JOURNAL OF VIROLOGY, Nov. 1984, p. 656–663
0022-538X/84/110656-08$02.00/0

Vol. 52, No. 2

Analysis of the Sendai Virus M Gene and Protein

BENJAMIN M. BLUMBERG,[1] KEITH ROSE,[2] MARCO G. SIMONA,[2] LAURENT ROUX,[1] COLOMBA GIORGI,[3] AND DANIEL KOLAKOFSKY[1]*

Department of Microbiology, University of Geneva Medical School, 1205 Geneva,[1] and Department of Medical Biochemistry, University of Geneva—CMU, 1211 Geneva 4,[2] Switzerland; and Istituto Superiore di Sanità Rome, Italy[3]

Received 9 April 1984/Accepted 27 June 1984

The nucleotide sequence of the Sendai virus M (matrix or membrane) gene region was determined from cloned genomic DNA, and the limits of the M mRNA were determined by S1 nuclease mapping. The M mRNA is 1,173 nucleotides long and contains a single long open reading frame coding for a protein of 348 amino acids. The amino acid sequences of the N- and C-terminal peptides of the M protein were obtained by mass spectrometric analysis and correspond to those predicted from the open reading frame, with the N terminus modified in vivo by cleavage of the initiating methionine and acetylation of the following amino acid. The amphiphilic nature of the M protein structure is discussed.

The paramyxovirus Sendai virus contains six structural proteins which participate in the formation of two substructures, the inner nucleocapsid core and the outer envelope. The nucleocapsid core consists of the 15-kilobase RNA genome tightly complexed with the NP protein, and less tightly with the L and P proteins. This helical structure is surrounded by the viral envelope, derived from the host cell plasma membrane containing the viral glycoproteins HN and F, which form the external spikes of the virion and are thought to span the lipid envelope, and the nonglycosylated M (matrix or membrane) protein, which is thought to form a sheath beneath the lipid bilayer (5).

The assembly of paramyxoviruses takes place by budding at the plasma membrane. When the surfaces of infected cells are examined by scanning electron microscopy, two distinct viral structures are seen: linear strands and spherical virion buds (3). The linear strands are thought to represent viral nucleocapsids lying beneath the infected cell surface, which already contains the envelope glycoproteins, and are considered to be the precursors of the buds. When the inner surfaces of infected-cell plasma membranes were examined by freeze-fracture techniques (1), the membrane sites to which viral nucleocapsids adhered were modified by virus-specific particles arranged in a crystalline orthogonal array. The presence of the same array in the hydrophobic domain of the membranes indicated that these particles were inserted into the inner lipid leaflet (4). The spatial association of both the surface glycoprotein spikes and the internal nucleocapsids with this crystalline array, and the striking resemblance of this array to previously demonstrated cylindrical structures formed by purified M protein (16), led Büechi and Bächi to suggest that the ordered structure at the inner surface of the plasma membrane was indeed composed of M protein (4). This hypothesis was further supported by the discovery that purified M protein in vitro could form not only cylindrical structures but also sheets containing the same periodicity found in the crystalline arrays (14).

The Sendai virus M protein in mature virions closely contacts the nucleocapsid core, since the M protein can be chemically cross-linked to the NP protein under conditions where no other heterodimers can be detected (22). M protein coats the nucleocapsid core (33, 36) and during virus maturation mediates the association of the core with modified areas of host cell plasma membrane containing the viral glycoproteins (24, 36). More recently, Heggeness et al. (14) have pointed out that the viral glycoproteins can initially be detected on the infected cell surface in the absence of underlying nucleocapsids only by immune electron microscopy, but they are seen as their characteristic spikes without the aid of antibodies once nucleocapsids have aligned beneath the modified plasma membrane. Therefore, interaction of M protein with nucleocapsids also seems important for the correct positioning of the viral HN and F protein spikes.

Considerable evidence therefore exists that the Sendai virus M protein is highly interactive and plays a crucial role in the association of the viral nucleocapsids with the modified host cell membrane during virus maturation. To help clarify some of these properties of the M protein, we have determined the amino acid sequence of the Sendai virus M protein by nucleotide sequencing of cloned genomic DNA and analyzed the N- and C-terminal peptides of the protein by gas-liquid chromatography-mass spectrometry (GLC-MS).

MATERIALS AND METHODS

Cells and viruses. The growth of the Harris strain of Sendai virus and the preparation of CsCl pellet RNA as a source of mRNA from infected and uninfected cells have been described by Leppert et al. (21).

Cloning and sequencing of the Sendai virus genome. The cloning of the Sendai virus (−) genome was reported by Dowling et al. (6). Both strands of the viral-specific DNA inserts were sequenced from the restriction sites shown in Fig. 1 by the method of Maxam and Gilbert (23). The 5′ ends of the restriction fragments were labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP after pretreatment with calf intestinal alkaline phosphatase (Worthington Diagnostics); the 3′ ends were labeled by a fill-in reaction with the Klenow fragment of DNA polymerase I (Boehringer-Mannheim Biochemicals) and the appropriate α-^{32}P-labeled deoxynucleotide triphosphate (Amersham Corp.). Labeled DNAs were then recut with appropriate restriction enzymes, and the fragments were isolated by electrophoresis on 6% nondenaturing polyacrylamide gels.

S1 nuclease mapping. CsCl pellet RNA from uninfected or Sendai virus-infected BHK cells was mixed with end-labeled double-stranded DNA probe, annealed at 54°C in buffer

* Corresponding author.

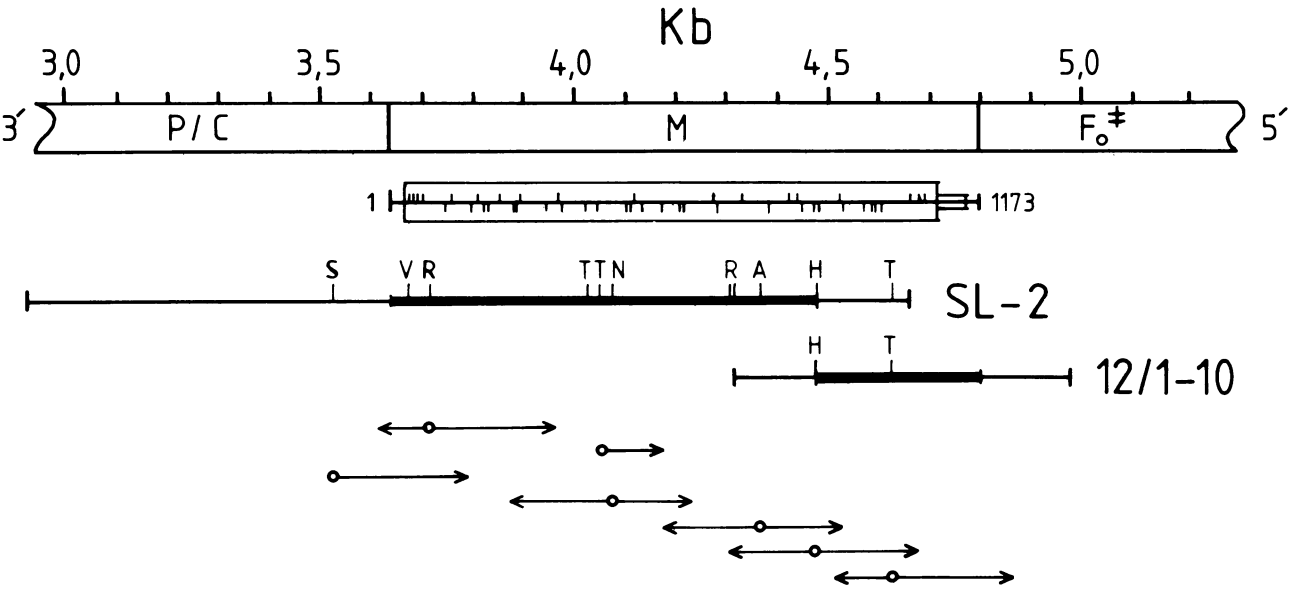


FIG. 1. Schematic representation of the Sendai virus M gene region. The layout of the (−) genome from the 3′ end (scale based on the sequence data of Shioda et al. [34]), is shown at the top, including limits of the M mRNA showing the ORF (large box), the stop signals in the other two reading frames (short bars), and the trailer peptide region (small box). Beneath the genome diagram is the layout of the clones used for S1 mapping and sequencing (SL-2 and 12/1-10), including only the relevant restriction sites. Arrows show the extent of the DNA sequence obtained from the various sites (both strands were sequenced from each site). The heavy bars in the plasmid inserts show the limits of the M mRNA as determined by S1 mapping (see Fig. 3). Restriction site code: A, *Ava*I; H, *Hin*dIII; N, *Nco*I; R, *Rsa*I; S, *Sma*I; T, *Taq*I; V, *Eco*RV. ‡ A tentative identification of this gene is given in the text.

containing 80% formamide, and treated with S1 nuclease as previously described (9). All samples, including restriction fragment markers, were ethanol precipitated before electrophoresis at 50°C on 0.5-mm-thick 8% polyacrylamide sequencing gels containing 7 M urea buffered with TBE.

Preparation and mass spectrometric analysis of Sendai virus M protein. Sendai virus M protein (ca. 2 mg) was prepared from egg-grown virus (1, 25). This protein was prepared for direct amino- and carboxy-terminal analysis by the GLC-MS micromethod of Rose et al. (27–29). For amino-terminal analysis, the basic technique (28) involves N^{α}-acetylation (if the N terminus is unblocked), cleavage of the protein with an appropriate protease, and permethylation to produce a volatile derivative of the N-terminal peptide suitable for mass spectrometric analysis. Since the Sendai virus M protein is thought to be amino blocked (32), 0.2 mg of the protein was suspended in 50 μl of 1% NH_4HCO_3 and treated twice for 2 h at 37°C with 2 μl of chymotrypsin solution (1 mg/ml) before lyophilization and permethylation (27). One-fourth of the resulting derivative (ca. 1 nmol) was subjected to analysis. For carboxy-terminal determination (29), 0.2 mg of protein was suspended in 50 μl of 60 atom% $H_2{}^{18}O$ and 5 μl of 11% NH_4HCO_3 in water, and 4 μl of a solution of *Armillaria mellea* protease (prepared by the method of P. L. Walton, R. W. Turner, and B. Broadbent [British patent, 1263956] and checked against insulin for activity and specificity just before use) was added. After incubation for 16 h at 37°C, the digest was lyophilized, trifluoroacetylated, and permethylated (28). Again, one-fourth of the product was used for mass spectrometric analysis.

RESULTS

Nucleotide sequence of the M gene region. We have previously reported the cloning of the Sendai virus genome and the ordering of a series of overlapping clones representing the 3′ proximal one-third (5 kilobases) of the (−) genome (6). The ordered clones were also used to hybrid select viral mRNAs whose translation in vitro led to the partial gene order 3′ NP-P/C-M · · · 5′. The clone farthest from the 3′ end of the (−) genome, SL-2, hybrid selected mRNAs for both the P and C proteins, which are expressed from a single mRNA that contains overlapping genes (9), and for the M protein. We have previously sequenced the P/C gene of Sendai virus (9), and the positions of the 3′ end of the P/C mRNA was determined to lie ca. 750 nucleotides from the distal end of pSL-2. Since the estimated size of the M protein (35 kilodaltons [kDa]) suggested that the M gene would be slightly more than 1 kilobase in length, it appeared unlikely that the entire M gene sequence would be contained in this plasmid. To obtain more distal clones, the library of Sendai clones was screened with the nick-translated insert of pSL-2, and a number of positive clones were then examined by restriction analysis for limited overlaps with pSL-2. One such clone, 12/1-10, contained a clear overlap with only the distal end of SL-2, and its size suggested that it would also contain the C-terminal end of the M gene, as described by Blumberg et al. (*in* D. H. L. Bishop and R. W. Compans, ed., *Nonsegmented Negative Strand Viruses*, in press). The relevant regions of pSL-2 and p12/1-10 were therefore sequenced as outlined in Fig. 1. The nucleotide and predicted amino acid sequence are shown in Fig. 2.

S1 nuclease mapping of the M mRNA. The limits of the M gene region were determined by S1 nuclease mapping in both ends of the M gene mRNA. Since the overlapping plasmids pSL-2 and p12/1-10 share a sequence containing the same unique *Hin*dIII site in their inserts, the *Hin*dIII sites in both plasmids were end labeled, and the resulting labeled DNA fragments were used to determine the limits of the entire M mRNA. Figure 3A shows the results of S1 mapping in which we used as a probe the 5′ end-labeled 1,600-base pair (bp)

*Hin*dIII-*Pst*I fragment of pSL-2 annealed with CsCl pellet RNA from both uninfected (lanes 5 and 7) and Sendai virus-infected (lanes 6 and 8) BHK cells. Single strong bands of ca. 820 bp were found only when infected cell RNA was used to protect the probe. Weaker bands that may have represented a small amount of P/C-M readthrough mRNA were found in lanes 6 and 8 at the level of the 1,600-bp probe. In addition, a series of weaker bands below the ca. 820-bp bands probably represented a small amount of breakdown in the CsCl pellet RNA, since this film was greatly overexposed to clearly show the markers in lanes 2 to 4. As chain length markers we used unique restriction fragments from pSL-2 itself to define the mRNA start. Lane 2 shows the 5′-labeled *Hin*dIII fragments from pSL-2 after redigestion with *Pst*I; the bands at 1,600 and 200 bp define the limits of the insert. The middle band (779 bp; arrow) represents the *Hin*dIII-*Pst*I fragment of

```
                                                       M
                                  .-15          .1             .15             .30
                                  AAGATTAAGAAAAACTTAGGGTGAAAGAAATTTCACCTAACACGGCGCA
                                             .60
ATG GCA GAT ATC TAT AGA TTC CCT AAG TTC TCA TAT GAG GAT AAC GGT ACT GTG GAG CCC
Met Ala Asp Ile Tyr Arg Phe Pro Lys Phe Ser Tyr Glu Asp Asn Gly Thr Val Glu Pro
                                             .120
CTG CCT CTG AGA ACT GGT TCA GAT AAG AAA GCC ATC CCC TAC ATC AGG ATT ATC AAG GTA
Leu Pro Leu Arg Thr Gly Ser Asp Lys Lys Ala Ile Pro Tyr Ile Arg Ile Ile Lys Val
                                             .180
GGA GAC CCT CCT AAA CAT GGA GTG AGA TAC CTA GAT TTA TTG CTC TTG GGT TTC TTT GAG
Gly Asp Pro Pro Lys His Gly Val Arg Tyr Leu Asp Leu Leu Leu Leu Gly Phe Phe Glu
                                             .240
ACA CCG AAA CAA ACA ACC AAT CTA GGG AGC GTA TCT GAC TTG ACA GAG CCG ACC AGC TAC
Thr Pro Lys Gln Thr Thr Asn Leu Gly Ser Val Ser Asp Leu Thr Glu Pro Thr Ser Tyr
                                             .300
TCA ATA TGC GGC TCC GGG TCG TTA CCC ATA GGT GTG GCC AAA TAC TAC GGG ACT GAT CAG
Ser Ile Cys Gly Ser Gly Ser Leu Pro Ile Gly Val Ala Lys Tyr Tyr Gly Thr Asp Gln
                                             .360
GAA CTC TTA AAG GCC TGC ACC GAT CTC AGA ATT ACG GTG AGG AGG ACT GTT CGA GCA GGA
Glu Leu Leu Lys Ala Cys Thr Asp Leu Arg Ile Thr Val Arg Arg Thr Val Arg Ala Gly
                                             .420
GAG ATG ATC GTA TAC ATG GTG GAT TCG ATT GGT GCT CCA CTC CTA CCA TGG TCA GGC AGG
Glu Met Ile Val Tyr Met Val Asp Ser Ile Gly Ala Pro Leu Leu Pro Trp Ser Gly Arg
                                             .480
CTG AGA CAG GGA ATG ATA TTT AAT GCA AAC AAG GTC GCA CTA GCT CCC CAA TGC CTC CCT
Leu Arg Gln Gly Met Ile Phe Asn Ala Asn Lys Val Ala Leu Ala Pro Gln Cys Leu Pro
                                             .540
GTG GAC AAG GAC ATA AGA TTC AGA GTG GTG TTT GTC AAT GGG ACA TCT CTA GGG GCA ATC
Val Asp Lys Asp Ile Arg Phe Arg Val Val Phe Val Asn Gly Thr Ser Leu Gly Ala Ile
                                             .600
ACC ATA GCC AAG ATC CCA AAG ACC CTT GCA GAC CTT GCA TTG CCC AAC TCT ATA TCC GTT
Thr Ile Ala Lys Ile Pro Lys Thr Leu Ala Asp Leu Ala Leu Pro Asn Ser Ile Ser Val
                                             .660
AAC CTA CTG GTG ACA CTC AAG ACC GGG ATC TCC ACA GAA CAA AAG GGG GTA CTC CCA GTA
Asn Leu Leu Val Thr Leu Lys Thr Gly Ile Ser Thr Glu Gln Lys Gly Val Leu Pro Val
                                             .720
CTT GAT GAT CAA GGG GAG AAA AAG CTC AAT TTT ATG GTG CAC CTC GGG TTG ATC AGG AGA
Leu Asp Asp Gln Gly Glu Lys Lys Leu Asn Phe Met Val His Leu Gly Leu Ile Arg Arg
                                             .780
AAG GTC GGG AAG ATA TAC TCT GTT GAG TAC TGC AAG AGC AAG ATT GAG AGA ATG CGG CTG
Lys Val Gly Lys Ile Tyr Ser Val Glu Tyr Cys Lys Ser Lys Ile Glu Arg Met Arg Leu
                                             .840
ATT TTC TCA CTT GGG TTA ATC GGC GGT ATA AGC TTC CAT GTT CAG GTT ACT GGG ACA CTA
Ile Phe Ser Leu Gly Leu Ile Gly Gly Ile Ser Phe His Val Gln Val Thr Gly Thr Leu
                                             .900
TCT AAG ACA TTC ATG AGT CAG CTC GCA TGG AAG AGG GCA GTC TGC TTC CCA TTA ATG GAT
Ser Lys Thr Phe Met Ser Gln Leu Ala Trp Lys Arg Ala Val Cys Phe Pro Leu Met Asp
                                             .960
GTG AAT CCC CAT ATG AAC CTG GTG ATT TGG GCG GCA TCT GTA GAA ATC ACA GGC GTC GAT
Val Asn Pro His Met Asn Leu Val Ile Trp Ala Ala Ser Val Glu Ile Thr Gly Val Asp
                                             .1020
GCG GTG TTC CAA CCG GCC ATC CCT CGT GAT TTC CGC TAC TAC CCT AAT GTT GTG GCT AAG
Ala Val Phe Gln Pro Ala Ile Pro Arg Asp Phe Arg Tyr Tyr Pro Asn Val Val Ala Lys
                                             .1080
AAC ATC GGA AGG ATC AGA AAG CTG TAA ATG TGC ACC CAT CAG AGA CCT GCG ACA ATG CCC
Asn Ile Gly Arg Ile Arg Lys Leu *Oc Met Cys Thr His Gln Arg Pro Ala Thr Met Pro
                                             .1140
CAA GCA GAC ACC ACC TGG CAG TCG GAG CCA CCG GGT CAC TCC TTG TCT TAA ATAAGAAAAAC
Gln Ala Asp Thr Thr Trp Gln Ser Glu Pro Pro Gly His Ser Leu Ser *Oc

TTAGGGATAAA
```

FIG. 2. Nucleotide sequence of the Sendai virus M gene region. Both strands of each of the restriction fragments shown in Fig. 1 were sequenced (see text). The nucleotide sequence is shown as mRNA (+) sense DNA, with the start of the M mRNA at position 1. The mRNA termination signals (12) are underlined. The translated amino acid sequence of the ORF is shown below the nucleotide sequence. Peptide sequences determined by MS (see Fig. 4) are underlined with dashes.

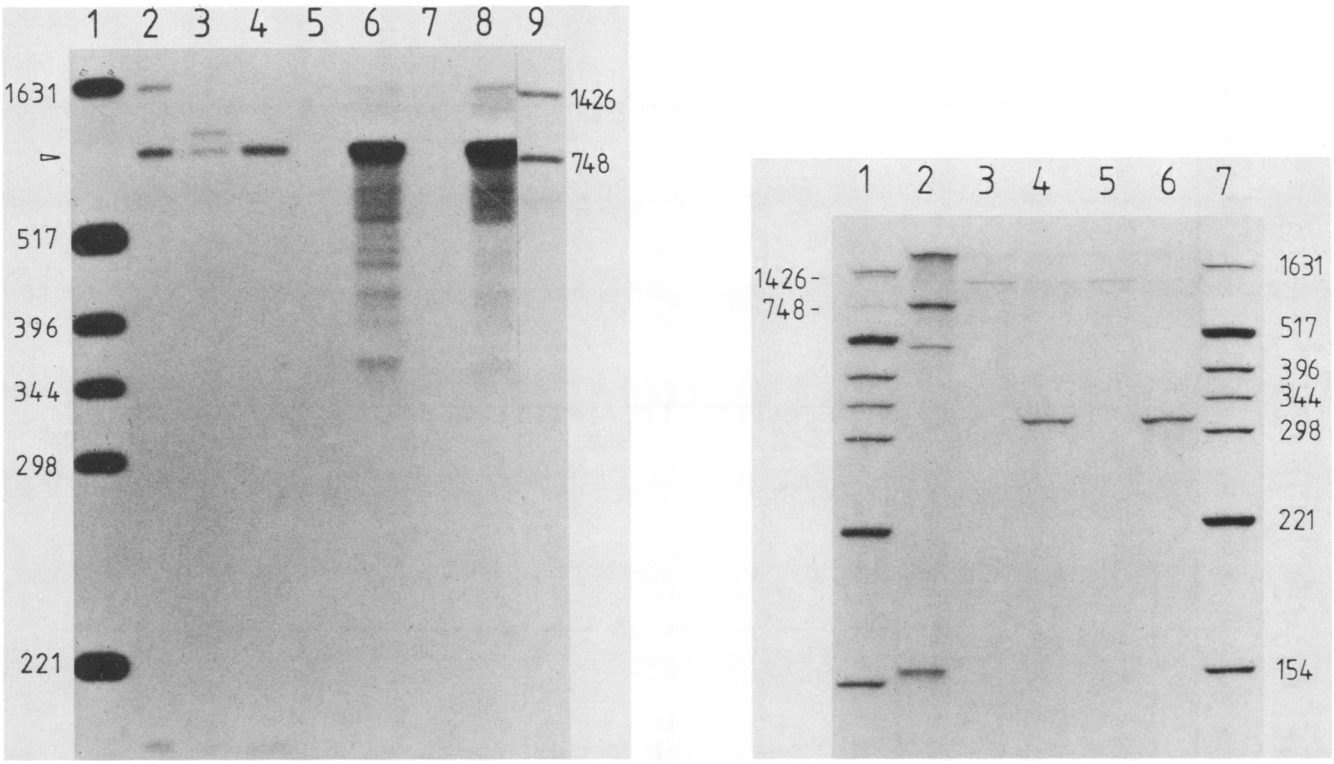


FIG. 3. Nuclease S1 mapping of the Sendai virus M mRNA. (A) Determination of the 5′ end of the M mRNA. Lanes: 2, 2,000 cpm of pSL-2, 5′ end labeled at the *Hin*dIII sites, and cut with *Pst*I (the 1,600-bp fragment was used as the probe); 3, as in lane 2, recut with *Sma*I; 4, as in lane 2, recut with *Eco*RV; 6 and 8, 200,000 cpm of the pSL-2 probe annealed with 0.2 and 0.6 optical density units, respectively, of CsCl pellet RNA from Sendai virus-infected BHK cells and treated with S1 nuclease (see text); 5 and 7, control experiments with 0.2 and 0.6 optical density units, respectively, of CsCl pellet RNA from mock-infected BHK cells annealed to the pSL-2 probe and S1 nuclease-treated; 1, *Hin*fI-cut pBR322 markers; 9, pBR322-labeled at the *Eco*RI site and recut with *Pst*I and *Ava*I. (B) Determination of the 3′ end of the M mRNA. Lanes: 2, 4,000 cpm of p12/1-10, 3′ end labeled at the *Hin*dIII sites, and recut with *Pst*I (the 500-bp fragment was used as the probe); 4 and 6, 200,000 cpm of the p12/1-10 probe annealed with 0.2 and 0.6 optical density units, respectively, of CsCl pellet RNA from Sendai virus-infected BHK cells and treated with S1 nuclease (see text) 3 and 5, control experiments with 0.2 and 0.6 optical density units, respectively, of CsCl pellet RNA from mock-infected BHK cells annealed to the p12/1-10 probe and S1 nuclease-treated; 1 and 7, markers as in lanes 1 and 9 in (A).

the vector. Lane 3 shows the above fragments after further digestion with *Sma*I. The unique *Sma*I site of pSL-2 (located at position 1813 of the P/C gene sequence [9]) is 917 bp upstream from the *Hin*dIII site. Lane 4 shows the fragments from lane 2 after further digestion with *Eco*RV. The unique *Eco*RV site in pSL-2 is 803 bp upstream from the *Hin*dIII site, and the resulting fragment runs just above the 779-bp vector band. Together, these markers establish the length of the S1 bands in lanes 6 and 8 as slightly more than 803 but considerably less than 917 bp and place the 5′ end of the M mRNA somewhat upstream from the *Eco*RV site.

The 3′ end of the M mRNA was mapped in an analogous experiment (Fig. 3B); as a probe we used the 500-bp *Hin*dIII-*Pst*I fragment of p12/1-10 3′-end-labeled at the same *Hin*dIII site. Strong bands just above the 298-bp markers, present only when Sendai-infected CsCl pellet RNA was used, place the 3′ end of the M mRNA ca. 310 bp downstream of the *Hin*dIII site in p12/1-10. The band at ca. 1,500 bp in lanes 3 to 6 appears to be an artifact, since it was also present when uninfected cell RNA was used.

High-resolution S1 mapping of the 3′ end of the P/C mRNA previously showed that this mRNA ended within the termination-polyadenylation sequence TAAGAAAAA found next to the polyadenylate tail of most if not all of the Sendai virus mRNAs (12). The 5′ end of the M mRNA thus maps precisely in this region. As previously noted (9), there is a perfect direct repeat of 22 bp at the NP-P/C and P/C-M gene boundaries, including the 9-base termination sequence, a 3-base intercistronic region (CTT), and the purine-rich sequence representing the 5′ end of the P/C mRNA (AGGGTGAAAG). Since, in addition, an almost identical purine-rich sequence, AGGGTCAAAG, was determined by S1 mapping to represent the 5′ end of the NP mRNA (2), it seems likely that the 5′ end of the M mRNA begins with the sequence AGGGTGAAAG, starting at position 1 in Fig. 2. Analogous considerations would then place the 3′ end of the M mRNA at the termination-polyadenylation sequence TAAGAAAAA at position 1165 to 1173 of Fig. 2.

As shown in Fig. 1, the M mRNA contains only one long open reading frame (ORF), which begins at the AUG position 33 to 35 and terminates at the UAA at position 1077 to 1079. This ORF would code for a protein of 348 amino acids (38,590 Da), in fair agreement with the estimated size of the M protein on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (35 kDa). A curious feature at the C-terminal end of the M mRNA is that the UAA which terminates the M protein ORF is immediately followed by an AUG (position 1080 to 1082), and the following reading frame remains open until the UAA (1161 to 1163), which just precedes the mRNA termination sequence. It seems unlikely that ribosomes

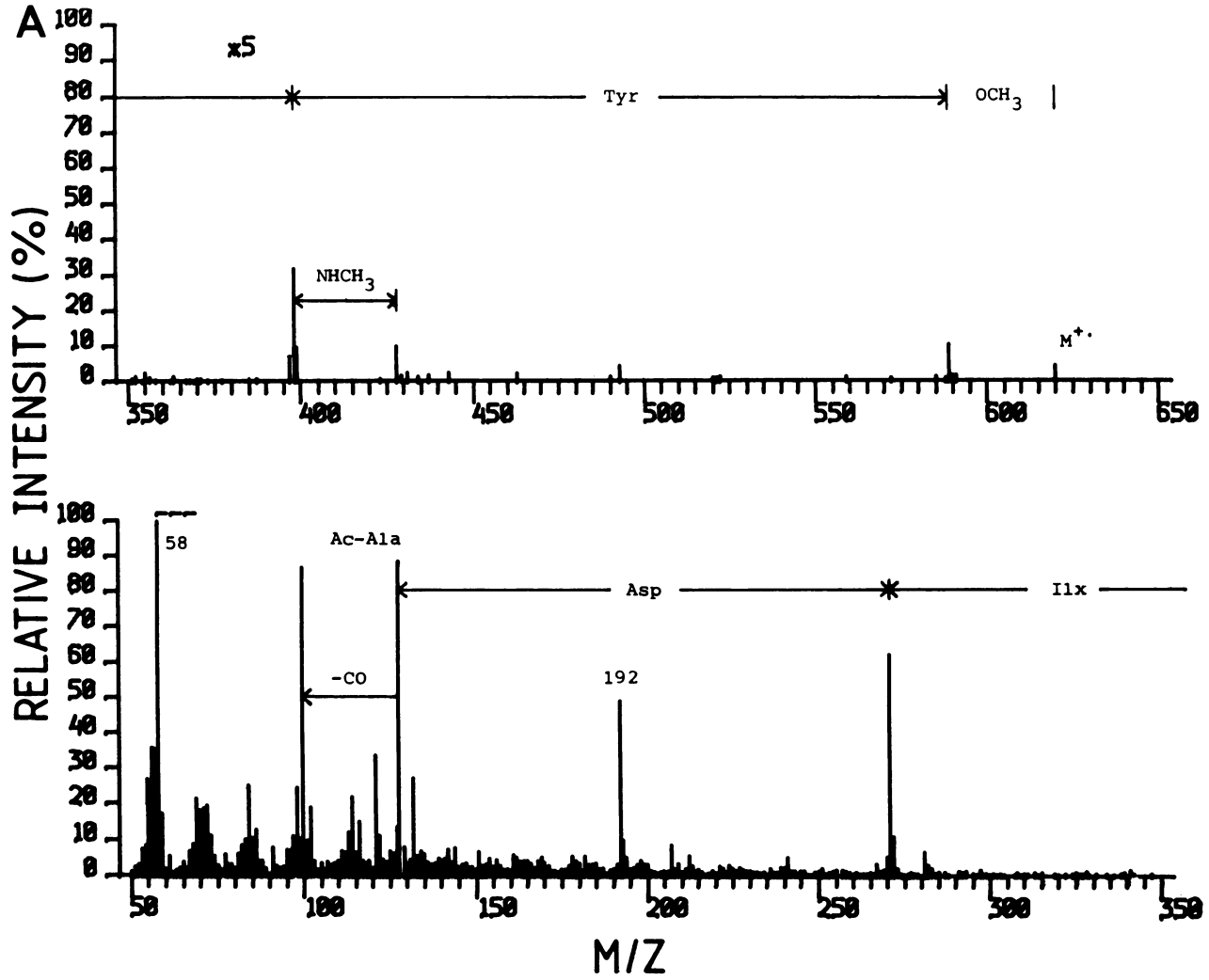
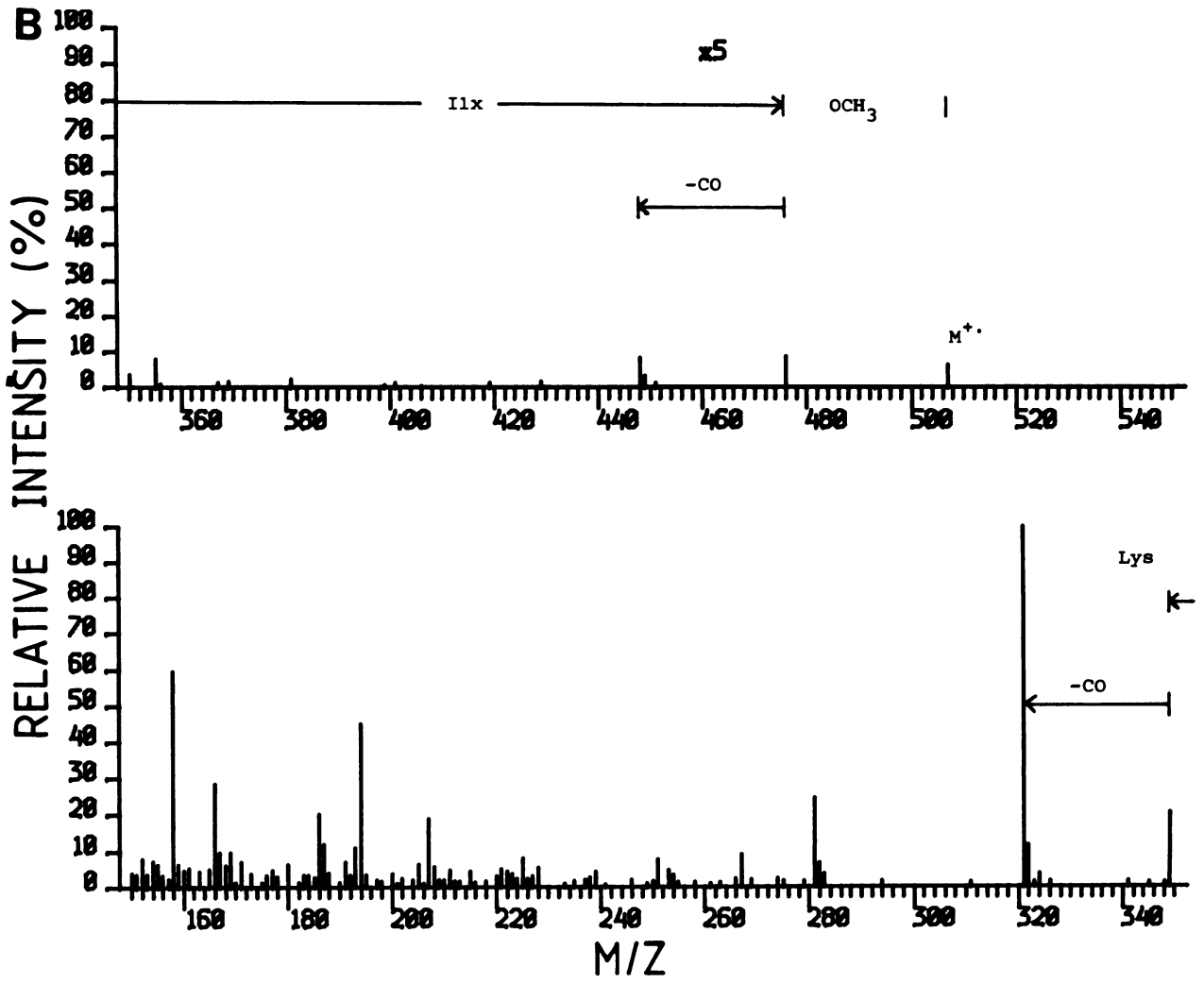


FIG. 4. Mass spectrometric analysis of the amino and carboxy termini of the Sendai virus M protein. Mass spectra were determined after treatment of the protein as described in the text. (A) Mass spectrum of the N-terminal peptide. For clarity, the conventional three-letter amino acid code is used in the figures and in the relevant discussion in the text to indicate *N,O*-permethylated amino acid residues. The silicon ion series *m/z* 207, 281, 341, 355 is due to column bleed; this is the raw mass spectrum with no averaging or background subtraction performed. Signals at *m/z* 58 and 192 show the presence of alanine (without indicating its position in the sequence) and tyrosine (in the C-terminal position), respectively. Strong sequence ions are present at *m/z* 128, 271, 398, and 589, and the molecular ion is visible at *m/z* 620. The sequence assignment is N^{α}-acetyl-Ala-Asp-Ilx-Tyr-OCH_3, where Ilx represents either leucine or isoleucine (28). (B) Mass spectrum of the C-terminal peptide. This is the raw mass spectrum with no averaging or background subtraction performed. The silicon ion series due to column bleed is present (see above). The sequence Lys-Ilx-OCH_3 is deduced from signals at *m/z* 349 and 476. Losses of carbon monoxide are evident, as is the molecular ion at *m/z* 507. In the low-mass region, *m/z* 158 is due to charge retention on the C-terminal (Ilx-OCH_3) fragment, and signals at *m/z* 166 and 194 are associated with the presence of N-terminal lysine as its N^{α}-trifluoroacetyl permethyl derivative.

could independently initiate at this AUG, but one possibility that cannot be excluded is that ribosomes which terminate at the first UAA can simply reinitiate protein synthesis on the succeeding AUG. The sequence of such a reinitiated trailer polypeptide is also shown in Fig. 2. The Sendai virus M mRNA is therefore 1,173 nucleotides long and contains a 32-base untranslated region at its 5′ end and either 93 or 10 untranslated nucleotides at its 3′ end, depending on whether the trailer polypeptide is expressed.

Mass spectrometric analysis of the M protein. To check the accuracy of our DNA sequence and to control possible processing of the protein in vivo, we subjected a preparation of Sendai virus M protein to direct N- and C-terminal analysis by a recently developed micromethod involving GLC-MS (27–29; see above). The mass spectrum obtained from the N-terminal peptide was clearly determined to be N^{α}-acetyl-Ala-Asp-Ilx-Tyr (Fig. 4A; Ilx represents either leucine or isoleucine, which are not differentiated by this technique). This blocked tetrapeptide corresponds only to the sequence Ala-Asp-Ile-Try immediately following the initiating methionine of the ORF in Fig. 2. The mass spectrum of the C terminus of the M protein was obtained after digestion in $H_2{}^{18}O$ with *Armillaria mellea* protease, an enzyme that cleaves N-terminal to lysine, so that any C-termini formed by cleavage would incorporate ^{18}O (29). The mass spectrum obtained from the digest was easily interpreted as Lys-Ilx-COOH (Fig. 4B). There was no evidence in this spectrum of incorporation of ^{18}O into the carboxyl



group; had the peptide Lys-Ilx not been C-terminal, we should have observed signals at *m/z* 160, 478, and 509 of an intensity approximately equal to that of the signals at *m/z* 158, 476, and 507, respectively. The M protein was therefore shown to terminate as either Lys-Ile or Lys-Leu.

DISCUSSION

Mass spectrometric analysis of peptides in conjunction with determination of amino acid sequences by DNA sequencing of genes is a very powerful strategy for determining the primary structure of proteins (15, 28). Only the GLC-MS technique used here is capable of directly providing N- and C-terminal amino acid sequence information from proteins as large as or larger than M without previous cleavage and isolation of smaller peptides. The mass spectra shown were each obtained from 5 nmol of sample. Reduction of the GLC-MS approach to a micromethod (27–29) should therefore be of general interest.

Many proteins are known to be amino blocked and otherwise processed in the cell cytoplasm (35). Such N-blocked proteins are refractory to sequencing by dansylation or Edman degradation. We have recently found by our GLC-MS technique that the nucleocapsid proteins of vesicular stomatitis virus and Sendai virus are processed in vivo by cleavage of the initiating methionine and *N*-acetylation of the following amino acid (2). The results shown in Fig. 4A show that the Sendai virus M protein undergoes similar modification. The C-terminal peptide determined in Fig. 4B corresponds to the sequence Lys-Leu (position 1071 to 1076) at the end of the ORF in Fig. 2. There is some ambiguity here, since four other peptides also correspond to the sequence Lys-Ilx. The one nearest the C-terminal is the sequence Lys-Ile at position 792 to 797 in Fig. 2. If this became the C-terminus as a result of processing, the protein would consist of 255 amino acids, with a molecular mass of less than 30 kDa, much smaller that that estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. In addition, C-terminal processing of the M protein has not been observed in pulse-chase studies (20; unpublished data). The primary translation product of the ORF is therefore very probably that shown in Fig. 2, with the methionine at position 33 to 35 in the most favored context for initiation on ribosomes (19). If the N-terminal modifications described above are included, the acetylated Sendai virus M protein contains 347 amino acids (38,501 Da) and is highly basic, with a charge of +19 at pH 6.5. The amino acid composition is in good agreement with that reported for the Z strain of Sendai virus (32, 34).

Ionic interactions appear to be important in the associa-

tion of the basic M protein with the nucleocapsid core, since disruption of purified paramyxovirions by nonionic detergents in low salt concentrations causes the M protein to sediment with the nucleocapsid core, whereas in the presence of high salt concentrations the M protein is found free in the supernatant (1, 25). The Sendai virus M protein is also hydrophobic since, similarly to the influenza virus M protein, it is the only protein extracted from virions by acidic chloroform-methanol (11), and hydrophobic interactions between the M and NP proteins have been suggested (26). It is noteworthy in this context that the homologous membrane protein of the paramyxovirus simian virus 5 has been reported to be acidic (25). This suggests that the spatial distribution of charged amino acids rather than the net charge is critical in the structure and activity of these amphiphilic proteins, which serve as a bridge between the lipophilic inner surface of the modified host cell plasma membrane and nascent viral nucleocapsids on the cytoplasmic side.

As a first approximation in examining the structural basis for these interactions, we made a computer prediction of the secondary structure of the Sendai virus M protein by using the algorithm of Garnier et al. (7). To set the decision constants, we used the circular dichroism data of Giuffre et al. (10), which suggest that in its native state the Sendai virus M protein contains only 11% α-helix and 20% β-sheet structure. Under these constraints, 18 of the 32 amino acid residues predicted as helical were found to lie in a single α-helix (positions 696 to 749 in Fig. 2). This region may be an important determinant of the overall secondary structure of the Sendai virus M protein. The primary structure of the M protein reported here should help in the further characterization of the various domains.

During persistent infections with Sendai virus in cell culture, the M protein continues to be made at normal rates, but its intracellular concentration is diminished because it is unstable and turns over at a greatly increased rate (30). Concomitant with the instability of the M protein, such cultures release few virus particles into the culture medium. This situation has a counterpart in subacute sclerosing panencephalitis (SSPE), a persistent infection of human brains by measles virus (MV), another paramyxovirus, in which MV nucleocapsids are present intracellularly, but measles M protein cannot be detected and virus particles are not released from the infected tissues (13). Furthermore, Johnson et al. (18) have used intracerebral inoculation of MV in hamsters as an experimental model for SSPE, and have shown that, whereas measles protein is present at the early acute stages of the infection when MV is also liberated, within 2 weeks the measles M protein can no longer be detected concomitant with the absence of MV maturation. The reasons for the absence or reduced level of M protein expression in SSPE are not clear, but the correlation between persistency with reduced maturation of virions and lack of M protein in both paramyxoviruses is obvious. From our present understanding of paramyxovirus replication, M protein would not be required intracellularly to maintain persistent infections and, in the absence of selective pressure for its function, would be expected to accumulate mutations at the high rate characteristic of RNA viruses (17). Given its large content of basic residues (arginine and lysine represent 13% of the protein on a molar basis), such mutations might alter the Sendai virus M protein structure so that the mutated protein would be particularly sensitive to cellular proteases with trypsinlike specificities.

The Sendai virus (−) genome is a linear array of presumably six genetic domains (excluding the leader regions), of which at least one, the P/C gene, codes for two proteins (9, 34). During the sequence determination of the M gene, we also unavoidably obtained information about the following gene. Figure 2 shows that the termination sequence of the M gene is followed by the intercistronic CTT and then by a purine-rich sequence similar to the 5′ ends of the preceding genes. We have continued the sequence of this following gene on both p12/1-10 and on other plasmids which extend to the next intercistronic boundary (manuscript in preparation); 117 amino acids from the start of this ORF is a region which corresponds to the unblocked hydrophobic N terminus of the F_1 protein produced upon cleavage-activation of F_0 (8, 31). This region is at the expected distance from the N terminus of the primary translation product of this gene and is preceded by an arginine residue (F_0 is cleaved by a cellular protease with trypsinlike specificity). It therefore seems likely that the gene following M is indeed F_0. Assuming by analogy to vesicular stomatitis virus that the L gene is last and that Sendai virus contains only the six genetic domains mentioned above, the order of the Sendai (−) genome would now read 3′ NP-P/C-M-F_0-HN-L 5′.

ACKNOWLEDGMENTS

We are indebted to Robin Offord for his continuing interest in and advice on the application of MS to our studies on Sendai viral proteins and for his kind gift of a sample of *Armillaria* protease. We thank Amos Bairoch for assistance with the computer studies, François Veuthey for plotting the M gene sequence, Rosette Bandelier, Colette Pasquier, and Pascale Beffy for expert technical assistance, and Rita Lenertz for aid in preparing the manuscript.

LITERATURE CITED

1. **Bächi, T.** 1980. Intramembrane structural differentiation in Sendai virus maturation. Virology **106**:41–49.
2. **Blumberg, B. M., C. Giorgi, K. Rose, and D. Kolakofsky.** 1984. Preparation and analysis of the nucleocapsid proteins of vesicular stomatitis and Sendai virus, and analysis of the Sendai virus leader-NP gene region. J. Gen. Virol. **65**:769–779.
3. **Bohn, W., G. Rutter, H. Hohenberg, and K. Mannweiler.** 1983. Inhibition of measles virus budding by phenothiazines. Virology **130**:44–55.
4. **Büechi, M., and T. Bächi.** 1982. Microscopy of internal structures of Sendai virus associated with the cytoplasmic surface of host membranes. Virology **120**:349–359.
5. **Choppin, P. W., and R. W. Compans.** 1975. Reproduction of paramyxoviruses, p. 95–178. *In* H. Fraenkel-Conrat and R. R. Wagner (ed.), Comprehensive virology, vol. 4. Plenum Publishing Corp., New York.
6. **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. U.S.A. **80**:5213–5216.
7. **Garnier J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
8. **Gething, M. J., J. M. White, and M. D. Waterfield.** 1978. Purification of the fusion protein of Sendai virus: analysis of the NH_2-terminal sequence generated during precursor activation. Proc. Natl. Acad. Sci. U.S.A. **75**:2730–2740.
9. **Giorgi, C., B. M. Blumberg, and D. Kolakofsky.** 1983. Sendai virus contains overlapping genes expressed from a single mRNA. Cell **35**:829–836.
10. **Giuffre, R. A., D. R. Tovell, C. M. Kay, and D. L. J. Tyrrell.** 1982. Evidence for an interaction between the membrane protein of a paramyxovirus and actin. J. Virol. **42**:963–968.
11. **Gregoriades, A.** 1973. The membrane protein of influenza virus: extraction from virus and infected cells with acidic chloroform-methanol. Virology **54**:369–383.

12. **Gupta, K. C., and D. W. Kingsbury.** 1982. Conserved polyadenylation signals in two negative-strand RNA virus families. Virology **120**:518–523.
13. **Hall, W. W., and P. W. Choppin.** 1979. Evidence for lack of synthesis of the M polypeptide of measles virus in brain cells in subacute sclerosing panencephalitis. Virology **99**:443–447.
14. **Heggeness, M. H., P. R. Smith, and P. W. Choppin.** 1982. *In vitro* assembly of the nonglycosylated membrane protein (M) of Sendai virus. Proc. Natl. Acad. Sci. U.S.A. **79**:6232–6236.
15. **Herlihy, W. C., N. J. Royal, K. Biemann, S. D. Putney, and P. R. Schimmel.** 1980. Mass spectra of partial protein hydrolysates as a multiple phase check for long polypeptides deduced from DNA sequences: NH_2-terminal segment of alanine tRNA synthetase. Proc. Natl. Acad. Sci. U.S.A. **77**:6531–6535.
16. **Hewitt, J. A., and M. V. Nermut.** 1977. A morphological study of the M-protein of Sendai virus. J. Gen. Virol. **34**:127–136.
17. **Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol.** 1982. Rapid evolution of RNA genomes. Science **215**:1577–1585.
18. **Johnson, K. P., E. Norrby, P. Swoveland, and D. R. Carrigan.** 1981. Experimental subacute sclerosing panencephalitis: selective disappearance of measles virus matrix protein from the central nervous system. J. Infect. Dis. **144**:161–169.
19. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.
20. **Lamb, R. A., R. W. Peluso, and P. W. Choppin.** 1978. SV5 and Sendai virus polypeptides: aspects of composition, synthesis and phosphorylation, p. 181–193. *In* B. W. J. Mahy and R. D. Barry (ed.), Negative strand viruses and the host cell. Academic Press, Inc., New York.
21. **Leppert, M., L. Rittenhouse, J. Perrault, D. F. Summers, and D. Kolakofsky.** 1979. Plus and minus strand leader RNAs in negative strand virus-infected cells. Cell **18**:735–747.
22. **Markwell, M. A. K., and C. F. Fox.** 1980. Protein-protein interactions within paramyxoviruses identified by native disulfide bonding or reversible chemical cross-linking. J. Virol. **33**:152–166.
23. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
24. **McSharry, J. J., R. W. Compans, and P. W. Choppin.** 1971. Proteins of vesicular stomatitis virus and of phenotypically mixed vesicular stomatitis virus-simian virus 5 virions. J. Virol. **8**:722–729.
25. **McSharry, J. J., R. W. Compans, H. Lackland, and P. W. Choppin.** 1975. Isolation and characterization of the nonglycosylated membrane protein and a nucleocapsid complex from the paramyxovirus SV5. Virology **67**:365–374.
26. **Mountcastle, W. E., R. W. Compans, H. Lackland, and P. W. Choppin.** 1974. Proteolytic cleavage of subunits of the nucleocapsid of the paramyxovirus simian virus 5. J. Virol. **14**:1253–1261.
27. **Rose, K., H. P. Kocher, B. M. Blumberg, and D. Kolakofsky.** 1983. An improved procedure, involving mass spectrometry, for N-terminal amino acid sequence determination of proteins which are N^{α}-blocked. Biochem. J. **217**:253–272.
28. **Rose, K., M. G. Simona, and R. E. Offord.** 1983. Amino acid sequence determination by g.l.c.-mass spectrometry of permethylated peptides. Biochem. J. **215**:261–272.
29. **Rose, K., M. G. Simona, R. E. Offord, C. P. Prior, B. Otto, and D. R. Thatcher.** 1983. A new mass-spectrometric C-terminal sequencing technique finds a similarity between γ-interferon and α_2-interferon and identifies a proteolytically clipped γ-interferon that retains full antiviral activity. Biochem. J. **215**:273–277.
30. **Roux, L., and F. A. Waldvogel.** 1982. Instability of the viral M protein in BHK-21 cells persistently infected with Sendai virus. Cell **28**:293–302.
31. **Scheid, A., M. C. Graves, S. M. Silver, and P. W. Choppin.** 1978. Studies on the structure and functions of paramyxovirus glycoproteins, p. 181–193. *In* B. W. J. Mahy and R. D. Barry (ed.), Negative strand viruses and the host cell. Academic Press, Inc., New York.
32. **Semba, T., Y. Hosaka, and F. Sakiyama.** 1977. Isolation of the M polypeptide of Sendai virus (HVJ) with column chromatography. Biken J. **20**:77–80.
33. **Shimizu, K., and N. Ishida.** 1975. The smallest protein of Sendai virus: its candidate function of binding nucleocapsid to envelope. Virology **67**:427–436.
34. **Shioda, T., Y. Hidaka, T. Kanda, H. Shibuta, A. Nomoto, and K. Iwasaki.** 1983. Sequence of 3687 nucleotides from the 3′ ends of Sendai virus genome RNA and the predicted amino acid sequences of viral NP, P and C proteins. Nucleic Acids Res. **11**:7317–7330.
35. **Wold, F.** 1981. *In-vivo* chemical modification of proteins (post-translational modification). Annu. Rev. Biochem. **50**:783–814.
36. **Yoshida, T., Y. Nagai, S. Yoshi, K. Maeno, T. Matsumoto, and M. Hoshino.** 1976. Membrane (M) protein of HVJ (Sendai virus): its role in virus. Virology **71**:143–161.