Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 4594–4598, July 1986
Biochemistry

Nucleotide sequences for the gene junctions of human respiratory syncytial virus reveal distinctive features of intergenic structure and gene order

(paramyxoviruses/transcription signals/negative strand RNA viruses)

PETER L. COLLINS*, LILLIAN E. DICKENS†, ALICIA BUCKLER-WHITE*, ROBERT A. OLMSTED*, MELANIE K. SPRIGGS*, ENA CAMARGO*, AND KATHLEEN V. W. COELINGH*

*Laboratory of Infectious Diseases, Room 100, Building 7, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892; and †Department of Microbiology and Immunology, University of North Carolina School of Medicine, Chapel Hill, NC 27514

Communicated by Robert M. Chanock, December 16, 1985

ABSTRACT **Complete sequences for the intergenic regions of the genome of human respiratory syncytial virus were obtained by dideoxynucleotide sequencing using synthetic oligonucleotides. These experiments established that the 10 respiratory syncytial viral genes are arranged, without additional intervening genes, in the order 3′ *1C-1B-N-P-M-1A-G-F-22K-L* 5′. For the first nine genes, the exact gene boundaries were identified by comparison of the genomic sequences with previously determined mRNA sequences. The intergenic regions varied in length from 1 to 52 nucleotides and lacked any obvious conserved features of primary or secondary structure except that each sequence ended (3′ to 5′) with an adenosine residue. The exact start site of the 10th gene, the *L* gene, was not determined. However, RNA blot hybridization using a synthetic oligonucleotide designed from the genomic sequence mapped the *L* gene to within 54 nucleotides of the end of the penultimate *22K* gene. The lack of conservation of chain length and nucleotide sequence for the respiratory syncytial viral intergenic regions, together with the complexity of the genetic map, contrasts with previous observations for other nonsegmented negative-strand viruses.**

Human respiratory syncytial (RS) virus, a pneumovirus of the paramyxovirus family, is a pleomorphic, enveloped, RNA-containing virus that is a primary cause of severe pediatric respiratory tract disease (1). Efforts to characterize the viral genome and gene products have been complicated by the instability of the virus and its poor growth in tissue culture.

Ten viral mRNAs were identified by cDNA cloning and RNA blot hybridization analysis of infected-cell mRNAs (2, 3). The polypeptide coding assignments of the 10 mRNAs were identified by their individual translation *in vitro* (3). Complete cDNAs and nucleotide sequences have been obtained for 9 of the 10 viral mRNAs, excluding only the L mRNA (3–12). Each mRNA appears to encode a single major polypeptide. The 10 RS viral proteins identified to date are the major nucleocapsid protein (N, 42.6 kDa), nucleocapsid phosphoprotein (P, 27.2 kDa), large nucleocapsid protein (L, approximately 200 kDa), fusion glycoprotein (F, 68–70 kDa), large glycoprotein (G, 84–90 kDa), nonglycosylated matrix protein (M, 28.7 kDa), a second nonglycosylated matrix or membrane protein (22K protein, 22 kDa), two nonstructural proteins (1B protein, 14.7 kDa; 1C protein, 15.6 kDa), and a small hydrophobic protein that is present in virions in trace amounts (1A protein, 7.5 kDa) (3–13). As described here and elsewhere (1–3), the pattern of RS viral gene products is more complex than that of the six or seven proteins encoded by other paramyxoviruses such as Sendai virus, Newcastle disease virus, and simian virus type 5.

The RS viral genome is a single negative-strand of RNA of approximately 15,000 nucleotides (14), approximately equal to the combined sizes of the 10 viral mRNAs. The viral genome appears to be transcribed sequentially from a single promoter site, based on the kinetics of UV-inactivation of gene transcription (15). A viral transcriptional map, 3′ *1C-1B-N-P-M-1A-G-F-22K-L* 5′, was deduced from the UV mapping studies and from analysis of the relationships between the 10 mRNAs and a number of polycistronic readthrough mRNAs identified as minor products of viral transcription (2, 3). However, these experimental approaches were inexact and provided no direct or detailed information on the structure of the viral genome.

In the work described here, the intergenic and flanking gene sequences in genomic RNA were determined by the dideoxynucleotide method. This approach was designed to (*i*) confirm the sequences of the gene termini, deduced previously from cDNAs of mRNAs, (*ii*) investigate the structure of the intergenic regions, (*iii*) investigate the possible existence of additional, unidentified genes, and (*iv*) determine unambiguously the 3′ to 5′ order of the RS viral genes.

MATERIALS AND METHODS

Viral Genomic RNA. Cloned stocks of RS virus were propagated in HEp-2 cells and purified by differential centrifugation and two cycles of banding in 20–60% (wt/wt) discontinuous sucrose gradients (14). Virions were disrupted by incubation at 56°C for 20 min in the presence of 500 μg of proteinase K/ml and 0.3% $NaDodSO_4$, and the RNA was extracted with phenol/chloroform/isoamyl alcohol (25:25:1, vol/vol) and recovered by ethanol precipitation. On average, cells from 40 tissue-culture flasks (175 cm^2) yielded 100 μg of RNA, of which 10% was genome-length.

Dideoxynucleotide Sequencing. Synthetic oligonucleotide primers were purchased from Pharmacia P-L Biochemicals or OCS Labs (Denton, TX), or were synthesized on an Applied Biosystems (Foster City, CA) DNA synthesizer. Gel-purified primer (50 ng) was labeled at the 5′ end with [γ^{32}P]ATP by polynucleotide kinase and was purified by phenol extraction and gel filtration through Sephadex G-25 in H_2O. The labeled primer was lyophilized, combined with 10 μg of RNA, heated at 95°C for 1 min, quick-chilled, and extended with reverse transcriptase in the presence of 100 μM each dNTP and 5 μM of the appropriate dideoxynucleotide (6, 16). Nucleic acids were precipitated with ethanol and analyzed on 6% sequencing gels (17). For the experiments

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

Abbreviations: RS virus, respiratory syncytial virus; VSV, vesicular stomatitis virus.

shown here, the following primers were used, designated according to gene and position in the mRNA sequence: 1C-485, 5′ TATCAACTAGCA; 1B-459, 5′ CTCCATAGTC; N-1148, 5′ TCAAACATCAGC; P-871, 5′ AACTAACCAC-CCGGA; M-928, 5′ GATCTAGTACTCAAAT; 1A-363, 5′ TCCCACCATGCAAACCACTATCCA; G-892, 5′ CCCAA-CACACCACGCC; F-1832, 5′ ATCGAAACTCTCATCTA-TAAACCA; 22K-934, 5′ AATGTTTATCTAACC.

RNA Blots. Viral mRNAs were purified by phenol extraction from cytoplasmic extracts of virus-infected HEp-2 cells and selected by oligo(dT)-cellulose chromatography (2). RNAs were electrophoresed on 1% agarose/formaldehyde gels, transferred to nitrocellulose, and analyzed by hybridization with [^{32}P]oligonucleotides prepared by 5′ end-labeling and [^{32}P]cDNA clones prepared by nick-translation (18).

RESULTS

Sequences of Genomic Intergenic Regions. Complete nucleotide sequences were available for 9 of the 10 RS viral mRNAs, excluding only the L mRNA (4–12). From these sequences, synthetic oligonucleotides were designed that were complementary to genomic RNA near the 5′ (downstream) end of each gene. The oligonucleotides were 5′ end-labeled, hybridized to viral genomic RNA extracted from purified virions, and used to direct dideoxynucleotide sequencing of intergenic and flanking gene regions (Fig. 1).

Direct analysis of genomic RNA rather than of cloned cDNAs provided a consensus sequence from a population of authentic molecules and specifically avoided the possible introduction of artifacts during cDNA synthesis and cloning. However, the sequencing ladders shown in Fig. 1 did contain several ambiguities at individual positions that were attributed to (*i*) termination by reverse transcription independent of dideoxynucleotides, which is often observed during reverse transcription of RNA templates, and (*ii*) priming by oligonucleotides at secondary sites, which was minimized but not eliminated by reducing the primer to template ratio and by using alternate primers. Our experience from dideoxynucleotide sequencing of influenza and parainfluenza viral genomes is that these artifacts reflected the low yield of intact RS viral genomic RNA and the high content of contaminating nucleic acid copurifying with virus. Specific ambiguities were resolved as follows: the sequencing ladder for the *1B-N* intergenic region (Fig. 1*B*) contained additional bands that were most prominent in lane C. These were concluded to have arisen from priming at a secondary site because, among several experiments, their intensity varied independently of the authentic signals (not shown). The authenticity of the sequence shown was substantiated by the finding of its exact

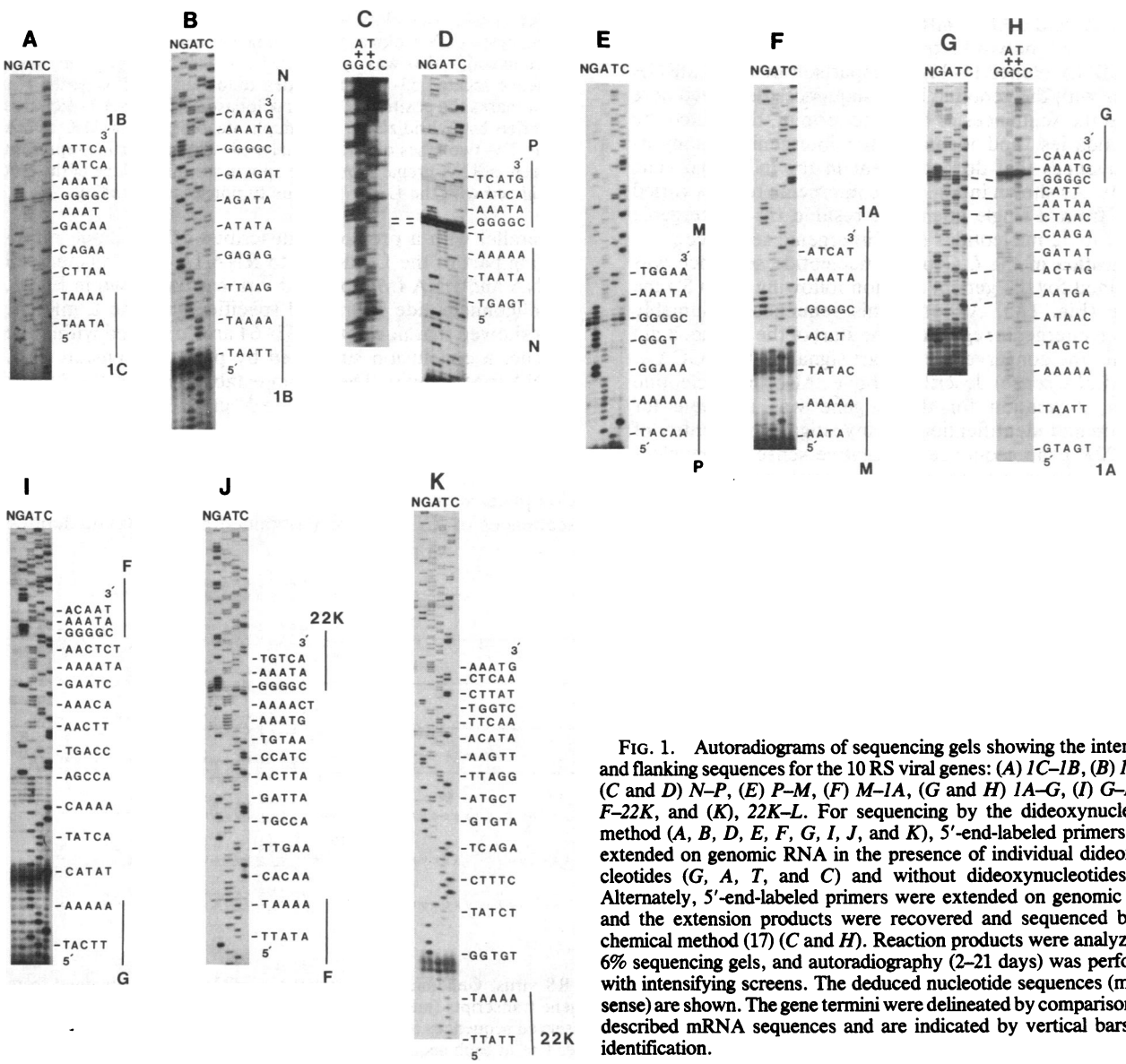

FIG. 1. Autoradiograms of sequencing gels showing the intergenic and flanking sequences for the 10 RS viral genes: (*A*) *1C–1B*, (*B*) *1B–N*, (*C* and *D*) *N–P*, (*E*) *P–M*, (*F*) *M–1A*, (*G* and *H*) *1A–G*, (*I*) *G–F*, (*J*) *F–22K*, and (*K*), *22K–L*. For sequencing by the dideoxynucleotide method (*A*, *B*, *D*, *E*, *F*, *G*, *I*, *J*, and *K*), 5′-end-labeled primers were extended on genomic RNA in the presence of individual dideoxynucleotides (*G*, *A*, *T*, and *C*) and without dideoxynucleotides (*N*). Alternately, 5′-end-labeled primers were extended on genomic RNA and the extension products were recovered and sequenced by the chemical method (17) (*C* and *H*). Reaction products were analyzed on 6% sequencing gels, and autoradiography (2–21 days) was performed with intensifying screens. The deduced nucleotide sequences (mRNA sense) are shown. The gene termini were delineated by comparison with described mRNA sequences and are indicated by vertical bars with identification.

complement at the intergenic junction of the 1B-N dicistronic mRNA, as determined by sequencing a cloned cDNA (4). The sequencing ladder of the *N-P* intergenic region (Fig. 1*D*) contained strong stops at two positions. The identities of these nucleotides were determined by chemical sequencing (17) of gel-purified extension products (Fig. 1*C*). For the *P-M* intergenic region (Fig. 1*E*), the identities of ambiguous nucleotides were resolved by using additional primers and by repeating the experiments several times (not shown): in each experiment there were strong stops at two or more positions, but the locations of the stops varied and consideration of all of the experiments yielded a complete sequence. For the *1A-G* intergenic region, the sequence was investigated both by the dideoxynucleotide method (Fig. 1*G*) and by chemical sequencing of gel-purified extension products (Fig. 1*H*). Both sequencing ladders contained ambiguities, but together they yielded a complete sequence. In several other intergenic and flanking sequences, strong stops obscured the identity of the initiating nucleotide of the downstream gene (5′ *G*GGG, in mRNA sense). However, in each case an assignment could be made by the observation (6) that the nine RS viral mRNAs sequenced to date initiate with the conserved sequence 5′ GGGGCAAAU.

The sequences of the intergenic and flanking gene regions are shown in Fig. 2. In previous work, the 5′ termini of 9 of the 10 viral mRNAs were mapped and sequenced precisely by primer extension and dideoxynucleotide sequencing of viral mRNA, and each 3′ mRNA terminus was identified by the location of poly(dA) tracts in several independently derived cDNA clones (4–12). Comparison of these mRNA sequences with the genomic RNA sequences described here confirmed the sequences of the gene termini, delineated the gene boundaries, and identified the intergenic regions as those sequences that do not appear in the individual gene transcripts. As shown in Fig. 2, the intergenic regions varied in length from a single adenosine residue (*N–P* intergenic sequence) to 52 nucleotides (*G–F* intergenic sequence).

Identification of the *L* Gene. A nucleotide sequence was also obtained for the genomic region following (3′ to 5′) the *22K* gene (Fig. 1*K*). Although this sequence presumably contains an intergenic region and the start of the *L* gene, it did not contain the conserved gene start signal (5′ GGGGCAAAU, in mRNA sense) described above. Also, no nucleotide sequence information for the *L* gene was available for comparison and identification. To investigate the identity of the post-*22K* gene sequence, a negative-sense oligonucleotide was synthesized to represent the genomic sequence of positions 53–64 (3′ CAGGAATAGAGT) following the *22K* gene. The oligonucleotide was 5′ end-labeled and hybridized,

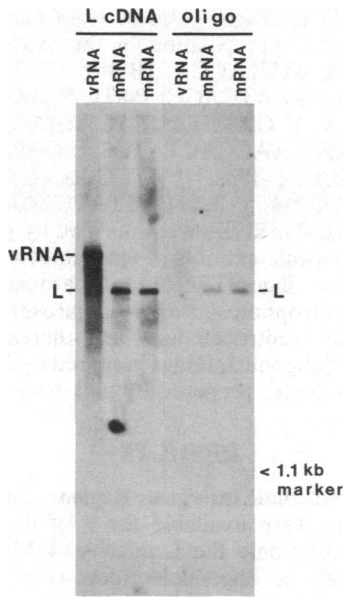

FIG. 3. Identification of the *L* gene. RNA extracted from purified virions and mRNA extracted from infected cells were separated by electrophoresis in a 1% agarose/formaldehyde gel and transferred to nitrocellulose. The blots were analyzed by hybridization with a radiolabeled cDNA clone of the L mRNA and with a radiolabeled oligonucleotide that was synthesized to contain the genomic RNA sequence located 53–64 nucleotides following the *22K* gene. The arrow marks the position of the 1.1-kilobase 22K mRNA, which was identified by hybridization of a duplicate blot with a 22K cDNA clone. The two lanes of mRNA analyzed with each probe represent different mRNA preparations. The spot in the middle lane of the blot hybridized with the L cDNA is due to nonspecific binding.

in parallel with a previously described cDNA clone representing part of the L mRNA, to RNA blots of infected-cell mRNA and RNA from purified virions. As shown in Fig. 3, the oligonucleotide hybridized specifically to the L mRNA. This showed that nucleotides 53–64 are contained within the *L* gene, a conclusion supported by primer extension on L mRNA (not shown). These data established that the *L* gene follows the *22K* gene in the 3′ to 5′ gene order.

DISCUSSION

Complete sequences for the intergenic and flanking gene sequences of RS virus were obtained by dideoxynucleotide

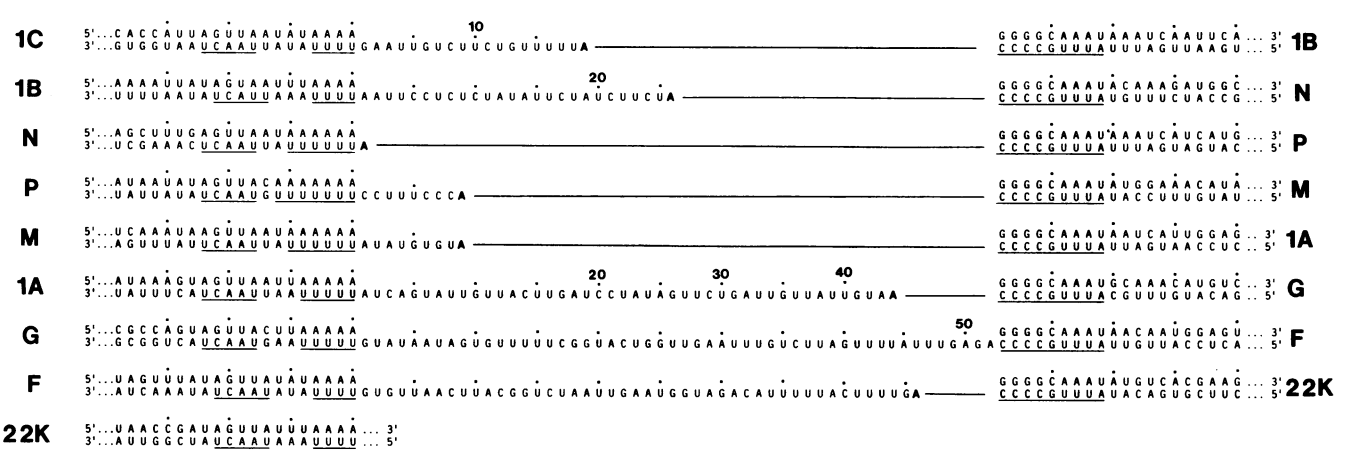

FIG. 2. Comparison of the intergenic and flanking gene sequences of RS virus. Genomic RNA sequences (vRNA) were obtained from experiments such as shown in Fig. 1, and the sequences of the individual gene transcripts (mRNA) were from published work and are shown to mark the gene boundaries. The intergenic sequences are numbered. Conserved sequences in the gene termini are underlined. The identities of the upstream (*Left*) and downstream (*Right*) genes are shown. The upper line in each sequence is mRNA, and the lower is vRNA.

sequencing of viral genomic RNA. Comparison of these sequences with the previously determined sequences of the termini of nine viral mRNAs delineated the intergenic regions and established the RS viral gene order as 3′ *1C-1B-N-P-M-1A-G-F-22K-L* 5′. These data provide a continuous sequence of 8593 nucleotides including the complete sequences of the first nine genes and their intergenic regions and a partial sequence of the *L* gene. No additional intervening genes were found. RNA blot hybridization showed that the *L* gene follows the *22K* gene in the 3′ to 5′ gene order. This observation was confirmed and extended by mapping and sequencing studies of the *L* gene and L transcripts that will be described in a subsequent report.

Comparison of the RS viral gene order with those of vesicular stomatitis virus (VSV) and the parainfluenza viruses (Fig. 4) showed apparent homology in the placements of the *N, P, M, F,* and *L* genes. On the other hand, numerous differences exist between the genome maps of RS virus and the parainfluenza viruses: (*i*) The genome sizes for RS virus and the parainfluenza viruses are similar (approximately 15,000 nucleotides), but the RS viral genome contains a larger number of relatively smaller, separate genes. (*ii*) The order of the RS viral glycoprotein genes (3′ *G–F* 5′) is the reverse of the parainfluenza order (3′ *F–HN* 5′). (*iii*) For Sendai and human type 3 viruses, the nonstructural C protein is encoded by a second, overlapping reading frame in the P mRNA (refs. 20 and 26; M.K.S. and P.L.C., unpublished data). In contrast, RS virus has two distinct nonstructural proteins (1C and 1B) that are encoded by nonoverlapping genes. (*iv*) The RS viral gene order is interrupted by the insertions of the *22K* and *1A* genes. The 22K protein lacks a known parainfluenza viral counterpart, but the small hydrophobic 1A protein might be analogous to the small hydrophobic SH protein of simian virus 5 (24). (*v*) The *N* gene is not first in the RS viral map, but instead is preceded by the nonstructural *1C* and *1B* protein genes. The RS viral N and P proteins are substantially smaller than their parainfluenza viral counterparts. Given the juxtaposition of these genes with the *1C* and *1B* genes, and postulating a common ancestor for the RS and parainfluenza viruses, it is interesting to speculate that the proteins encoded by the 3′-proximal genes of the RS virus (the *1C, 1B, N,* and *P* genes, totaling 3198 nucleotides) and of the parainfluenza viruses (the *N* and *P* +*C* genes, totaling 3578 nucleotides for Sendai virus) might have analogous enzymatic activities and functional domains that have become distributed among different numbers of polypeptide chains.

The first nine RS viral genes are nonoverlapping, separated by intergenic regions that by definition do not appear in individual gene transcripts. The eight intergenic regions described here total 206 nucleotides and, like the RS viral genome in general, are uridine-rich (33 G, 46 A, 98 U, and 27 C). The intergenic regions vary in length from a single nucleotide (the *N–P* intergenic region) to 52 nucleotides (the *G–F* intergenic region). Although several runs of 4–6 nucleotides were shared among several intergenic regions, no generally conserved sequence was apparent except that each intergenic region ended (3′ to 5′) with an adenosine residue. The lack of conservation of intergenic sequence or length described here contrasts with the conserved dinucleotide (3′ CA or 3′ GA) and trinucleotide (3′ GAA or 3′ GGG) intergenic sequences described for VSV and Sendai virus, respectively (19, 27). Computer-assisted analysis of the RS viral intergenic and flanking gene regions did not detect any significant conserved features of local secondary structure in the template or transcriptive products.

Intergenic structures have also been sequenced for five dicistronic readthrough mRNAs, the 1C-1B, IB-N, N-P, M-1A and F-22K mRNAs (4, 9–12). In each case, the mRNA intergenic sequence was the exact complement of the genomic sequence determined here, showing that the polytranscripts are generated by precise transcriptional readthrough of adjacent genes and the intervening intergenic region.

As shown in Fig. 2, the 3′ (upstream) termini of the nine RS viral genes sequenced to date initiate with the conserved nine-nucleotide sequence 3′ CCCCGUUUA that encodes the common sequence of the 5′ termini of the mRNAs. The genes of VSV and Sendai virus initiate with analogous 10-nucleotide sequences, termed gene-start signals. These sequences are conserved within each virus but do not share close sequence identity between viruses. As described above, the RS viral intergenic regions all end with a single adenosine residue, which might be a nontranscribed element of the start-signal for each following gene.

Examination of the complete sequences for the 5′ (downstream) termini of the nine RS viral genes (Fig. 2) revealed two variations of a conserved gene-end sequence:

3′ UCA$^{U}_{A}$UA$^{U}_{A}$AUUUU (genes *1B, F,* and *22K*), and

3′ UCAAU$^{UUUU}_{GAAA}$UUUU (genes *1C, N, P, M, 1A,* and *G*).

The principal difference between these two sequences is that the second contains an additional, intervening nucleotide (italicized). The gene-end sequences terminate with oligo(U) tracts of four to seven residues (Fig. 2). Analogous conserved gene-end sequences containing terminal oligo(U) tracts have been described for VSV, 3′ AUACUUUUUUU, and Sendai virus, 3′ UNAUUCUUUUU (19, 28, 29). The poly(A) tails of

RHABDOVIRUS-VSV			N	NS	M			G				L
PARAINFLUENZA-SENDAI			NP	P+C	M			F		HN		L
HUMAN TYPE 3			NP	P+C	M			F		HN		L
SV5			(NP	P+V)	M			F	SH	HN		(L)
PNEUMOVIRUS-RS VIRUS	1C	1B	N	P	M	1A	G	F			22K	L

FIG. 4. 3′ to 5′ gene orders for five nonsegmented negative-strand RNA viruses: VSV, a rhabdovirus (19); Sendai virus, a murine type 1 parainfluenza virus (20–22); human type 3 parainfluenza virus (M.K.S. and P.L.C., unpublished data); simian virus 5, a canine type 2 parainfluenza virus (23, 24); and the pneumovirus RS virus. Gene orders were determined by nucleotide sequence analysis except where indicated with parentheses, in which case the order is tentative. The gene orders have been aligned, with gaps introduced, to maximize the matching of analogous proteins. The VSV G protein has both attachment and fusogenic activities (25), indicating that it has analogy to both the HN and F proteins: it was placed arbitrarily in the F protein column because, like the F protein and unlike the HN protein, it has a COOH-terminal membrane anchor. Two proteins could not be aligned appropriately because of the constraints of the linear orders: the RS viral G protein is analogous to the HN protein because it probably is an attachment protein and is proposed to have an NH_2 terminal membrane anchor (7) and the RS viral 1A protein is a small, hydrophobic protein (12), probably analogous to the simian virus 5 SH protein (24). As described elsewhere the P mRNA of Sendai virus (20, 26) and human type 3 (M.K.S. and P.L.C., unpublished data) contains a second overlapping reading frame that encodes the nonstructural C protein. The nonstructural V protein of simian virus 5 is also encoded by the P mRNA, but is related by peptide mapping to the P protein and, therefore, probably is encoded by the same reading frame (23).

the mRNAs are thought to be synthesized by reiterative copying of the oligo(U) tracts in the gene-end sequences (29).

The 3′ to 5′ physical map described here is colinear with the viral transcriptional map. This indicates that transcription initiates at or near the 3′ genome end and proceeds linearly along the template. The two proposed mechanisms for sequential transcription of negative-strand genomes involve a precursor-processing mechanism or a stop/start mechanism (30, 31). Our data do not prove or disprove either model, but are more consistent with a stop/start mechanism. A processing model predicts that the polycistronic mRNAs are the substrates for an endonuclease, which would excise the plus-sense intergenic sequences. However, the lack of conserved sequence or secondary structure for the intergenic regions make them seem poorly suited as signals for processing. Also, because the polycistronic mRNAs lack long intervening poly(A) tracts, as described above, a processing mechanism does not provide for the polyadenylylation of mRNAs. For a possible stop/start mechanism, the considerable variability in the intergenic chain lengths suggests that the termination and reinitiation steps might not be tightly linked reactions for the RS viral polymerase. Instead, the polymerase presumably translocates without synthesis across the intergenic regions. Also, the lack of common sequences in the intergenic regions suggests that these sequences do not contain putative signals for transcriptive events such as polyadenylylation or capping.

For nonsegmented negative-strand viruses in general, attenuation of sequential transcription at the intergenic regions appears to be a principal factor determining the intracellular molar ratios of the different mRNAs and proteins. Attenuation was examined in detail for VSV and was found to occur to the same degree at each intergenic region (32), perhaps reflecting the high degree of conservation of intergenic and flanking structures. Preliminary analyses of the RS viral mRNAs suggest that their abundances also do generally reflect their map position (2, 3). The variation in the RS viral intergenic regions might be a mechanism for obtaining differing degrees of attenuation. Detailed quantitation of the intracellular viral mRNAs will be required to assess this hypothesis.

For several negative strand viruses, the intergenic regions share sequence identity both within each viral genome and among different viruses. For example, the intergenic sequence of Sendai virus, 3′ GAA (27), is shared with measles virus (33) and human parainfluenza type 3 (M.K.S. and P.L.C., unpublished data). As described above, the intergenic sequences of Sendai virus and VSV also are similar. The striking degree of identity of these sequences implied that they, in concert with less highly conserved flanking gene sequences, have functional or structural significance, most probably in transcriptive events at the gene junctions. Consistent with this, the intergenic regions appear to be the sites of transcriptional attenuation (32) and might contain the signal for termination of transcription (34). Also, nucleotide differences in the intergenic sequences of VSV and Sendai virus correlated with marked differences in the production of readthrough mRNAs (35, 36) suggestive of interaction between the intergenic sequences and the transcriptase. The striking homology among the intergenic regions of the different viruses mentioned above also suggested that these viruses share a common ancestor. Preliminary observations on the mode of RS viral transcription and replication were suggestive of general similarities with other nonsegmented negative-strand viruses (2, 14, 15), and the partial homology of the genetic maps (Fig. 4) was suggestive of a shared ancestor. It, therefore, is interesting that the RS viral intergenic structures are diverse in sequence and length. It will be important to reassess and obtain more detailed information on mode of the RS viral RNA synthesis. Interestingly, preliminary evidence indicates that the intergenic sequences of a second paramyxovirus, simian virus 5, also are nonconserved in sequence and length (24, 37). Further analyses of various paramyxoviruses will continue to provide new insights into structural, functional, and evolutionary relationships in this diverse virus family.

This work was performed in the laboratory of Drs. Brian R. Murphy and Robert M. Chanock. We thank them for their support and interest. We thank Charles E. Buckler for help with the computer work. L.E.D. was supported by Public Health Grants A112464 and A115134 to Gail W. Wertz.

1. McIntosh, K. & Chanock, R. M. (1985) in *Virology*, ed. Fields, B. N. (Raven, New York), pp. 1285–1304.
2. Collins, P. L. & Wertz, G. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3208–3212.
3. Collins, P. L., Huang, Y. T. & Wertz, G. W. (1984) *J. Virol.* **49**, 572–578.
4. Collins, P. L., Anderson, K., Langer, S. J. & Wertz, G. W. (1985) *Virology* **146**, 69–77.
5. Satake, M., Elango, N. & Venkatesan, S. (1984) *J. Virol.* **52**, 991–994.
6. Collins, P. L., Huang, Y. T. & Wertz, G. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7683–7687.
7. Wertz, G. W., Collins, P. L., Huang, Y., Gruber, C., Levine, S. & Ball, L. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4075–4079.
8. Satake, M. & Venkatesan, S. (1984) *J. Virol.* **50**, 92–99.
9. Collins, P. L. & Wertz, G. W. (1985) *J. Virol.* **54**, 65–71.
10. Collins, P. L. & Wertz, G. W. (1985) *Virology* **143**, 442–451.
11. Elango, N., Satake, M. & Venkatesan, S. (1985) *J. Virol.* **55**, 101–110.
12. Collins, P. L. & Wertz, G. W. (1985) *Virology* **141**, 283–291.
13. Huang, Y. T., Collins, P. L. & Wertz, G. W. (1985) *Virus Res.* **2**, 157–173.
14. Huang, Y. T. & Wertz, G. W. (1982) *J. Virol.* **43**, 150–157.
15. Dickens, L. E., Collins, P. L. & Wertz, G. W. (1984) *J. Virol.* **52**, 364–369.
16. Zimmern, D. & Kaseberg, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4257–4261.
17. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–559.
18. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
19. Rose, J. K. (1980) *Cell* **19**, 415–421.
20. Shioda, T., Hidaka, Y., Kanda, T., Shibuta, H., Nomoto, A. & Iwasaki, K. (1983) *Nucleic Acids Res.* **21**, 7317–7330.
21. Dowling, P. C., Giorgi, C., Roux, L., Dethlefsen, L. A., Galantowicz, M. E., Blumberg, B. M. & Kolakofsky, D. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5213–5216.
22. Blumberg, B., Giorgi, C., Roux, L., Raju, R., Dowling, P., Chollet, A. & Kolakofsky, D. (1985) *Cell* **41**, 269–276.
23. Paterson, R. G., Harris, T. J. R. & Lamb, R. A. (1984) *Virology* **138**, 310–323.
24. Hiebert, S. W., Paterson, R. G. & Lamb, R. A. (1985) *J. Virol.* **55**, 744–751.
25. Florkiewicz, R. & Rose, J. (1984) *Science* **225**, 721–723.
26. Giorgi, C., Blumberg, B. M. & Kolakofsky, D. (1983) *Cell* **35**, 829–836.
27. Gupta, K. C. & Kingsbury, D. W. (1984) *Nucleic Acids Res.* **12**, 3829–3841.
28. Gupta, K. C. & Kingsbury, D. W. (1982) *Virology* **120**, 518–523.
29. Schubert, M., Keene, J. D., Herman, R. C. & Lazzarini, R. A. (1980) *J. Virol.* **34**, 550–559.
30. Banerjee, A. K., Abraham, G. & Colonno, R. J. (1976) *J. Gen. Virol.* **34**, 1–8.
31. Ball, L. A. & Wertz, G. W. (1981) *Cell* **26**, 143–144.
32. Iverson, L. E. & Rose, J. K. (1981) *Cell* **23**, 477–484.
33. Richardson, C. D., Berkovich, A., Rozenblatt, S. & Bellini, W. J. (1985) *J. Virol.* **54**, 186–193.
34. Hsu, C.-H., Re, G. G., Gupta, K. C., Portner, A. & Kingsbury, D. W. (1985) *Virology* **146**, 38–45.
35. Masters, P. S. & Samuel, C. E. (1984) *Virology* **134**, 277–286.
36. Gupta, K. C. & Kingsbury, D. W. (1985) *Virology* **141**, 102–109.
37. Paterson, R. G., Harris, T. J. R. & Lamb, R. A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6706–6710.