Volume 14 Number 4 1986 Nucleic Acids Research

Determination of the complete nucleotide sequence of the Sendai virus genome RNA and the predicted amino acid sequences of the F, HN and L proteins

Tatsuo Shioda, Kentaro Iwasaki[1] and Hiroshi Shibuta*

Department of Viral Infection, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108 and [1]Department of Physiological Chemistry, Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113, Japan

Received 14 January 1986; Accepted 24 January 1986

ABSTRACT

We previously determined the 3' proximal 5,824 nucleotides of the Sendai virus genome RNA (Nucleic Acids Res. 11, 7317-7330, 1983; Nucleic Acids Res. 12, 7965-7973, 1984), and present here the sequence of the remaining 5' proximal 9,559 nucleotides. Thus, this is the first paramyxovirus to have its genome organization elucidated. The set of complementary DNA clones used was prepared by the method of Okayama and Berg from polyadenylylated viral genome RNA. We sequenced the region containing the 5' proximal half of the F gene, and the subsequent HN and L genes, and predicted the complete amino acid sequence of the products of these genes. Sequence analyses confirmed that all the genes are flanked by consensus sequences and suggest that the viral mRNAs are capable of forming stem-and-loop structures. Comparison of the F and HN glycoproteins of Sendai virus with those of simian virus 5 strongly suggests that the cysteine residues are highly important for maintenance of the molecular structures of these glycoproteins.

INTRODUCTION

Recently, the molecular cloning of complementary DNA (cDNA) to RNA has greatly contributed to the elucidation of the gene and genome structures of several RNA viruses. Analyses of the viral gene and genome structures provide essential information for the solving of several important problems such as the mechanisms of viral replication and transcription, the nature of each viral protein, biological events caused by viruses and genetical relationships among viruses.

Of paramyxoviruses, Sendai virus has been most extensively studied as a prototype, its genome being a single continuous RNA of about 15 kilobases long with negative polarity. Sendai virus contains six structural proteins, i.e. large (L), RNA polymerase (P), nucleocapsid (NP), hemagglutinin-neuraminidase (HN), fusion (F) and membrane (M) proteins. In addition, at least one non-structural viral protein designated as C (1) and a small transcript called leader RNA (2) have been identified in infected cells. In previous communications (3,4), we reported the sequence of the 5,824

nucleotides from the 3' end of the Sendai virus genome RNA and determined the complete primary structures of the first three genes and a part of the fourth gene. Our results as well as data presented by others established that the gene order is 3'-leader-NP-P+C-M-F- (3,4,5,6,7,8,9), and the subsequent gene order has been proposed to be -HN-L-5' (10,11). In order to fully elucidate the genome structure of Sendai virus, we continued to construct cDNA clones toward the 5' end, and present here the full sequence of the remaining region, that is, the 5' proximal 9,559 nucleotides, which contains the 5' proximal half of the F gene and the subsequent HN and L genes. Thus, we have determined the complete primary structure of the genome of Sendai virus strain Z, and have predicted the amino acid sequence of each gene product. We also compare the predicted structures of the F and HN glycoproteins of Sendai virus with those reported for another paramyxovirus, simian virus 5 (SV5) (12,13).

MATERIALS AND METHODS

Preparation of viral RNA

Sendai virus strain Z was used. Viral 50S genome RNA was prepared from virions purified from infectious allantoic fluids of chicken eggs as described previously (3). Sendai virus mRNAs were prepared in a similar way to as described previously (3) using BHK-21 cells as the host cells, and RNA was extracted by the guanidium thiocyanate method (14). The mRNAs were selected from crude RNA by oligo(dT) cellulose column chromatography.

Synthesis and cloning of the complementary DNA (cDNA)

Sendai virus genome RNA was polyadenylylated according to Inokuchi et al. (15) except that the amount of ATP:RNA adenyltransferase (poly A polymerase) was increased to 20-fold. Synthesis and molecular cloning of cDNA from the polyadenylylated genome RNA were performed according to Okayama and Berg using a pSV7186-derived vector-primer (16). *E. coli* K12 strain HB101 was transformed with the resulting recombinant plasmids by the standard method (17).

Sequence determination of cDNA

CDNAs were cleaved into fragments with appropriate restriction endonucleases, and after subcloning of the fragments into M13 phage (18), their nucleotide sequences were determined by the method of Sanger et al.(19).

Colony hybridization

Bacterial colonies grown on nitrocellulose filters were lysed and fixed

according to Grunstein and Hogness (20). The nitrocellulose filters were used for hybridization (21) with the viral 50S RNA probe labeled with [γ-^{32}P] ATP (22) or the cDNA probe labeled with [α-^{32}P] dGTP (23).

Northern blot hybridization

Viral mRNA resolved in an agarose gel was transferred to a nitrocellulose filter as described previously (3). Hybridization of this filter with ^{32}P-labeled cDNA was performed as described above.

Enzymes and other materials

ATP:RNA adenyltransferase, E. coli DNA ligase, terminal deoxynucleotidyl transferase and ribonuclease H were purchased from P-L Biochemicals, Milwaukee, U.S.A.; avian myeloblastosis virus reverse transcriptase from Seikagaku Kogyo, Tokyo, Japan; M13 cloning and sequencing kits, nick-translation kits and all the radioactive compounds from Amersham Incorporation plc, Amersham, England; HAWP nitrocellulose membrane filters from Millipore, Bedford, U.S.A.; and bovine alkaline phosphatase, T4 polynucleotide kinase, E. coli DNA polymerase I, the Klenow fragment of DNA polymerase I, T4 DNA ligase and all the restriction endonucleases from Takara Shuzo, Kyoto, Japan.

RESULTS

Cloning and sequence determination of cDNA

We previously described the molecular cloning and sequencing of the Sendai virus genome cDNAs which cover the 3' proximal 5,824 nucleotides (3,4). In order to obtain cDNA clones which represent the remaining genome region, we tried to clone cDNAs which were prepared by reverse-transcription of partially digested and then in vitro polyadenylylated viral genome RNA. For this, 6 micrograms of Sendai virus genome RNA was incubated with 4 units of poly A polymerase, which is about 20-fold the amount used for the standard polyadenylation reaction. After the incubation, the 50S genome RNA was found to have been digested to an average size of about 28S (data not shown), which was most likely caused by the ribonuclease activity contaminating the poly A polymerase preparation. Subsequent synthesis and molecular cloning of cDNAs from the polyadenylylated RNA were performed according to Okayama and Berg (16).

Transformation of E. coli HB101 with the recombinant plasmids yielded about 500 ampicillin-resistant colonies, from which 41 were selected on the basis of the intense hybridization signals with the Sendai virus genome RNA probe. Out of these clones, 27, which did not hybridize with probes of cDNA

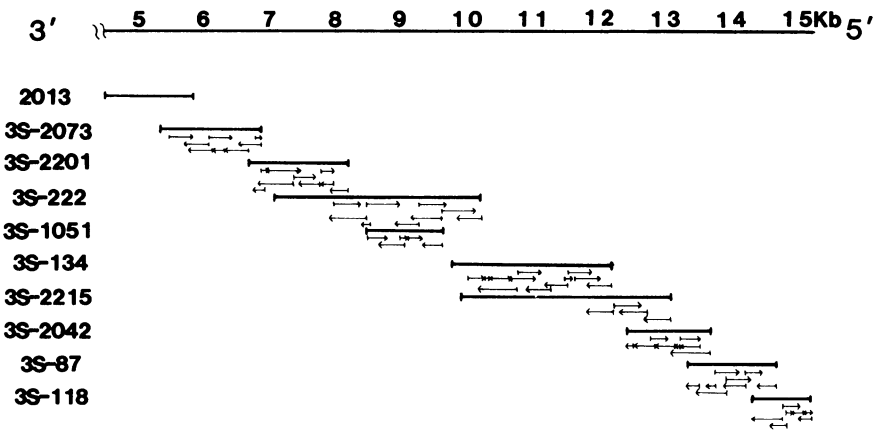


Fig. 1. Locations of previously obtained cDNA clone 2013 and newly prepared cDNA clones 3S-2073, 3S-2201, 3S-222, 3S-134, 3S-1051, 3S-2215, 3S-2042, 3S-87 and 3S-118 in the genome RNA and the sequencing strategy. Arrows indicate the cDNA fragments sequenced and the direction of sequencing.

fragments corresponding to the NP, P+C or M genes (3,4), were selected by the colony hybridization technique. The mutual relationships among the 27 clones as well as their relationships to the previously reported clones were determined by colony hybridization tests using appropriate cDNA probes in combination with restriction map analyses, and it was found that these clones could be lined without any gap in order from the 3' proximity to the 5' proximity. We selected 9 clones for sequencing. The sequential relationships between these 9 clones and to the previously reported clone 2013 together with the sequencing strategy are shown in Fig. 1. Clone 3S-2073 showed an overlap of about 400 nucleotides with clone 2013, which had been shown to cover the 5' proximal 331 nucleotides of the M gene and the 3' proximal 1,013 nucleotides of the F gene (4). Clone 3S-118, which was identified as the most 5' proximal clone, had an identical sequence with the reported 5' terminal 70 nucleotides of the Sendai virus genome RNA (24) except for 3 substitutions, which was followed by 6 guanine residues and a PstI site derived from the linker DNA fragment. This indicates that 3S-118 included the very 5' end portion of the genome, and that we had obtained a set of cDNA clones that covered the entire genome RNA. By analyzing these 9 clones, we determined the sequence of the 9,559 nucleotides of the Sendai virus genome, which followed the 3' proximal 5,824 nucleotides that we had previously determined (3,4). Thus, we concluded that the total length of

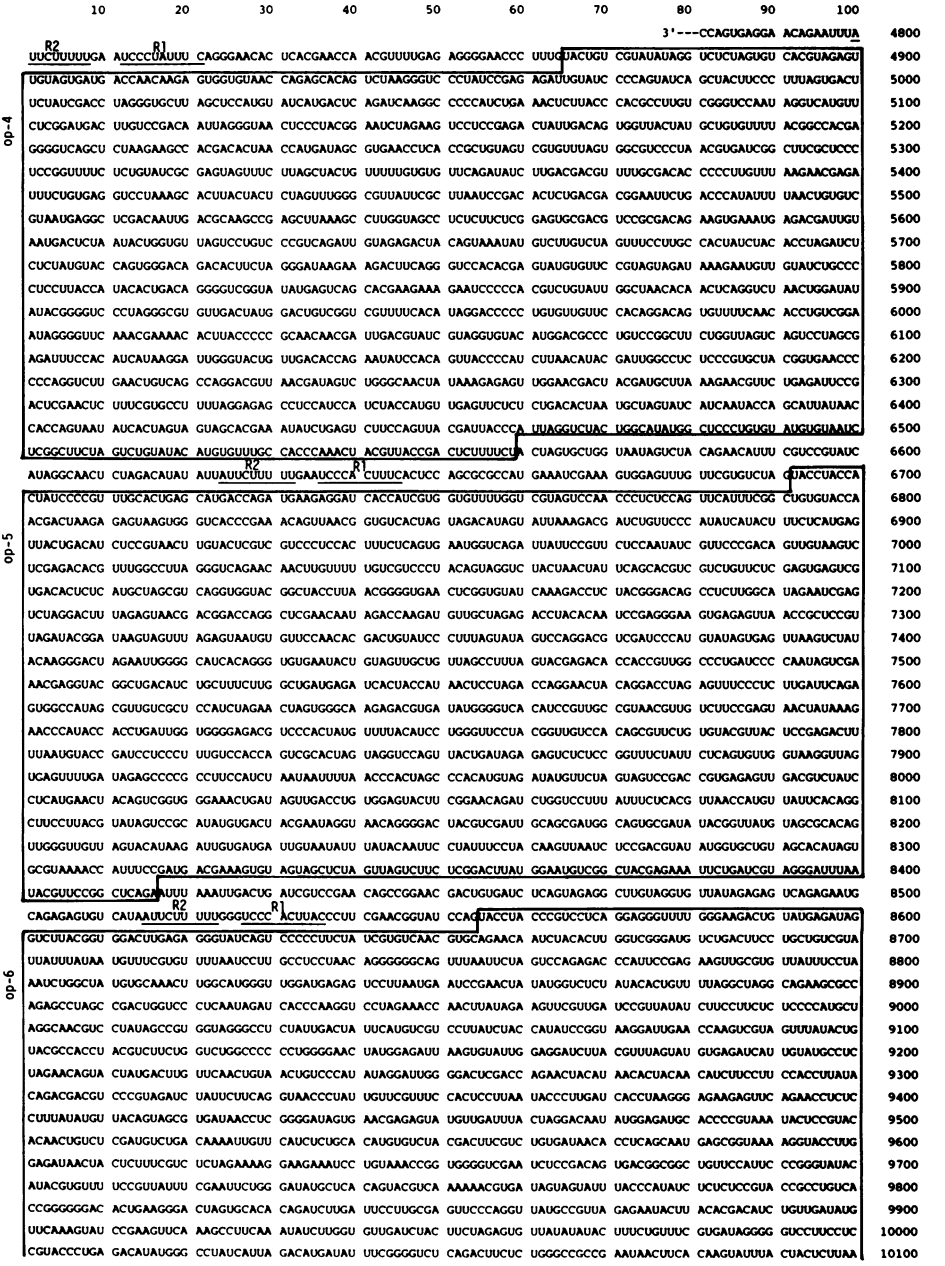
the Sendai virus genome RNA is 15,383 nucleotides, which is in good agreement with the previous value estimated from the sedimentation coefficient obtained with a sucrose density gradient (25,26). The nucleotide sequence from position 4,781, the 5' end portion of the M gene, to 15,383, the 5' end of the genome RNA, is shown in Fig. 2.

Analysis of the nucleotide sequence

Distribution of the translation termination codons in the plus strand (complementary strand to the genome) of the above region is shown in Fig.3, indicating the presence of three large open reading frames. Open reading frame op-4, which starts from nucleotide position 4,865 and encodes the F protein as reported previously (4), was found to terminate at nucleotide position 6,559, encoding a polypeptide of 565 amino acids with a molecular weight of 61,495. Downstream of op-4, two open reading frames, designated as op-5 (from 6,692 to 8,416) and op-6 (from 8,555 to 15,238), were observed. Op-5 corresponds to 575 amino acids, the calculated molecular weight of which is 63,408, and op-6 encodes a giant polypeptide of 2,228 amino acids with a calculated molecular weight of 252,864.

Previous studies (3,4,5,6,7,8) showed that the Sendai virus NP, P+C and M genes are flanked by consensus sequence R1 (UCCCACUUUC or UCCCAGUUUC) at the 3' end and by R2 (AUUCUUUUU) at the 5' end, and that the junctiuon between these genes is composed of R2-GAA-R1. This structure was also found for op-4, op-5 and op-6, i.e. they are preceded by UCCCUAUUUC, UCCCACUUUC and UCCCACUUAC, respectively, which were assigned for R1, and followed by AUUCUUUUU (R2), and the junction between R1 and R2 was GAA or GGG. Thus, we assigned the sequences from R1 starting at position 4,812 to R2 ending at 6,632, from 6,636 to 8,523 and from 8,527 to 15,326 as the fourth, fifth and sixth genes, respectively, and GAA as the intergenic sequence with one exception between the fifth and sixth genes where it was -GGG-. It is of interest to note that the sequence, -GAA-, was again found after R2 of the last gene.

Since we have already assigned the first to fourth genes for NP, P+C, M and F, respectively, the fifth and sixth genes correspond to the HN and L genes. The coding capacity of op-5, 63,408, is very close to the reported value for the molecular weight of the unglycosylated form of the HN protein of 63K (27,28), while that of op-6, 252,864, is very similar to the estimated molecular weight of the L protein of about 200K (29). In order to confirm that op-6 really produces a large transcript, northern blot hybridization was performed between mRNAs from infected cells and a probe,


10
20
30
40
50
60
70
80
90
100
3'---CCAGUGAGGA ACAGAAUUUA 4800
R2
R1
UUCUUUUUGA AUCCCUAUUU CAGGGAACAC UCACGAACCA ACGUUUUGAG AGGGGAACCC UUUGUACUGU CGUAUAUAGG UCUCUAGUGU CACGUAGAGU 4900
UGUAGUGAUG ACCAACAAGA GUGGUGUAAC CAGAGCACAG UCUAAGGGUC CCUAUCCGAG AGAUUGUAUC CCCAGUAUCA GCUACUUCCC UUUAGUGACU 5000
UCUAUCGACC UAGGGUGCUU AGCUCCAUGU AUCAUGACUC AGAUCAAGGC CCCCAUCUGA AACUCUUACC CACGCCUUGU CGGGUCCAAU AGGUCAUGUU 5100
CUCGGAUGAC UUGUCCGACA AUUAGGGUAA CUCCCUACGG AAUCUAGAAG UCCUCCGAGA CUAUUGACAG UGGUUACUAU GCUGUGUUUU ACGGCCACGA 5200
GGGGUCAGCU CUAAGAAGCC ACGACACUAA CCAUGAUAGC GUGAACCUCA CCGCUGUAGU CGUGUUUAGU GGCGUCCCUA ACGUGAUCGG CUUCGCUCCC 5300
UCCGGUUUUC UCUGUAUCGC GAGUAGUUUC UUAGCUACUG UUUUUGUGUG UUCAGAUAUC UUGACGACGU UUUGCGACAC CCCCUUGUUU AAGAACGAGA 5400
UUUCUGUGAG GUCCUAAAGC ACUUACUACU CUAGUUUGGG CGUUAUUCGC UUAAUCCGAC ACUCUGACGA CGGAAUUCUG ACCCAUAUUU UAACUGUGUC 5500
GUAAUGAGGC UCGACAAUUG ACGCAAGCCG AGCUUAAAGC CUUGGUAGCC UCUCUUCUCG GAGUGCGACG UCCGCGACAG AAGUGAAAUG AGACGAUUGU 5600
AAUGACUCUA AUACUGGUGU UAGUCCUGUC CCGUCAGAUU GUAGAGACUA CAGUAAAUAU GUCUUGUCUA GUUUCCUUGC CACUAUCUAC ACCUAGAUCU 5700
CUCUAUGUAC CAGUGGGACA GACACUUCUA GGGAUAAGAA AGACUUCAGG GUCCACACGA GUAUGUGUUC CGUAGUAGAU AAAGAAUGUU GUAUCUGCCC 5800
CUCCUUACCA UACACUGACA GGGGUCGGUA UAUGAGUCAG CACGAAGAAA GAAUCCCCCA CGUCUGUAUU GGCUAACACA ACUCAGGUCU AACUGGAUAU 5900
AUACGGGGUC CCUAGGGCGU GUUGACUAUG GACUGUCGGU CGUUUUCACA UAGGACCCCC UGUGUUGUUC CACAGGACAG UGUUUUCAAC ACCUGUCGGA 6000
AUAGGGGUUC AAACGAAAAC ACUUACCCCC GCAACAACGA UUGACGUAUC GUAGGUGUAC AUGGACGCCC UGUCCGGCUU CUGGUUAGUC AGUCCUAGCG 6100
AGAUUUCCAC AUCAUAAGGA UUGGGUACUG UUGACACCAG AAUAUCCACA GUUACCCCAU CUUAACAUAC GAUUGGCCUC UCCCGUGCUA CGGUGAACCC 6200
CCCAGGUCUU GAACUGUCAG CCAGGACGUU AACGAUAGUC UGGGCAACUA UAAAGAGAGU UGGAACGACU ACGAUGCUUA AAGAACGUUC UGAGAUUCCG 6300
ACUCGAACUC UUUCGUGCCU UUUAGGAGAG CCUCCAUCCA UCUACCAUGU UGAGUUCUCU CUGACACUAA UGCUAGUAUC AUCAAUACCA GCAUUAUAAC 6400
CACCAGUAAU AUCACUAGUA GUAGCACGAA AUAUCUGAGU CUUCCAGUUA CGAUUACCCA UUAGGUCUAC UGGCAUAUGG CUCCCUGUGU AUGUGUAAUC 6500
UCGGCUUCUA GUCUGUAUAC AUGUGUUUGC CACCCAAACU ACGUUACCGA CUCUUUUCUA CUAGUGCUGG UAAUAGUCUA CAGAACAUUU CGUCCGUAUC 6600
R2
R1
AUAGGCAACU CUAGACAUAU AUUAUUCUUU UUGAAUCCCA CUUUCACUCC AGCGCGCCAU GAAAUCGAAA GUGGAGUUUG UUCGUGUCUA CUACCUACCA 6700
CUAUCCCCGU UUGCACUGAG CAUGACCAGA UGAAGAGGAU CACCAUCGUG GUGUUUUGGU CGUAGUCCAA CCCUCUCCAG UUCAUUUCGG CUGUGUACCA 6800
ACGACUAAGA GAGUAAGUGG GUCACCCGAA ACAGUUAACG GUGUCACUAG UAGACAUAGU AUUAAAGACG AUCUGUUCCC AUAUCAUACU UUCUCAUGAG 6900
UUACUGACAU CUCCGUAACU UGUACUCGUC GUCCCUCCAC UUUCUCAGUG AAUGGUCAGA UUAUUCCGUU CUCCAAUAUC GUUCCCGACA GUUGUAAGUC 7000
UCGAGACACG UUUGGCCUUA GGGUCAGAAC AACUUGUUUU UGUCGUCCCU ACAGUAGGUC UACUAACUAU UCAGCACGUC GUCUGUUCUC GAGUGAGUCG 7100
UGACACUCUC AUGCUAGCGU CAGGUGGUAC GGCUACCUUA ACGGGGUGAA CUCGGUGUAU CAAAGACCUC UACGGGACAG CCUCUUGGCA UAGAAUCGAG 7200
UCUAGGACUU UAGAGUAACG ACGGACCAGG CUCGAACAAU AGACCAAGAU GUUGCUAGAG ACCUACACAA UCCGAGGGAA GUGAGAGUUA ACCGCUCCGU 7300
UAGAUACGGA UAAGUAGUUU AGAGUAAUGU GUUCCAACAC GACUGUAUCC CUUUAGUAUA GUCCAGGACG UCGAUCCCAU GUAUAGUGAG UUAAGUCUAU 7400
ACAAGGGACU AGAAUUGGGG CAUCACAGGG UGUGAAUACU GUAGUUGCUG UUAGCCUUUA GUACGAGACA CCACCGUUGG CCCUGAUCCC CAAUAGUCGA 7500
AACGAGGUAC GGCUGACAUC UGCUUUCUUG GCUGAUGAGA UCACUACCAU AACUCCUAGA CCAGGAACUA CAGGACCUAG AGUUUCCCUC UUGAUUCAGA 7600
GUGGCCAUAG CGUUGUCGCU CCAUCUAGAA CUAGUGGGCA AGAGACGUGA UAUGGGGUCA CAUCCGUUGC CGUAACGUUG UCUUCCGAGU AACUAUAAAG 7700
AACCCAUACC ACCUGAUUGG UGGGGAGACG UCCCACUAUG UUUUACAUCC UGGGUUCCUA CGGUUGUCCA CAGCGUUCUG UGUACGUUAC UCCGAGACUU 7800
UUAAUGUACC GAUCCUCCCU UUGUCCACCA GUCGCACUAG UAGGUCCAGU UACUGAUAGA GAGUCUCUCC GGUUUCUAUU CUCAGUGUUG GUAAGGUUAG 7900
UGAGUUUUGA UAGAGCCCCG CCUUCCAUCU AAUAAUUUUA ACCCACUAGC CCACAUGUAG AUAUGUUCUA GUAGUCCGAC CGUGAGAGUU GACGUCUAUC 8000
CUCAUGAACU ACAGUCGGUG GGAAACUGAU AGUUGACCUG UGGAGUACUU CGGAACAGAU CUGGUCCUUU AUUUCUCACG UUAACCAUGU UAUUCACAGG 8100
CUUCCUUACG UAUAGUCCGC AUAUGUGACU ACGAAUAGGU AACAGGGGAC UACGUCGAUU GCAGCGAUGG CAGUGCGAUA UACGGUUAUG UAGCGCACAG 8200
UUGGGUUGUU AGUACAUAAG AUUGUGAUGA UUGUAAUAUU UAUACAAUUC CUAUUUCCUA CAAGUUAAUC UCCGACGUAU AUGGUGCUGU AGCACAUAGU 8300
GCGUAAAACC AUUUCCGAUG ACGAAAGUGU AGUAGCUCUA GUUAGUCUUC UCGGACUUAU GGAAUGUCGG CUACGAGAAA UUCUGAUCGU AGGGAUUUAA 8400
R2
R1
UACGUUCCGG CUCAGAAAUU AAAUUGACUG AUCGUCCGAA CAGCCGGAAC GACUGUGAUC UCAGUAGAGG CUUGUAGGUG UUAUAGAGAG UCAGAGAAUG 8500
CAGAGAGUGU CAUAAUUCUU UUUGGGUCCC ACUUACCCUU CGAACGGUAU CCAGUACCUA CCCGUCCUCA GGAGGGUUUU GGGAAGACUG UAUGAGAUAG 8600
GUCUUACGGU GGACUUGAGA GGGUAUCAGU CCCCCUUCUA UCGUGUCAAC GUGCAGAACA AUCUACACUU GGUCGGGAUG UCUGACUUCC UGCUGUCGUA 8700
UUAUUUAUAA UGUUUCGUGU UUUAAUCCUU GCCUCCUAAC AGGGGGGCAG UUUAAUUCUA GUCCAGAGAC CCAUUCCGAG AAGUUGCGUG UUAUUUCCUA 8800
AAUCUGGCUA UGUGCAAACU UGGCAUGGGU UGGAUGAGAG UCCUUAAUGA AUCCGAACUA UAUGGUCUCU AUACACUGUU UUAGGCUAGG CAGAAGCGCC 8900
AGAGCCUAGC CGACUGGUCC CUCAAUAGAU CACCCAAGGU CCUAGAAACC AACUUAUAGA AGUUCGUUGA UCCGUUAUAU CUUCCUUCUC UCCCCAUGCU 9000
AGGCAACGUC CUAUAGCCGU GGUAGGGCCU CUAUUGACUA UUCAUGUCGU CCUUAAUCUC CAUAUCCGGU AAGGAUUGAA CCAAGUCGUA GUUUAUACUG 9100
UACGCCACCU ACGUCUUCUG GUCUGGCCCC CCUGGGGAAC UAUGGAGAUU AAGUGUAUUG GAGGAUCUUA CGUUUAGUAU GUGAGAUCAU UGUAUGCCUC 9200
UAGAACAGUA CUAUGACUUG UUCAACUGUA ACUGUCCCAU AUAGGAUUGG GGACUCGACC AGAACUACAU AACACUACAA CAUCUUCCUU CCACCUUAUA 9300
CAGACGACGU CCCGUAGAUC UAUUCUUCAG GUAACCCUAU UGUUCGUUUC CACUCCUUAA UACCCUUGAU CACCUAAGGG AGAAGAGUUC AGAACCUCUC 9400
CUUUAUAUGU UACAGUAGCG UGAUAACCUC GGGGAUAGUG AACGAGAGUA UGUUGAUUUA CUAGGACAAU AUGGAGAUGC ACCCCGUAAA UACUCCGUAC 9500
ACAACUGUCU CGAUGUCUGA CAAAAUUGUU CAUCUCUGCA CAUGUGUCUA CGACUUCGUC UGUGAUAACA CCUCAGCAAU GAGCGGUAAA AGGUACCUUG 9600
GAGAUAACUA CUCUUUCGUC UCUAGAAAAG GAAGAAAUCC UGUAAACCGG UGGGGUCGAA UCUCCGACAG UGACGGCGGC UGUUCCAUUC CCGGGUAUAC 9700
AUACGUGUUU UCCGUUAUUU CGAAUUCUGG GAUAUGCUCA CAGUACGUCA AAAAACGUGA UAGUAGUAUU UACCCAUAUC UCUCUCCGUA CCGCCUGUCA 9800
CCGGGGGGAC ACUGAAGGGA CUAGUGCACA CAGAUCUUGA UUCCUUGCGA GUUCCCAGGU UAUGCCGUUA GAGAAUACUU ACACGACAUC UGUUGAUAUG 9900
UUCAAAGUAU CCGAAGUUCA AAGCCUUCAA AUAUCUUGGU GUUGAUCUAC UUCUAGAGUG UUAUAUAUAC UUUCUGUUUC GUGAUAGGGG GUCCUUCCUC 10000
CGUACCCUGA GACAUAUGGG CCUAUCAUUA GACAUGAUAU UUCGGGGUCU CAGACUUCUC UGGGCCGCCG AAUAACUUCA CAAGUAUUUA CUACUCUUAA 10100
op-4
op-5
op-6

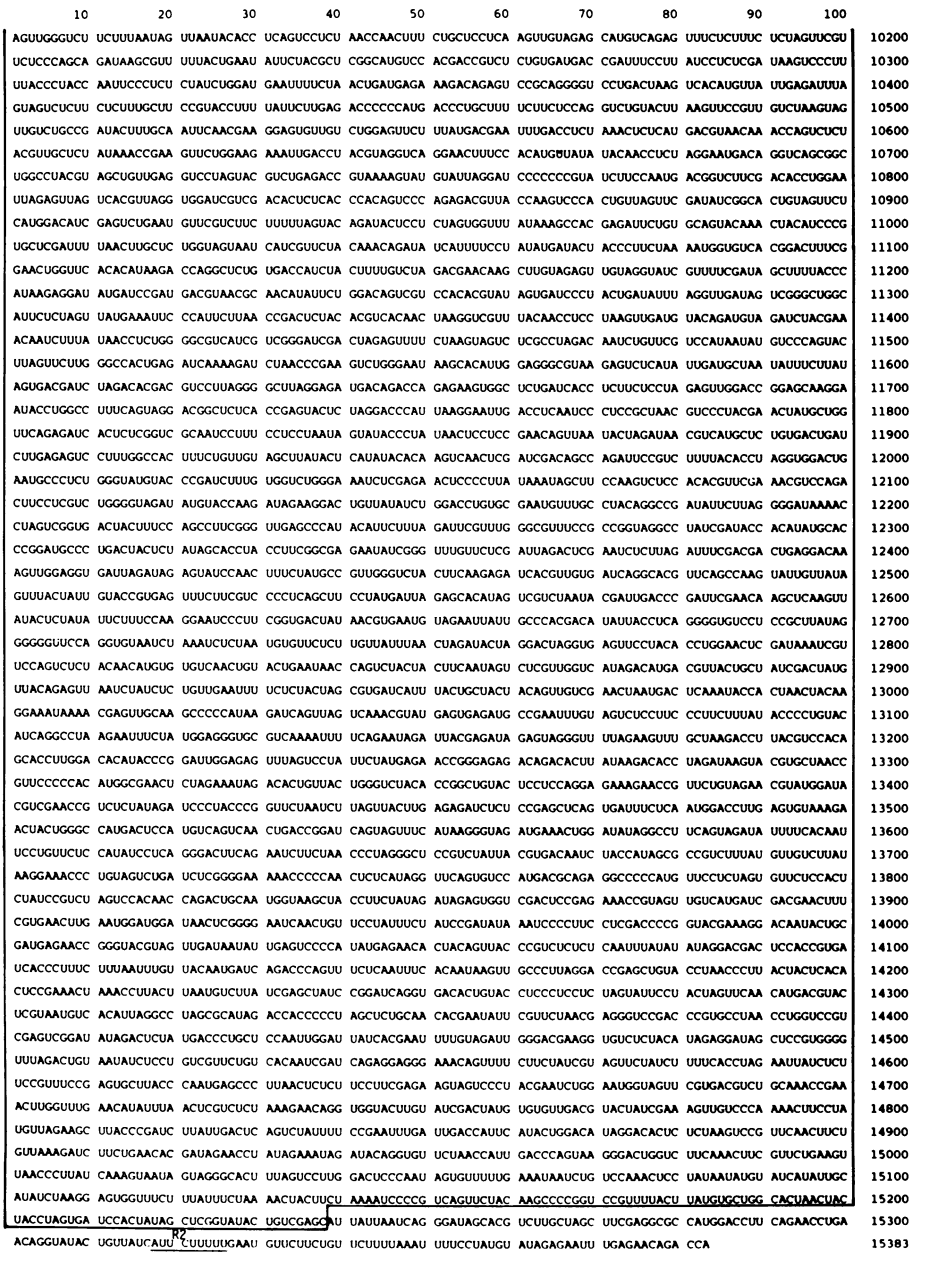


Fig. 2. The RNA sequence of the Sendai virus genome (Z strain) from nucleotide position 4,781 to 15,383, the 5' end of the genome. R1 and R2 are the repeating consensus sequences. Op-4, op-5 and op-6 denote the large open reading frames.

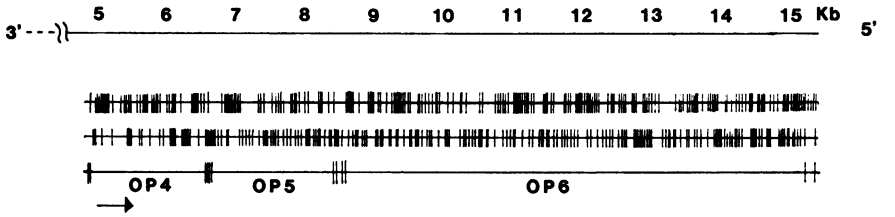


Fig. 3. Distribution of translation termination codons (vertical bars) in the plus strand RNA of the genome region presented in Fig. 2. An arrow indicates the direction of translation.

namely a cDNA fragment correponding to the 3' end portion of op-6 (from 8,540 to 9,719), which was prepared from clone 3S-222 by digestion with HindIII, followed by labeling in vitro by nick-translation. This probe hybridized with the largest virus specific poly(A) RNA as shown in Fig. 4. From these observations, we concluded that op-5 and op-6 encode the HN and L proteins, respectively.

No open reading frame was detected within the sequence of the 5' terminal 54 nucleotides following -GAA after the last gene, and the 3' half of this sequence was found to be U-rich and the 5' half to be A-rich. Thus, this region may be considered as the 5' leader region as reported for vesicular stomatitis virus (VSV). It is interesting to note that the most 5' terminal 12 nucleotides of this region are complementary (including one wobbling base pair) to the most 3' terminal 12 nucleotides of the genome,

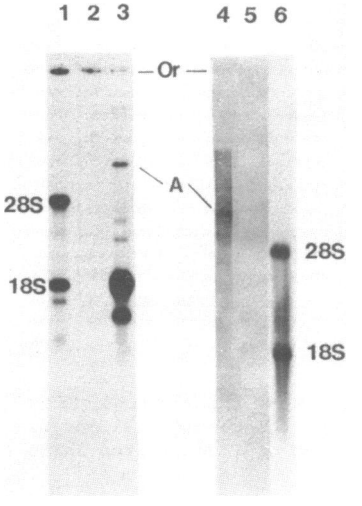


Fig.4. Left: Agarose gel electrophoresis of ^{32}P-labeled mRNAs from Sendai virus infected BHK-21 cells (lane 3) and those from uninfected cells (lane 2). Right: Hybridization of ^{32}P-labeled cDNA corresponding to the 3' end portion of OP-6 (nucleotide position 8,540 to 9,721) with mRNAs from infected cells (lane 4) and uninfected cells (lane 5). ^{32}P-labeled ribosomal RNA from uninfected cells served as size markers (lanes 1 and 6). "A" indicates the largest mRNA. "Or" means the origin.

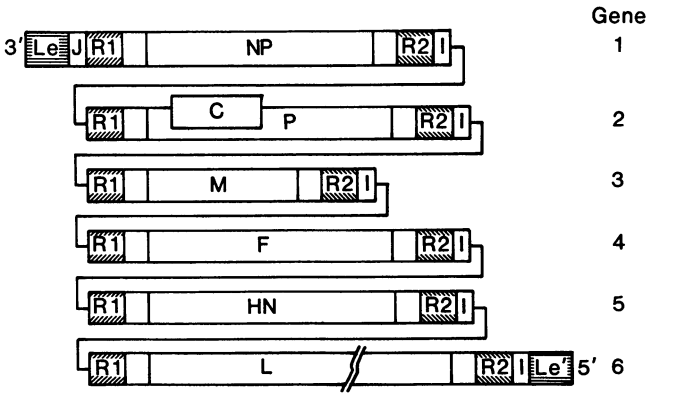


Fig. 5. A schematic illustration of the entire Sendai virus genome RNA. Le and Le' indicate the 3' leader region and the putative 5' leader region, respectively. I is the sequence, GAA or GGG, and J the sequence, AAAA.

which is in good agreement with a previous observation by Re et al.(24). This observation suggests the possibility that both ends of the genome RNA construct a stable secondary structure, giving the genome RNA a panhandle structure.

Combining the present results with the previous ones (3,4), we present

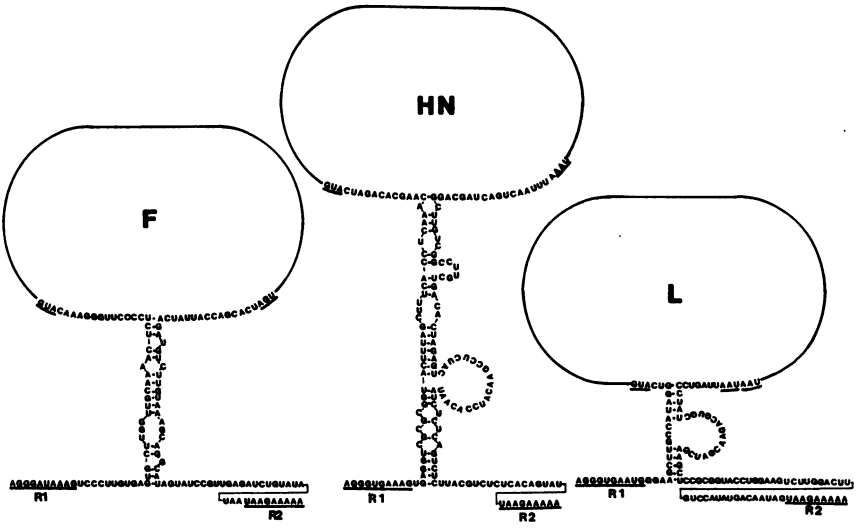


Fig. 6. Proposed secondary structures of the mRNAs for the F, HN and L proteins. The initiation and termination codons are underlined. R1 and R2 are the consensus sequences.

F protein

```
        10         20         30         40         50         60         70         80
MTAYIQRSQC ISTSLLVVLT TLVSCQIPRD RLSNIGVIVD EGKSLKIAGS HESRYIVLSL VPGVDFENGC GTAQVIQYKS   80
LLNRLLIPLR DALDLQEALI TVTNDTTQNA GAPQSRFFGA VIGTIALGVA TSAQITAGIA LAEAREAKRD IALIKESMTK  160
THKSIELLQN AVGEQILALK TLQDFVNDEI KPAISELGCE TAALRLGIKL TQHYSELLTA FGSNFGTIGE KSLTLQALSS  240
LYSANITEIM TTIRTGQSNI SDVIYTEQIK GTVIDVDLER YMVTLSVKIP ILSEVPGVLI HKASSISYNI DGEEWYVTVP  320
SHILSRASFL GGADITDCVE SRLTYICPRD PAQLIPDSQQ KCILGDTTRC PVTKVVDSLI PKFAFVNGGV VANCIASTCT  400
CGTGRRPISQ DRSKGVVFLT HDNCGLIGVN GVELYANRRG HDATWGVQNL TVGPAIAIRP VDISLNLADA TNFLQDSKAE  480
LEKARKILSE VGRWYNSRET VITIIVVMVV ILVVIIVIII VLYRLRRSML MGNPDDRIPR DTYTLEPKIR HMYTNGGFDA  560
MAEKR                (565)
```

HN protein

```
        10         20         30         40         50         60         70         80
MDGDRGKRDS YWSTSPSGST TKPASGWERS SKADTWLLIL SFTQWALSIA TVIICIIISA RQGYSMKEYS MTVEAINMSS   80
REVKESLTSL IRQEVIARAV NIQSSVQTGI PVLLNKNSRD VIQMIDKSCS RQELTQHCES TIAVHHADGI APLEPHSFWR  160
CPVGEPYLSS DPEISLLPGP SLLSGSTTIS GCVRLPSLSI GEAIYAYSSN LITQGCADIG KSYQVLQLGY ISLNSDMFPD  240
LNPVVSHTYD INDNRKSCSV VATGTRGYQL CSMPTVDERT DYSSDGIEDL VLDVLDLKGR TKSHRYRNSE VDLDHPFSAL  320
YPSVGNGIAT EGSLIFLGYG GLTTPLQGDT KCRTQGCQQV SQDTCNEALK ITWLGGKQVV SVIIQVNDYL SERPKIRVTT  400
IPITQNYLGA EGRLLKLGDR VYIYTRSSGW HSQLQIGVLD VSHPLTINWT PHEALSRPGN KECNWYNKCP KECISGVYTD  480
AYPLSPDAAN VATVTLYANT SRVNPTIMYS NTTNIINMLR IKDVQLEAAY TTTSCITHFG KGYCFHIIEI NQKSLNTLQP  560
MLFKTSIPKL CKAES        (575)
```

L protein

```
        10         20         30         40         50         60         70         80
MDGQESSQNP SDILYPECHL NSPIVRGKIA QLHVLLDVNQ PYRLKDDSII NITKHKIRNG GLSPRQIKIR SLGKALQRTI   80
KDLDRYTFEP YPTYSQELLR LDIPEICDKI RSVFAVSDRL TRELSSGFQD LWLNIFKQLG NIEGREGYDP LQDIGTIPEI  160
TDKYSRNRWY RPFLTWFSIK YDMRWMQKTR PGGPLDTSNS HNLLECKSYT LVTYGDLVMI LNKLTLTGYI LTPELVLMYC  240
DVVEGRWNMS AAGHLDKKSI GITSKGEELW ELVDSLFSSL GEEIYNVIAL LEPLSLALIQ LNDPVIPLRG AFMRHVLTEL  320
QTVLTSRDVY TDAEADTIVE SLLAIFHGTS IDEKAEIFSF FRTFGHPSLE AVTAADKVRA HMYAQKAIKL KTLYECHAVF  400
CTIIINGYRE RHGGQWPPCD FPDHVCLELR NAQGSNTAIS YECAVDNYTS FIGFKFRKFI EPQLDEDLTI YMKDKALSPR  480
KEAWDSVYPD SNLYYKAPES EETRRLIEVF INDENFNPEE IINYVESGDW LKDEEFNISY SLKEKEIKQE GRLFAKMTYK  560
MRAVQVLAET LLAKGIGELF RENGMVKGEI DLLKRLTTLS VSGVPRTDSV YNNSKSSEKR NEGMENKNSG GYWDEKKRSR  640
HEFKATDSST DGYETLSCFL TTDLKKYCLN WRFESTALFG QRCNEIFGFK TFFNWMHPVL ERCTIYVGDP YCPVADRMHR  720
QLQDHADSGI FIHNPRGGIE GYCQKLWTLI SISAIHLAAV RVGVRVSAMV QGDNQAIAVT SRVPVAQTYK QKKNHVYEEI  800
TKYFGALRHV MFDVGHELKL NETIISSKMF VYSKRIYYDG KILPQCLKAL TKCVFWSETL VDENRSACSN ISTSIAKAIE  880
NGYSPILGYC IALYKTCQQV CISLGMTINP TISPTVRDQY FKGKNWLRCA VLIPANVGGF NYMSTSRCFV RNIGDPAVAA  960
LADLKRFIRA DLLDKQVLYR VMNQEPGDSS FLDWASDPYS CNLPHSQSIT TIIKNITARS VLQESPNPLL SGLFTETSGE 1040
EDLNLASFLM DRKVILPRVA HEILGNSLTG VREAIAGMLD TTKSLVRASV RKGGLSYGIL RRLVNYDLLQ YETLTRTLRK 1120
PVKDNIEYEY MCSVELAVGL RQKMWIHLTY GRPIHGLETP DPLELLRGIF IEGSEVCKLC RSEGADPIYT WFYLPDNIDL 1200
DTLTNGCPAI RIPYFGSATD ERSEAQLGYV RNLSKPAKAA IRIAMVYTWA YGTDEISWME AALIAQTRAN LSLENLKLLT 1280
PVSTSTNLSH RLKDTATQMK FSSATLVRAS RFITISNDNM ALKEAGESKD TNLVYQQIML TGLSLFEFNM RYKKGSLGKP 1360
LILHLHLNNG CCIMESPQEA NIPPRSTLDL EITQENNKLI YDPDPLKDVD LELFSKVRDV VHTVDMTYWS DDEVIRATSI 1440
CTAMTIADTM SQLDRDNLKE MIALVNDDDV NSLITEFMVI DVPLFCSTFG GILVNQFAYS LYGLNIRGRE EIWGHVVRIL 1520
KDTSHAVLKV LSNALSHPKI FKRFWNAGVV EPVYGPNLSN QDKILLALSV CEYSVDLFMH DWQGGVPLEI FICDNDPDVA 1600
DMRRSSFLAR HLAYLCSLAE ISRDGPRLES MNSLERLESL KSYLELTFLD DPVLRYSQLT GLVIKVFPST LTYIRKSSIK 1680
VLRTRGIGVP EVLEDWDPEA DNALLDGIAA EIQQNIPLGH QTRAPFWGLR VSKSQVLRLR GYKEITRGEI GRSGVGLTLP 1760
FDGRYLSHQL RLFGINSTSC LKALELTYLL SPLVDKDKDR LYLGEGAGAM LSCYDATLGP CINYYNSGVY SCDVNGQREL 1840
NIYPAEVALV GKKLNNVTSL GQRVKVLFNG NPGSTWIGND ECEALIWNEL QNSSIGLVHC DMEGGDHKDD QVVLHEHYSV 1920
IRIAYLVGDR DVVLISKIAP RLGTDWTRQL SLYLRYWDEV NLIVLKTSNP ASTEMYLLSR HPKSDIIEDS KTVLASLLPL 2000
SKEDSIKIEK WILIEKAKAH EWVTRELREG SSSSGMLRPY HQALQTFGFE PNLYKLSRDF LSTMNIADTH NCMIAFNRVL 2080
KDTIFEWARI TESDKRLKLT GKYDLYPVRD SGKLKTISRR LVLSWISLSM STRLVTGSFP DQKFEARLQL GIVSLSSREI 2160
RNLRVITKTL LDRFEDIIHS ITYRFLTKEI KILMKILGAV KMFGARQNEY TTVIDDGSLG DIEPYDSS       (2228)
```

here the primary structure of the entire Sendai virus genome, which is schematically illustrated in Fig. 5. As a whole, 99.17% of the Sendai virus genome is transcribed into mRNA and 93.63% is translated into proteins, indicating that the structure of the genome is utilized quite efficiently.

It is noteworthy that an open reading frame corresponding to 249 amino acids was detected in the genome sense strand within the L gene region (from nucleotide position 9,588 to 8,842), which is longer than that of the Sendai virus C protein, and the only one long open reading frame capable of coding for more than 150 amino acids in the genome strand. However, this frame is not flanked by R1 and R2, nor could a single stranded cDNA probe complementary to this open reading frame detect any subgenomic transcript from the infected cells, either poly (A) plus or poly (A) minus (data not shown). From these observations, it seems unlikely that this open reading frame is transcribed, although further studies should be carried out before a definite conclusion is drawn.

Proposed structures of mRNAs for the F, HN and L proteins

From the nucleotide sequence of the genome RNA, the nucleotide sequences of mRNAs for the F, HN and L proteins were deduced. On detailed examination of these sequences, it was found that in every mRNA, a part of the 5' noncoding sequence was complementary to that of the 3' noncoding sequence, suggesting that these ends might form a double-stranded structure, which gives the mRNA a panhandle structure (Fig. 6). These secondary structures seems to be fairly stable on the basis of the free energy levels (30).

Characteristics of the F, HN and L gene products

The amino acid sequences of the F, HN and L gene products are shown in Fig. 7. The deduced amino acid sequence of the F protein indicates that the F protein is highly hydrophobic overall. As we reported previously (4), a signal peptide of 24 to 27 amino acids was detected in its N terminus, and the cleavage site for the F1 and F2 proteins was assumed to be the arginine residue at position 116. The most hydrophobic region of this protein is located near its C terminus (from amino acid position 500 to 523), which is followed by a hydrophilic region of 42 amino acids. The F protein of Sendai virus has been reported to penetrate the viral envelope, leaving an at least 3K portion exposed to the inside of the envelope (31), which is similar to

Fig. 7. The predicted amino acid sequences of the F, HN and L proteins. The putative N-linked carbohydrate attachment sites are boxed. The underlining indicates the most hydrophobic regions.

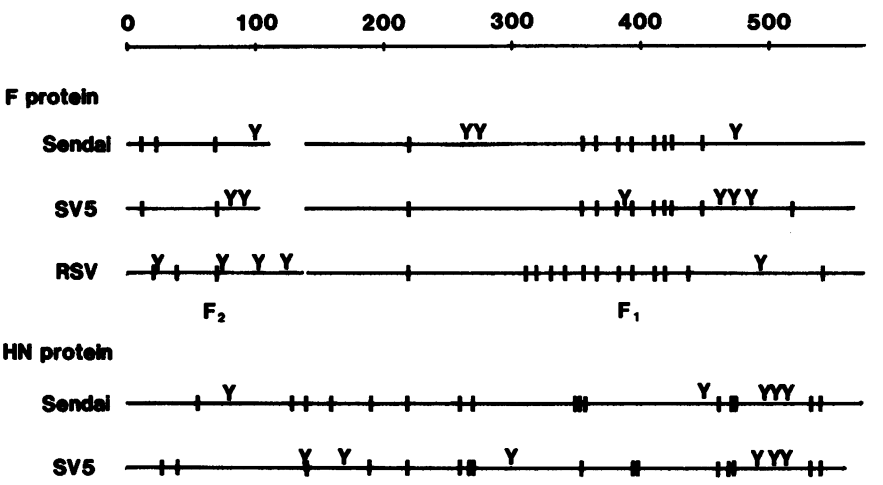


Fig. 8. Locations of the cysteine residues (vertical bars) and the putative N-linked carbohydrate attachment sites (Y) within the F proteins of Sendai virus, simian virus 5 (SV5) and respiratory syncytial virus (RSV), and the HN proteins of Sendai virus and SV5. To align the N termini of the F1 subunits, small gaps are left between the F2 and F1 subunits of Sendai virus and SV5.

in the case of hemagglutinin of influenza virus (32). Thus, it is probable that the hydrophobic region near the C terminus may anchor the F protein within the viral envelope and the very C terminal hydrophilic region may be located at the inner surface of the viral envelope, interacting with the viral M protein. As shown in Fig.8, the F protein has 12 cysteine residues, of which 2 are found in the signal peptide, 1 in the F2 portion and 9 in the F1 portion. Eight of the 9 cysteine residues in the F1 portion are clustered in its middle part, from position 338 to 424. Since the results of secondary structure analysis (33) indicated many reverse turns within the cysteine-rich portion of the F1 protein, this cluster may stabilize the tertiary structure of the F protein by forming intramolecular disulphide bonds. We could detect four putative N-linked carbohydrate attachment sites (Asn-X-Ser or Asn-X-Thr) (34) in the F protein, i.e. from position 104 to 106, 245 to 247, 259 to 261, and 449 to 451, respectively, which is in good agreement with the results reported by Kohama et al. (35) showing that the F protein of Sendai virus has four N-linked carbohydrate chains.

The most hydrophobic region of the HN protein is, in contrast to in the F protein, located near the N terminus, i.e. from position 36 to 58, which is preceded by a hydrophilic domain of 35 amino acids. This structure

resembles that of influenza virus neuraminidase (36,37), in which the hydrophobic region near the N terminus serves as an signal for membrane translocation and as an anchor in the membrane. Thus, it is plausible that this hydrophobic region near the N terminus of the HN protein may act as a signal and an anchor, and the preceding hydrophilic region of 35 amino acids is exposed to the inside of the viral envelope, since the HN protein was reported to penetrate the viral envelope and the molecular weight of the portion exposed to the inside of the envelope was estimated to be at least 2K (31). Within the deduced amino acid sequence of the HN protein, we found five N-linked carbohydrate attachment sites, i.e. 77 to 79, 448 to 450, 499 to 501, 504 to 506, and 511 to 513, which is in accordance with previous suggestion that the HN protein of Sendai virus has at least four N-linked carbohydrate chains (35).

Analyses of the deduced amino acid sequence of the L gene product provided us with little special information as to the structure of the L protein or its functional domains, except that Ser-Asp-Asp found from position 1,429 to 1,431 or Leu-Asp-Asp from 1,649 to 1,651 might be the active site of RNA synthesis as in the case of reverse transcriptases of retroviruses and RNA polymerases of picornaviruses (38). It is interesting to note that there is a possibility that another L gene product might be present which starts from the sixth AUG codon in mRNA, corresponding to the methionine at position 249 of the L protein or nucleotide position 9,299 to 9,301 with a calculated molecular weight of 224,005, since this codon is preceded by A 3 nucleotides upstream while the first five AUG codons are preceded by G or U, and according to Kozak (39), A might be preferentially recognized by ribosomes.

Comparison of the Z and Harris strains as to the F and HN genes

Recently, Blumberg et al. (8) reported the complete nucleotide sequence of the F gene of the Harris strain of Sendai virus as well as the sequence of the tripeptide of the N terminus of the F2 protein purified from virions of the same strain. On comparison of the results presented in this paper with theirs neither insertion nor deletion could be detected, and 98.63% of the nucleotides and 97.88% of the amino acids were found to be conserved between these two strains. The observation that the N terminus of the F2 protein of the Harris strain was glutamic acid at amino acid position 26 is in good agreement with our previous prediction as to the cleavage site of the signal peptide (4), thus we concluded that the peptide from 1 to 25 is the signal peptide and that from 26 to 116 is the F2 protein.

During preparation of this manuscript, the nucleotide sequences of the HN genes of the Harris strain (40) and Z strain (41) of Sendai virus were also presented. Comparing our present results for the Z strain with those for the Harris strain, we found that the latter is longer than the former by one amino acid residue, namely, the genome of the latter strain has an additional -AAA- between nucleotide positions 8,288 and 8,289 of that of the former, which led to an insertion of a serine residue. This insertion seems to make no significant difference as to the structure of the HN protein between these strains. Except for this insertion, most of the nucleotide sequence as well as the amino acid sequence was conserved between these two strains, giving 98.52% and 97.57% homology, respectively. On comparison of our results and those presented by Miura et al. (41) on the HN protein of the Z strain, four substitutions of amino acid residues were found while neither insertion nor deletion was detected.

Comparison of Sendai virus and SV5 as to the F and HN proteins

The nucleotide sequences of the F and HN genes of SV5, another paramyxovirus, were also reported recently (12,13). When the amino acid sequence of the F protein of the Sendai virus Z strain described above was aligned with that of SV5 to give minimal gaps for comparison, 133 amino acids of the F proteins were found to coincide with each other, but the overall homology was estimated to be only 23.5%. Interestingly, however, certain portions of the proteins show more than 50% homology, which were found from amino acid position 116 to 135 and from 458 to 477 of the F protein of Sendai virus (Fig. 9). This suggests the importance of these sequences for the function of the protein, since the N terminal portion of the F1 protein seemed to act as a functional domain during membrane fusion (42). It is noteworthy that the distribution of cysteine residues within the F proteins was well conserved between these two viruses. Both the F protein of Sendai virus and that of SV5 have 12 cysteine residues, out of which 10 could be aligned at the same positions, 1 in the F2 portion and 9 in the F1 portion, as shown in Fig. 8. The eight cysteine residues clustered in the middle part of the F1 protein of Sendai virus are all conserved at the corresponding positions of the F protein of SV5. In spite of these similarities, however, there is no indication that the carbohydrate attachment sites are distributed similarly in the F proteins of the two viruses.

When the amino acid sequences of the HN proteins of these viruses were aligned, 138 amino acids coincided and the overall homology was about 24%.

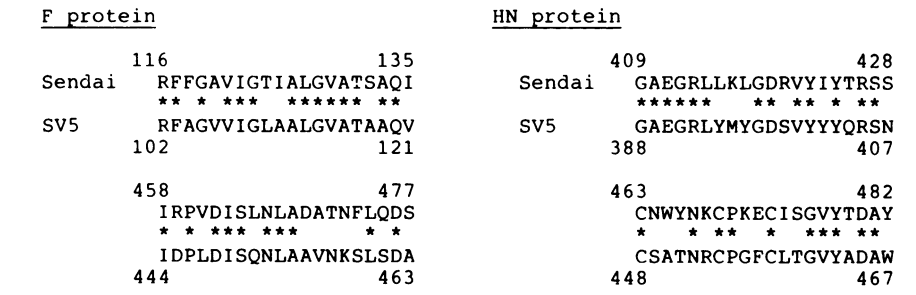


Fig. 9. Highly conserved regions of the F and HN proteins of Sendai virus and SV5. Numbers indicate the amino acid positions. Asterisks indicate the identity of amino acid residues.

As in the case of the F proteins, however, highly conserved portions were found from amino acid position 409 to 428 and from amino 463 to 483 (Fig.9), which may be included in the active sites for hemagglutinin and neuraminidase. Similarities could also be found between these two viruses in the locations of the cysteine residues in the HN proteins and those of carbohydrate attachment sites, three of which were found in the C terminal regions of the proteins.

DISCUSSION

At the beginning of our work concerning the determination of the nucleotide sequence of the Sendai virus genome RNA, we obtained relatively long cDNAs, of about 3,000 to 4,000 nucleotides in length, starting from its 3' end, which could be sequenced satisfactorily (3). However, attempts to elongate these cDNAs by the primer extension method generally only yielded short cDNA clones of about 600 to 1,000 nucleotides in length (4), which greatly hampered the determination of the entire nucleotide sequence of the Sendai virus genome RNA. Thus, we decided to adopt a new cDNA cloning strategy, which involved starting cDNA synthesis from multiple sites in the genome RNA in combination with the cloning method of Okayama and Berg (16). It was nesessary for this purpose to cut the genome RNA into a few fragments and to add a poly(A)-tail to the resulting fragments. We found that this could be achieved by using an excess amount of poly A polymerase (P-L Biochemicals, lot no. 206-7) during the polyadenylylation reaction of the genome RNA, since we had found that a trace amount of ribonuclease activity contaminated the poly A polymerase preparation. Thus, we succeeded in establishing a set of cDNA copies that completely covered the Sendai virus

genome RNA, and could determine the whole sequence of the genome RNA of 15,383 nucleotides, which revealed that the gene structure of the Sendai virus is 3'-NP-P+C-M-F-HN-L-5'.

One of the characteristic features of the genome is that each gene is flanked by consensus sequences at both ends, that is, R1 at the 3' end and R2 at the 5' end. Since R1 shows minor variations from gene to gene, i.e. UCCCAGUUUC for the NP gene (3), UCCCACUUUC for the P+C (3), M (4) and HN genes, UCCCUAUUUC for the F gene (4), and UCCCACUUAC for the L gene, its common structure was deduced to be UCCC-A/U-C/G/A-UU-U/A-C or UCCCNNUUNC. On the other hand, R2 was found to be AUUCUUUUU for all genes. It is highly possible that R1 is the recognition sequence for viral RNA polymerase, minor differences in which may play a role in the control of the expression of each gene, while R2 is a polyadenylation signal. Consensus sequences similar to R1 and R2 were also reported for non-segmented negative stranded RNA viruses, i.e. VSV (43,44), respiratory syncytial virus (RSV) (45) and measles virus (46), suggesting their common importance in the transcription and/or replication process of these viruses. It is interesting to note that the sequence, GAA or GGG, which was found between two adjacent genes (or between R2 and R1) and thought to be an intergenic sequence (3,4), was detected after R2 of the L gene. This strongly indicates that R2 together with this trinucleotide may constitute a signal sequence for the termination of transcription as well as for polyadenylation. The finding that the 12 nucleotides of the very 5' end of the genome are complementary to the 12 nucleotides of the very 3' end of the genome is very important because it supports the prediction that the genome would form a very stable panhandle structure, which will provide the signal sequences for recognition by viral RNA polymerase and for association between the RNA and nucleocapsid proteins (47).

In VSV, it has been reported that about 50 nucleotides of the very 3' end of both the genome and antigenome RNA are transcribed to small RNAs designated as plus and minus leader RNAs, respectively (2,48). In Sendai virus, however, plus leader RNA but not minus leader RNA was detected in infected cells (2). Accordingly, it is of interest to investigate whether the 3' terminal 54 nucleotides of the Sendai virus antigenome, which is complementary to the 5' terminal 54 nucleotides of the genome RNA, can be transcribed to produce minus leader RNA, and to determine the function of minus leader RNA.

It is noteworthy that a stable secondary structure could be formed

within the non-coding portions of all the mRNAs for the F, HN and L proteins, giving each mRNA a panhandle structure, and similar structures are also possible for the construction of mRNAs for the NP, P+C and M genes (not shown). This type of secondary structure has been proposed for the mRNA of gene 10 of human rota virus (49), although the panhandle structure involves coding sequences in this case. The model presented in the present paper is of special interest, since the panhandle is constructed from only non-coding sequences, leaving the initiation as well as the termination codon within the single-stranded loop structure. According to this model, this secondary structure of the mRNA may be very important for its translation, because ribosomes may select the first AUG in the loop as the initiation codon, and bind directly to it with the aid of the initiation factors. In this regard, it is interesting to note that two forms of the secondary structure are possible for the construction of the mRNA of the P+C gene, which has two open reading frames that overlap (3). In one form, the initiation codon as well as the termination codon for both the P and C proteins are present within the loop structure, whereas in the other form, both the initation and termination codons for the P protein are buried in the stem structure, while those for the C protein remain in the loop structure. Details of these structures will be published elsewhere.

Recently, the nucleotide sequence of the F gene of respiratory syncytial virus (RSV) was reported (45,50). As expected from the fact that this virus is classified as a pneumovirus different from paramyxoviruses (51), there is little homology between the deduced amino acid sequence of this gene and that of the Sendai virus F gene. It is noteworthy, however, that the F protein of RSV also has a cluster of cysteine residues in the middle portion of the F1 protein (Fig. 8) as the F proteins of Sendai virus and SV5 do, indicating that this structure may be very important in the determination of the tertiary structure of biologically active F proteins.

The L protein of paramyxoviruses is expected to exhibit multifunctional activities as to viral transcription and replication, including the initiation, elongation, termination, polyadenylation, methylation and capping reactions, as in the case of the L protein of VSV. However, we could not find any significant homology in the amino acid sequence between the L gene product of Sendai virus and that of VSV (52), and it is too early to infer the functional domains of L proteins, and it is necessary to have further information on the structures of the L proteins of other

paramyxovirus. Therefore, we have started analyzing the L gene of bovine parainfluenza type 3 virus.

ACKNOWLEDGEMENT

We wish to express our thanks to Dr. A. Nomoto for his helpful discussions. This work was supported by grants from the Ministry of Education, Science and Culture and the Ministry of Health and Welfare of Japan.

*To whom correspondence should be addressed

REFERENCES

1. Lamb,R.A. and Choppin,P.W. (1978) Virology 84, 469-478.
2. Leppert,M., Rittenhouse,L. Perraut,J., Summers,D.F. and Kolakofsky,D. (1979) Cell 18, 735-747.
3. Shioda,T., Hidaka,Y., Kanda,T., Shibuta,H., Nomoto,A. and Iwasaki,K. (1983) Nucleic Acids Res. 11, 7317-7333.
4. Hidaka,Y., Kanda,T., Iwasaki,K., Nomoto,A., Shioda,T. and Shibuta,H. (1984) Nucleic Acids Res. 12, 7965-7972.
5. Giorgi,C., Blumberg,B.M. and Kolakofsky,D. (1983) Cell 35, 829-936.
6. Blumberg,B.M., Giorge,C., Rose,K. and Kolakofsky,D. (1984) J.gen.Virol. 65, 769-779.
7. Blumberg,B.M., Rose,K., Simona,M.G., Roux,L., Giorgi,C. and Kolakofsky,D. (1984) J.Virol. 52, 656-663.
8. Blumberg,B.M., Gorgi,C., Rose,K. and Kolakofsky,D. (1985) J.gen.Virol. 66, 317-331.
9. Morgan,E.M., Re,G.G. and Kingsbury,D.W. (1984) Virolgy 135, 279-287.
10. Dowling,P.C., Giorgi,C., Roux,L., Dethlesen,L.A., Galantowitz,M.E., Blumberg,B.M. and Kolakofsky,D. (1983) Proc.Natl.Acad.Sci. USA. 80, 5213-5216.
11. Glazier,K., Raghow,R. and Kingsbury,D.W. (1977) J.Virol. 21, 863-871.
12. Paterson,R.G., Harris,T.J.R. and Lamb,R.A. (1984) Proc.Natl.Acad.Sci. USA. 81, 6706-6710.
13. Hiebert,S.W., Paterson,R.G. and Lamb,R.A. (1985) J.Virol. 54, 1-6.
14. Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) Biochmistry 18, 5294-5299.
15. Inokuchi,Y. Hirashima,A. and Watanabe,I. (1982) J.Mol.Biol. 158, 711-730.
16. Okayama,H. and Berg,P. (1982) Mol.Cel.Biol. 2, 161-170.
17. Boliver,F. and Backman,K. (1979) Methods Enzymol. 68, 245-267.
18. Messing,J. (1983) Methods Enzymol. 101, 20-78.
19. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc.Natl.Acad.Sci. USA. 74, 5463-5467.
20. Grunstein,M. and Hogness,D.S. (1975) Proc.Natl.Acad.Sci. USA. 72, 3961-3965.
21. Wahl,G.M., Stern,M. and Stark,G.R. (1979) Proc.Natl.Acad.Sci. USA. 76, 3683-3687.
22. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol. 65, 499-560.
23. Rigby,P.W.J., Deckman,M., Rhodes,C. and Berg,P. (1977) J.Mol.Biol. 113, 237-259.
24. Re,G.G., Gupta,K.C. and Kingsbury,D.W. (1983) Virology 130, 390-396.

25. Kolakofsky,D., Boy de la Tour,E. and Delius,H. (1974) J.Virol. 13, 261-268.
26. Shibuta,H., Kanda,T., Adachi,A. and Yogo,Y. (1979) Microbiol. Immunol. 23, 617-628.
27. Nakamura,K., Hommma,M. and Compans,R.W. (1982) Virology 119, 474-487.
28. Hsu, M.-C. and Choppin,P.W. (1984) Proc.Natl.Acad.Sci.USA. 81, 7732-7736.
29. Lamb,R.A., Mahy,B.W.J. and Choppin,P.W. (1976) Virology 69, 116-131.
30. Tinoco Jr.I., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Chorothers,D.M. and Gralla,J. (1973) Nature New Biol. 246, 40-41.
31. Lyle,D.S. (1979) Proc.Natl.Acad.Sci.USA. 76, 5621-5625.
32. Gething,M.J., Bye,J., Skehel,J. and Waterfield,M. (1980) Nature 287, 301-306.
33. Chou,P.Y. and Fasman,G.D. (1978) Adv.Enzymology 47, 45-148.
34. Neuberger,A., Gottschalk,A., Marshall,R.D. and Spiro,R.G. (1972) in The Glycoproteins: Their composition, structure and function, Gottschalk,A. Ed., pp. 450-490, Elsevier, Amsterdam.
35. Kohama,T., Shimizu,K. and Ishida,N. (1978) Virology 90, 226-234.
36. Blok,J., Air,G.M., Laver,W.G., Ward,C.W., Lilley,G.G., Woods,E.F., Roxburgh,C.M. and Inglis,A.S. (1982) Virology 119, 109-121.
37. Bos,T.J., Davis,A.R. and Nayak,D.P. (1984) Proc.Natl.Acad.Sci.USA. 81, 2327-2331.
38. Kamer,G. and Agros,P. (1984) Nucleic Acids Res. 12, 7269-7282.
39. Kozak,M. (1984) Nature 308, 241-246.
40. Blumberg,B.M., Giorgi,C., Roux,L., Ramaswamy,R., Dowling,P., Chollet,A. and Kolakofsky,D. (1985) Cell 41, 269-278.
41. Miura,N., Nakatani,Y., Ishimura,M., Uchida.T. and Okada,Y. (1985) FEBS lett. 188, 112-116.
42. Richardson,C.D., Sceid,A. and Choppin,P.W. (1980) Virology 105, 205-222.
43. Rowlands,D.J. (1979) Proc.Natl.Acad.Sci. USA. 76, 4793-4797.
44. Rose,J.K. (1980) Cell 19, 415-421.
45. Collins,P.L., Huang,Y.T. Wertz,G.W. (1984) Proc.Natl.Acad.Sci. 81, 7683-7687.
46. Bellini,W.J., Englund,G., Rozenblatt,S., Arnheiter,H. and Richardson,C.D. (1985) J.Virol. 53, 908-919.
47. Blumberg,B.M., Giorgi,C. and Kolakofsky,D. (1983) Cell 32, 559-567.
48. Colonno,R.J. and Banerjee,A.K. (1976) Cell 8, 197-204.
49. Okada,Y., Richardson,M.A., Ikegami,N., Nomoto,A. and Furuichi,Y. (1984) J.Virol. 51, 856-859.
50. Elango,N., Satake,M., Coligan,J.E., Norrby,E., Camargo,E. and Venkatesen,S. (1985) Nucleic Acids Res. 13, 1559-1574.
51. Matthews,R.E.F. (1982) Intervirology 17, 1-200.
52. Schubert,M., Harmison,G.G. and Meier,E. (1984) J.Virol. 51, 505-514.