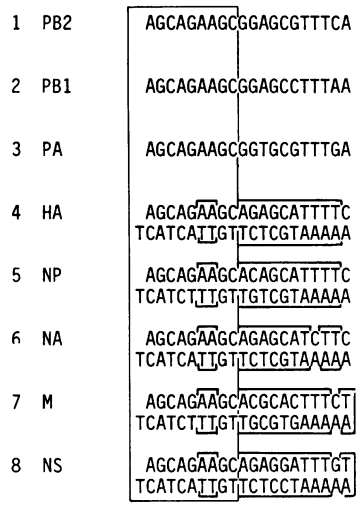

FIG. 5. Inverted complementarity of the segment-specific sequences adjacent to the common termini of influenza B virus RNAs. To demonstrate complementarity of the segment-specific sequences, the 3′ termini are indented one nucleotide over the 5′ termini. Smaller boxes enclose the complementary nucleotides in each segment. The common 3′ and 5′ termini are enclosed in the large box. All sequences are of the ND/83 isolate except segment 1 (MD/59). Abbreviations are as in Fig. 1.

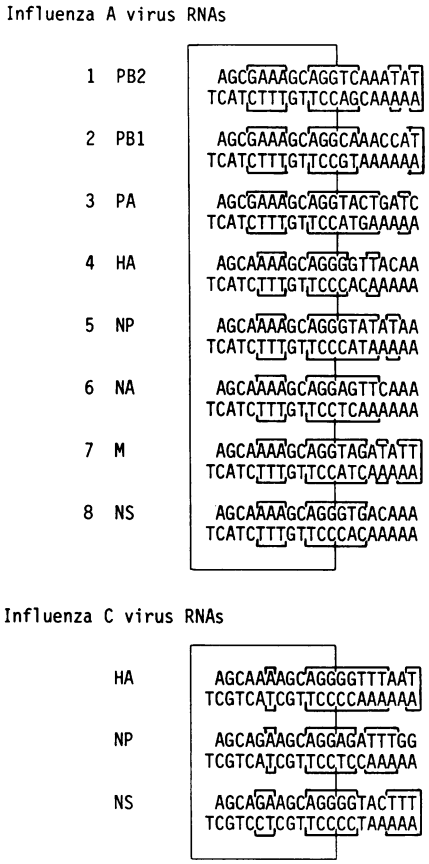

FIG. 6. Inverted complementarity of segment-specific sequences in influenza A (9) and C (19–21) virus RNAs. As in Fig. 5, all sequences are shown in message sense with the nucleotide sequence representing 3′ terminal viral RNA indented 1 base over the 5′ terminal sequence. Smaller boxes enclose the complementary nucleotides found in each segment. The common 3′ and 5′ termini are enclosed in a large box. Abbreviations are as in Fig. 1.

have a T-rich sequence interspersed by one or two Gs or Cs at positions 50 and 70. The possible function of this repeated sequence is not known. One can speculate that the T-rich regions are binding sites for nucleoprotein during viral capsid assembly. Another possibility is that the T-rich sequences upstream of the polyadenylylation site play a role in triggering polyadenylylation of mRNA. One observation apparent from Fig. 3 is that the 5′ noncoding region of segment 8 is much shorter than that in segments 4–7, at least in the Lee/40 strain. Further sequencing is needed to determine whether the length of the segment 8 5′ noncoding region is conserved in other strains.

In addition to these common sequences, precisely conserved segment-specific nucleotide sequences adjacent to the common 3′ and 5′ termini of each segment were found. The only exception was that segments 4 and 6 showed identical sequences in these regions. These conserved segment-specific sequences adjacent to the common termini showed perfect inverted complementarity in all five segments in which both termini were examined. In addition, we found that the nucleotide at position 10 distinguishes segments 1–3 from segments 4–8; in the former it is a G and in the latter it is an A.

Comparison of the present data with published reports on other influenza viruses reveals some interesting similarities. Segment-specific sequences adjacent to the common termini that show inverted complementarity were noted in published sequences of influenza A virus RNAs (8–10). We have examined published sequences of influenza C virus RNAs (19–21) and found segment-specific sequences with inverted complementarity adjacent to the common termini (Fig. 6). We have found that the inverted complementarity of the segment-specific nucleotides in all three virus types is revealed when the termini are aligned with the 3′ terminal sequence indented one nucleotide over the 5′ terminal sequence (Figs. 5 and 6). Thus, we have noted that influenza A, B, and C virus RNAs all have segment-specific sequences adjacent to the common termini that show inverted complementarity in addition to the previously recognized partial inverted complementarity of the common termini. It is not known whether base-pairing of the termini of influenza virus RNAs actually occurs during viral replication or assembly. Several theoretical base-pairing interactions between and within segments have been suggested (8). The results obtained here indicate that interaction between the termini of each segment may be more energetically favorable than previously recognized. A similar pattern of organization of the terminal noncoding regions is found in other segmented negative-strand RNA viruses. Inverted complementarity of common and segment-specific terminal noncoding nucleotide sequences has been noted in bunyaviruses (22) and may also be present in arenaviruses (23).

One consequence of segment-specific inverted complementarity is that inspection of a short region (3–6 nucleotides) near the 3′ terminus is sufficient to identify the segment, and the exact location of the sequence indicates whether it is a plus or minus strand. Such an identifying sequence might be important in viral replication. Two processes that appear to require differentiation or sorting of segments are RNA transcription and virion assembly. An identifying sequence present at the 3′ termini of the plus and minus strands, which are the presumed polymerase entry sites (8–10), could be important in regulating rates of transcription. In this regard, it may be significant that segments 1–3 differ from the others by having a G in position 10, where other segments have an A. The P proteins encoded by these segments are the least abundant viral proteins in the infected cell, and this paucity correlates with low amounts of mRNAs encoding these proteins (24, 25). One can speculate that a G at this position may influence the level of mRNA synthesis from these segments. Some reports indicate that influenza A virus segments 1–3 differ from the other segments by having

a G at position 4, where other segments have an A (9). In most of the studies in which this difference was not found, the terminal sequence was not determined directly. It thus appears that influenza A and B virus RNA segments 1–3 share this nucleotide difference. Each has a G in the sequence AGCG, where the other segments have an A in the sequence AGCA, although the position of this tetranucleotide differs in the two virus types. It is also interesting that influenza B virus segments 4 and 6, which encode the glycoproteins hemagglutinin and neuraminidase, are the only two segments that have an identical sequence in the region between the common 3′ terminus and the T-rich region. This identity may lead to similar handling of their RNAs, perhaps by directing hemagglutinin and neuraminidase mRNAs to the appropriate ribosomal fraction for eventual targeting to the endoplasmic reticulum and Golgi apparatus. Alternatively, it is possible that the segment-specific sequences function somehow in virion assembly in facilitating the ordered packaging of RNA segments. For this to occur, the segment-specific sequences may be recognized by another factor, presumably a viral protein(s). Interaction of different segments with different domains on the same protein would obviate the need in this model for a separate recognition protein for each segment.

In summary, we have identified both common and segment-specific features in the terminal noncoding regions of influenza B virus RNAs. The findings suggest hypothetical functions for these noncoding regions. Further experiments with chimeric RNAs combining the noncoding and coding regions of different segments and cell-free transcription and translation systems may help in exploring the functions of the noncoding regions of influenza virus RNAs and may be applicable to segmented negative-strand RNA viruses in general.

We thank Elizabeth Esposito for excellent technical assistance. B/Maryland/1959 was provided by Dr. Brian Murphy and B/North Dakota/83 was provided by Dr. Alan Kendal. This work was supported by Grant AI-05600 from the National Institutes of Health. M.Y.S. is the recipient of a Physician's Research Training Fellowship from the American Cancer Society.

1. Lamb, R. A. & Choppin, P. W. (1983) *Annu. Rev. Biochem.* **52,** 467–506.
2. Briedis, D. J., Lamb, R. A. & Choppin, P. W. (1982) *Virology* **116,** 581–588.
3. Briedis, D. J. & Lamb, R. A. (1982) *J. Virol.* **42,** 186–193.
4. Briedis, D. J. & Tobin, M. (1984) *Virology* **133,** 448–455.
5. Kendirim, S., Palefsky, J. & Briedis, D. J. (1986) *Virology* **152,** 126–135.
6. Krystal, M., Elliott, R. M., Benz, E. W., Jr., Young, J. F. & Palese, P. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4800–4804.
7. Shaw, M. W., Lamb, R. A., Erickson, B. W., Briedis, D. J. & Choppin, P. W. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6817–6821.
8. Desselberger, U., Racaniello, V. R., Azara, J. J. & Palese, P. (1980) *Gene* **8,** 315–328.
9. Robertson, J. S. (1979) *Nucleic Acids Res.* **6,** 3745–3756.
10. Skehel, J. J. & Hay, A. J. (1978) *Nucleic Acids Res.* **5,** 1207–1219.
11. Air, G. M. & Compans, R. W., eds. (1983) in *Genetics of Influenza Viruses* (Springer, New York), pp. 280–304.
12. Palese, P. & Schulman, J. L. (1976) *J. Virol.* **17,** 876–884.
13. Glisin, V., Crkvenjakov, R. & Byus, C. (1974) *Biochemistry* **13,** 2633–2637.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
16. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. (1981) *Cell* **23,** 847–858.
17. Robertson, J. S., Schubert, M. & Lazzarini, R. A. (1981) *J. Virol.* **38,** 157–163.
18. Buonagurio, D. A., Naka, S., Parvin, J. D., Krystal, M., Palese, P. & Fitch, W. M. (1986) *Science* **232,** 980–982.
19. Nakada, S., Creager, R. S., Krystal, M., Aaronson, R. P. & Palese, P. (1984) *J. Virol.* **50,** 118–124.
20. Nakada, S., Creager, R. S., Krystal, M. & Palese, P. (1984) *Virus Res.* **1,** 433–441.
21. Nakada, S., Graves, P. N., Desselberger, U., Creager, R. S., Krystal, M. & Palese, P. (1985) *J. Virol.* **56,** 221–226.
22. Obijeski, J. F., McCauley, J. & Skehel, J. J. (1980) *Nucleic Acids Res.* **8,** 2431–2438.
23. Auperin, D. D., Romanowski, V., Falinski, M. & Bishop, D. H. L. (1984) *J. Virol.* **52,** 897–904.
24. Barrett, T., Wolstenholme, A. J. & Mahy, B. W. J. (1979) *Virology* **98,** 211–225.
25. Smith, G. L. & Hay, A. J. (1982) *Virology* **118,** 96–108.