JOURNAL OF VIROLOGY, Sept. 1986, p. 646–654
0022-538X/86/090646-09$02.00/0

Vol. 59, No. 3

Human Parainfluenza Virus Type 3: Messenger RNAs, Polypeptide Coding Assignments, Intergenic Sequences, and Genetic Map

MELANIE K. SPRIGGS AND PETER L. COLLINS*

Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892

Received 31 January 1986/Accepted 14 May 1986

cDNA clones of mRNAs for the major nucleocapsid protein (NP), the nucleocapsid P protein plus the nonstructural C protein (P+C), and the matrix protein (M) of human parainfluenza virus type 3 (PF3) were identified by hybrid arrest and hybrid selection of in vitro translation. Previously, cDNA clones were identified and sequenced for the hemagglutinin-neuraminidase glycoprotein (HN) and the fusion glycoprotein (F) mRNAs (N. Elango, J. E. Coligan, R. C. Jambou, and S. Venkatesan, J. Virol. 57:481–489, 1986; M. K. Spriggs, R. A. Olmsted, S. Venkatesan, J. E. Coligan, and P. L. Collins, Virology 152:241–251, 1986). Synthetic oligonucleotides, designed from nucleotide sequences of the cDNAs, were used to direct dideoxynucleotide sequencing of gene junctions in PF3 genomic RNA (vRNA). From sequencing of vRNA, a sixth viral gene was detected and identified as the large nucleocapsid protein (L) gene by hybridization of a synthetic oligonucleotide to intracellular PF3 mRNAs separated by gel electrophoresis. The order of the six PF3 genes on vRNA was 3′-NP-P+C-M-F-HN-L-5′. The five intergenic regions consisted of the trinucleotide 3′-GAA. The PF3 genes initiated with semiconserved 10-nucleotide gene-start sequences and terminated with semiconserved 12-nucleotide gene-end sequences. The M gene terminated with an aberrant gene-end sequence; analysis of intracellular mRNA showed that this aberrant sequence correlated with a disproportionately high accumulation of readthrough mRNA. These studies showed that PF3 encodes six unique mRNAs (NP, P+C, M, F, HN, and L) that encode seven proteins (NP, P, C, M, F, HN, and L) and provided evidence of a close relationship between PF3 and Sendai (murine parainfluenza type 1) viruses.

Human parainfluenza type 3 (PF3), a paramyxovirus, is an important agent of pediatric respiratory tract disease (4). The PF3 genome (vRNA) is a single negative strand of RNA of approximately 15,000 nucleotides that has been reported to encode between 5 and 11 mRNAs (10, 19, 32, 33, 35, 37, 38). Analyses of proteins isolated from PF3-infected cells, purified virions, and subviral components identified six viral structural proteins. These were a large nucleocapsid protein, L (M_r, 200,000); the major nucleocapsid protein, NP (M_r, 66,000–68,000); a nucleocapsid phosphoprotein, P (M_r, 83,000–90,000); the matrix protein, M (M_r, 35,000–40,000); the fusion glycoprotein, F (M_r, 60,000–63,000); and the hemagglutinin-neuraminidase glycoprotein, HN (M_r, 69,000–72,000) (10, 19, 32, 33, 34a, 35, 37, 38). Three other putative viral proteins have also been described; two nonstructural proteins with M_rs of 35,000 and 25,000 and a virion-associated protein (VP8) with an M_r of 21,000 (32, 33, 37). However, the viral specificity and significance of these additional proteins were not established, and other workers specifically reported the absence of putative nonstructural proteins (38). Thus, the gene products of PF3 remained to be defined completely.

All paramyxoviruses examined to date encode six structural proteins analogous to the six decribed above for PF3, namely the L, NP, P, M, F, and HN (or H or G) proteins (4, 5, 8, 15, 16, 20, 22, 27). However, recent cDNA cloning, sequencing, and mapping studies for several paramyxoviruses identified additional gene products and coding strategies that are represented in some paramyxoviruses but not in others. Most strikingly, human respiratory syncytial virus has four additional genes that encode a second nonglycosylated matrix protein (the 22,000-molecular-weight protein, a small hydrophobic protein (SH or 1A) and two nonstructural proteins (NS1 or 1C and NS2 or 1B) (6, 7; P. L. Collins, L. E. Dickens, A. Buckler-White, R. A. Olmsted, M. K. Spriggs, E. Camargo, and K. V. W. Coelingh, Proc. Natl. Acad. Sci. USA, in press). Simian virus 5 (SV5, canine parainfluenza virus type 2) also was shown recently to contain an additional gene encoding an SH protein (17), and there is indirect evidence for an additional small gene for Newcastle disease virus (39). In contrast, comprehensive mapping and sequencing studies of Sendai virus confirmed the existence of one additional unique protein, the nonstructural C protein (see below), but demonstrated the absence of additional genes, such as an SH gene (3, 9, 11, 14, 22, 34). C proteins have also been demonstrated for measles and canine distemper viruses; in all cases the C protein is encoded by a second, overlapping open reading frame in the P mRNA (1, 2, 15). In contrast, Newcastle disease virus, SV5, and mumps virus lack C proteins; instead, they encode nonstructural proteins that are related to their respective P proteins and therefore are not unique gene products (8, 16, 27). Thus, the genetic organization of paramyxoviruses exhibits elements of unexpected diversity.

To characterize the mRNAs, vRNA, and proteins of PF3, our laboratory constructed and cloned cDNAs of intracellular viral mRNAs. In the work described here, we used cDNA clones and synthetic oligonucleotides as hybridization probes and as primers for dideoxynucleotide sequencing of vRNA. These experiments (i) identified six viral genes (NP, P+C, M, F, HN, and L), (ii) identified mRNA and protein coding assignments, (iii) determined the sequences for five intergenic and flanking gene regions in vRNA, and

* Corresponding author.

(iv) determined the 3′ to 5′ physical order of the six viral genes.

MATERIALS AND METHODS

Virus and cells. Plaque-purified stocks of PF3 (strain 47885) were propagated in monolayer cultures of HEp-2 cells. In some experiments, CV-1 and BSC-1 cells were also used, yielding similar patterns of viral mRNAs and proteins. For the preparation of vRNA, PF3 was grown in roller bottle cultures of LLC-MK2 cells (34a).

RNA preparation. PF3 vRNA was isolated from purified virions by solubilization with proteinase K and sodium dodecyl sulfate (SDS) and by extraction with phenol-chloroform (Collins et al., in press). For the preparation of mRNA, HEp-2 cells were infected at a multiplicity of 1 to 5 PFU per cell and incubated for 16 h, with 4 μg of actinomycin D per ml present for the final 6 h. When cell fusion was evident but not extensive, cytoplasmic extracts were prepared (6, 7), RNA was isolated by extraction with phenol-chloroform, and mRNA was selected by chromatography on columns of oligo(dT)-cellulose.

Nucleotide sequencing. Dideoxynucleotide sequencing of viral vRNA with synthetic oligonucleotide primers was as previously described (34a), with the sequences of specific primers noted below. Sequencing was performed by combining the chemical method (25) with dideoxynucleotide sequencing of purified plasmid, vRNA, and mRNA by using synthetic oligonucleotide primers (34a).

Oligonucleotides. Synthetic oligonucleotides were purchased from OCS Laboratories (Denton, Tex.) or were synthesized on an Applied Biosystems synthesizer. The mRNA-sense primers for sequencing gene junctions, identified by gene and position in the complete 5′ to 3′ mRNA sequence, were as follows: NP-1603, ACTGAGTCAACA; P+C-1980, CAATAGACATCA; M-1077, ATCAAACAATG GAAC; F-1768, TCGAGTGGATCA; and HN-1766, ACA GAGATTCCA. The vRNA-sense primers for mapping the 5′ ends of the mRNAs, numbered according to distance of the hybridized primer from the 5′ mRNA end, were as follows: P+C-66, ACTCTCTATTGA; M-43, TTGCAGAGTTAG; F-40, TTCCAGCGAGTG; HN-59, TCTCGGATTTGG; and L-86, AGAGGATAGTAATTCCC. The vRNA-sense oligonucleotide used for hybrid arrest of translation of P+C mRNA was P+C-61, GCATCGCTTTCCATCAACTC TCTATTGATTGAG. The vRNA-sense oligonucleotide used to identify the L mRNA by in situ gel hybridization was L-111, GTGTTAATGTGTGATAATACTCAGATGGT.

RNA blot hybridization and in situ gel hybridization. vRNA and intracellular mRNA were separated by electrophoresis on 0.8% agarose-formaldehyde gels (23). Following conventional procedures (23), the RNAs were transferred to nitrocellulose by blotting and were analyzed by hybridization with cDNA clones that had been radiolabeled by nick translation. The final wash of hybridized filters was in 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 55°C (23). For in situ gel hybridization, gels were washed after electrophoresis for 1 h in 10 mM Tris hydrochloride (pH 7.4), dried in vacuo, rehydrated with water, and incubated for 3 h at 37°C in 5× SSC containing 0.1% (wt/vol) SDS, 10 μg of denatured calf thymus DNA per ml, and 25 ng of ^{32}P-labeled oligonucleotide (36). The hybridized gels were washed two times, for 15 min each, in 5× SSC at room temperature and then washed for 5 min in 5× SSC at 37°C, after which the gels were analyzed by autoradiography.

Radiolabeling of PF3 virions and infected-cell proteins. To prepare [^{35}S]methionine-labeled PF3 virions, cell monolayers were infected with a multiplicity of 2 PFU per cell. At 8 h postinfection, the medium was replaced with minimal essential medium containing 2% calf serum, 10% of the normal methionine content, and 25 μCi of [^{35}S]methionine per ml. Incubation was continued until 30 h postinfection, when cell fusion was extensive but the monolayer was intact. The virus was collected by differential centrifugation and subjected to two cycles of banding in discontinuous sucrose gradients (Collins et al., in press). The virus was solubilized by boiling for 2 min in gel sample buffer (21). To prepare [^{35}S]methionine-labeled proteins from PF3-infected and mock-infected cells, monolayers were infected as described above, and radiolabeling was performed for 1 h beginning 10 h postinfection. Cells were subsequently washed with cold phosphate-buffered saline, solubilized in gel sample buffer, immediately boiled for 5 min, and analyzed by electrophoresis (SDS-PAGE) (21).

In vitro protein synthesis and SDS-PAGE. mRNA was translated in the presence of [^{35}S]methionine in retriculocyte lysates that were purchased from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.) and used according to the instructions of the manufacturer. Hybrid arrest and hybrid selection of mRNAs for in vitro translation was as previously described (26, 29). [^{35}S]methionine-labeled proteins were analyzed with SDS-PAGE (8) by the discontinuous gel system of Laemmli (21).

cDNA clones. A library of cDNA clones was constructed by using mRNA from PF3-infected cells as a template and pBR322 as a vector (10). Previously, a cDNA clone of HN mRNA was identified and sequenced (10), but viral specificity and identities of the other cDNA clones were unclear. In this study, the library was screened by colony hybridization with three separate radiolabeled probes: (i) [^{32}P]cDNAs of uninfected-cell mRNA, synthesized with oligo(dT) as primer; (ii) [^{32}P]cDNAs of mRNA from PF3-infected cells, synthesized with oligo(dT) as primer; and (iii) [^{32}P]cDNAs of vRNA isolated from purified virions, synthesized with random oligomers derived from calf thymus DNA as primer. Hybridization of probes ii and iii and not i was presumptive evidence of viral specificity, which was confirmed by hybridization to blots of vRNA and intracellular mRNA (see Results). These procedures followed conventional protocols (23).

RESULTS

PF3 cDNA clones and polypeptide coding assignments. In previous work, a cDNA clone library was constructed by using mRNA from PF3-infected cells as a template; in this way, a complete cDNA clone of the HN mRNA was isolated and sequenced (10). In this study, additional cDNA clones were identified as virus specific (see Materials and Methods) and were segregated by reciprocal cross-hybridization into four additional families distinct from HN cDNA. As described below, these were identified as cDNA clones of the NP, P+C, M, and F mRNAs.

cDNA clones of the NP, P+C, and M mRNAs were identified by hybrid arrest and hybrid selection of in vitro translation. Figure 1 shows the SDS-PAGE patterns of [^{35}S]methionine-labeled proteins extracted from purified virions (lane k) and from PF3-infected (lane b) and uninfected (lane a) cells, as well as proteins that were synthesized in vitro in reponse to mRNA from PF3-infected cells (lane c). Comparison with studies by our laboratory and by others (19, 32, 33, 34a, 35, 37, 38) provided unambiguous identification of the HN, F, NP, P, and M proteins. Extracts

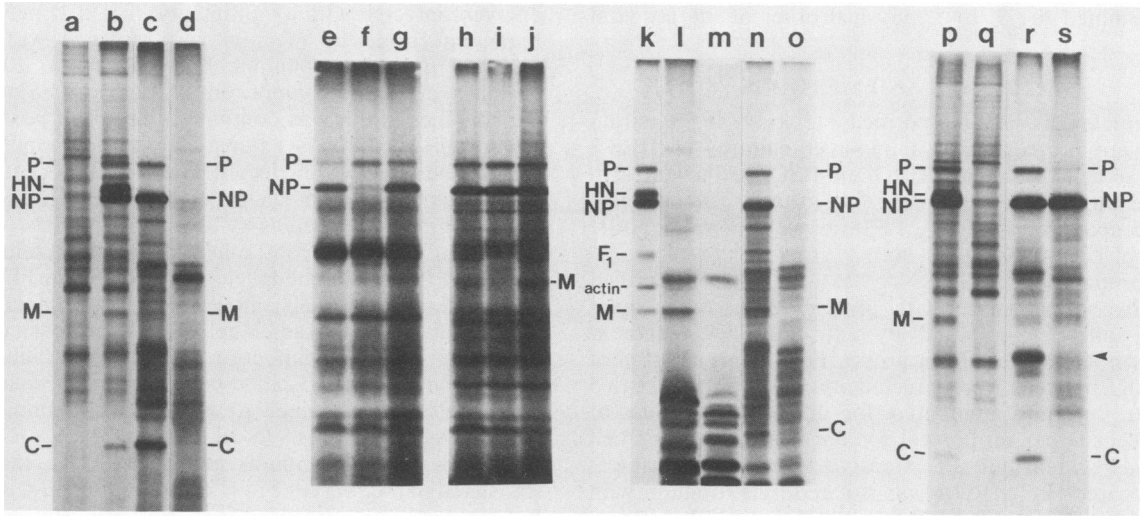

FIG. 1. SDS-PAGE of [^{35}S]methionine-labeled proteins of PF3: identification of cDNAs of the NP, P+C, and M mRNAs by hybrid arrest and hybrid selection of in vitro translation. Proteins labeled during a 1 h pulse were extracted from (a) uninfected and (b) PF3-infected HEp-2 cells and compared with proteins synthesized in reticulocyte lysates in response to mRNA from (c) PF3-infected and (d) uninfected cells. To identify cDNAs of the NP mRNA by hybrid arrest of in vitro translation, intracellular PF3 mRNA was hybridized to the cloned cDNA insert and analyzed by translation in vitro (e) after boiling to dissociate hybrids (a control to verify intactness of mRNAs) or (f) in the hybridized form. A parallel lane (g) contains proteins synthesized in vitro with untreated PF3 mRNA. To identify cDNAs of the M mRNA by hybrid-arrested translation, intracellular PF3 mRNA was hybridized with a cDNA clone and analyzed by translation in vitro (h) after boiling to dissociate hybrids or (i) in the hybridized form. A parallel lane (j) contains proteins synthesized in vitro with untreated PF3 mRNA. To confirm the identification of the M cDNA by hybrid-selected translation, mRNA from (l) PF3-infected and (m) uninfected cells was selected by hybridization to filter-bound cDNA, dissociated, and translated in vitro in parallel with mRNA from (n) PF3-infected and (o) uninfected cells. A parallel lane (k) contains proteins from purified virions. To identify cDNAs of the P+C mRNA by hybrid-arrested translation, intracellular PF3 mRNA was analyzed by translation in vitro (r) without treatment or (s) after hybridization with a synthetic oligonucleotide (designed to be complementary to nucleotide 61 through 93 of the corresponding mRNA). Parallel lanes contain pulse-labeled proteins from (p) PF3-infected and (q) uninfected cells. All gels contained 10% polyacrylamide and were analyzed by fluorography. The positions of major PF3 proteins are marked. The arrow indicates an additional, unidentified protein in lane r whose synthesis was inhibited in parallel with the P and C proteins (see text). In the virion pattern (lane k), the identification of cellular actin was based on previous reports (37), on its size, and on its comigration with a major cellular protein (not shown).

of PF3-infected cells contained an additional protein with an M_r of approximately 22,000 (Fig. 1). We concluded that this protein, absent in purified virions, was nonstructural. It was therefore designated the C protein.

In hybrid-arrest experiments, cDNAs from one family reversibly blocked in vitro synthesis of the M protein (Fig. 1, lanes h through j), thus identifying this family as cDNAs of the M mRNA. This was confirmed by hybrid selection and in vitro translation of the M mRNA (Fig. 1, lanes k through o). A second family consisted of clones identified as NP cDNAs as a result of hybrid-arrest experiments in which these cDNAs reversibly blocked synthesis of NP (Fig. 1, lanes e through g).

For a third group of cDNAs, a synthetic oligonucleotide, P+C-61 (see Materials and Methods), was designed from the complete mRNA sequence (unpublished data) to be complementary to nucleotides 61 through 93 of the corresponding mRNA. This oligonucleotide covered the first ATG triplet in the mRNA sequence (positions 80 through 82; nucleotide sequence not shown), the likely translation start site of the P protein, as well as the second ATG triplet at positions 90 through 92, the likely translation start site of the C protein. The use of this oligonucleotide in hybrid arrest experiments blocked the synthesis of both P and C proteins (Fig. 1, lanes p through s) as well as that of an additional, unidentified protein (Fig. 1, arrow) that was not analyzed further. This additional protein lacked a counterpart of the same size in infected cells and purified virions and was presumed to be a product of premature translation termination or a degradation product of P, such as was described previously (33). These results established that the P and C proteins are encoded by the same genomic locus. Elsewhere, we will present the nucleotide sequence of the P+C mRNA, together with evidence indicating that the C protein is a seventh unique PF3 protein encoded by a second open reading frame in the P+C mRNA.

A fourth family consisted of clones identified as cDNAs of the F mRNA, because the amino acid sequence deduced from the nucleotide sequence was an exact match with a partial amino acid sequence determined directly from purified F_1 protein (34a). Thus, cDNAs have been identified for the NP, P+C, M, F, and HN mRNAs.

Analysis of intracellular PF3 mRNAS. Previous reports described 5 to 11 mRNAs for PF3 (10, 33). Researchers assumed that some of theses species were unique mRNAs representing individual genes, whereas others were polycistronic readthrough mRNAs (polytranscripts) that each represent two or more adjacent genes, as was described for other paramyxoviruses (6–8, 13, 14, 17, 27, 28, 39, 40). Recent reports identified several intracellular PF3 mRNAs by hybridization with NP, P, and HN cDNAs (10, 33). In this study, to complete the identifications and make side-by-side comparison of five intracellular PF3 mRNAs, cDNAs of the NP, P+C, M, F, and HN genes were radiolabeled and hybridized individually to blots of RNA from purified virions and PF3-infected and uninfected cells (Fig. 2).

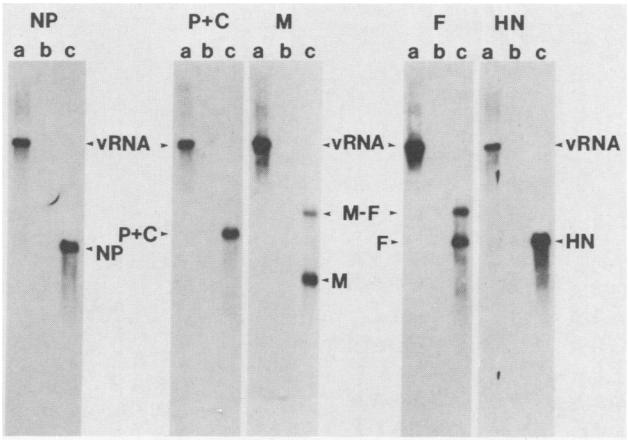

FIG. 2. Characterization of five PF3 cDNAs by RNA blot hybridization. RNA isolated from (a) purified virions and mRNA from (b) uninfected and (c) PF3-infected HEp-2 cells were electrophoresed in 0.8% agarose-formaldehyde gels and transferred to nitrocellulose. Blots were hybridized to radiolabeled cDNAs of the NP, P+C, M, F, and HN mRNAs (as indicated at the top of each blot) and analyzed by autoradiography. Identities and positions of the mRNAs are marked with arrows. Identities (polypeptide-coding assignments) indicated here for cDNAs and mRNAs were determined as described in the text and the legend to Fig. 1. The sizes of the mRNAs, determined by nucleotide sequencing, are listed in the Results. The radiolabeled hybridization probes were as follows. The NP mRNA was hybridized with pPF5 (10), which contains the 3'-terminal 1,100 nucleotides of the 1,641-nucleotide mRNA (Jambou et al., submitted); entirely similar results were obtained with pPF5-17, which contains the complete mRNA sequence (unpublished results). The P+C mRNA was hybridized with pP-2, which contains the 3'-terminal 944 nucleotides of the 2,009-nucleotide P+C mRNA; entirely similar results were obtained with an oligonucleotide containing nucleotides 379 through 393 from the 5' mRNA terminus (unpublished results). The M mRNA was hybridized with pM-3, which contains the 3'-terminal 820 nucleotides of the 1,150-nucleotide M mRNA. The F mRNA was hybridized with pF-16 (34a), which contains the 3'-terminal 1,530 nucleotides of the 1,845-nucleotide F mRNA; entirely similar results were obtained with pF-7 (not shown), which contains the 5' F mRNA end (34a). The HN mRNA was hybridized with pPF4, which contains the complete HN mRNA sequence (10).

All five cDNAs hybridized to vRNA, which was proof of viral specificity. The NP, P+C, and HN mRNAs each hybridized to a single major species of PF3 mRNA, indicating that possible corresponding polytranscripts were of low abundance. As evidence that such polytranscripts do exist, a cDNA clone representing a dicistronic P-M mRNA was identified in the process of screening the cDNA library (data not shown). The F and M cDNAs each hybridized to an mRNA of an appropriate length to be the corresponding unique mRNA (~2,100 and 1,250 nucleotides, respectively, including polyadenylate). In addition, the F and M cDNAs both hybridized to an abundant mRNA of an appropriate length to be a dicistronic M-F mRNA (~3,200 nucleotides). In contrast to the low levels of accumulation of the other polycistronic mRNAs, the M-F mRNA appeared to be almost equimolar to the F mRNA. This finding was reproducible for mRNAs extracted from HEp-2, BSC-1, and CV-1 cells (unpublished data).

Thus, these data identified six major species of PF3 mRNA. These included five unique mRNAs (NP, P+C, M, F, and HN) and one uniquely abundant polytranscript (M-F mRNA). The lengths of the five mRNAs, determined from nucleotide sequencing of cDNAs, mRNAs, and vRNA and excluding polyadenylate, are as follows: NP mRNA, 1,641 nucleotides (R. C. Jambou, N. Elango, S. Venkatesan, and P. L. Collins, J. Gen. Virol., in press); P + C mRNA, 2,009 nucleotides (M. K. Spriggs and P. L. Collins, submitted for publication); M mRNA, 1,150 nucleotides (unpublished data); F mRNA, 1,845 nucleotides (34a); and HN mRNA, 1,882 nucleotides (10 [as amended; see below]).

Gene order. From the nucleotide sequences of the NP, P+C, M, F, and HN mRNAs (10, 34a; Jambou et al., in press; Spriggs and Collins, submitted; unpublished data), synthetic oligonucleotides were constructed to be complementary to vRNA near the 5' (downstream) end of each gene. The oligonucleotides were hybridized to vRNA and used to direct dideoxynucleotide sequencing across gene junctions. Sequencing gels illustrating the five intergenic and flanking gene regions are shown in Fig. 3, and the five sequences are compared in Fig. 4. Each downstream gene was identified by comparison of the vRNA sequence with the sequences determined from cDNAs and mRNAs. This comparison established the gene order for the five PF3 genes as 3' NP-P+C-M-F-HN. Detailed features of the sequences of the gene junctions and the mapping of the L gene are described below.

Consensus gene-start, gene-end, and intergenic sequences. In the sequences shown in Fig. 4, the exact boundaries of the viral genes were identified by mapping and sequencing the 3' and 5' ends of the viral mRNAs. The 5' ends of the mRNAs were mapped and sequenced by primer extension on mRNA under conditions of dideoxynucleotide sequencing (Fig. 5); the mapping and sequencing of the 5' end of the NP mRNA will be described elsewhere (Jambou et al., in press). These results showed that the PF3 mRNAs initiate with a semiconserved 10-nucleotide sequence; the vRNA-sense complement of this sequence, 3' UCCUNNUUUC, is called the gene-start sequence (indicated in Fig. 4 with dotted lines).

The 3' termini of each PF3 mRNA were identified by the locations of 3'-terminal poly(dA) tracts in the mRNA-sense strands of several independently derived cDNA clones, except that the 3' end of the published HN mRNA sequence was determined from a single cDNA clone (10). Alignment of the mRNA and vRNA sequences showed that the NP and P+C genes terminate with identical gene-end sequences, i.e., UUUAUUCUUUUU (vRNA sense) (indicated in Fig. 4 with dashed lines). Interestingly, the M gene terminates with the elements of a gene-end sequence that is identical to those of the NP and P+C genes, except that the M gene-end sequence contains the insertion of eight additional nucleotides (italicized): 3' UUUAUUC*UCUAUUAG*UUUUU (vRNA sense). The F gene terminates with a gene-end sequence, 3' UU*A*AU*A*UUUUUU, that differs at three positions (italicized) from that described above for the NP and P+C genes. For the HN mRNA, the final two nucleotides preceding the poly(A) tract in the previously published sequence, 5' UCAGACAA*AG*-poly(A) (mRNA sense) (10), were not confirmed here by sequencing vRNA. Because the published sequence was determined from a single cDNA clone, a likely explanation is that the cDNA clone sequenced in the earlier work contains two extraneous nucleotides added during mRNA synthesis, cDNA synthesis, or plasmid repair and replication in bacteria. The vRNA sequence obtained here represents the consensus of a population of molecules rather than a single cloned transcript and therefore is considered authentic. The vRNA sequence contains a 12-nucleotide segment that is identical to the F gene-end sequence except for one nucleotide difference, 3'

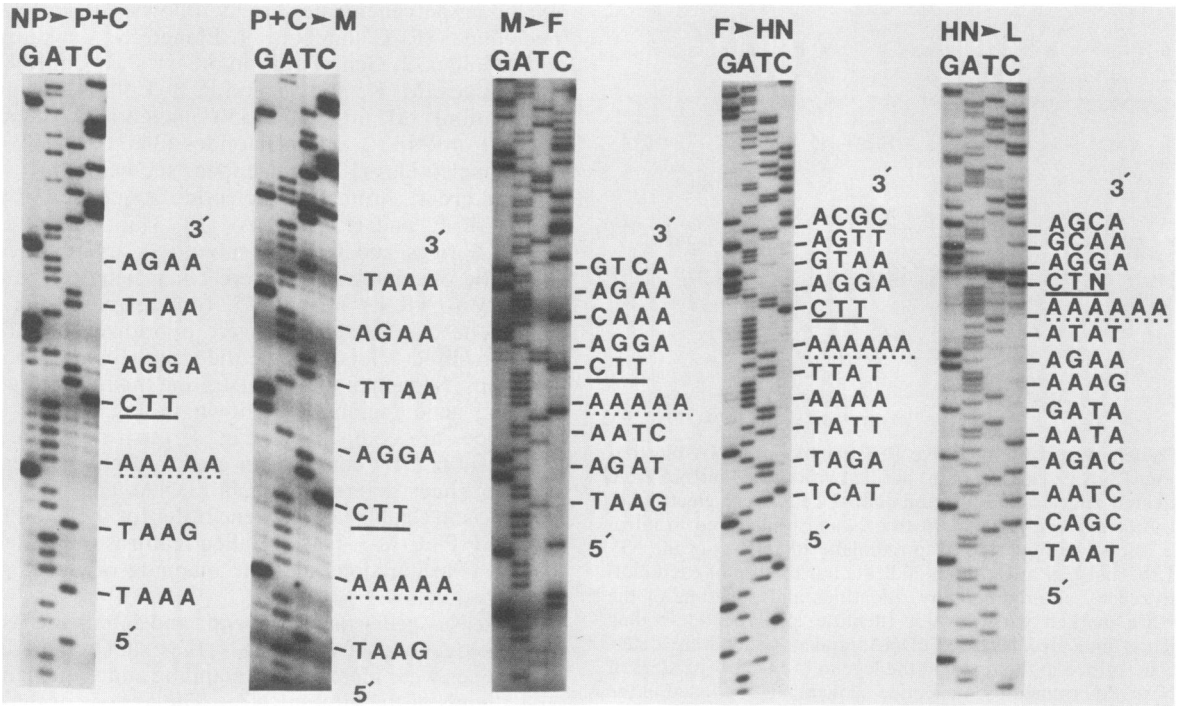

FIG. 3. Identification of PF3 gene order and sequence analysis of intergenic regions in PF3 vRNA. Synthetic oligonucleotides (mRNA sense) were constructed and used to direct dideoxynucleotide sequencing on vRNA across gene junctions (NP → P+C, P+C → M, M → F, F → HN, HN → L). Autoradiograms of 6% sequencing gels are shown, with specificities of the sequencing lanes marked (G, A, T, C). Deduced sequences (mRNA sense) are shown; direction of sequencing (bottom to top) is from each upstream gene toward its downstream neighbor. Oligo(dA) tracts that terminate each mRNA-sense gene sequence (-----) and intergenic trinucleotides (——) indicated. Identification of the L gene is described in the text and in the legend to Fig. 6.

UU*U*AUAUUUUUU (vRNA sense) (Fig. 4). We suggest that this is the HN gene-end sequence. If so, the authentic sequence for the HN mRNA contains 18 additional nucleotides, 5′ *UAGAUAAAAGAGAAAUAU*-poly(A), in place of the two aberrant nucleotides in the published sequence, 5′ *AG*-poly(A). With the exception of the M gene, the gene-end sequences shown in Fig. 4 conform to the following consensus gene-end sequence (vRNA sense): 3′ UU$^{A}_{U}$AU$^{AU}_{UC}$UUUUU.

Demarcation of the gene boundaries identified the intergenic regions. The five intergenic regions shown in Fig. 4 are conserved in length and sequence and consist of the trinucleotide 3′ GAA (vRNA sense). In the case of the intergenic region following the HN gene, 3′ GAN, the third nucleotide was not identified unambiguously, but numerous repetitions of the experiment suggested that it is an A (not shown).

Mapping and identification of the L gene. A nucleotide sequence also was obtained for the vRNA segment following (3′ to 5′) the HN gene (Fig. 4). This sequence contained a putative intergenic region, 3′ GAN (vRNA sense), and a putative gene-start sequence, 3′ UCCUCGUUUC (vRNA sense). A sequence of 630 nucleotides was obtained for this putative gene by successive primer extensions (not shown). In mRNA sense, the sequence contains a potential translational start site at nucleotides 23 through 25 (numbering the first nucleotide of the gene-start sequence) that initiates an open reading frame that remains open for the length of the available sequence. To identify this putative gene, a vRNA-sense oligonucleotide was constructed that contained positions 111 through 139 of the putative gene. This oligonucleotide was 5′ end labeled and hybridized in situ to agarose gels containing separated PF3 mRNAs. The oligonucleotide hybridized specifically to a unique viral mRNA (Fig. 6). On the basis of its large size (6,000 to 7,000 nucleotides) and by analogy to other paramyxoviruses, this mRNA was identified as the L mRNA.

The 5′ end of the L mRNA was mapped and sequenced by primer extension on mRNA (Fig. 5). This confirmed that the HN and L genes are separated by the trinucleotide 3′ GAN (vRNA sense) and that the L gene initiates with the consensus gene-start signal.

DISCUSSION

Identifications and polypeptide coding assignments were made for five PF3 mRNAs (NP, P+C, M, F, and HN) based on analyses of cDNA clones in work described here and elsewhere (Fig. 1) (10, 19, 33, 34a). A sixth viral mRNA, the L mRNA, was identified by hybridization of mRNA blots with a synthetic oligonucleotide (Fig. 6) that was designed from sequence analysis of vRNA (Fig. 3 and 4). Dideoxynucleotide sequencing from one gene to the next on vRNA established that the PF3 gene order is 3′-NP-P+C-M-F-HN-L-5′.

The six PF3 mRNAs (NP, P+C, M, F, HN, and L) encode seven proteins (NP, P, C, M, F, HN, and L). Translational hybrid-arrest experiments established that the P and C proteins are encoded by the same genomic locus (Fig. 1), and hybridization probes spanning most of this genomic locus hybridized to a single 2,200-nucleotide (including polyade-

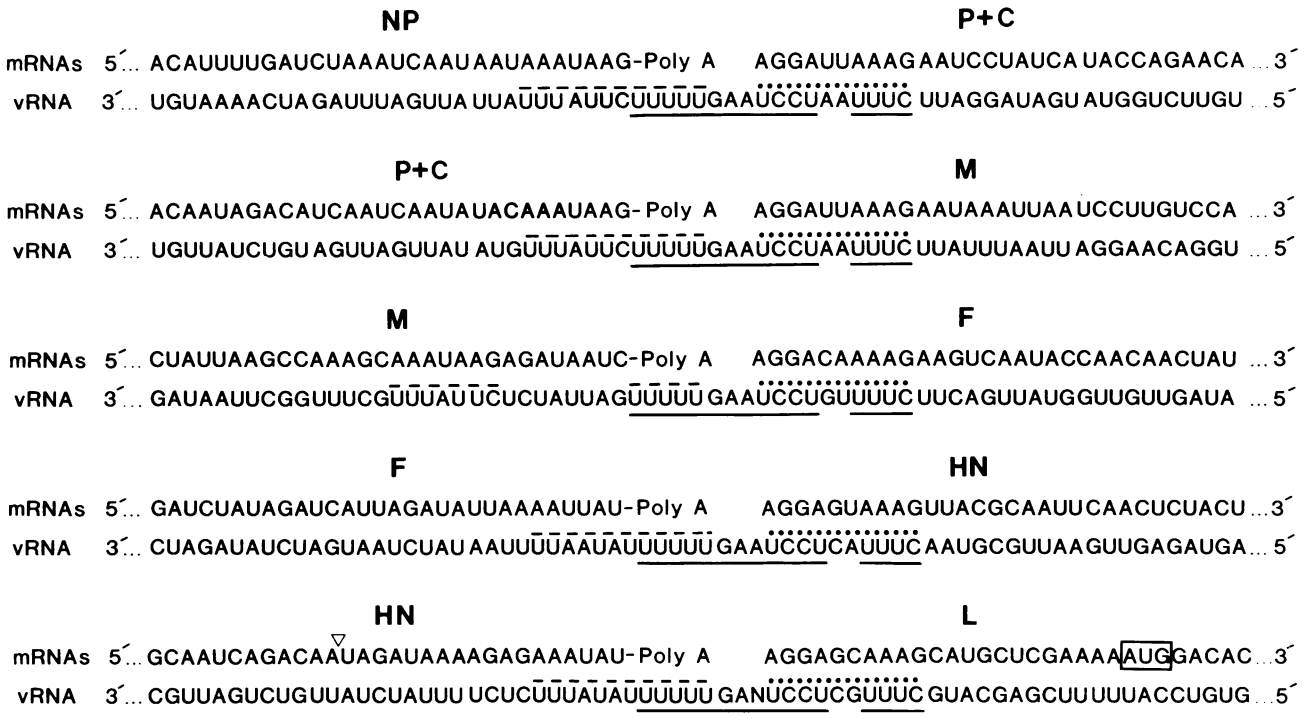

FIG. 4. Comparison of intergenic and flanking gene sequences of PF3 vRNA. vRNA sequences were obtained from experiments such as those described in the legend to Fig. 3. mRNA sequences were obtained from experiments such as those described in the legend to Fig. 5 and from sequencing cDNAs (10, 34a; unpublished data). Identification of the L gene and mRNA is described in the text and the legend to Fig. 6. Exactly conserved sequences (——), semiconserved gene-end vRNA sequences (-----), gene-start vRNA sequences (·····), and a divergence (▽) between the published HN mRNA sequence (10) and the HN vRNA sequence (see the text) are indicated. The proposed translational start site of L mRNA is boxed.

nylate) mRNA (Fig. 2; unpublished data). Experiments to be described elsewhere (Spriggs and Collins, submitted) showed that the P and C proteins are unrelated by peptide mapping and are encoded by separate, overlapping reading frames in the P+C mRNA, a coding strategy that is closely analogous to that described previously for Sendai, measles, and canine distemper viruses (1, 2, 11).

There is general agreement that L, NP, P, M, F, and HN are PF3 structural proteins. It has been suggested that three additional proteins are PF3-specific; these include two nonstructural proteins with M_rs of 35,000 and 25,000 and a virion-associated protein, VP8, with an M_r of 21,000 (32, 33). However, the viral specificity and uniqueness of these proteins were not determined. In this study, analysis of viral polypeptides extracted from PF3-infected cells and purified virions identified a single major additional protein with an M_r of 22,000, the C protein, which appears to be nonstructural. The significance of the other putative viral proteins mentioned above is unclear. The gene mapping experiments described here (Fig. 3 and 4) exclude the existence of additional intervening genes, although the formal possibility exists that one or more additional genes might precede the NP gene or follow the L gene in the 3′ to 5′ gene order. However, it seems likely that the additional proteins do not represent additional mRNAs; instead, they might be fragments or electrophoretically distinct forms of the seven major proteins or secondary translation products of their mRNAs. The available evidence indicates that PF3 does not have additional mRNAs that encode proteins such as the NS1, NS2, SH, and 22,000-molecular-weight proteins of respiratory syncytial virus and the SH protein of SV5.

The sequences for the five PF3 gene junctions were determined by dideoxynucleotide sequencing of vRNA (Fig. 3 and 4). The nontranscribed intergenic sequences were identified by comparison of the vRNA sequences with sequences representing the exact 5′ and 3′ ends of the viral mRNAs (Fig. 4). This showed that each of the intergenic regions is the trinucleotide 3′ GAA. Thus, the intergenic regions within the PF3 genome are conserved exactly in sequence and length. Furthermore, they are identical to four intergenic regions of Sendai virus and two intergenic regions of measles virus (3, 11, 13, 34). In contrast, eight intergenic regions for respiratory syncytial virus (Collins et al., in press) and two for SV5 (17, 28) exhibit no conservation of length and sequence and are structurally distinct from the sequences described here for PF3, Sendai, and measles virus (for a discussion, see Collins et al., in press).

Sendai viral genes initiate with the conserved gene-start sequence 3′ UCCCANUUUC and terminate with the conserved gene-end sequence 3′ UNAUUCUUUUU (3, 11–13, 34). Analogous sequences have been described for the rhabdovirus vesicular stomatitis virus, another prototype of nonsegmented negative strand viruses (31). These sequences, perhaps in concert with the conserved intergenic sequences, are thought to be signals for the viral transcriptase. For example, one hypothesis is that the U-rich gene-end sequences act as both a signal and a template for adding 3′ poly(A) tails to mRNAs by a reiterative copying mechanism. The intergenic sequence could act as a signal to terminate transcription, and the gene-start sequence could act as a signal for the polymerase to resume transcription. Consistent with the Sendai virus model, the genes of PF3 contain the conserved decanucleotide gene-start signals 3′-UCCUNNUUUC that encode the exact 5′ termini of the

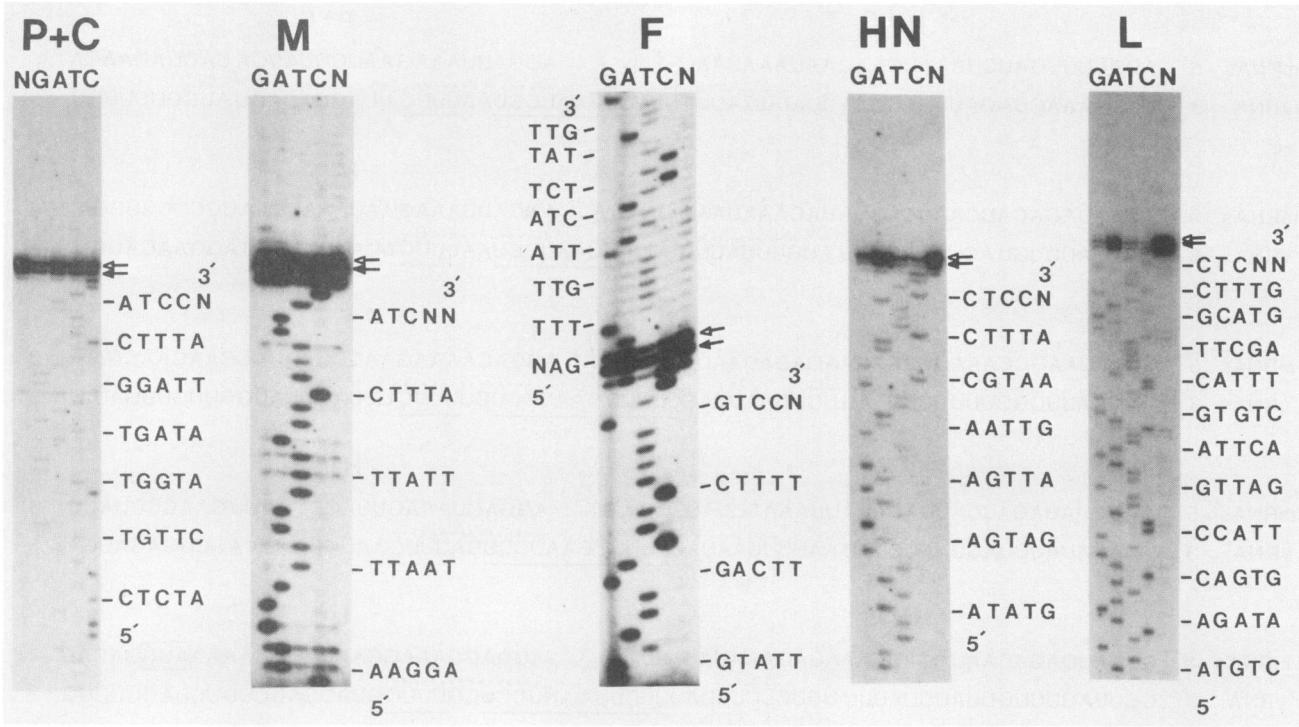

FIG. 5. Mapping and sequencing of 5′ ends of P+C, M, F, HN, and L mRNAs. Synthetic oligonucleotides (vRNA sense) were constructed and used to direct primer extension on mRNA in the absence of dideoxynucleoside triphosphates (N) and in the presence of the appropriate dideoxynucleoside triphosphate (G, A, T, and C). Reaction products were separated on 6% sequencing gels and analyzed by autoradiography. Deduced sequences are shown in vRNA sense. In each case, the position of the extension product representing authentic 5′ mRNA end is indicated (◁—); the artifactual end-addition product characteristic of primer extension on mRNA (13) is also indicated (◀—). For the F sequencing ladder, the sequence above the strong stops represents primer extension on the abundant M-F dicistronic mRNA described in the text. This sequence (shown at the left) contains the intergenic trinucleotide and the M gene-end sequence.

mRNAs (Fig. 4). The PF3 and Sendai virus gene-start sequences have 80% identity.

The PF3 genes also contain gene-end sequences which conform to the consensus sequence 3′ UU$\overset{A}{U}$AU$\overset{AU}{UC}$UUUUU, except in the case of the M gene (Fig. 4; see below). The variability of the PF3 gene-end sequence differs from the situation for Sendai virus and vesicular stomatitis virus, where the gene-end sequence is invariant within each virus. The gene-end sequences of the PF3 NP and P+C mRNAs (3′ UUUAUUCUUUUU) are identical to one another and have 83.3% sequence identity with the Sendai virus gene-end sequences. The PF3 F and HN gene-end sequences are almost identical to one another and are somewhat distinct from the NP and P+C sequences; the last 10 nucleotides of the F and HN genes (3′ $\overset{A}{U}$AUAUUUUUU) are almost identical to the two available gene-end sequences for measles virus (3′ AAUAUUUUUU) (30).

Interestingly, the gene-end sequence of the M gene contains an insertion of eight nucleotides (Fig. 4; see Results); apart from this insertion, the M gene-end sequence is identical to its counterparts in the NP and P+C genes. Significantly, the M-F polytranscript was uniquely abundant, whether detected by RNA blot hybridization (Fig. 2) or primer extension on mRNA (Fig. 5). Thus, the aberrant M gene-end sequence correlated with a high incidence of readthrough trancription. This is evidence, albeit circumstantial, that the gene-end sequences are involved in transcriptive events at the gene junctions. Previous studies showed that nucleotide differences in the intergenic sequences for Sendai virus and vesicular stomatitis virus also resulted in changes in the abundances of the corresponding polytranscripts (14, 24), although the differences in abundance were not as great as those described here. In other work, an internal-deletion mutant of Sendai virus was shown to contain an NP gene that lacked the final two U residues of the gene-end sequence, as well as the downstream intergenic and gene-start sequence; this NP gene was expressed exclusively as a readthrough mRNA (18). Thus, gene-end and intergenic sequences, and possibly also the contiguous gene-start sequences, appear to influence the efficiency of polyadenylation and termination. It will be of interest to examine other naturally occurring strains of human PF3, as well as the bovine strain, to see if this aberrant M gene-end sequence is ubiquitous.

Overall, the results described here demonstrate close parallels between PF3 and Sendai (murine parainfluenza type 1) viruses in (i) genetic map, (ii) constellation of gene products, and (iii) sequences of intergenic and flanking gene regions. Comparison of the sequences of the HN, F, and NP genes and proteins of the two viruses also demonstrated a high degree of relatedness (10, 34a; Jambou et al., in press). In contrast, by all of these criteria, SV5 appears to be related less closely.

ACKNOWLEDGMENTS

This work was performed in the laboratory of Brian R. Murphy and Robert M. Chanock. We thank them for advice and support. We thank Alicia Buckler-White for synthesizing oligonucleotides, Ena Camargo for preparing mRNA and virus, A.B.W. and Charles E.

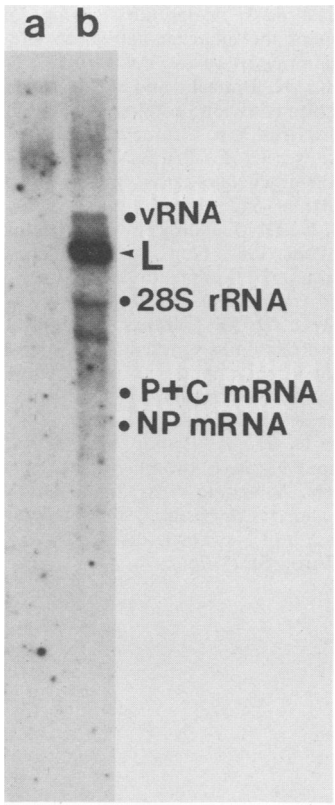

FIG. 6. Identification of L gene. Intracellular mRNA from (a) uninfected and (b) PF3-infected cells was separated by electrophoresis on an 0.8% agarose–formaldehyde gel. The gel was analyzed by in situ hybridization with a radiolabeled vRNA-sense oligonucleotide that was designed to contain nucleotides 111 through 139 of the putative L gene. The hybridized gel was analyzed by autoradiography. Positions are shown for marker RNAs that were electrophoresed in adjacent lanes and detected by staining with ethidium bromide (PF3 vRNA from purified virions and 28S rRNA from HEp-2 cells) or by blot transfer and hybridization with cDNA clones (PF3 NP and P+C mRNAs).

Buckler for help with computer analysis, and Linda Jordan for editorial assistance.

LITERATURE CITED

1. **Barrett, T., S. B. Shrimpton, and S. E. H. Russell.** 1985. Nucleotide sequence of the entire protein coding region of canine distemper virus polymerase-associated (P) protein mRNA. Virus Res. **3**:367–372.
2. **Bellini, W. J., G. Englund, S. Rozenblatt, H. Arnheiter, and C. D. Richardson.** 1985. Measles virus P gene codes for two proteins. J. Virol. **53**:908–919.
3. **Blumberg, B. M., C. Giorgi, L. Roux, R. Raju, P. Dowling, A. Chollet, and D. Kolakofsky.** 1985. Sequence determination of the Sendai virus HN gene and its comparison to the influenza virus glycoproteins. Cell **41**:269–278.
4. **Chanock, R. M., and K. McIntosh.** 1985. Parainfluenza viruses, p. 1241–1253. *In* B. N. Fields (ed.), Virology. Raven Press, New York.
5. **Collins, P. L., L. E. Hightower, and L. A. Ball.** 1980. Transcriptional map for Newcastle disease virus. J. Virol. **35**:682–693.
6. **Collins, P. L., Y. T. Huang, and G. W. Wertz.** 1984. Identification of a tenth mRNA of respiratory syncytial virus and assignment of polypeptides to the 10 viral genes. J. Virol. **49**:572–578.
7. **Collins, P. L., and G. W. Wertz.** 1983. cDNA cloning and transcriptional mapping of nine polyadenylated RNAs encoded by the genome of human respiratory syncytial virus. Proc. Natl. Acad. Sci. USA **80**:3208–3212.
8. **Collins, P. L., G. W. Wertz, L. A. Ball, and L. E. Hightower.** 1982. Coding assignments of the five smaller mRNAs of Newcastle disease virus. J. Virol. **43**:1024–1031.
9. **Dowling, P. C., C. Giorgi, L. Roux, L. A. Dethlefsen, M. E. Galantowicz, B. M. Blumberg, and D. Kolakofsky.** 1983. Molecular cloning of the 3′-proximal third of Sendai virus genome. Proc. Natl. Acad. Sci. USA **80**:5213–5216.
10. **Elango, N., J. E. Coligan, R. C. Jambou, and S. Venkatesan.** 1986. Human parainfluenza type 3 virus hemagglutinin-neuraminidase glycoprotein: nucleotide sequence of mRNA and limited amino acid sequence of the purified protein. J. Virol. **57**:481–489.
11. **Giorgi, C., B. M. Blumberg, and D. Kolakofsky.** 1983. Sendai virus contains overlapping genes from a single mRNA. Cell **35**:829–836.
12. **Gupta, K. C., and D. W. Kingsbury.** 1982. Conserved polyadenylation signals in two negative-strand RNA virus families. Virology **120**:518–523.
13. **Gupta, K. C., and D. W. Kingsbury.** 1984. Complete sequences of the intergenic and mRNA start signals in the Sendai virus genome—homologies with the genome of vesicular stomatitis virus. Nucleic Acids Res. **12**:3829–3841.
14. **Gupta, K. C., and D. W. Kingsbury.** 1985. Polytranscripts of Sendai virus do not contain intervening polyadenylate sequences. Virology **141**:102–109.
15. **Hall, W. W., R. A. Lamb, and P. W. Choppin.** 1980. Polypeptides of canine distemper virus: synthesis in infected cells and relatedness to the polypeptides of other morbilliviruses. Virology **100**:433–449.
16. **Herrler, G., and R. W. Compans.** 1982. Synthesis of mumps virus polypeptides in infected Vero cells. Virology **119**:430–438.
17. **Hiebert, S. W., R. G. Paterson, and R. A. Lamb.** 1985. Identification and predicted sequence of a previously unrecognized small hydrophobic protein, SH, of the paramyxovirus simian virus 5. J. Virol. **55**:744–751.
18. **Hsu, C. H., G. G. Re, K. C. Gupta, A. Portner, and D. W. Kingsbury.** 1985. Expression of Sendai virus defective-interfering genomes with internal deletions. Virology **146**:38–49.
19. **Jambou, R. C., N. Elango, and S. Venkatesan.** 1985. Proteins associated with human parainfluenza virus type 3. J. Virol. **56**:298–302.
20. **Kingsbury, D. W., M. A. Bratt, P. W. Choppin, R. P. Hanson, Y. Hosaka, V. ter Meulen, E. Norrby, W. Plowright, R. Rott, and W. H. Wunner.** 1978. Paramyxoviridae. Intervirology **10**: 137–152.
21. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
22. **Lamb, R. A., and P. W. Choppin.** 1978. Determination by peptide mapping of the unique polypeptides in Sendai virions and infected cells. Virology **84**:469–478.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Masters, P. S., and C. E. Samuel.** 1984. Detection of *in vivo* synthesis of polycistronic mRNAs of vesicular stomatitis virus. Virology **134**:277–286.
25. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
26. **Paterson, B. M., B. E. Roberts, and E. L. Kuff.** 1977. Structural gene identification and mapping by DNA-mRNA hybrid-arrested cell-free translation. Proc. Natl. Acad. Sci. USA **74**:4370–4374.
27. **Paterson, R. G., T. J. R. Harris, and R. A. Lamb.** 1984. Analysis and gene assignment of mRNAs of a paramyxovirus, simian virus 5. Virology **138**:310–323.
28. **Paterson, R. G., T. J. R. Harris, and R. A. Lamb.** 1984. Fusion protein of the paramyxovirus simian virus 5: nucleotide sequence of mRNA predicts a highly hydrophobic glycoprotein. Proc. Natl. Acad. Sci. USA **81**:6706–6710.
29. **Ricciardi, R. P., J. S. Miller, and B. E. Roberts.** 1979. Purifica-

tion and mapping of specific mRNAs by hybridization-selection and cell-free translation. Proc. Natl. Acad. Sci. USA **76:** 4927–4931.

30. **Richardson, C. D., A. Berkovich, S. Rozenblatt, and W. J. Bellini.** 1985. Use of antibodies directed against synthetic peptides for identifying cDNA clones, establishing reading frames, and deducing the gene order of measles virus. J. Virol. **54:**186–193.
31. **Rose, J. K.** 1980. Complete intergenic and flanking sequences from the genome of vesicular stomatitis virus. Cell **19:**415–421.
32. **Sanchez, A., and A. K. Banerjee.** 1985. Studies on human parainfluenza virus 3: characterization of the structural proteins and *in vitro* synthesized proteins coded by mRNAs isolated from infected cells. Virology **143:**45–54.
33. **Sanchez, A., and A. K. Banerjee.** 1985. Coding and gene assignments of mRNAs of human parainfluenza virus 3. Virology **147:**177–186.
34. **Shioda, T., Y. Hidaka, T. Kanda, T. Shibuta, A. Namoto, and K. Iwasaki.** 1983. Sequence of 3687 nucleotides from the 3′ end of Sendai virus genomic RNA and the predicted amino acid sequences of viral NP, P and C proteins. Nucleic Acids Res. **11:**7317–7331.

34a. **Spriggs, M. K., R. A. Olmsted, S. Venkatesan, J. E. Colligan, and P. L. Collins.** 1986. Fusion glycoprotein of human parainfluenza virus type 3: nucleotide sequence of the gene, direct identification of the cleavage-activation site, and comparison with other paramyxoviruses. Virology **152:**241–251.

35. **Storey, D. G., K. Dimock, and C. Y. Kang.** 1984. Structural characterization of virion proteins and genomic RNA of human parainfluenza virus 3. J. Virol. **52:**761–766.
36. **Tsao, S. G. S., C. F. Brunk, and R. E. Pearlman.** 1983. Hybridization of nucleic acids directly in agarose gels. Anal. Biochem. **131:**365–372.
37. **Wechsler, S. L., D. M. Lambert, M. S. Galinski, B. E. Heineke, and M. W. Pons.** 1985. Human parainfluenza virus 3: purification and characterization of subviral components, viral proteins and viral RNA. Virus Res. **3:**339–351.
38. **Wechsler, S. L., D. M. Lambert, M. S. Galinski, and M. W. Pons.** 1985. Intracellular synthesis of human parainfluenza type 3 virus-specified polypeptides. J. Virol. **54:**661–664.
39. **Wilde, A., and T. Morrison.** 1984. Characterization of polycistronic transcripts in Newcastle disease virus-infected cells, p. 161–166. *In* D. H. L. Bishop and R. W. Compans (ed.), Nonsegmented negative strand viruses: paramyxoviruses and rhabdoviruses. Academic Press, Inc., New York.
40. **Wilde, A., and T. Morrison.** 1984. Structural and functional characterization of Newcastle disease virus polycistronic RNA species. J. Virol. **51:**71–76.