Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 7732–7736, December 1984
Biochemistry

Analysis of Sendai virus mRNAs with cDNA clones of viral genes and sequences of biologically important regions of the fusion protein

(paramyxovirus/membrane fusion/virus penetration/polycistronic mRNA/gene order)

MING-CHU HSU AND PURNELL W. CHOPPIN

The Rockefeller University, New York, NY 10021

Contributed by Purnell W. Choppin, August 20, 1984

ABSTRACT **cDNA clones representing five of the genes of Sendai virus (*P*, *HN*, *NP*, *F*, and *M*) were isolated and used to identify the viral mRNAs by hybridization. Five mRNAs that were monocistronic transcripts of these genes were identified. A sixth transcript, which was identified on the basis of size and of hybridization to viral RNA but not to the cDNA of the other five genes, is thought to represent the message for the L protein. In addition, polycistronic transcripts of the *NP* and *P* genes and of the *M* and *F* genes were also found. The latter establishes the position of the *F* gene adjacent to the *M* gene; these results confirm and extend the previously reported partial gene order of the virus. Nucleotide sequences and derived amino acid sequences of two biologically important regions of the F protein—approximately 25% of F proximal to its COOH terminus and the region spanning the site of the proteolytic cleavage that activates the fusion activity of the protein—are presented. The F protein has an unusually large "cytoplasmic domain" of 42 amino acids beyond the hydrophobic region by which it is anchored in the viral membrane. A single possible trypsin cleavage site was found at the junction of the F_1 and F_2 polypeptides, and 26 hydrophobic amino acids extend from this cleavage site at the NH_2 terminus of the F_1 polypeptide.**

The proteins of Sendai virus, a member of the paramyxovirus group, consist of two surface glycoproteins (HN and F), a nonglycosylated membrane protein (M), the structural protein of the nucleocapsid (NP), two proteins associated with the viral transcriptase (L and P), and a nonstructural protein (C) (1–3). The two glycoproteins, HN and F, mediate viral–cell membrane interactions. HN is responsible for adsorption of the virus to the receptors and has neuraminidase activity (4, 5). The F protein is involved in virus penetration, which occurs through fusion of viral and cell membranes, and in virus-induced cell fusion and hemolysis (6). The active F protein consists of two disulfide-linked polypeptides, F_1 and F_2, which are derived by proteolytic cleavage of an inactive precursor, F_0, by a host-cell enzyme (7–10). Much effort has been directed to understanding the mechanism of F-protein-induced membrane fusion. A hydrophobic interaction between the F protein and the target membrane has been proposed on the basis of the following findings: (*i*) There is a long hydrophobic amino acid sequence at the NH_2 terminus of the F_1 polypeptide that is generated by the activating cleavage (11, 12). (*ii*) This sequence is highly conserved among paramyxoviruses (12, 13). (*iii*) Oligopeptides with sequences resembling this NH_2-terminal sequence specifically inhibit the fusion action of the F protein (13) and act at the cell membrane, presumably by competition with the F_1 NH_2 terminus for a specific site(s) (14). A fusion mechanism involving direct interaction between the F protein and the cell membrane was further supported by our finding that the activating cleavage of the F protein involves a conformation change that exposes a hydrophobic region on the protein (15) and that Sendai virions with a cleaved F protein, but not those with uncleaved F, can fuse with liposomes containing phospholipids and cholesterol (16). The precise nature of this interaction of the NH_2 terminus of the F_1 polypeptide with the target membrane is still unknown.

An understanding of the mechanism of action of the F protein and of the other viral proteins would be enhanced by further knowledge of the structure of the proteins. The primary structures of the proteins can be derived most easily from cDNAs of the viral genes. We have cloned five Sendai protein genes (*P*, *HN*, *F*, *NP*, and *M*) from their mRNAs. The mRNAs from Sendai-infected cells were analyzed with the cDNAs by hybridization, and the previously proposed gene order on the Sendai virus genome (17) was confirmed and extended by examination of polycistronic mRNAs. The cDNA and amino acid sequences of two functionally important regions of the F protein—25% of F proximal to the carboxyl end of the molecule and the sequence which spans the activating cleavage site—are presented.

MATERIALS AND METHODS

Preparation of Sendai Virus-Specific RNA. mRNAs were purified from CV-1 cells, 16 hr after infection with the RU strain of Sendai virus, by Dounce homogenization and phenol/chloroform extraction of the postnuclear supernatant (18), followed by oligo(dT)-cellulose chromatography (19); 50S viral RNA was prepared by disruption of egg-grown virus with 2% $NaDodSO_4$ and purification on a sucrose gradient (15–30% wt/wt).

Cloning and Identification of Virus-Specific DNA. Poly(A)$^+$ mRNAs from infected cells were used as templates for reverse transcription primed with oligo(dT)$_{10-12}$. The single-stranded cDNAs were given oligo(dC) tails and primed with oligo(dG)$_{10-12}$ for second-strand synthesis by reverse transcriptase. The double-stranded cDNAs were inserted into the *Pst* I site of pBR322 plasmid DNA by $(dG)_n \cdot (dC)_n$ hybridization. The resulting recombinant DNA molecules were used to transform *Escherichia coli* strain HB101 (ref. 20, pp. 229–251). Tetracycline-resistant colonies were screened for virus-specific DNA by hybridization to 50S viral RNA that had been subjected to alkaline hydrolysis (0.1 M NaOH, 30 min on ice) and ^{32}P-labeled with T4 polynucleotide kinase (21). The cDNAs in the positive colonies were identified by hybrid-arrested translation in a wheat germ cell-free system (22). The unglycosylated HN and F proteins obtained by cell-free translation were identified by tryptic-map comparisons with proteins from Sendai virus-infected cells (23). A second cDNA library specific for the *F* gene

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: VS, vesicular stomatitis; ND, Newcastle disease.

was obtained using a restriction fragment of clone A21, which maps at the 3′ end of the mRNA of the *F* gene (see *Results*), to prime and reverse-transcribe the mRNA from infected cells. The order among the clones of the *F* gene was established by hybridization with Southern transfer blots (24) and restriction endonuclease mapping.

Analysis of Viral mRNAs by Transfer Blotting. Poly(A)$^+$ mRNAs were electrophoretically separated on a 1% agarose gel containing 10 mM methylmercury, 100 mM Tris, 100 mM borate and 2 mM EDTA and transferred to diazobenzyloxymethyl paper (25). Cloned DNA inserts for each viral gene were ^{32}P-labeled by nick-translation and hybridized to mRNAs (26).

Nucleotide Sequence Analysis. Restriction endonuclease analyses were carried out on each cDNA. Restriction fragments were ^{32}P-labeled at their 5′ ends with T4 polynucleotide kinase and sequenced by the procedure of Maxam and Gilbert (27), except that formic acid was used for chemical modification of purine residues (28).

RESULTS

***In Vitro* Translation of Sendai Virus mRNAs.** A wheat germ cell-free system was used for *in vitro* translation of viral mRNAs isolated from infected cells (Fig. 1). The P, NP, and M proteins of the *in vitro* translated products were identified by their migration on a polyacrylamide gel with marker proteins obtained from infected cells. Since the two glycoproteins, F and HN, are not glycosylated *in vitro* and, therefore, their migration in gels would differ from the proteins synthesized *in vivo*, positive identification of these two proteins was achieved by trypic peptide mapping (Fig. 2). Proteins 2 and 4 (see Fig. 1) from the *in vitro* translation were shown to be HN and F, respectively. When tunicamycin was used to block glycosylation in infected cells, the unglycosylated F protein from those cells migrated as a slightly smaller protein than the *in vitro* translation product (Fig. 1), presumably due to the lack of cleavage of the signal peptide from the F protein in the cell-free system. However, the unglycosylated HN protein from tunicamycin-treated cells had the same mobility as the *in vitro* translation product. This could indicate that HN retains its signal peptide and, thus, that its NH_2 terminus remains membrane-associated.

Clones Containing Virus-Specific Sequences. After (dC)·(dG) hybridization of 0.1 μg of double-stranded cDNA reverse-transcribed from poly(A)$^+$ mRNA and 0.2 μg of pBR322 plasmid DNA and transformation of *E. coli*. 669 tetracycline-resistant colonies were screened for virus-specific DNA sequences by hybridization to alkaline-hydrolyzed, ^{32}P-labeled 50S viral RNA. Positive colonies were analyzed for the size of virus-specific sequences in their plasmid DNA by alkaline lysis of 1.5 ml of cells, followed by *Pst* I excision of the virus-specific sequences from the plasmids (ref. 20, pp. 368–369). The viral genes represented in each of the positive colonies were identified by hybrid-arrested translation (22). The cloned DNA excised from $CsCl_2$-gradient-purified plasmid hybridized to the corresponding viral mRNA and inhibited the translation of the protein for which it codes (Fig. 3). Clone B393 inhibited the synthesis of NP, identifying it as containing sequences specific for NP; B273 inhibited synthesis of P; A21 and S4, of F; B96, of M; and B446, of HN. In some instances, a smaller polypeptide appeared, due to the premature termination of its translation by the cDNA hybridized to the mRNA—e.g., hybridization with B273 resulted in an abbreviated P polypeptide migrating between F_0 and M.

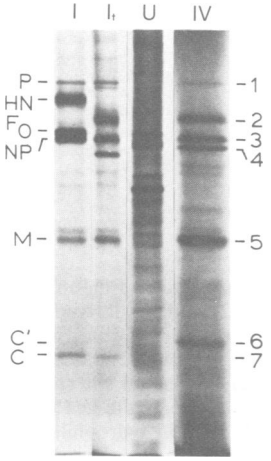

FIG. 1. Synthesis of Sendai virus proteins in infected CV-1 cells and *in vitro* translation of poly(A)$^+$ mRNAs isolated from infected cells. At 16 hr after infection, cells were labeled with [^{35}S]cysteine for 2 hr, lysed, and subjected to $NaDodSO_4$/PAGE. I, infected cells; I_t, infected cells with tunicamycin (0.5 μg/ml) added to the medium at the time of infection; U, uninfected cells. At 16 hr post infection poly(A)$^+$ mRNAs were isolated from infected cells and translated in wheat germ extract with [^{35}S]cysteine (lane IV). F_0 is the uncleaved F protein.

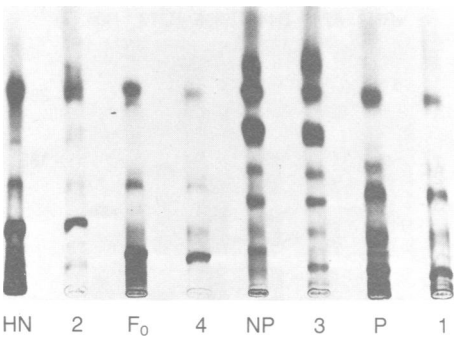

FIG. 2. Tryptic peptide maps of Sendai virus proteins synthesized *in vivo* and *in vitro*. Proteins were separated by gel electrophoresis. The stained protein bands (see Fig. 1) were cut out, eluted, and digested with trypsin, and the digested products were separated by thin-layer chromatography.

Analysis of Viral mRNAs with cDNA Specific for Each Gene. ^{32}P-labeled cDNAs to genes for P, HN, F, NP, and M were hybridized to poly(A)$^+$ mRNAs from infected cells on an RNA transfer blot (Fig. 4). When alkaline-hydrolyzed Sendai 50S viral RNA was used as a hybridization probe, three bands were seen (Fig. 4 *Upper*). However, five Sendai proteins were translated *in vitro* from the mRNAs. Hybridization to cDNA specific for each gene revealed comigration of the mRNAs for the P and HN proteins in band 1 and of the F and NP mRNAs in band 2. Band 3 is the mRNA for the M protein. The estimated sizes of the mRNAs in bands 1, 2, and 3 are 1850, 1670, and 1150 nucleotides, respectively. This is in agreement with the reported sequences of the *P*

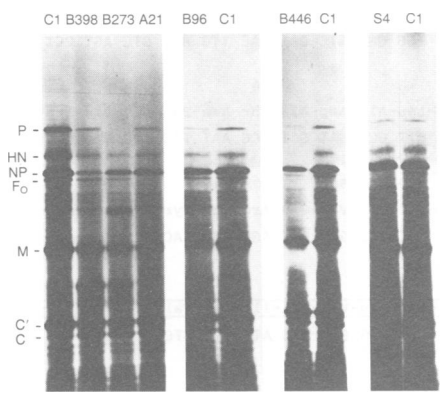

FIG. 3. Hybrid-arrested translation of proteins from Sendai mRNAs by specific cDNA clones. Purified insert DNA (2–3 μg) from each clone (identified at the top of each lane) was hybridized with mRNA purified from a 10-cm plate of Sendai-infected CV-1 cells. Lanes C1: *in vitro* translation of Sendai mRNA without added DNA. The proteins and the clones that inhibited their synthesis are as follows: NP (B398), P (B273), F (A21 and S4), HN (B446), and M (B96).

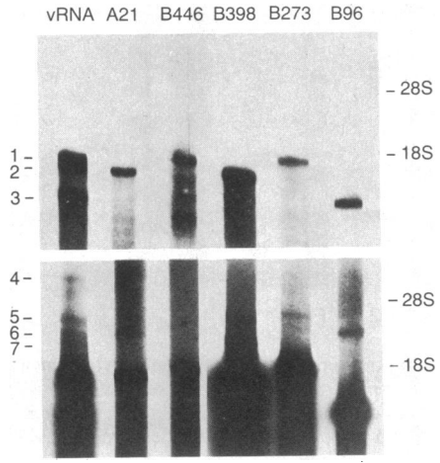

FIG. 4. Separation of Sendai mRNAs on methylmercury/agarose gel and identification by hybridization with virus-specific cDNA probes. Poly(A)$^+$ mRNAs from Sendai-infected CV-1 cells were separated on a methylmercury/agarose gel and transferred to diazobenzyloxymethyl paper, and individual lanes were hybridized with ^{32}P-labeled specific cDNA probes characterized in Fig. 3. Specificities: A21, F; B446, HN; B398, NP; B273, P; B96, M. ^{32}P-labeled ribosomal RNAs from HeLa cells were used as size markers, indicated on the right. RNAs 1–3 (*Upper*) and 4–7 (*Lower*) were detected by short and long autoradiographic exposures, respectively.

and *NP* genes, 1894 and 1683 nucleotides, respectively (29, 30, 56). The smaller RNAs hybridized to each cDNA in Fig. 4 could be breakdown products.

Longer autoradiographic exposure revealed four higher molecular weight mRNAs (designated 4–7 in Fig. 4 *Lower*). From the pattern of hybridization to cDNA specific for each gene and the size of the RNA, RNAs 5 and 6 are polycistronic—i.e., RNA 5 contains sequences specific for NP and P, and RNA 6 contains sequences for F and M. RNA 4 (≈7800 nucleotides) did not hybridize to any of the five cDNAs and is therefore probably a transcript of the *L* gene. RNA 7 hybridized to only the cDNA of the *NP* gene. The origin of this RNA is not clear.

Nucleotide Sequences Proximal to the 3′ End of the mRNA. Clone A21 has a poly(A) stretch of 21 nucleotides (Fig. 5). Therefore, it corresponds to the region of the *F* gene encoding the 3′ end of the mRNA. The sequence contains 437 nucleotides, excluding the poly(A) sequence and the oligo(dG) tail. The sequence 5′-G-T-A-T-A-T-A-A-T-3′ next to the poly(A) sequence corresponds to the sequence at the end of one of the five Sendai viral genes reported by Gupta and Kingsbury (31); the identity of this gene was not reported. These authors also found a common polyadenylylation signal on the viral genome; i.e., 3′-A-$(U)_2$-C-$(U)_5$-5′, which is similar to the sequence 3′-A-U-A-C-$(U)_7$-5′ in vesicular stomatitis (VS) virus (32, 33). The cytidine residue preceding the five uridine residues is not present in clone A21. This is probably due to the use of oligo(dT) to prime first-strand synthesis.

A second library specific for the *F* gene was constructed by priming and reverse-transcribing the mRNAs with a single-stranded fragment of 408 nucleotides of clone A21 (*Ava* II/*Bgl* II fragment, positions 15–423). The order of the cDNA starting at the 3′ end of the F mRNA was established by hybridization and restriction endonuclease mapping.

Derived Protein Sequence of Clone A21. Of the 3 reading frames of clone A21, reading frame 1 encodes a protein sequence of 124 amino acid (Fig. 5). The following characteristics indicate it is the correct reading frame for the F protein: (*i*) It is the longest open reading frame and is followed by three successive termination codons, at positions 373, 400, and 421. (*ii*) It encodes a stretch of 24 hydrophobic amino acids, residues 59–82, which could span a lipid bilayer and anchor the F protein into the viral membrane. This is in agreement with previous work from this laboratory showing that the COOH terminus of the F protein is embedded in the membrane (9). An unexpected finding is the unusually long sequence, 42 amino acids, between the hydrophobic sequence and the COOH terminus, a region which is presumably on the cytoplasmic surface of the cell membrane during virus assembly. There is one potential glycosylation site at residues 8–10.

The Amino Acid Sequence Spanning the Site of the Cleavage That Activates the F Protein. Clone S4 contains ≈770 nucleotides. Fig. 6 shows the sequence of 166 of these nucleotides and the derived amino acid sequence that spans the activating cleavage site on the F protein. A portion of this sequence was determined previously by amino acid sequencing of the NH_2 terminus of the F_1 polypeptide (11–13). The F protein is activated by cleavage with trypsin or trypsin-like proteases (7–10). Among the 18 amino acid residues upstream from the known NH_2 terminus of the F_1 polypeptide there is no site

```
                                        10                                  20
Asp-Ala-Thr-Trp-Gly-Val-Gln-Asn-Leu-Thr-Val-Ser-Pro-Ala-Ile-Ala-Ile-Arg-Pro-Val-Asp-Ile-
GAT GCC ACT TGG GGG GTC CAG AAC TTG ACA GTC AGT CCT GCA ATT GCT ATC AGA CCC GTT GAT ATT

                                30                                  40
Ser-Leu-Asn-Leu-Ala-Asp-Ala-Thr-Asn-Phe-Leu-Gln-Asp-Ser-Lys-Ala-Glu-Leu-Glu-Lys-Ala-Arg-
TCT CTC AAC CTT GCT GAT GCT ACG AAT TTC TTG CAA GAC TCT AAG GCT GAG CTT GAG AAA GCA CGG

                            50                                  60
Lys-Ile-Leu-Ser-Glu-Val-Gly-Arg-Trp-Tyr-Asn-Ser-Arg-Glu-Thr-Val-Ile-Thr-Ile-Ile-Val-Val-
AAA ATC CTC TCT GAG GTA GGT AGA TGG TAC AAC TCA AGA GAG ACT GTG ATT ACG ATC ATA GTA GTT

            70                                  80
Met-Val-Val-Ile-Leu-Val-Val-Ile-Ile-Val-Ile-Val-Ile-Val-Leu-Tyr-Arg-Leu-Lys-Arg-Ser-Met-
ATG GTC GTA ATA TTG GTG GTC ATT ATA GTG ATC GTC ATC GTG CTT TAT AGA CTC AAA AGG TCA ATG

    90                                  100                                 110
Leu-Met-Gly-Asn-Pro-Asp-Asp-Arg-Ile-Pro-Arg-Asp-Thr-Tyr-Thr-Leu-Glu-Pro-Lys-Ile-Arg-His-
CTA ATG GGT AAT CCA GAT GAC CGT ATA CCG AGG GAC ACA TAT ACA TTA GAG CCG AAG ATC AGA CAT

                                120
Met-Tyr-Thr-Asn-Gly-Gly-Phe-Asp-Ala-Met-Ala-Glu-Lys-Arg
ATG TAC ACA AAC GGT GGG TTT GAT GCG ATG GCT GAG AAA AGA TGATCACGACCATTATCAGATGTCTTGTAAA

GCAGGCATGGTATCCGTTGAGATCTGTATATAAT poly(A)
```

FIG. 5. Nucleotide sequence of the mRNA sense-strand of clone A21 and the derived amino acid sequence. The hydrophobic, membrane-anchoring region is enclosed in horizontal lines. A potential glycosylation site (amino acids 8–10) is boxed, and the acidic and basic residues at the COOH-terminal hydrophilic domain are underlined with dashed lines.

```
          15                  10                  5          ← F2 ⇩ F1 →
Leu-Ile-Thr-Val-Thr-Asn-Asp-Thr-Thr-Gln-Asn-Ala-Gly-Val-Pro-Gln-Ser-Arg-Phe-Phe-Gly-Ala-
CTG ATA ACT GTC ACC AAT GAT ACG ACA CAA AAT GCC GGT GTT CCA CAG TCG AGA TTC TTC GGT GCT

5                  10                  15                  20                  25
Val-Ile-Gly-Thr-Ile-Ala-Leu-Gly-Val-Ala-Thr-Ser-Ala-Gln-Ile-Thr-Ala-Gly-Ile-Ala-Leu-Ala-
GTG ATT GGT ACT ATC GCA CTT GGA GTG GCG ACA TCA GCA CAG ATC ACC GCA GGG ATT GCA CTA GCC

          30                  35
Glu-Ala-Arg-Glu-Ala-Lys-Arg-Asp-Ile-Ala-Leu
GAA GCG AGG GAG GCC AAA AGA GAC ATA GCG CTC
```

FIG. 6. Nucleotide sequence of the mRNA sense-strand of clone S4 and the derived amino acid sequence. Only part of the S4 sequence, which spans the activating cleavage site (indicated by an open arrow) is shown. The NH_2 terminus of the F_1 polypeptide and the COOH terminus of the F_2 polypeptide are indicated by horizontal arrows, and the amino acids are numbered in each direction starting from these termini. The hydrophobic sequence at the F_1 NH_2 terminus is enclosed in horizontal lines. A possible glycosylation site is boxed.

for trypsin cleavage, except between arginine and phenylalanine (Fig. 6). Since there is no loss of amino acids upon activation cleavage of the F protein large enough to be detected by $NaDodSO_4$/PAGE (7, 9), the activation of the F protein apparently involves only a single trypsin cleavage.

The hydrophobic sequence at the NH_2 terminus of F_1 includes a stretch of 26 uncharged residues, followed by 6 charged residues in the next 8. There is a possible glycosylation site at residues 11–13 from the F_2 COOH terminus.

DISCUSSION

In vitro translation of the P, NP, M, C, and C′ proteins of Sendai virus has been reported previously (17, 34, 35); however, unequivocal translation of the glycoproteins *in vitro* has been elusive. Using a cell-free wheat germ extract and mRNAs prepared by phenol/chloroform extraction of postnuclear supernatant of infected cells, we were able to obtain, in addition to the nonglycosylated proteins, translation of the F and HN proteins. The authenticity of these *in vitro* translation products was shown by tryptic mapping (Fig. 2). This enabled us to identify unambigously cDNAs to five of the Sendai virus genes. In addition, RNA 4 in Fig. 4 could be the transcript of the *L* gene because it hybridized to the viral RNA, but not to the cDNA of the other five Sendai genes. The size (≈7800 nucleotides) and low abundance of this transcript may explain the lack of a detectable *in vitro* translation product.

RNAs 5 and 6 in Fig. 4 appear to be polycistronic on the basis of their size and hybridization pattern. Polycistronic transcripts, resulting from transcriptional read-through, have been reported for VS virus (36, 37), respiratory syncytial virus (38), and Newcastle disease ND virus (39–41). Transcriptional read-through at each gene junction was shown for VS (37) and ND viruses (41). We have detected such read-through in Sendai virus between NP and P (RNA 5) and between M and F (RNA 6). The relative abundance of these read-through transcripts to the single-gene transcripts appears to be <1%, which is much less than what was found with the VS and ND viruses.

A partial gene order, 3′-*NP*-(*P*+*C*)-*M*, has been established by mapping and direct sequencing of cDNAs derived from the Sendai 50S genome (17, 29, 30). The polycistronic transcript of *NP* and *P* in the present studies is consistent with this gene order, the polycistronic transcript of *M* and *F* indicates that the gene next to *M* is *F*. The *L* gene has been reported to be at the 5′ end of the genome (42). Therefore, the complete gene order on the Sendai genome would be 3′-*NP*-(*P*+*C*)-*M*-*F*-*HN*-*L*-5′.

An unusual feature of the COOH terminus of the Sendai virus F protein is the long sequence of 42 amino acids beyond the membrane-spanning hydrophobic sequence. This "cytoplasmic domain" of the surface glycoproteins of envelope viruses is believed to interact with the membrane protein M during virus assembly and to play a role in the intracellular transport. Eleven or fewer amino acids have been found in the cytoplasmic domain of the HA protein of different strains of influenza virus (43), and 24 amino acids in that of respiratory syncytial virus (44). The G protein of VS virus has a cytoplasmic domain of 28 and 29 amino acids for the New Jersey and Indiana serotypes, respectively (45, 46), longer than that of the HA of influenza or respiratory syncytial virus, but smaller than that of the Sendai F. The cytoplasmic domain of the F protein is highly charged and basic, with 11 basic and 5 acidic residues among the 42 amino acids, a composition somewhat similar to that of the cytoplasmic domain of the VS virus G protein. Further studies are required to elucidate the biological role of the cytoplasmic domain of the F protein and to find out whether differences in sequence, size, or charge in this region of various viral proteins have any functional significance.

Lyles (47) reported that 2–3 kDa of the F protein were digested with proteases when sealed, inside-out ghosts of erythrocytes were prepared after fusion of Sendai virus with the cells, suggesting that ≈20–30 amino acids were removed from the cytoplasmic surface by the protease. The larger cytoplasmic domain found in the present studies may indicate that a portion of the protein is not susceptible to protease due to its secondary structure.

A stretch of 20 hydrophobic amino acids at the NH_2 termini of the F_1 polypeptides of Sendai, simian virus 5 (SV5), and ND viruses was found previously by amino acid sequencing (12, 13). Work from this laboratory has also indicated that this hydrophobic sequence is directly involved in the membrane fusion reaction (12–16, 48). The amino acid sequence of the Sendai F_1 derived from the cDNA sequence of clone S4 shows that the first charged residue, arginine, is at position 27 from the NH_2 terminus of F_1. This length of hydrophobic sequence is comparable to or longer than the membrane-embedding sequence of some membrane proteins; e.g., the F protein reported here has a hydrophobic sequence of 24 amino acids near the carboxyl end of the protein to anchor it in a lipid bilayer. Despite its hydrophobic nature, the following evidence suggests that at least part of the hydrophobic sequence at the NH_2 terminus of F_1 is exposed to the aqueous surroundings: (*i*) the F protein of tissue-culture-grown virus can be cleaved *in vitro* by trypsin (7, 10) and (*ii*) the NH_2 terminus of F_1 can be chemically modified in intact virus by addition of a dansyl group (unpublished data); and is accessible to aminopeptidase (49). How a hydrophobic sequence of this length is transported across the membrane of rough endoplasmic reticulum during its biosynthesis and precisely how it is involved in the membrane fusion reaction are two intriguing questions. There is some evidence to suggest that it interacts with specific site(s) on the target-cell membrane during fusion (14).

An amino acid variation was found in the hydrophobic sequence of the NH_2 terminus of F_1 derived from clone S4, as compared to the previously published amino acid sequence from this laboratory (13): serine at position 16 instead of alanine. This change involves a single nucleotide substitution. This is not surprising because the sequence reported here is derived from a single clone, and the virus has undergone several passages between the amino acid sequencing and the cloning. In addition, the amino acid substitution does not alter the hydrophobic nature of this part of the F protein.

The amino acid sequence derived from clone S4 shows

only a single possible cleavage site for trypsin or trypsin-like protease for the activation of the F protein. There is no other site within 18 amino acids of the COOH terminus of F_2. Although no loss of amino acids has been detected by gel electrophoresis after cleavage, it is possible that the arginine is removed from the COOH terminus of the F_2 polypeptide by carboxypeptidase, as occurs to the HA protein of influenza virus (43, 50, 51) and the F protein of ND virus (52). It is of interest in this regard that the COOH terminus of the Sendai F_2 (Gln-Ser-Arg) is similar to those of the HA_1 of the influenza A H10 strain (Gln-Ser-Arg) (50) and H3 strain (Gln-Thr-Arg) (51).

Very recently (44) the nucleotide sequence of the fusion protein of respiratory syncytial virus, a member of another genus of the paramyxovirus family, has been obtained. Although amino acid analyses of the termini of the polypeptides of this virus are not available, there is a hydrophobic stretch of 19 amino acids preceded by several arginine and lysine residues. By analogy to other paramyxoviruses, this represents the site of cleavage that activates this protein. This region of the respiratory syncytial virus protein is similar to the NH_2-terminal region of Sendai, simian virus 5 (SV5), and ND virus F proteins (12, 13) in being hydrophobic and in having a phenylalanine at the NH_2 terminus and a glycine at residue 3, but the remainder of the sequence is not the same. The NH_2 terminus of the HA_2 polypeptide of influenza viruses, which is generated by the proteolytic cleavage that activates membrane fusion and penetration of these viruses (53–55), is highly conserved among different strains and consists of 10 hydrophobic amino acids, with a glycine at the NH_2-terminus and in position 4 (43). Thus, the primary structure of the NH_2 terminus of the polypeptides of the different viruses that are involved in membrane fusion and virus penetration are conserved within virus genera and show some similarities between virus families. The exact mechanism whereby these proteins cause membrane fusion remains to be determined.

The authors thank Dr. Robert A. Lamb for helpful discussions regarding cloning procedures, Drs. Michael W. Shaw and Pravinkumar B. Sehgal for discussions on some experimental procedures, and Debbie Azaula and Rose Dolezar for excellent technical assistance. Supported by Grants AI 05600 and CA 18213 from the National Institutes of Health and PCM 13464 from the National Science Foundation.

1. Mountcastle, W. E., Compans, R. W. & Choppin, P. W. (1971) *J. Virol.* **7,** 47–52.
2. Choppin, P. W. & Compans, R. W. (1975) in *Comprehensive Virology*, eds. Fraenkel-Conrat, H. & Wagner, R. R. (Plenum, New York), Vol. 4, pp. 95–178.
3. Kingsbury, D. W. (1977) in *The Molecular Biology of Animal Viruses*, ed. Nayak, D. P. (Dekker, New York), Vol. 1, pp. 349–382.
4. Scheid, A., Caliguiri, L. A., Compans, R. W. & Choppin, P. W. (1972) *Virology* **50,** 640–652.
5. Tozawa, H., Watanabe, M. & Ishida, N. (1973) *Virology* **55,** 242–253.
6. Choppin, P. W. & Scheid, A. (1980) *Rev. Infect. Dis.* **2,** 40–61.
7. Scheid, A. & Choppin, P. W. (1974) *Virology* **57,** 475–490.
8. Scheid, A. & Choppin, P. W. (1976) *Virology* **69,** 265–277.
9. Scheid, A. & Choppin, P. W. (1977) *Virology* **80,** 54–66.
10. Homma, M. & Ohuchi, M. (1973) *J. Virol.* **12,** 1457–1465.
11. Gething, M.-J., White, J. M. & Waterfield, M. D. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2737–2740.
12. Scheid, A., Graves, M. C., Silver, S. M. & Choppin, P. W. (1978) in *Negative Strand Viruses and the Host Cell*, eds. Mahy, B. W. J. & Barry, R. D. (Academic, New York), pp. 183–191.
13. Richardson, C. D., Scheid, A. & Choppin, P. W. (1980) *Virology* **105,** 205–222.
14. Richardson, C. D. & Choppin, P. W. (1983) *Virology* **131,** 518–532.
15. Hsu, M.-C., Scheid, A. & Choppin, P. W. (1981) *J. Biol. Chem.* **256,** 3557–3563.
16. Hsu, M.-C., Scheid, A. & Choppin, P. W. (1983) *Virology* **126,** 361–369.
17. Dowling, P. C., Giorgi, C., Roux, L., Dethlefsen, L. A., Galantowicz, M., Blumberg, B. M. & Kolakofsky, D. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5213–5216.
18. Lamb, R. A. & Choppin, P. W. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4908–4912.
19. Inglis, S. C. & Mahy, B. W. J. (1979) *Virology* **95,** 154–164.
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
21. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3961–3965.
22. Lamb, R. A., Choppin, P. W., Chanock, R. M. & Lai, C.-J. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1857–1861.
23. Lamb, R. A., Etkind, P. R. & Choppin, P. W. (1978) *Virology* **91,** 60–78.
24. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
25. Alwine, J. C., Kemp, D. J. & Stark, G. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5350.
26. Rigly, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
27. Maxam, A. W. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
28. Huddleston, J. A. & Brownlee, G. G. (1982) *Nucleic Acids Res.* **10,** 1029–1038.
29. Shioda, T., Hidaka, Y., Kanda, T., Shibuta, H., Nomoto, A. & Iwasaki, K. (1983) *Nucleic Acids Res.* **11,** 7317–7330.
30. Giorgi, C., Blumberg, B. M. & Kolakofsky, D. (1983) *Cell* **35,** 829–836.
31. Gupta, K. C. & Kingsbury, D. W. (1982) *Virology* **120,** 518–523.
32. Rose, J. K. (1980) *Cell* **19,** 415–421.
33. Schubert, M., Keene, J. D., Herman, R. C. & Lazzarini, R. A. (1980) *J. Virol.* **34,** 550–559.
34. Etkind, P. R., Cross, R. K., Lamb, R. A., Merz, D. C. & Choppin, P. W. (1980) *Virology* **100,** 22–33.
35. Gupta, K. C., Morgan, E. M., Kitchingman, D. & Kingsbury, D. W. (1983) *J. Gen. Virol.* **64,** 1679–1688.
36. Herman, R. C., Schubert, M., Keene, J. D. & Lazzarini, R. A. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 4662–4665.
37. Masters, P. S. & Samuel, C. E. (1984) *Virology* **134,** 277–286.
38. Collins, P. L. & Wertz, G. W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3208–3212.
39. Varich, N. L., Lukashevich, I. S. & Kaverin, N. W. (1979) *Acta Virol.* **23,** 273–283.
40. Collins, P. L., Wertz, G. W., Ball, L. A. & Hightower, L. E. (1982) *J. Virol.* **43,** 1024–1031.
41. Wilde, A. & Morrison, T. (1984) in *Nonsegmented Negative Strand Viruses*, eds. Bishop, D. H. L. & Compans, R. W. (Academic, New York), pp. 161–166.
42. Glazier, K., Raghow, R. & Kingsbury, D. W. (1977) *J. Virol.* **21,** 863–871.
43. Laver, G. & Air, G., eds. (1980) *Structure and Variation in Influenza Virus* (Elsevier/North-Holland).
44. Collins, P. L., Huang, Y. T. & Wertz, G. W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7683–7687.
45. Rose, J. K. & Gallione, C. J. (1981) *J. Virol.* **39,** 519–528.
46. Gallione, C. J. & Rose, J. K. (1983) *J. Virol.* **46,** 162–169.
47. Lyles, D. S. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5621–5625.
48. Hsu, M.-C., Scheid, A. & Choppin, P. W. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5862–5866.
49. Asano, K., Murachi, T. & Asano, A. (1983) *J. Biochem.* **93,** 733–741.
50. Bosch, F. X., Garten, W., Klenk, H.-D. & Rott, R. (1981) *Virology* **113,** 725–735.
51. Garten, W., Bosch, F. X., Linder, D., Rott, R. & Klenk, H.-D. (1981) *Virology* **115,** 361–374.
52. Kohama, T., Garten, W. & Klenk, H.-D. (1981) *Virology* **111,** 364–376.
53. Lazarowitz, S. G., Compans, R. W. & Choppin, P. W. (1971) *Virology* **46,** 830–843.
54. Klenk, H.-D., Rott, R., Orlich, M. & Blödorn, J. (1975) *Virology* **68,** 426–439.
55. Lazarowitz, S. G. & Choppin, P. W. (1975) *Virology* **68,** 440–454.
56. Morgan, E. M., Re, G. G. & Kingsbury, D. W. (1984) *Virology* **135,** 279–287.