JOURNAL OF VIROLOGY, Apr. 1984, p. 92–99
0022-538X/84/040092-08$02.00/0

Vol. 50, No. 1

Nucleotide Sequence of the Gene Encoding Respiratory Syncytial Virus Matrix Protein

MASANOBU SATAKE AND SUNDARARAJAN VENKATESAN*

Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20205

Received 9 September 1983/Accepted 12 December 1983

The amino acid sequence of the matrix protein of the human respiratory syncytial virus (RS virus) was deduced from the sequence of a cDNA insert in a recombinant plasmid harboring an almost full-length copy of this gene. It specifically hybridized to a single 1,050-base mRNA from infected cells. The recombinant containing 944 base pairs of RS viral matrix protein gene sequence lacked five nucleotides corresponding to the 5′ end of the mRNA. The nucleotide sequence of the 5′ end of the mRNA was determined by the dideoxy sequencing method and found to be 5′ NGGGC, wherein the C residue is one nucleotide upstream of the cloned viral sequence. The initiator ATG codon for the matrix protein is embedded in an AATATGG sequence similar to the canonical PXXATGG sequence present around functional eucaryotic translation initiation codons. There is no conserved sequence upstream of the polyadenylate tail, unlike vesicular stomatitis virus and Sendai virus, in which four nucleotides upstream of the polyadenylate tail are conserved in all genes. There is no equivalent of the eucaryotic polyadenylation signal AAUAAA upstream of the polyadenylate tail. The matrix protein of 28,717 daltons has 256 amino acids. It is relatively basic and moderately hydrophobic. There are two clusters of hydrophobic amino acid residues in the C-terminal third of the protein that could potentially interact with the membrane components of the infected cell. The matrix protein has no homology with the matrix proteins of other negative-strand RNA viruses, implying that RS virus has undergone extensive evolutionary divergence. A second open reading frame potentially encoding a protein of 75 amino acids and partially overlapping the C terminus of the matrix protein was also identified.

Human respiratory syncytial virus (RS virus), a pleomorphic enveloped virus, is the most important etiological agent of severe lower respiratory illness among infants and children (5). Although originally classified as a paramyxovirus, it has been placed in a separate genus, *Pneumovirus*, because of morphological differences and lack of hemagglutinin and neuraminidase activities. It grows poorly in tissue culture and presents formidable problems in purification because of virus instability. This has hindered definitive biological and molecular studies with this virus. It is only recently that the general polypeptide composition and genetic organization of this virus have been described. It is a negative-strand RNA virus containing an encapsidated 5,000 kilodalton (kdal) genomic RNA (13). Seven or eight virus-encoded polypeptides have been identified (2, 24, 34, 40) consistent with the genetic analysis of *ts* mutants that identified seven or eight nonoverlapping complementation groups (10, 39). There are two surface glycoproteins of 84 and 68 kdal responsible for recognition of cellular receptors and entry and cell fusion. The internal viral proteins include a 160-kdal protein (probably the viral polymerase by analogy with paramyxoviruses), a 46-kdal nucleocapsid protein, a 36-kdal phosphoprotein, a 28-kdal matrix protein, and two nonstructural proteins (NS_1 and NS_2) of 22 and 18 kdal, respectively. The functional roles of the nonstructural proteins are presently not understood.

Negative-strand RNA viruses mature in the cytoplasm of infected cells. During viral morphogenesis, they incorporate host cell membranes as the outer envelope; within this envelope are embedded viral glycoproteins responsible for viral penetration and cell-to-cell spread. The structure of the budding virions is maintained by an abundant viral protein referred to as membrane or matrix protein that forms an inner shell beneath the viral envelope scaffolding the viral nucleocapsid within the virus-altered plasma membrane. In addition to a purely structural role, the M protein has also been demonstrated to have a regulatory function affecting transcription. In the case of vesicular stomatitis virus (VSV), a well-studied prototype of unsegmented negative-strand RNA viruses, purified matrix protein has been shown to inhibit transcription of viral nucleocapsids and to cause a morphological change in the structure of the nucleocapsids (4, 8). In addition, *ts* mutants of VSV containing a lesion in the M protein gene have an enhanced rate of transcription at nonpermissive temperatures (6). Influenza virus M protein may have a similar role in regulating transcription. It has been shown that the susceptibility of influenza virus to the antiviral drug amantadine is mediated by the M protein influencing transcription (21, 41).

RS virus M protein would be expected to share some of the properties described above. Knowledge of the primary structure of this protein would provide information regarding the possible polypeptide domains involved in the interactions described above. In addition, we are interested in understanding the genetic organization of this important human pathogen at the nucleotide level. Toward this goal, we have employed recombinant DNA techniques to analyze the different transcriptional units of this virus. Recently, cDNA clones encoding four RS viral genes (nucleocapsid, matrix, phosphoprotein, or nonstructural protein 2) were identified and characterized (34). One of these, containing an almost complete copy of the RS viral nucleocapsid gene, was sequenced and the amino acid sequence of the encoded protein was deduced (9). In this report, we describe the

* Corresponding author.

sequence of the RS viral M protein gene and discuss its significance in relation to other known viral M proteins.

MATERIALS AND METHODS

Isolation of mRNA. HEp-2 cell monolayers were infected by RS virus, strain A2, at a multiplicity of infection of 5 PFU per cell and maintained in Eagle minimum essential medium supplemented with 2% heat-inactivated fetal calf serum. Actinomycin D (2 μg/ml) was added 8 h after infection. At 16 h after infection, cells were harvested and cytoplasmic RNA was isolated as described previously (11). Polyadenylate [poly(A)]-containing RNA was purified by two cycles of oligodeoxythymidylate-cellulose chromatography (1).

Characterization of recombinant cDNA plasmids. Recombinants $pRSB_7$ and $pRSD_3$ encoding the RS viral M protein gene were previously identified in a RS cDNA library constructed in *Escherichia coli* (HB101), using plasmid pBR322 as a vector (34). Recombinants were grown in medium containing tetracycline (12.5 μg/ml), and plasmid DNA was prepared by using a scaled-up version of the modified alkali-sodium dodecyl sulfate procedure (3). Both $pRSB_7$ and $pRSD_3$ lacked the *Pst*I site used for cloning: however, on cleavage with *Hpa*II, fragments of 950 and 1,020 base pairs (bp) containing RS viral sequence were released. The 1,020-bp *Hpa*II fragment derived from $pRSD_3$ was digested with *Dde*I, and a 300-bp *Dde*I fragment containing the coding sequence of RS viral M protein gene was isolated by electroelution (37). This fragment was nick translated (26) and used to probe a sublibrary of RS cDNA clones by the colony hybridization procedure (31). Colonies reacting positively with the probe were selected, and the recombinant plasmids were purified by a mini-isolation procedure (3). *Hpa*II fragments containing RS viral sequence were isolated and screened by restriction nuclease digestion. A recombinant ($pRSA_{11}$) contained a 1,112-bp *Hpa*II fragment that was similar to the *Hpa*II fragment by $pRSD_3$ based on isomorphous restriction enzyme cleavage patterns and limited DNA sequencing. Restriction mapping was accomplished by partial digestion of asymmetrically end-labeled fragments (30) and by double digestions (18).

End labeling of DNA. Restriction fragments were treated with calf intestinal alkaline phosphatase and labeled at the 5′ end with [γ-^{32}P]ATP and T4 polynucleotide kinase. The Klenow fragment of DNA polymerase or avian myeloblastosis reverse transcriptase was used to add a single complementary [α-^{32}P]dNTP at the recessed 3′ end of the fragments (22).

DNA sequencing. End-labeled DNA fragments were cleaved by secondary restriction enzymes, and asymmetrically labeled fragments were purified by agarose gel electrophoresis. A partial chemical degradation procedure (22) was used for DNA sequencing. Sequencing reaction products were resolved on gels (0.4 mm thick) with 6, 8, or 20% acrylamide gels (30 by 160 and 40 by 80 cm). DNA sequence was analyzed by the computer program of Queen and Korn (25). The program developed by Wilbur and Lipman (36) was used to search for homologies among different proteins.

RNA blotting and filter hybridizations. Poly(A)-containing RNA from infected cells treated with actinomycin D (2 μg/ml) or from mock-infected cells was denatured in 6% formaldehyde and electrophoresed on formaldehyde agarose (1.5%) gels as described by Lehrach et al. (20). After electrophoresis, the RNA was quantitatively transferred to nitrocellulose paper and hybridized to ^{32}P-labeled plasmid DNA by the method of Thomas (32), with the modification that the hybridization medium also contained poly(A) (50 μg/ml) and polyuridylate (50 μg/ml). Additionally, recombinant DNAs were partially digested with *Hpa*II, denatured, and immobilized on nitrocellulose filters. The DNA filters were used to hybrid select virus specific mRNAs that were then translated in vitro in a reticulocyte lysate system as described before (34).

Mapping the 5′ end of the mRNA. Poly(A)-containing RNA purified from virus-infected, actinomycin D-treated cells was hybridized to a DNA fragment containing a label at the 5′ end of the coding strand (33). The RNA-DNA hybrids were purified after hybridization, and the DNA primer was extended on the RNA template essentially as described elsewhere (33 and references therein). The 5′ sequence of the mRNA was determined by the dideoxy sequencing method, using the same DNA primer.

Materials. Restriction endonucleases were purchased from Bethesda Research Laboratories (Gaithersburg, Md.) or New England Biolabs (Beverly, Mass.). Calf intestinal alkaline phosphatase, DNA polymerase holoenzyme, and the Klenow fragment were from Boehringer Mannheim Biochemicals (Indianapolis, Ind.). Avian myeloblastosis virus reverse transcriptase was obtained from Life Sciences Corp. (Coral Gables, Fla.). T4 polynucleotide kinase and nucleoside triphosphates were from P-L Biochemical (Milwaukee, Wis.). Oligodeoxynucleotides were from Collaborative Research, Inc. (Waltham, Mass.). Radioactive nucleotides were from Amersham Corp. (Arlington Heights, Ill.).

RESULTS

A library of recombinant cDNA plasmids containing sequences representative of poly(A)-containing RNAs from RS virus-infected cells was constructed. A cloning strategy was employed to optimize yields of nearly full-length cDNA clones encoding RS virus-specific sequences (17). This included (i) isolation of discrete-sized, single-stranded cDNAs and (ii) oligodeoxycytidylic acid [oligo(dC)] tailing of the first cDNA strand and oligodeoxyguanylic acid [oligo(dG)] priming for synthesis of the second strand. This strategy avoided the self-priming reaction for second-strand synthesis and the need for S1 nuclease treatment to remove the hairpin caused by self priming. Double-stranded cDNAs tailed at the 3′ end by oligo(dC) were inserted into the oligo(dG)-tailed *Pst*I site of plasmid pBR322 before transformation of *E. coli* (HB101). Recombinant plasmids containing RS viral cDNA inserts were identified by several screening procedures (34). Many of the cDNA clones lacked one or both *Pst*I sites; however, large inserts were released by digestion with *Hpa*II which cleaves the vector DNA approximately 50 bp on either side of the *Pst*I site. The absence of *Hpa*II sites within several cloned cDNA inserts allowed us to recover *Hpa*II fragments encoding entire RS viral sequences. Two recombinant plasmids ($pRSB_7$ and $pRSD_3$) previously identified as containing the M protein gene (34) were initially selected for DNA sequence analysis. Neither had a string of A residues corresponding to the 3′ end of the mRNA, probably resulting from abortive second-strand synthesis of cDNA. However, they shared extensive sequence homology at one end. In addition, the RS cDNA inserts from these plasmids hybridized to an mRNA from infected cells that was larger than the insert as determined by RNA blotting and filter hybridization. A restriction fragment containing RS viral sequence was isolated from $pRSD_3$, radioactively labeled, and used to identify other clones that contained M protein gene sequences. One recombinant plasmid ($pRSA_{11}$) bearing a larger insert was thus identified. It shared extensive sequence homology with $pRSB_7$ and

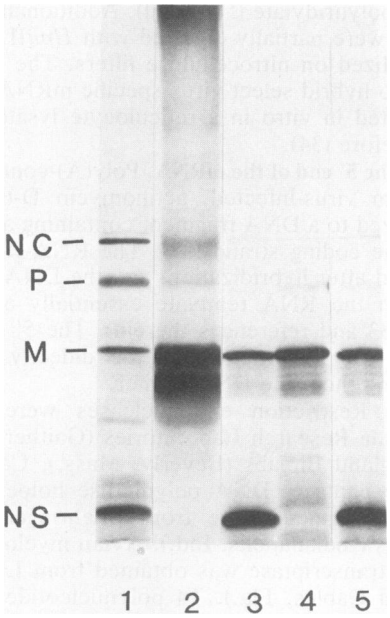

FIG. 1. In vitro translation of RS viral mRNAs. Poly(A)-containing RNA from RS virus-infected cells (5 μg) was hybridized to nitrocellulose filters containing immobilized recombinant RS viral plasmid DNA (40 μg). The conditions of hybridization, subsequent washing, and elution of specific mRNAs were described elsewhere (34). The eluted mRNAs were translated in a rabbit reticulocyte messenger dependent system. Lanes 1 through 5 represent the gel electrophoretic profiles of the translation products of unselected mRNA from infected cells or RNA hybrid selected by the recombinant DNAs $pRSA_{11}$, $pRSB_7$, $pRSC_{12}$, and $pRSD_3$.

$pRSD_3$ within the relevant *Hpa*II fragment containing the RS cDNA sequence. Hybrid selection of mRNA from infected cells, using plasmid DNA from $pRSA_{11}$, and subsequent cell-free translation of selected mRNA yielded M protein (Fig. 1). The polypeptide migrating faster than the authentic viral M protein has been shown to be the result of premature termination of translation (34). A translation product comigrating with a viral nonstructural protein was present on translation of mRNAs hybrid selected with certain plasmids although these plasmids had extensive sequence homologies with $pRSA_{11}$ (Fig. 1).

DNA sequencing. The plasmid $pRSA_{11}$ was cleaved with *Hpa*II, and a 1,112-bp *Hpa*II fragment containing RS viral sequence was isolated. Figure 2 illustrates the restriction map of the *Hpa*II insert derived from $pRSA_{11}$ and the sequencing strategy. An RS viral cDNA sequence of 944 nucleotides is shown in Fig. 3 in the messenger sense. There was a single long open reading frame starting at the fifth nucleotide after the G:C tails up to position 772 in the cDNA insert. This encoded a protein of 28,717 daltons containing 256 amino acids rich in basic amino acids and moderately hydrophobic. The termination codon for this open reading frame was 169 nucleotides upstream of the 6A residues derived from the poly(A) tail of the mRNA. Figure 4 is a schematic illustration of translation of the DNA sequence in three different reading frames. In reading frame +1 the protein mentioned above is illustrated. In frame +2 there is an open reading frame starting at position 714 and ending at position 938, potentially encoding a protein of 75 amino acids. Reading frame 0 has numerous stop codons, whereas regions devoid of stop codons have no methionines. The initiator codon ATG for the protein of 256 amino acids is embedded in an AATATGG sequence similar to the canonical PXXATGG sequence that is present around functional eucaryotic translational initiator codons (16). On the other hand, the initiator codon for the protein potentially encoded in the +2 frame is within an ATTATGT sequence. There is no counterpart of the canonical eucaryotic polyadenylation sequence AAUAAA upstream of the putative poly(A) tail. As mentioned in an earlier report (9), no homologous sequences upstream of the poly(A) tail are present in four RS viral transcripts. This is quite unlike the case of VSV or Sendai virus, wherein an AUAC or an AUUC sequence is present in all genes preceding the seven or five U residues that are reiteratively transcribed to yield the poly(A) tail of transcripts (12, 19, 28). In the case of influenza virus, transcription terminates at a string of eight U residues in all segments, 15 to 21 nucleotides upstream of the 5′ end of the template. The eight U residues are reiteratively transcribed to yield the poly(A) tail of the mRNAs, but the transcripts

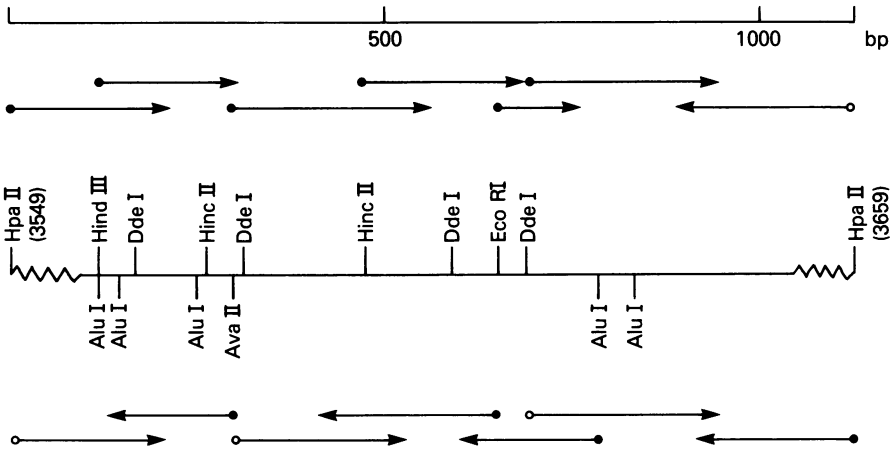

FIG. 2. Restriction map and sequencing strategy. The 944 bp of RS viral cDNA sequence contained within the *Hpa*II fragment of $pRSA_{11}$ is illustrated by a straight line. The wavy lines on either side represent the pBR322 sequence and the G:C tails generated during cloning. Restriction sites relevant to DNA sequencing are illustrated. The arrows with closed and open circles denote 5′ and 3′ end-labeled DNA fragments. The arrowheads indicate the extent of sequence determination.

```
                                                    10                                                  20
AAAT ATG GAA ACA TAC GTG AAC AAG CTT CAC GAA GGC TCC ACA TAC ACA GCT GCT GTT CAA TAC
     MET GLU THR TYR VAL ASN LYS LEU HIS GLU GLY SER THR TYR THR ALA ALA VAL GLN TYR
                                                    30                                                  40
     AAT GTC TTA GAA AAA GAC GAT GAC CCT GCA TCA CTT ACA ATA TGG GTG CCC ATG TTC CAA
     ASN VAL LEU GLU LYS ASP ASP ASP PRO ALA SER LEU THR ILE TRP VAL PRO MET PHE GLN
                                                    50                                                  60
     TCA TCT ATG CCA GCA GAT TTA CTT ATA AAA GAA CTA GCT AAT GTC AAC ATA CTA GTG AAA
     SER SER MET PRO ALA ASP LEU LEU ILE LYS GLU LEU ALA ASN VAL ASN ILE LEU VAL LYS
                                                    70                                                  80
     CAA ATA TCC ACA CCC AAG GGA CCT TCA CTA AGA GTC ATG ATA AAC TCA AGA AGT GCA GTG
     GLN ILE SER THR PRO LYS GLY PRO SER LEU ARG VAL MET ILE ASN SER ARG SER ALA VAL
                                                    90                                                  100
     CTA GCA CAA ATG CCC AGC AAA TTT ACC ATA TGC GCT AAT GTG TCC TTG GAT GAA AGA AGC
     LEU ALA GLN MET PRO SER LYS PHE THR ILE CYS ALA ASN VAL SER LEU ASP GLU ARG SER
                                                    110                                                 120
     AAA CTA GCA TAT GAT GTA ACC ACA CCC TGT GAA ATC AAG GCA TGT AGT CTA ACA TGC CTA
     LYS LEU ALA TYR ASP VAL THR THR PRO CYS GLU ILE LYS ALA CYS SER LEU THR CYS LEU
                                                    130                                                 140
     AAA TCA AAA AAT ATG TTG ACT ACA GTT AAA GAT CTC ACT ATG AAG ACA CTC AAC CCT ACA
     LYS SER LYS ASN MET LEU THR THR VAL LYS ASP LEU THR MET LYS THR LEU ASN PRO THR
                                                    150                                                 160
     CAT GAT ATT ATT GCT TTA TGT GAA TTT GAA AAC ATA GTA ACA TCA AAA AAA GTC ATA ATA
     HIS ASP ILE ILE ALA LEU CYS GLU PHE GLU ASN ILE VAL THR SER LYS LYS VAL ILE ILE
                                                    170                                                 180
     CCA ACA TAC CTA AGA TCC ATC AGT GTC AGA AAT AAA GAT CTG AAC ACA CTT GAA AAT ATA
     PRO THR TYR LEU ARG SER ILE SER VAL ARG ASN LYS ASP LEU ASN THR LEU GLU ASN ILE
                                                    190                                                 200
     ACA ACC ACT GAA TTC AAA AAT GCT ATC ACA AAT GCA AAA ATC ATC CCT TAC TCA GGA TTA
     THR THR THR GLU PHE LYS ASN ALA ILE THR ASN ALA LYS ILE ILE PRO TYR SER GLY LEU
                                                    210                                                 220
     CTA TTA GTC ATC ACA GTG ACT GAC AAC AAA GGA GCA TTC AAA TAC ATA AAG CCA CAA AGT
     LEU LEU VAL ILE THR VAL THR ASP ASN LYS GLY ALA PHE LYS TYR ILE LYS PRO GLN SER
                                                    230                                  **  *          240
     CAA TTC ATA GTA GAT CTT GGA GCT TAC CTA GAA AAA GAA AGT ATA TAT TAT GTT ACC ACA
     GLN PHE ILE VAL ASP LEU GLY ALA TYR LEU GLU LYS GLU SER ILE TYR TYR VAL THR THR
                                                    250
     AAT TGG AAG CAC ACA GCT ACA CGA TTT GCA ATC AAA CCC ATG GAA GAT TAA CCTTTTTCCT
     ASN TRP LYS HIS THR ALA THR ARG PHE ALA ILE LYS PRO MET GLU ASP END

     CTACATCAGT GTGTTAATTC ATACAAACTT TCTACCTACA TTCTTCACTT CACCATCACA ATCACAAACA CTCTG

     TGGTTCAACC AATCAAACAA AACTTATCTG AAGTCCCAGA TCATCCCAAG TCATTGTTTA TCAGATCTAG TACTC
        ***
     AAATAAGTTA ATAAAAAA

     MOLECULAR WEIGHT = 28717
```

FIG. 3. DNA sequence of RS virus M protein gene. The DNA sequence is given in the messenger sense. The derived amino acid sequence in reading frame +1 is numbered starting with the N-terminal methionine. The sequence contained within the asterisks represents the second open reading frame encoding 75 amino acids.

lack homologous sequences upstream of the poly(A) tail (27).

mRNA for the RS virus M protein. Other workers have demonstrated that poly(A) RNAs from RS virus infected cells could be resolved by electrophoresis on acid urea agarose gels into eight or nine discrete species (7, 13). One of these, approximately 1,050 bases in length, was translated in a cell-free system yielding a 28-kdal protein, the putative M protein of RS virus (14). When ^{32}P-labeled pRSA$_{11}$ was hybridized to poly(A) RNAs from infected cells that had been electrophoresed under denaturing conditions and transferred to nitrocellulose, we observed a single radioactive band of ca. 1,050 bases (Fig. 5). RNA from uninfected cells failed to hybridize to the recombinant under these conditions. This implied that the plasmid is a nearly full-length copy of the transcript, making an allowance for an average of 100 A residues at the 3′ end of the viral transcripts.

The 5′ end of the mRNA was mapped by the primer extension method (33). A *Hin*dIII-*Dde*I fragment downstream from the 5′ end of the mRNA was 5′ end labeled at the *Dde*I site (Fig. 6, inset). The labeled primer was then annealed to mRNA from infected cells and extended on the template by using reverse transcriptase. Primer extension occurred only when the label was at the *Dde*I site, thus confirming the orientation of the cloned DNA. To locate precisely the 5′ end of the mRNA within a few nucleotides, a *Dde*I-*Hha*I fragment 5′ end labeled at the *Dde*I site was used to generate a chemical sequence ladder. The *Hha*I site lies within the pBR322 sequence between the *Hpa*II site and the original *Pst*I site (Fig. 6). When extension products of *Hin*dIII-*Dde*I primer 5′ end labeled at the *Dde*I site were compared with the chemical DNA sequence ladder, it was clear that the 5′ end of the mRNA lay seven nucleotides beyond the end of the RS viral sequence (Fig. 6, lanes 3 to 5). Thus, this plasmid appeared to lack only the seven nucleotides coresponding to the 5′ end of the mRNA. Artifactual extension products seen in a complete reaction (lane 3) or when RNA was omitted (lane 2) were eliminated in the presence of actinomycin D (lane 4) or sodium PP$_i$ (lane 5) in the reaction. As discussed in an earlier paper, these addition-

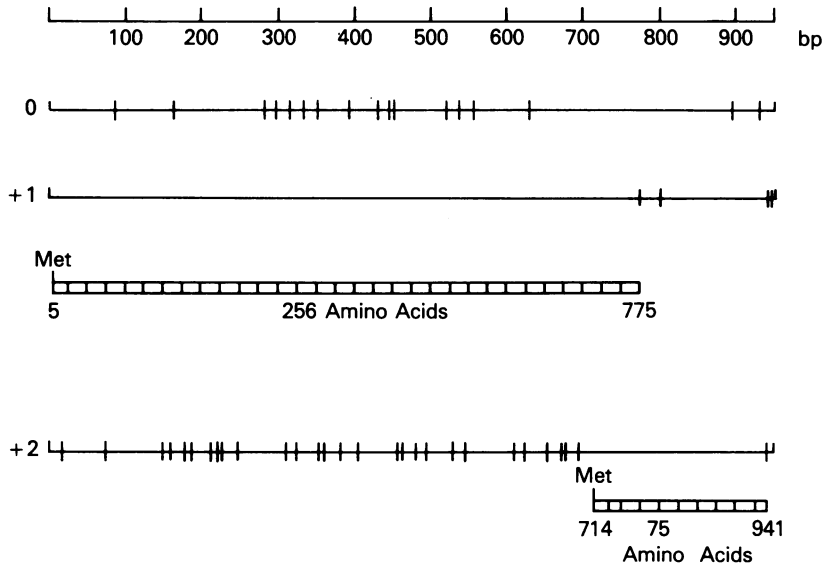

FIG. 4. Translation of the RS viral cDNA sequence in three different reading frames. The length of the cloned sequence is given in bp. The termination codons in all three reading frames are denoted by vertical lines. The potential polypeptides encoded in frames +1 and +2 are shown by hatched rectangles, with the number of amino acids within each protein given underneath.

al smaller extension products represent self-copying reactions (9). The 5′ end sequence of the mRNA was further determined by the dideoxy sequencing method. The GGGCAAAT sequence at the 5′ end of mRNA, wherein the underlined nucleotides GGGC were missing in the cloned cDNA, was deduced by this procedure (Fig. 7). However, the initiating nucleotide could not be determined by this method. The identity of the initiating nucleotide could only be established by RNA sequencing of the 5′-capped oligonucleotides derived from the mRNA or by sequencing the viral genomic RNA across the intercistronic boundaries. However, our analysis provided a more accurate mapping of the 5′ end of the mRNA than the one described earlier (Fig. 6), since a 1- to 1.5-base downward correction had to be made when primer extension products were run alongside the chemical sequencing reactions (35 and references therein).

DISCUSSION

A recombinant plasmid ($pRSA_{11}$) containing 944 bp of RS viral cDNA sequence was selected for DNA sequence analysis. The plasmid DNA specifically hybrid selected mRNA from infected cells that yielded M protein on cell-free translation. In addition, on Northern blot analysis, the ^{32}P-labeled DNA reacted with a single mRNA ca. 1,050 bases in length. Additional faint, higher-molecular-weight bands were only occasionally observed. They have been interpreted to be polycistronic transcripts (7), and, as mentioned in an earlier report (9), their significance is not clear to us. We observed such faint bands only when mRNA was prepared from cells infected at a high multiplicity. They might represent transcription from defective interfering genomes lacking intercistronic regulatory sequences. The plasmid lacked 7 bases corresponding to the 5′ end of the mRNA. Dideoxy sequencing of the mRNA revealed this recombinant plasmid to be lacking only a five-nucleotide NGGGC sequence. This error can be attributed to visual comparisons between enzymatic and chemical sequence ladders. The fact that the DNA insert hybridized to an mRNA approximately 100 nucleo-

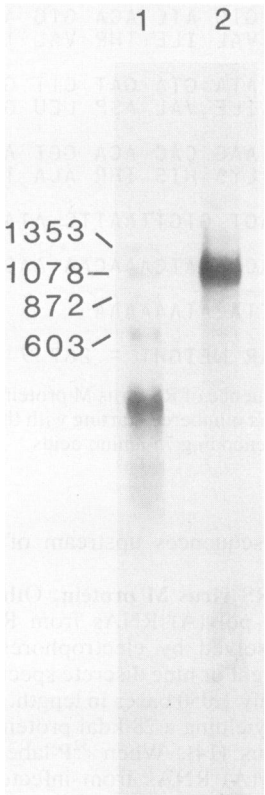

FIG. 5. Hybridization of ^{32}P-labeled $pRSA_{11}$ to poly(A) RNA from RS virus-infected cells. Poly(A)-containing RNA from RS virus-infected cells was resolved by formaldehyde agarose (1.5%) gel electrophoresis and transferred to nitrocellulose paper. Hybridization was performed as described in the text. Lane 2 illustrates the results of hybridization to viral mRNA. End-labeled *Hae*III restriction fragments of ϕX174 replicative form DNA were electrophoresed alongside the RNA and transferred to nitrocellulose paper; they served as molecular weight markers (lane 1). The sizes of relevant fragments in bases are indicated on the left.

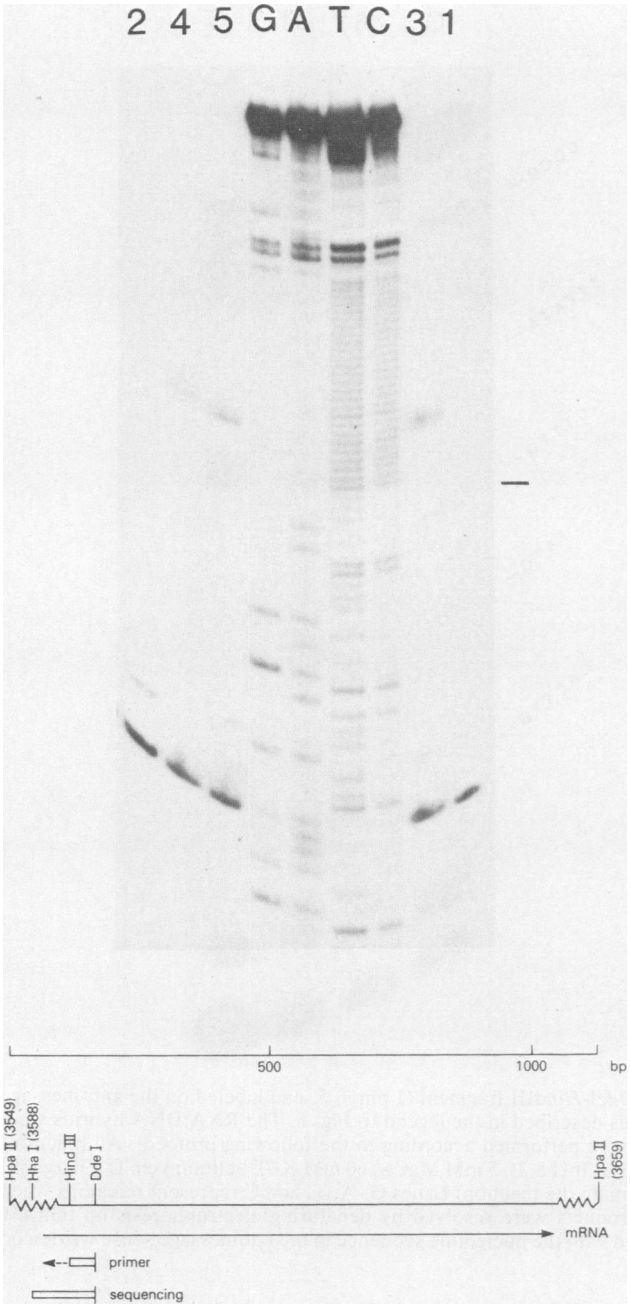

FIG. 6. Mapping of the 5′ end of RS virus M protein mRNA. The inset is a schematic illustration of the map coordinates of the *Hpa*II insert of $pRSA_{11}$ with respect to the mRNA. The RS virus specific sequence is illustrated by a straight line, and the wavy lines on either side represent the pBR322 sequence and the oligo(dG:dC) tails. A 159-bp *Dde*I-*Hpa*II fragment wherein the *Dde*I site lies downstream of the putative 5′ end of the mRNA was 5′ end labeled and digested with *Hha*I. The *Hha*I site lies within the pBR322 sequence flanking the RS viral cDNA. The 120-bp *Dde*I-*Hha*I fragment labeled at the *Dde*I site was used to generate a chemical DNA sequence ladder. A 47-bp *Dde*I-*Hin*dIII fragment 5′ end labeled at the *Dde*I site was used as the DNA primer. One picomole of this DNA primer was denatured and annealed to poly(A) RNA (6 μg) from RS virus-infected cells in 40 μl of hybridization buffer containing 80% formamide, 40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)](pH 6.2) and 0.4 M NaCl for 16 h at 37°C. The RNA-DNA hybrids were recovered after several cycles of ethanol precipitation and dissolved in 50 μl of 50 mM Tris-hydrochloride (pH 8.3), 10 mM $MgCl_2$, 80 mM KCl, and 1 mM dithiothreitol. Reverse transcriptase (400 U/ml) was added, and the reaction was incubated at 37°C for 30 min. After deproteinization by phenol:chloroform extraction, the products were recovered by ethanol precipitation, dissolved in 5 μl of 90% formamide containing 10 mM Tris-hydrochloride (pH 7.4), and electrophoresed on an 8% acrylamide-urea gel alongside DNA sequence reactions. Lanes 1 and 2 represent reaction results when no enzyme was added or when no RNA was present during hybridization. Lane 3 illustrates the results of a complete reaction, and the results when actinomycin D (50 μg/ml) or sodium PP_i (4 mM) was added to the reaction are shown in lanes 4 and 5. Lanes G, A, T, and C represent the DNA sequencing reactions. The bar to the right of the figure denotes the end of cloned RS viral sequence. The primer extension product is seven nucleotides beyond this position as judged by the number of C residues.

tides longer suggested that the plasmid might lack a few nucleotides present at the 3′ end of the mRNA. However, eucaryotic and many viral mRNAs contain on the average 100 A residues at the 3′ end, and if this is taken into account, the recombinant might well contain the entire 3′ end of the mRNA within the limitations of electrophoretic analysis.

The DNA sequence revealed an open reading frame capable of encoding a protein of 28,717 daltons containing 256 amino acids. The protein was relatively basic and moderately hydrophobic. It did not share any homology with other viral M proteins. The M protein of VSV is synthesized on free polyribosomes and rapidly accumulates in the plasma membrane fraction (15, 23); consistent with this observation is the lack of a signal peptide (29). Inspection of the N terminus of RS virus M protein showed that it also lacked a string of hydrophobic residues at the N terminus. M proteins of enveloped negative-strand RNA viruses have dual functions. They interact with virus-altered plasma membranes and the nucleocapsids. In the case of influenza virus, this interaction with the membrane is thought to occur via a central hydrophobic domain (38). Within the RS virus M protein, there were two regions rich in hydrophobic residues. Of the 22 amino acids between residues 185 and 206, there were 11 hydrophobic and 2 charged amino acids. Similarly, between positions 213 and 238, there were 13 hydrophobic and 5 charged residues. These two domains might play a role in membrane interactions. The other function of the M protein in the case of VSV is related to modulation of transcription by interaction with the transcriptionally active nucleocapsids (8). Although clustering of basic residues was not reported for VSV M protein, the effect of this protein on transcription has been attributed to its extremely high content of basic amino acids (29). In the case of influenza virus, however, clustering of basic amino acids on one side of the α helix is thought to be responsible for this interaction (38). The M protein of RS virus is relatively basic, but there is no clustering of basic residues.

A second open reading frame potentially encoding a protein of 75 amino acids was also present. This overlapped with the one encoding the putative M protein by 62 bp at the 5′ end. Since RS virus, like other unsegmented negative-strand RNA viruses, replicates in anucleate cells and in the presence of high levels of actinomycin D, splicing mechanisms are unlikely. Interestingly, certain recombinant plasmids encoding M gene hybrid-selected mRNAs that yielded on translation M protein and another small protein of ca. 15 kdal (34). Computer search of the RS viral cDNA sequence

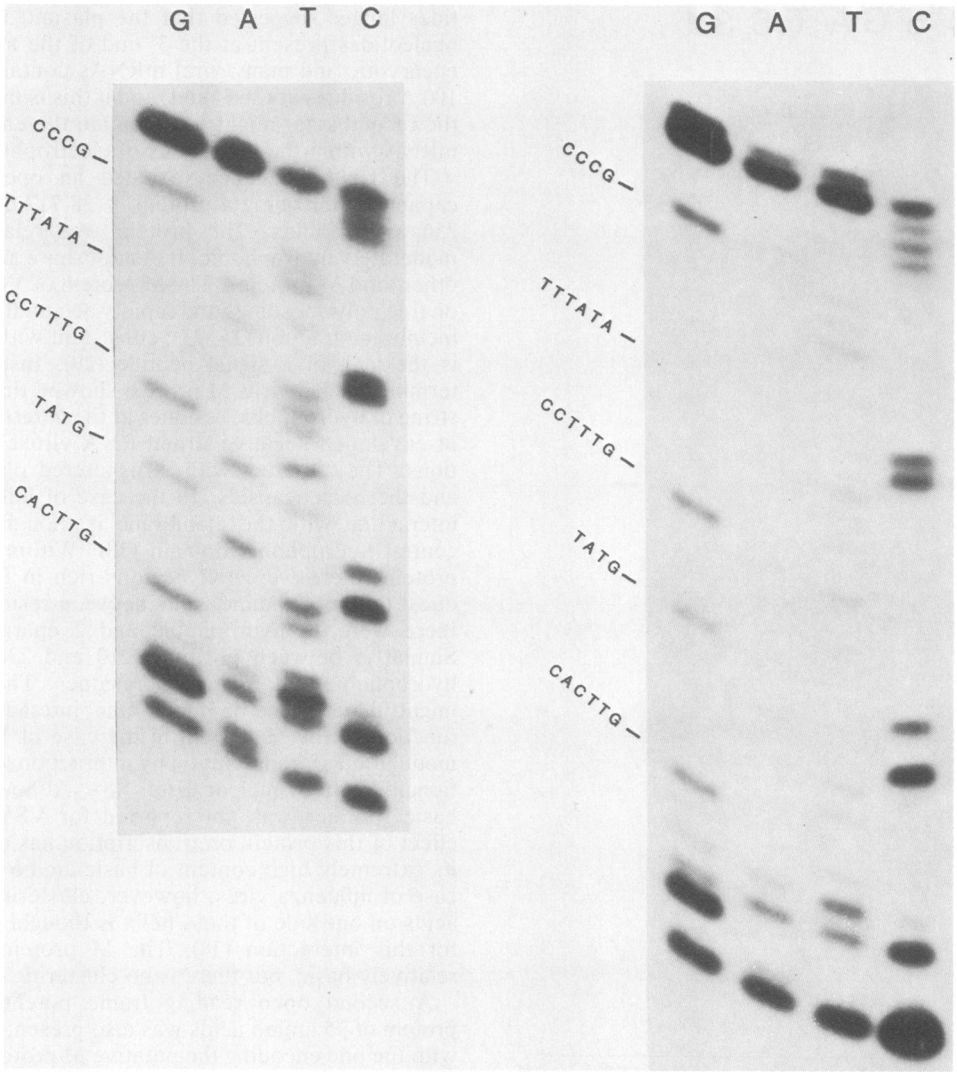

FIG. 7. Determination of the 5′ end sequence of the mRNA. A 47-bp *Dde*I-*Hin*dIII fragment (1 pmol) 5′ end labeled on the anti-message strand was hybridized to poly(A) RNA (6 μg) from infected cells exactly as described in the legend to Fig. 6. The RNA:DNA hybrids were then divided into four reactions, and partial chain termination reactions were performed according to the following protocol. All reactions contained 100 μM dATP, 50 μM each of dCTP, dGTP, and TTP in 50 mM Tris (pH 8.3), 5 mM $MgCl_2$, 60 mM KCl, actinomycin D (50 μg/ml), and 1 mM dithiothreitol. Reverse transcriptase (100 U/ml) was added to initiate the reaction. Lanes G, A, T, and C represent reactions when 50 μM ddGTP, ddATP, ddTTP, or ddCTP was present. The reaction products were resolved by denaturing electrophoresis on thin 6% polyacrylamide urea gels. Results of two different experiments are displayed with the nucleotide sequence in the antimessage sense written on the left side of each panel.

in $pRSA_{11}$ for eucaryotic splice signals revealed several acceptor sites but no competent donor site that would allow a spliced mRNA encoding a 15-kdal protein. Therefore, the biological significance of this second open reading frame is not clear. Recombinant plasmids $pRSA_{11}$ and $pRSC_{12}$ had extensive sequence homology at both ends and hybrid-selected mRNAs that yielded on translation only the M protein. In contrast, recombinants $pRSB_7$ and $pRSD_3$ had sequence homology with $pRSA_{11}$ at either the 5′ or 3′ end. However, it should be pointed out that all four of these clones possessed the coding sequence for the M protein gene based on isomorphous restriction enzyme cleavages and DNA sequencing. It is possible that the plasmids $pRSB_7$ and $pRSD_3$ represent cDNA copies of polycistronic RNAs or RNAs derived from the deletion type of defective interfering genomes that were cloned fortuitiously.

Inspection of the sequence around the initiator methionine showed the presence of a canonical eucaryotic initiation sequence PXXAUGG (16) wherein the initiator triplet AUG is underlined. There was no sequence homology upstream of the putative poly(A) tract with three other RS viral genes that all possessed long tracts of poly(A) (9; unpublished data). Codon usage for this protein showed a remarkable deficiency of codons of the types CGN and NCG. There was also an overall deficiency of the CG dinucleotide (0.6 versus 3.2% expected on a random basis). Similar observations have also been made with RS virus nucleocapsid protein (9), and such an inherent bias against CG dinucleotides has been reported before for VSV, influenza virus, and eucaryotic genomes (29).

In summary, this study is an extension of our efforts to understand the genomic organization of RS virus. Extension

of sequencing studies to other genes should ultimately provide information on the exact gene order for this virus. In addition, sequence information of the viral genes should allow us to study their functional roles by expressing these genes in eucaryotic cells.

ACKNOWLEDGMENTS

We are grateful to R. M. Chanock for his constant support and encouragement. B. Moss and R. Lazzarini were helpful in critical reading of the manuscript. We also thank C. E. Buckler and D. J. Lipman for many patient hours at the computer. Jean Worsley was most helpful in preparing this manuscript.

LITERATURE CITED

1. **Aviv, H., and P. Leder.** 1972. Purification of a biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. U.S.A. **69**:1408–1412.
2. **Bernstein, J. M., and J. F. Hruska.** 1981. Respiratory syncytial virus proteins: identification by immunoprecipitation. J. Virol. **38**:278–285.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
4. **Carroll, A. R., and R. R. Wagner.** 1979. Role of the membrane (M) protein in endogenous inhibition of in vitro transcription by vesicular stomatitis virus. J. Virol. **29**:134–142.
5. **Chanock, R. M., H. W. Kim, C. D. Brandt, and R. H. Parrott.** 1982. Respiratory syncytial virus, p. 471–489. *In* A. S. Evans (ed.), In Viral infections of humans: epidemiology and control. Plenum Publishing Corp., New York.
6. **Clinton, G. M., S. P. Little, F. S. Hagen, and A. S. Huang.** 1978. The matrix (M) protein of vesicular stomatitis virus regulates transcription. Cell **15**:1455–1462.
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. U.S.A. **80**:3205–3212.
8. **De, B. P., G. B. Thornton, D. Luk, and A. K. Banerjee.** 1982. Purified matrix protein of vesicular stomatitis virus blocks viral transcription in vitro. Proc. Natl. Acad. Sci. U.S.A. **79**:7137–7147.
9. **Elango, N., and S. Venkatesan.** 1983. Amino acid sequence of respiratory syncytial virus capsid protein. Nucleic Acids Res. **11**:5941–5951.
10. **Gimenez, H. B., and C. R. Pringle.** 1978. Seven complementation groups of respiratory syncytial virus temperature-sensitive mutants. J. Virol. **27**:459–464.
11. **Glisin, V., R. Crkvenjakov, and C. Byus.** 1974. Ribonucleic acid isolation by cesium chloride centrifugation. Biochemistry **13**:2633–2637.
12. **Gupta, K. C., and D. W. Kingsbury.** 1982. Conserved polyadenylation signals in two negative strand RNA virus families. Virology **120**:415–421.
13. **Huang, Y. T., and G. W. Wertz.** 1982. The genome of respiratory syncytial virus is a negative-stranded RNA that codes for at least seven mRNA species. J. Virol. **43**:150–157.
14. **Huang, Y. T., and G. W. Wertz.** 1983. Respiratory syncytial virus mRNA coding assignments. J. Virol. **46**:667–672.
15. **Knipe, D. M., D. Baltimore, and H. F. Lodish.** 1977. Separate pathways of maturation of the major structural proteins of vesicular stomatitis virus. J. Virol. **21**:1128–1139.
16. **Kozak, M.** 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. **9**:5233–5252.
17. **Land, H., M. Gretz, H. Hauser, W. Lindenmaier, and G. Schutz.** 1981. 5′-Terminal sequences of eukaryotic mRNA can be cloned with high efficiency. Nucleic Acids. Res. **9**:2251–2266.
18. **Lawn, R. M., E. F. Fritsch, R. C. Parker, G. Blake, and T. Maniatis.** 1978. The isolation and characterization of a linked β and δ globin gene from a cloned library of human DNA. Cell **15**:1157–1174.
19. **Lazzarini, R. A., J. D. Keene, and M. Schubert.** 1981. On the origin of defective interfering particles by the negative strand RNA virus. Cell **26**:145–154.
20. **Lehrach, H., D. Diamond, J. Wozney, and H. Boedtker.** 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry **16**:4743–4751.
21. **Lubeck, M. D., J. L. Schulman, and P. Palese.** 1978. Susceptibility of influenza A viruses to amantadine is influenced by the gene coding for M protein. J. Virol. **28**:710–716.
22. **Maxam, A., and W. Gilbert.** 1980. Sequencing end labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
23. **Morrison, T. G., and H. F. Lodish.** 1975. Site of synthesis of membrane and nonmembrane proteins of vesicular stomatitis virus. J. Biol. Chem. **250**:6955–6962.
24. **Peeples, M., and S. Levine.** 1979. Respiratory syncytial virus polypeptides: their location in the virion. Virology **95**:137–145.
25. **Queen, C. L., and L. J. Korn.** 1980. Computer analysis of nucleic acids and protein. Methods Enzymol. **65**:595–609.
26. **Rigby, P. W. J., M. Dieckman, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
27. **Robertson, J. S., M. Schubert, and R. Lazzarini.** 1981. Polyadenylation sites for influenza virus mRNA. J. Virol. **38**:157–163.
28. **Rose, J. K.** 1980. Complete intergenic and flanking sequences from the genome of vesicular stomatitis virus. Cell **19**:415–421.
29. **Rose, J. K., and C. J. Gallione.** 1981. Nucleotide sequences of the mRNA's encoding the vesicular stomatitis virus G and M proteins determined from cDNA clones containing the complete coding regions. J. Virol. **39**:519–528.
30. **Smith, H. O., and M. L. Birnstiel.** 1976. A simple method for DNA restriction site mapping. Nucleic Acids. Res. **3**:2387–2398.
31. **Thayer, R. E.** 1979. An improved method for detecting foreign DNA in plasmids of Escherichia coli. Anal. Biochem. **98**:60–63.
32. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. U.S.A. **77**:5201–5205.
33. **Venkatesan, S., B. M. Baroudy, and B. Moss.** 1981. Distinctive nucleotide sequences adjacent to multiple initiation and termination sites of an early vaccine virus gene. Cell **25**:805–813.
34. **Venkatesan, S., N. Elango, and R. M. Chanock.** 1983. Construction and characterization of cDNA clones for four respiratory syncytial viral genes. Proc. Natl. Acad. Sci. U.S.A. **80**:1280–1284.
35. **Venkatesan, S., A. Gershowitz, and B. Moss.** 1982. Complete nucleotide sequences of two adjacent early vaccinia virus genes located within the inverted terminal repetition. J. Virol. **44**:637–646.
36. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. U.S.A. **80**:726–730.
37. **Winberg, G., and M. L. Hammarskjold.** 1980. Isolation of DNA from agarose gels using DEAE-paper. Application to restriction site mapping of adenovirus type 16. Nucleic Acids. Res. **8**:253–264.
38. **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.
39. **Wright, P. F., M. A. Gharpure, D. S. Hodes, and R. M. Chanock.** 1973. Genetic studies of respiratory syncytial virus temperature-sensitive mutants. Arch. Gesamte Virusforch. **41**:238–247.
40. **Wunner, W. H., and C. R. Pringle.** 1976. Respiratory syncytial virus proteins. Virology **73**:228–243.
41. **Zvonarjev, A. Y., and Y. Z. Ghendon.** 1980. Influence of membrane (M) protein on influenza A virus virion transcriptase activity in vitro and its susceptibility to rimantadine. J. Virol. **33**:583–586.