JOURNAL OF VIROLOGY, Sept. 1988, p. 3348–3355
0022-538X/88/093348-08$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 62, No. 9

Evidence that the Matrix Protein of Influenza C Virus Is Coded for by a Spliced mRNA

MAKOTO YAMASHITA,† MARK KRYSTAL, AND PETER PALESE*

Department of Microbiology, Mount Sinai School of Medicine, City University of New York, Fifth Avenue and 100th Street, New York, New York 10029

Received 6 April 1988/Accepted 2 June 1988

In contrast to influenza A and B viruses, which encode their matrix (M) proteins via an unspliced mRNA, the influenza C virus M protein appears to be coded for by a spliced mRNA from RNA segment 6. Although an open reading frame in RNA segment 6 of influenza C/JJ/50 virus could potentially code for a protein of 374 amino acids, a splicing event results in an mRNA coding for a 242-amino-acid M protein. The message for this protein represents the major M gene-specific mRNA species in C virus-infected cells. Despite the difference in coding strategies, there are sequence homologies among the M proteins of influenza A, B, and C viruses which confirm the evolutionary relationship of the three influenza virus types.

Influenza viruses are classified into three types, A, B, and C, on the basis of antigenic differences among the internal proteins of the viruses. The genomes of influenza A and B viruses consist of eight segments of negative-sense RNA, whereas influenza C viruses possess only seven negative-sense RNA segments (1, 16, 17, 27). Although influenza C viruses lack a protein equivalent to the neuraminidase of A and B viruses, a receptor-destroying enzyme is found to be associated with the HE surface glycoprotein (35). The latter protein thus possesses both receptor-binding and receptor-destroying activities (14, 15, 26, 28, 35). Previously, we have analyzed RNA segments 4, 5, and 7 of influenza C viruses, which code for the NP, HE, and NS proteins, respectively (22–24). The coding strategies of these RNA segments were found to be similar to those of the related influenza A and B viruses (1, 17).

In the present study, RNA segment 6 of influenza C/JJ/50 virus was cloned, sequenced, and shown to code for the matrix (membrane; M) protein of the virus. The M RNA segment possesses one open reading frame which could code for a protein of over 40 kilodaltons (kDa) (374 amino acids). However, the M protein of C viruses is smaller than 40 kDa when analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (approximately 27 kDa). This apparent discrepancy is explained by evidence reported here, which suggests that the mRNA coding for the matrix protein undergoes a splicing event that removes a portion of the open reading frame. This finding contrasts with the situation in the influenza A and B viruses, which possess matrix proteins that are translated from unspliced mRNA molecules (2, 7, 36). In the case of influenza A viruses, a spliced mRNA is used to code for a small integral membrane protein, M2 (18–20).

MATERIALS AND METHODS

Viruses and cells. Influenza C/JJ/50 virus was grown in the amniotic sacs of 11-day-old embryonated chicken eggs at 35°C. Virus purification and RNA extraction have been described previously (30). Influenza C virus proteins were labeled in MDCK cells under hypertonic medium conditions as reported previously (25, 35). Extraction of RNA from purified virus and from virus-infected MDCK cells (17 h postinfection) was accomplished according to the procedures described previously (3). Selection of poly(A)$^+$ RNA was achieved through multiple passages over oligo(dT) columns (3).

Cloning and sequencing of virus-specific RNA. Double-stranded cDNA was obtained from influenza C/JJ/50 virus RNA by using mouse mammary tumor virus-cloned reverse transcriptase according to the recommendations of the manufacturer (Bethesda Research Laboratories, Inc., Gaithersburg, Md.). The DNA primer used for first-strand cDNA synthesis was a mixed oligonucleotide [AGCA(GA)AAG CAGG] corresponding to positions 1 through 12 at the 3′ end of all influenza C virus RNAs (11). Virus-specific double-stranded cDNA was ligated into the *Bam*HI site of plasmid pIBI30 (International Biotechnologies, Inc., New Haven, Conn.) after addition of synthetic *Bam*HI linkers (CGCG GATCCGCG). *Escherichia coli* DH5α cells were then transformed with the DNA, and plasmids containing RNA segment 6-specific cDNA were identified by Northern blot (RNA blot) analysis. The plasmid containing the longest segment 6-specific cDNA clone was identified (pCM144), and the insert was recloned into M13 bacteriophage. Sequencing of the cDNA insert was done on the original phage DNA and deletion subclones, which were created by nonrandom cloning of *Bal* 31-digested insert DNA (29). Deletion subclones were created by first digesting pCM144 with *Eco*RI, which only cuts in the multiple cloning site. After digestion with *Bal* 31 for various times, insert DNA was excised with *Bam*HI, purified, and inserted into the *Bam*HI-*Sma*I window of M13mp19 phage. Sequencing was accomplished by using the modified T7 polymerase, sequenase (United States Biochemical Corp., Cleveland, Ohio), according to the specifications of the manufacturer. The 5′ region of the viral RNA (positions 1103 through 1180) not present in pCM144 was directly sequenced by the dideoxy chain termination method by using the synthetic oligonucleotide primer GATGGTGTGGAGATATAAAGACCAC (positions 990 through 1014) and reverse transcriptase (7, 32). All oligonucleotides were chemically synthesized by using an Applied Biosystems 380A DNA synthesizer (Applied Biosystems, Foster City, Calif.).

* Corresponding author.

† Visiting research scientist from Bioscience Research Laboratories, Sankyo Co. Ltd., 1-2-58, Hiromachi, Shinagawa-ku, Tokyo, Japan.

Northern blot analysis of vRNA and mRNA. Total RNA was denatured by using glyoxal and dimethyl sulfoxide (3), separated on 3% polyacrylamide gels containing 7.6 M urea, and electrophoretically transferred to Zetabind paper (CUNO Inc., Meriden, Conn.). Filters were hybridized by standard procedures with either ^{32}P-labeled nick-translated probes derived from pCM144 DNA or ^{32}P-labeled cDNA transcripts derived from virion RNA (vRNA). The latter were obtained through reverse transcription of total vRNA with the mixed oligonucleotide primer described earlier.

Translation of hybrid-selected RNA. Hybrid selection of RNA followed established procedures (9). Briefly, alkaline-denatured DNA (pCM144) immobilized on a nitrocellulose filter was used to select mRNA segment 6-specific poly(A)$^+$ RNA obtained from C virus-infected cells. RNA was hybridized to the filter at 47°C for 2 h in PIPES [piperazine-*N*-*N*′-bis(2-ethanesulfonic acid)] buffer (pH 7.5) containing 30% formamide, 0.5 M NaCl, 0.4% SDS, and 2 mM EDTA. After extensive washing of the filter, bound poly(A)$^+$ RNA was released by boiling and then translated in a rabbit reticulocyte lysate (Promega Biotec, Madison, Wis.) containing [^{35}S]methionine (800 Ci/mmol; New England Nuclear Corp., Boston, Mass.). ^{35}S-labeled proteins were analyzed by SDS-polyacrylamide gel electrophoresis (14% polyacrylamide).

S1 nuclease assay. ^{32}P-labeled RNA probes were made by using T7 polymerase and either *Bal*I-, *Pvu*II-, or *Xba*I-digested pCM144 DNA (see Fig. 5). Following purification on 3% polyacrylamide gels containing 6 M urea, the different probes were hybridized with poly(A)$^+$ RNA from infected cells at 45°C for 16 h in PIPES buffer (pH 6.4) containing 1 mM EDTA, 0.4 M NaCl, and 80% formamide. Hybrids were digested with 1 U of S1 nuclease per ml at 37°C for 1 h and analyzed on a 6.5% polyacrylamide gel containing 6 M urea (21).

Sequencing of splice site. A 2.88-μg quantity of total poly(A)$^+$ RNA from C/JJ/50 virus-infected cells was first transcribed into cDNA by using reverse transcriptase and the oligonucleotide primer CCATCGAGTCAATTTCAGG CAT (corresponding to positions 1040 through 1019 in the vRNA sense). This cDNA template then was used for a polymerase chain reaction using *Taq* polymerase (New England BioLabs, Inc., Beverly, Mass.) and oligonucleotide primers CCATCGAGTCAATTTCAGGCAT (positions 1040 through 1019 in the vRNA sense) and GCCAGCACAG CAATTAACGA (positions 557 through 576 in the mRNA sense) (31). The amplified double-stranded cDNA fragment of approximately 250 base pairs was purified on an 8% polyacrylamide gel and sequenced by the dideoxy termination method by using the oligonucleotide CCATCGAGT CAATTTCAGGCAT (positions 1040 through 1019 in the vRNA sense) as a primer.

RESULTS

Cloning and sequencing of influenza C/JJ/50 virus RNA segment 6. Reverse transcription of influenza C/JJ/50 virus RNA and cloning of double-stranded cDNA into pIBI30 yielded recombinant plasmids with virus-specific sequences. Segment 6-specific clones were identified by Northern blot analysis, and the plasmid containing the longest cDNA insert (pCM144) was isolated (Fig. 1, lanes 1 and 2). Sequencing of the insert revealed that the clone contained sequences common to the conserved 3′ ends of influenza C virus RNAs (11). The insert was 1,101 nucleotides long but did not appear to be a full-length copy of RNA 6, because it was lacking 5′-terminal sequences shown to be conserved in influenza C viruses. This was not surprising, as second-strand cDNA synthesis was accomplished through partial RNase H cleavage of the vRNA and subsequent use of the RNA as a primer for second-strand cDNA synthesis (12). In order to complete the sequence of segment 6 vRNA, a synthetic primer close to the end of the insert (positions 990 through 1014) was used for direct RNA sequencing on purified influenza C virus vRNA. From these experiments, a total length of 1,180 nucleotides was deduced for RNA segment 6 of influenza C/JJ/50 virus (Fig. 2). This segment codes for one long open reading frame, which could potentially produce a protein of 374 amino acids with a predicted molecular weight of 41,700. In the two alternative reading frames, the longest open reading frame codes for a peptide of only 41 amino acids (Fig. 3).

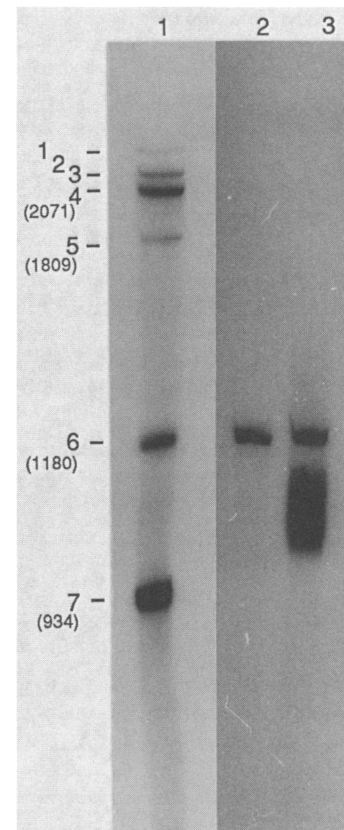

FIG. 1. Northern blot analysis of C/JJ/50 virus-specific vRNA and mRNA. Lane 1, Purified vRNA was separated by electrophoresis, transferred to a Zetabind filter, and hybridized with radioactive cDNA prepared from total vRNA; lane 2, purified vRNA hybridized with nick-translated pCM144 DNA; lane 3, poly(A)-enriched RNA from C/JJ/50 virus-infected cells hybridized with nick-translated pCM144 DNA. The numbers of the influenza C virus RNA segments are indicated to the left of the gel, and the nucleotide lengths of RNA segments 4, 5, 6, and 7 are shown in parentheses.

Coding assignment for RNA segment 6. In order to answer the question of which viral protein is encoded by RNA segment 6, mRNA from influenza C virus-infected cells was purified by hybrid selection by using pCM144 DNA bound to nitrocellulose filters. The segment 6-specific mRNA was then translated in a rabbit reticulocyte lysate (Fig. 4), and the products were analyzed on polyacrylamide gels. The only RNA 6-specific protein detected under these experimental conditions migrated in the similar position to the matrix

AGCAGAAGCAGGGGATTTCTAAACA ATG GCA CAT GAA ATA CTG ATT GCT GAA ACA GAG GCA TTT CTA AAA AAT GTT GCT CCT GAG ACC AGG ACA GCA ATA 100
Met Ala His Glu Ile Leu Ile Ala Glu Thr Glu Ala Phe Leu Lys Asn Val Ala Pro Glu Thr Arg Thr Ala Ile 25

ATT TCA GCA ATA ACA GGA GGA AAA TCA GCC TGC AAA TCA GCA GCT AAA CTG ATT AAG AAT GAA CAT CTT CCC CTA ATG TCT GGA GAA GCC ACC ACA ATG 199
Ile Ser Ala Ile Thr Gly Gly Lys Ser Ala Cys Lys Ser Ala Ala Lys Leu Ile Lys Asn Glu His Leu Pro Leu Met Ser Gly Glu Ala Thr Thr Met 58

CAC ATC GTT ATG AGG TGC TTA TAT CCT GAA ATA AAA CCA TGG AAA AAG GCA AGC GAC ATG CTG AAT AAA GCA ACT TCT AGT TTG AAA AAA TCA GAA GGA 298
His Ile Val Met Arg Cys Leu Tyr Pro Glu Ile Lys Pro Trp Lys Lys Ala Ser Asp Met Leu Asn Lys Ala Thr Ser Ser Leu Lys Lys Ser Glu Gly 91

Pvu II
AGA GAC ATA AGA AAG CAA ATG AAA GCA GCT GGA GAC TTC TTG GGA GTG GAG TCA ATG ATG AAA ATG AGG GCC TTC AGA GAT GAC CAA ATA ATG GAA ATG 397
Arg Asp Ile Arg Lys Gln Met Lys Ala Ala Gly Asp Phe Leu Gly Val Glu Ser Met Met Lys Met Arg Ala Phe Arg Asp Asp Gln Ile Met Glu Met 124

GTT GAA GAA GTA TAT GAT CAC CCA GAC GAC TAC ACA CCA GAC ATC CGA ATA GGA ACA ATC ACA GCT TGG TTG AGA TGC AAA AAC AAG AAA AGT GAA AGA 496
Val Glu Glu Val Tyr Asp His Pro Asp Asp Tyr Thr Pro Asp Ile Arg Ile Gly Thr Ile Thr Ala Trp Leu Arg Cys Lys Asn Lys Lys Ser Glu Arg 157

TAC AGG AGT AAT GTC TCA GAA AGT GGA AGA ACA GCT TTA AAA ATT CAT GAA GTA AGA AAA GCC AGC ACA GCA ATT AAC GAG ATT GCT GGT ATT ACT GGC 595
Tyr Arg Ser Asn Val Ser Glu Ser Gly Arg Thr Ala Leu Lys Ile His Glu Val Arg Lys Ala Ser Thr Ala Ile Asn Glu Ile Ala Gly Ile Thr Gly 190

Bal I
CTT GGA GAA GAA GCA CTA TCT CTC CAA AGA CAA ACA GAA AGT TTG GCC ATA TTA TGT AAT CAC ACT TTT GGA AGT AAT ATA ATG AGA CCC CAC TTG GAA 694
Leu Gly Glu Glu Ala Leu Ser Leu Gln Arg Gln Thr Glu Ser Leu Ala Ile Leu Cys Asn His Thr Phe Gly Ser Asn Ile Met Arg Pro His Leu Glu 223

AAA GCA ATA AAA GGA GTT GAA GGC AGA GTT GGA GAG ATG GGA CGA ATG GCA ATG AAA TGG TTA GTT GTT ATA ATA TGT TTC TCT ATA ACA AGT CAA CCT 793
Lys Ala Ile Lys Gly Val Glu Gly Arg Val Gly Glu Met Gly Arg Met Ala Met Lys (End) 242
Trp Leu Val Val Ile Ile Cys Phe Ser Ile Thr Ser Gln Pro

GCT TCT GCT TGC AAT CTA AAG ACC TGT CTA AAA CTA TTT AAC AAT ACT GAT GCG GTA ACT GTT CAT TGT TTT AAT GAA ACC CAA GGA TAC ATG CTA ACA 892
Ala Ser Ala Cys Asn Leu Lys Thr Cys Leu Lys Leu Phe Asn Asn Thr Asp Ala Val Thr Val His Cys Phe Asn Glu Thr Gln Gly Tyr Met Leu Thr

TTA GCC TCT TTG GGA TTA GGG ATA ATT ACT ATG TTG TAT TTA TTA GTA AAA ATC ATA ATT GAA CTT GTC AAT GGT TTT GTG CTC GGC AGA TGG GAG AGA 991
Leu Ala Ser Leu Gly Leu Gly Ile Ile Thr Met Leu Tyr Leu Leu Val Lys Ile Ile Ile Glu Leu Val Asn Gly Phe Val Leu Gly Arg Trp Glu Arg

TGG TGT GGA GAT ATA AAG ACC ACA ATT ATG CCT GAA ATT GAC TCG ATG GAA AAA GAT ATT GCC CTT TCT AGG GAG AGA CTT GAC CTG GGA GAG GAT GCT 1090
Trp Cys Gly Asp Ile Lys Thr Thr Ile Met Pro Glu Ile Asp Ser Met Glu Lys Asp Ile Ala Leu Ser Arg Glu Arg Leu Asp Leu Gly Glu Asp Ala

CCT GAC GAA ACC GAC ATC TCA CCA ATT CCT TTT TCC AAT GAT GGT ATT TTT GAA ATT TAA TTACCTTGAAAAAATCCCCTTGCTACTGCT 1180
Pro Asp Glu Thr Asp Ile Ser Pro Ile Pro Phe Ser Asn Asp Gly Ile Phe Glu Ile End

FIG. 2. Complete nucleotide sequence of segment 6 of influenza C/JJ/50 virus. The sequence is presented in cDNA form (mRNA sense). The pCM144 clone encompasses bases 2 through 1102. The uncloned 3′ specific nucleotide sequences were obtained by direct RNA sequencing of vRNA. Restriction enzyme sites used for the preparation of specific RNA probes (see Fig. 5) are indicated. Vertical arrows indicate the splice site positions for the M protein-specific mRNA. Horizontal arrows indicate the sequence of the mRNA-derived cDNA piece that was amplified by the polymerase chain reaction (see Fig. 6). The nucleotides TG and A (boxed) represent the newly generated termination codon in the spliced mRNA. The deduced amino acid sequence of the M protein is shown in roman type. The amino acid sequences which could be encoded by the 3′-terminal region of the full-length RNA are shown in italic type.

protein of C virus-infected cells. These results suggested that RNA 6 of influenza C virus codes for the matrix protein.

Evidence that the matrix protein is encoded by a spliced mRNA. As described above, RNA segment 6 of influenza C virus has the coding capacity for a protein of more than 41 kDa. However, analysis of influenza C virus proteins in virions, as well as in infected cells, did not reveal the presence of a major polypeptide species of 41 kDa (10, 13, 37). In order to rule out the possibility that the M protein migrates anomalously in SDS-polyacrylamide gels, an RNA transcript of the insert from pCM144 was transcribed in vitro by using T3 polymerase. In a rabbit reticulocyte translation system, this transcript directed the synthesis of a product that migrated as predicted from the sequence in the position of a 40-kDa protein on polyacrylamide gels (data not shown).

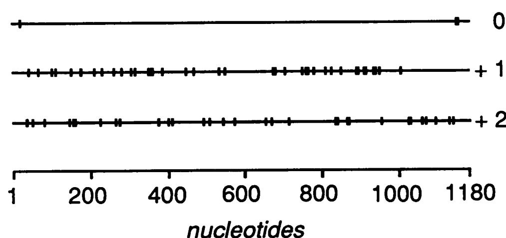

FIG. 3. Open reading frames of RNA 6 of influenza C/JJ/50 virus (mRNA sense). Termination codons are shown by vertical lines.

Since the open reading frame in RNA 6 potentially codes for a protein approximately 15 kDa larger than the viral M protein, attempts were made to analyze the M gene-specific mRNA species in virus-infected cells. A Northern blot analysis of poly(A)-enriched infected-cell RNA using pCM144 DNA as a probe is shown in Fig. 1 (lane 3). An intense diffuse band is present which is 100 to 200 nucleotides shorter than full-length segment 6. In addition, a sharp band migrating identically with full-length segment 6 (lane 1) is present. This latter band is probably the result of contam-

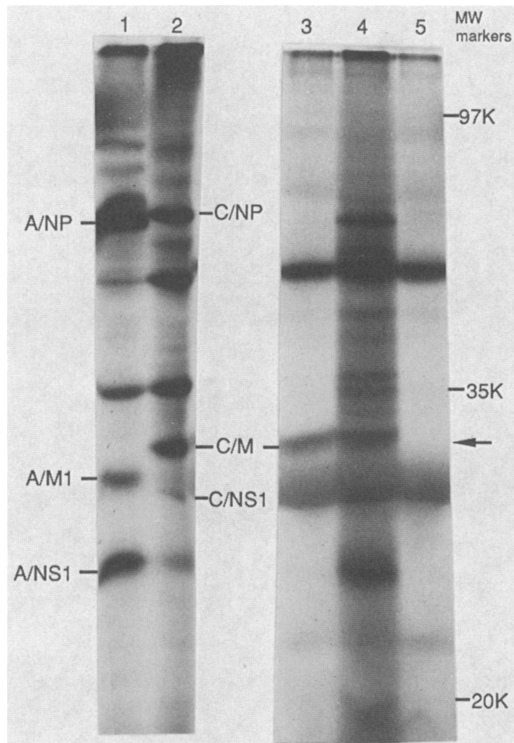

FIG. 4. In vitro translation of pCM144 hybrid-selected mRNA. Poly(A)$^+$ selected mRNA of influenza C virus-infected cells was enriched for RNA segment 6-specific sequences by binding to nitrocellulose filters containing pCM144 DNA, followed by elution at 100°C. Translation of the eluted mRNA preparation in the rabbit reticulocyte lysate was examined by polyacrylamide gel electrophoresis of the [^{35}S]methionine-labeled product. Lane 1, Influenza A/PR/8/34 virus-infected MDCK cells labeled with [^{35}S]methionine under hypertonic conditions; lane 2, influenza C/JJ/50 virus-infected MDCK cells labeled with [^{35}S]methionine under hypertonic conditions; lane 3, hybrid-selected mRNA preparation of C/JJ/50 virus-infected MDCK cells translated in the reticulocyte lysate; lane 4, total poly(A)$^+$ RNA preparation of C/JJ/50 virus-infected MDCK cells translated in the reticulocyte lysate; lane 5, reticulocyte lysate with no exogenous RNA added. The gel positions of the influenza A or C virus NP, M, and NS proteins are indicated. The arrow highlights the gel position of the influenza C virus M protein. MW, Molecular weight; K, molecular weight in thousands.

ination by vRNA or full-length template RNA in samples selected over oligo(dT) columns.

It was speculated that the smaller RNA seen in Fig. 1 may be the result of a splicing event. In order to test this hypothesis, S1 nuclease protection assays were done using probes derived from pCM144. Plasmid pCM144 was digested with either *Bal*I, *Pvu*II, or *Xba*I. Transcription with T7 polymerase yielded minus-sense probes of 499, 814, and 1,149 nucleotides, respectively (Fig. 5B). After gel purification, these probes were hybridized to total poly(A)$^+$ RNA (the RNA sample which was used in Fig. 1) and subjected to S1 digestion and polyacrylamide gel electrophoresis analysis. S1 analysis of each of the three probes after hybridization with the poly(A)$^+$ RNA preparation from infected cells resulted in an identical intense band of approximately 100 nucleotides (band labeled a; Fig. 5A, lanes 1, 5, and 9). The *Bal*I probe produced an additional intense band with an approximate length of 90 nucleotides (band b), while the *Pvu*II and *Xba*I probes produced bands of approximately 400 (band c) and 750 (band d) nucleotides, respectively. As a control, probe RNA was hybridized with unlabeled RNA produced by T3 transcription of *Eco*RI-digested pCM144 (Fig. 5A, lanes 2, 6, 10). These results are consistent with the model shown in Fig. 5B. The band labeled a corresponds to the fragment located 3′ to the splice junction and is protected by all three probes. Bands labeled b, c, and d correspond to fragments located 5′ to the splice junction; they vary due to the different lengths of the RNA probes. In this model for the mRNA of the matrix protein, approximately 250 bases are spliced out near the 3′ end of the message.

In addition to the intense bands seen in lanes 1, 5, and 9 of Fig. 5A, minor bands are also apparent. For each probe, a minor band is seen at a position corresponding to the size of an unspliced transcript. Although this may be due to contamination of the poly(A)$^+$ RNA with trace amounts of full-length template RNA (Fig. 1, lane 3), we cannot rule out the presence of small quantities of unspliced RNA segment 6-specific mRNA.

Identification of the splice site of RNA segment 6-specific mRNA. The data in Fig. 5 suggest that the 5′ and 3′ splicing sites for the mRNA of M protein are located around positions 750 and 1000, respectively. For precise identification of the splicing site, direct sequencing of the poly(A)$^+$ RNA preparation with a primer (positions 1040 through 1019 in the vRNA sense) was attempted. In this way, the splicing sites were tentatively identified at nucleotide 753 (5′ splicing site) and nucleotide 981 (3′ splicing site) (data not shown). In a confirmatory experiment, cDNA was produced by the polymerase chain reaction and directly sequenced. First, cDNA was synthesized from poly(A)$^+$ RNA isolated from infected cells using reverse transcriptase and an oligonucleotide primer (positions 1040 through 1019 in the vRNA sense). Polymerase chain reaction-amplified DNA was then obtained by using the cDNA as the template and *Taq* polymerase and two oligonucleotide primers (positions 1040 through 1019 in the vRNA sense and positions 557 through 576 in the mRNA sense). The amplified fragment of expected size was purified on an 8% polyacrylamide gel, and the cDNA was sequenced directly by using an oligonucleotide (positions 1040 through 1019 in the vRNA sense) as a primer in a dideoxy sequencing reaction. The results shown in Fig. 6 confirm that splicing occurred between nucleotide 753 (5′ splicing site) and nucleotide 981 (3′ splicing site). Although the position of the splicing site would not shift the reading frame of the protein, it is interesting that the splice junction creates a termination codon (Fig. 2). Therefore, the spliced mRNA of the matrix protein encodes an open reading frame of only 242 amino acids.

DISCUSSION

The entire sequence of RNA segment 6 of influenza C/JJ/50 virus was determined. Analysis of the sequence (1,180 nucleotides) revealed only one long open reading frame which could potentially code for a protein of 374 amino acids with an estimated molecular weight of 41,700. Although the matrix (M) protein of influenza C/JJ/50 virus was shown by hybrid selection experiments to be coded for by RNA segment 6, the apparent size of the matrix protein (approximately 28 kDa) was much smaller than that predicted from the deduced sequence. Initial evidence for a processed mRNA species came from Northern blots, which showed a diffuse RNA substantially shorter than full-length segment 6 vRNA (Fig. 1). S1 nuclease analysis then confirmed the splicing event, and sequencing of an amplified cDNA derived from an mRNA preparation allowed localization of the

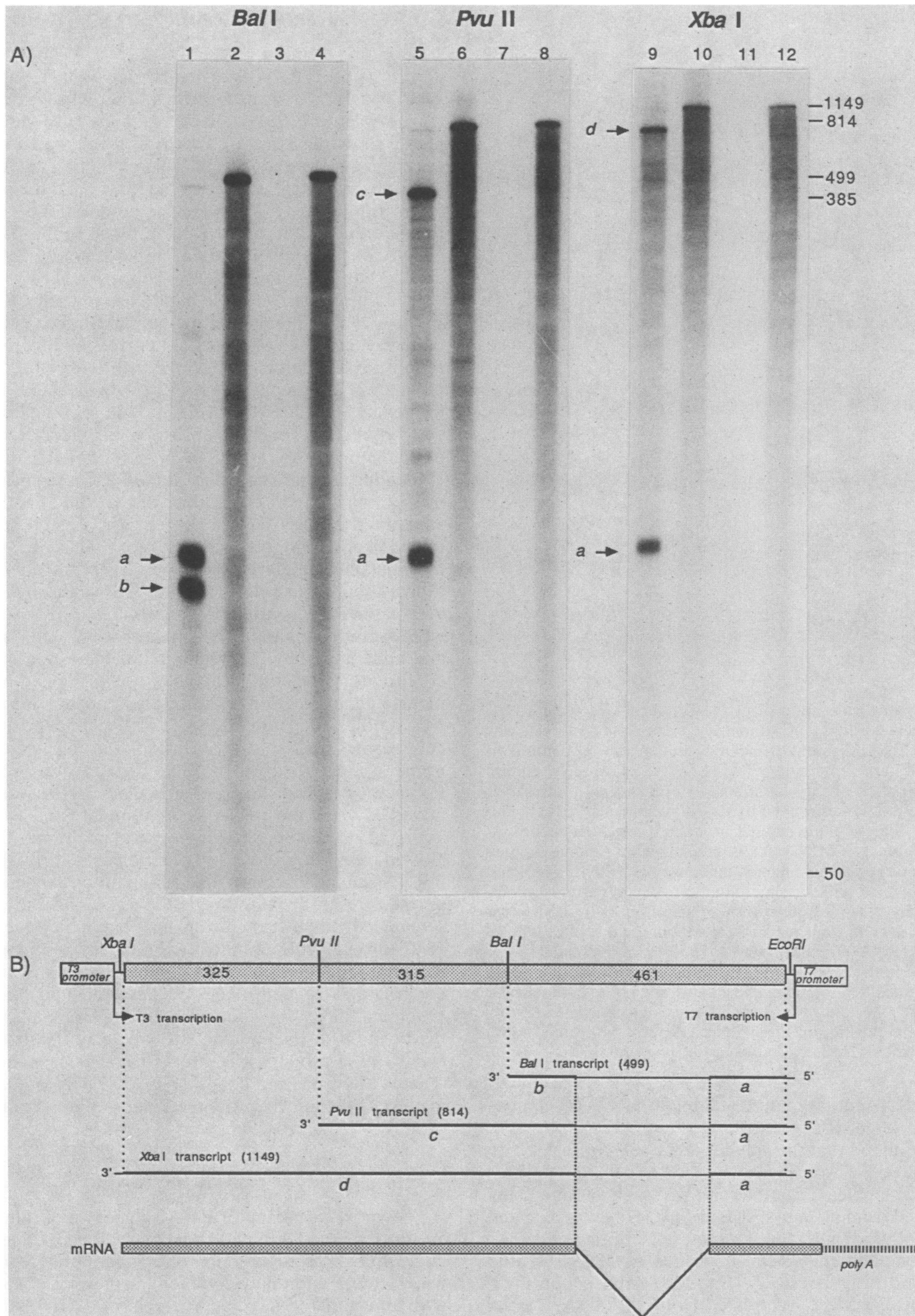

FIG. 5. S1 nuclease analysis of RNA segment 6-specific spliced mRNA. (A) Radioautogram of S1 nuclease analysis. Three different RNA segment 6-specific probes were prepared from *Bal*I-, *Pvu*II-, or *Xba*I-digested pCM144 DNA by using T7 polymerase (as shown at the top of the gel). Probes were hybridized with poly(A)$^+$ RNA from infected cells (lanes 1, 5, and 9) or with message-sense RNA, which was transcribed by T3 polymerase from *Eco*RI-digested pCM144 (lanes 2, 6, and 10). Sample and control (lanes 3, 7, and 11) were then digested with S1 nuclease. Lanes 4, 8, and 12 show undigested probe. Samples were analyzed on 6.5% polyacrylamide gels. Numbers shown on the right indicate nucleotide length of markers. (B) Diagrammatic representation of results from S1 analysis. The sites used to digest the DNA are illustrated within the pCM144 cDNA insert. Numbers in stippled boxes indicate nucleotide length. Direction of polymerase transcription is shown, and numbers in parentheses identify the length of the probes used. Bold lines of transcripts indicate S1 nuclease-resistant regions as shown in the analysis. Letters under the lines correspond to bands identified in panel A. The proposed structure of the spliced RNA 6-specific mRNA is shown on the bottom of the panel.

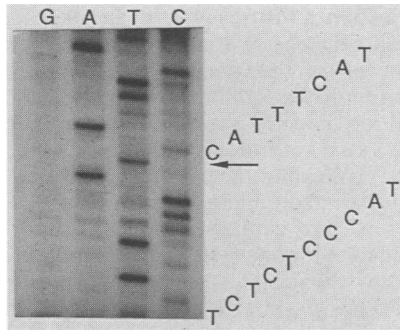

FIG. 6. Sequencing of splice site of mRNA. A double-stranded cDNA copy of the region around the splice site of the RNA 6-specific mRNA was obtained by a polymerase chain reaction described in Materials and Methods. The amplified double-stranded DNA was separated by polyacrylamide gel electrophoresis, isolated, and sequenced directly. Nucleotides around the splice junction (arrow) are shown. Sequencing was in the vRNA-sense direction.

5′ splicing site and 3′ splicing site (nucleotides 753 and 981, respectively). This spliced mRNA could code for a protein of 26,970 daltons, which is in agreement with the measured size of the M protein. Interestingly, the splicing event itself introduced a stop codon (TGA), with the TG coming from the 5′ splicing site and the A coming from the 3′ splicing site. In addition to the size argument mentioned above, several lines of evidence suggest that the spliced mRNA (rather than the unspliced form) gives rise to the M protein of influenza C virus. (i) Northern blotting shows that the major RNA segment 6-specific mRNA species in infected cells is an RNA which is shorter than full-length RNA segment 6 (Fig. 1). (ii) S1 analysis reveals that the major signal derives from the spliced mRNA species and not from full-length transcripts (Fig. 5A). (iii) Direct sequencing of poly(A)$^+$ RNA and polymerase chain reaction-amplified cDNA show that the spliced mRNA is the major form (Fig. 6).

It is interesting that the splicing pattern of the influenza C virus M gene is so different from that of influenza A and B viruses. With respect to the latter virus types, the matrix proteins are coded for by unspliced mRNAs; a spliced mRNA, which is not the major M gene-specific species, directs the synthesis of the M2 protein in influenza A virus-specific cells (4, 18, 20). This contrasts with the situation in influenza C viruses, in which the major mRNA species is spliced and appears to code for the structural matrix protein.

It should be noted that small quantities of an mRNA colinear with RNA 6 may exist, since a faint signal for such a species is observed following S1 analysis (Fig. 5). The protein translated from an unspliced mRNA would contain the entire M protein sequence, with an additional 132 amino acids tacked on to the carboxy terminus. The additional amino acids, as deduced from the sequence, are mostly hydrophobic and could change the properties of the M protein in a dramatic way. However, at this point, no evidence for the expression of a second long M protein of influenza C viruses has been obtained.

Comparing the M proteins of all influenza virus types, it is interesting that they have a similar length of approximately 250 amino acids (2, 4, 36). However, sequence homologies between the M proteins are low (Fig. 7). The greatest homology is between A and B type M proteins and amounts

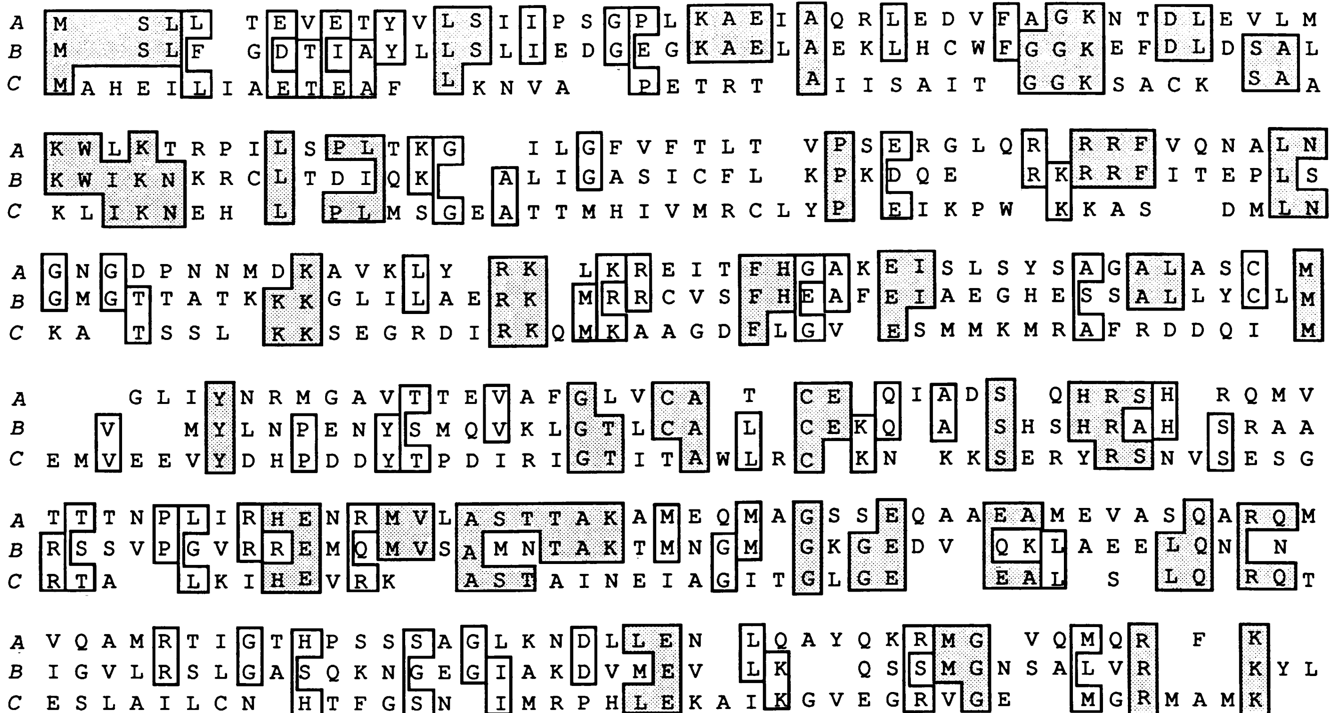

FIG. 7. Comparison of the deduced amino acid sequences of the matrix proteins of influenza A/PR/8/34 (2, 36), B/Lee/40 (4), and C/JJ/50 viruses. Single amino acids conserved between matrix proteins of any two types are in open boxes. Amino acids conserved in the matrix proteins of all three virus types and two or more amino acids conserved in the matrix proteins of two virus types are presented in stippled boxes. The alignment was done using the MicroGenie (Beckman Instruments, Inc., Fullerton, Calif.) program.

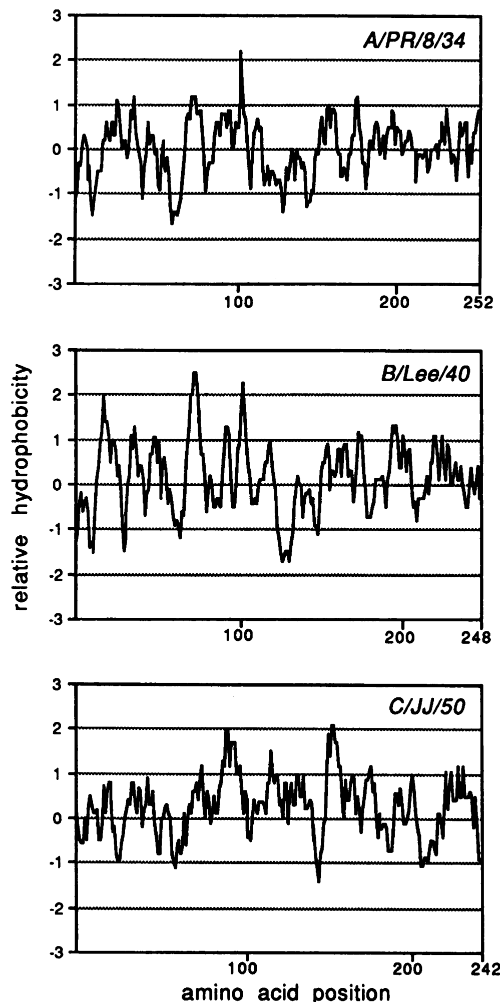

FIG. 8. Comparison of hydrophobicity plots of matrix proteins of influenza A/PR/8/34 (2, 36), B/Lee/40 (4), and C/JJ/50 viruses. The MicroGenie (Beckman Instruments, Inc.) program was used for the analysis.

to approximately 30%. Between A and C type and between B and C type M proteins, homologies are only about 20%. The conserved amino acids are spaced throughout the molecules, but no stretch of more than three amino acids is conserved between A and C or B and C type M proteins. In addition, hydrophobicity patterns of the M proteins are not superimposable (Fig. 8). Nevertheless, there is a clear evolutionary relationship between the M genes of influenza C virus and those of influenza A and B viruses. For example, there are eight different regions in the influenza C virus M protein which are homologous (three out of four amino acids identical) with sequences in the influenza A or B virus M proteins.

It is not known whether the splicing pattern of the M mRNA is common to all isolates of influenza C viruses. However, Sugawara et al. (34) reported that the M proteins of seven different influenza C virus isolates, including that of C/JJ/50 virus, have similar migration patterns on SDS-polyacrylamide gels. Furthermore, peptide maps of these proteins were similar, suggesting that the structure of these proteins (and possibly the splicing pattern of their mRNAs) is well conserved among the influenza C viruses.

Little is known about the precise function of matrix proteins in influenza viruses. It is suggested, however, that matrix proteins may play a structural role in the assembly and packaging of the virus and that they possess a regulatory function during RNA synthesis (5, 6, 8, 33). Therefore, attempts will be made to study possible interactions of the influenza C virus M protein with other viral proteins or RNAs during replication. Experiments will also be designed to determine whether analogous to influenza A viruses, M2-like proteins are expressed by influenza C viruses. A possible candidate for this activity is the protein which could be produced by the unspliced M gene-specific mRNA. This protein would possess, in addition to the 242 amino acids of the M protein, an extra carboxy-terminal domain of 132 amino acids. From the deduced sequence, this domain contains hydrophobic amino acids which may represent an M2-like membrane-spanning segment.

ACKNOWLEDGMENTS

We thank S. Minato, director of Bioscience Research Laboratories, Sankyo Co. Ltd., for supporting the sabbatical leave of M.Y. at the Mount Sinai School of Medicine.

This work was also supported by grants from the Alexandrine and Alexander Sinsheimer Foundation, the Charles H. Revson Foundation, and the National Foundation for Infectious Disease (M.K.) and by Public Health Service grants AI-11823 and AI-18998 (P.P.) from the National Institutes of Health.

LITERATURE CITED

1. **Air, G. M., and R. W. Compans.** 1983. Influenza B and C viruses, p. 280–304. *In* P. Palese and D. W. Kingsbury (ed.), Genetics of influenza viruses. Springer-Verlag, Vienna.
2. **Allen, H., J. McCauley, M. Waterfield, and M.-J. Gething.** 1980. Influenza virus RNA segment 7 has the coding capacity for two polypeptides. Virology **107**:548–551.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moor, J. A. Smith, J. G. Seidman, and K. S. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
4. **Briedis, D. J., R. A. Lamb, and P. W. Choppin.** 1982. Sequence of RNA segment 7 of influenza B virus genome: partial amino acid homology between the membrane proteins (M1) of influenza A and B viruses and conservation of a second open reading frame. Virology **116**:581–588.
5. **Bukrinskaya, A. G., N. K. Vorkunova, G. V. Kornilayeva, R. A. Normanbetova, and G. K. Vorkunova.** 1982. Influenza virus uncoating in infected cells and effect of rimantadine. J. Gen. Virol. **60**:49–59.
6. **Bukrinskaya, A. G., N. K. Vorkunova, and N. L. Pushkarskaya.** 1982. Uncoating of a rimantadine-resistant variant of influenza virus in the presence of rimantadine. J. Gen. Virol. **60**:60–66.
7. **Buonagurio, D. A., S. Nakada, J. D. Parvin, M. Krystal, P. Palese, and W. M. Fitch.** 1986. Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. Science **232**:980–982.
8. **Choppin, P. W., R. W. Compans, A. Scheid, J. McSharry, and S. Lazarowitz.** 1972. Structure and assembly of viral membranes, p. 163–179. *In* C. F. Fox (ed.), Membrane research. Academic Press, Inc., New York.
9. **Cochet, M., F. Permin, F. Gannon, A. Krust, and P. Chambon.** 1979. Cloning of an almost full-length chicken conalbumin double-stranded cDNA. Nucleic Acids Res. **6**:2435–2452.
10. **Compans, R. W., D. H. Bishop, and H. Meier-Ewert.** 1977. Structural components of influenza C virions. J. Virol. **21**:658–665.
11. **Desselberger, U., V. R. Racaniello, J. J. Zazra, and P. Palese.** 1980. The 3′ and 5′ terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. Gene **8**:315–328.
12. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.

13. **Herrler, G., R. W. Compans, and H. Meier-Ewert.** 1977. A precursor glycoprotein in influenza C virus. Virology **99:**49–56.
14. **Herrler, G., R. Rott, H.-D. Klenk, H.-P. Mueller, A. K. Shukla, and R. Schauer.** 1985. The receptor-destroying enzyme of influenza C virus is neuraminate-O-acetylesterase. EMBO J. **4:** 1503–1506.
15. **Kendal, A. P.** 1975. A comparison of influenza C virus with prototype myxovirus: receptor destroying activity (neuraminidase) and structural polypeptide. Virology **65:**87–89.
16. **Kilbourne, E. D.** 1975. The influenza viruses and influenza. Academic Press, Inc., New York.
17. **Lamb, R. A.** 1983. The influenza virus RNA segments and their encoded proteins, p. 21–69. *In* P. Palese and D. W. Kingsbury (ed.), Genetics of influenza viruses. Springer-Verlag, Vienna.
18. **Lamb, R. A., and P. W. Choppin.** 1981. Identification of a second protein (M2) encoded by RNA segment 7 of influenza virus. Virology **112:**729–737.
19. **Lamb, R. A., and C.-J. Lai.** 1981. Conservation of the influenza virus membrane protein (M1) amino acid sequence and an open reading frame of RNA segment 7 encoding a second protein (M2) in H1N1 and H3N2 strains. Virology **112:**746–751.
20. **Lamb, R. A., S. L. Zebedee, and C. D. Richardson.** 1985. Influenza virus M2 protein is an integral membrane protein expressed on the infected cell surface. Cell **40:**627–633.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Nakada, S., R. S. Creager, M. Krystal, R. P. Aaronson, and P. Palese.** 1984. Influenza C virus hemagglutinin: comparison with influenza A and B virus hemagglutinins. J. Virol. **50:**118–124.
23. **Nakada, S., R. S. Creager, M. Krystal, and P. Palese.** 1984. Complete nucleotide sequence of the influenza C/California/78 virus nucleoprotein gene. Virus Res. **1:**433–441.
24. **Nakada, S., P. N. Graves, U. Desselberger, R. S. Creager, M. Krystal, and P. Palese.** 1985. Influenza C virus RNA 7 codes for a nonstructural protein. J. Virol. **56:**221–226.
25. **Nuss, D. L., H. Oppermann, and G. Koch.** 1975. Selective blockage of initiation of host protein synthesis in RNA-virus-infected cells. Proc. Natl. Acad. Sci. USA **72:**1258–1262.
26. **Ohuchi, M., R. Ohuchi, and K. Mifune.** 1982. Demonstration of hemolytic and fusion activities of influenza C virus. J. Virol. **42:** 1076–1079.
27. **Palese, P., and J. F. Young.** 1983. Variation of influenza A, B and C viruses. Science **215:**1468–1474.
28. **Pfeifer, J. B., and R. W. Compans.** 1984. Structure of influenza C glycoprotein gene as determined from cloned DNA. Virus Res. **1:**281–296.
29. **Poncz, M., M. Solowiejczyk, E. Ballantine, E. Schwartz, and S. Surrey.** 1982. "Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc. Natl. Acad. Sci. USA **79:**4298–4302.
30. **Ritchey, M. B., P. Palese, and E. D. Kilbourne.** 1976. RNAs of influenza A, B, and C viruses. J. Virol. **18:**738–744.
31. **Saiki, R. K., D. H. Gelfand, S. Stoffel, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:**487–491.
32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
33. **Schultze, I. T.** 1972. The structure of influenza virus. II. A model based on morphology and composition of subviral particles. Virology **47:**181–196.
34. **Sugawara, K., K. Nakamura, and M. Homma.** 1983. Analyses of structural polypeptides of seven different isolates of influenza C virus. J. Gen. Virol. **64:**579–587.
35. **Vlasak, R., M. Krystal, M. Nacht, and P. Palese.** 1987. The influenza C virus glycoprotein (HE) exhibits receptor-binding (hemagglutinin) and receptor-destroying (esterase) activities. Virology **160:**419–425.
36. **Winter, G., and S. Fields.** 1980. Cloning of influenza cDNA into M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. Nucleic Acids Res. **8:**1965–1974.
37. **Yokota, M., K. Nakamura, K. Sugawara, and M. Homma.** 1983. The synthesis of polypeptides in influenza C virus-infected cells. Virology **130:**105–117.