Sequence of 3,687 nucleotides from the 3' end of Sendai virus genome RNA and the predicted amino acid sequences of viral NP, P and C proteins

Tatsuo Shioda, Yuji Hidaka, Tadahito Kanda, Hiroshi Shibuta*, Akio Nomoto¹ and Kentaro Iwasaki²

Dept. Viral Infection, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108,
¹Dept. Microbiology, Faculty of Medicine, University of Tokyo, and ²Dept. Physiological
Chemistry, Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113, Japan

Received 19 September 1983; Accepted 12 October 1983

ABSTRACT

The sequence of 3,687 nucleotides from the 3' end of the Sendai virus genome (Z strain) was determined by a molecular cloning technique followed by rapid sequence analysis. Two large open reading frames, one consisting of 1,572 nucleotides and the other of 1,704 nucleotides, were observed in the region, that is OP-1 and OP-2 from the 3' end of the genome. The amino acid sequences of the gene products were predicted from the observed sequence. Determination of amino acid compositions of viral proteins, P, HN, Fo, NP and M, led us to conclude that NP and P are the gene products of OP-1 and OP-2, respectively. An additional open reading frame consisting of 612 nucleotides (OP-3) was discovered in the 3' most proximal region of OP-2. The predicted product of OP-3 was considered to be viral non-structural protein C. The leader sequence of 51 nucleotides at the 3' terminal of the genome and consensus sequences at 3' and 5' ends of each gene for proteins NP and P were identified.

INTRODUCTION

The genome of Sendai virus, a prototype of paramyxoviruses, is a single continuous RNA molecule of approximately 15 kilobases long (1,2) with negative polarity. This negative strand RNA functions as the template for the synthesis of multiple monocistronic messenger RNA (mRNA) species in the cytoplasm of infected cells (3,4). To date, six viral structural proteins, i.e., large (L), RNA polymerase (P), hemagglutinin-neuraminidase (HN), fusion (Fo) and membrane (M) proteins, have been clearly identified. In addition to these, non-structural protein C has been reported to be virus-specific (5). Mapping of the genes which encode these proteins, however, has not been accomplished yet. The mechanisms for the gene expression as well as RNA replication of the Sendai virus genome are not well understood.

Here we present the sequence of 3,687 nucleotides from the 3' end of the Sendai virus genome (Z strain) and the precise location of the NP and P genes in this genome region. We also describe the 3' and 5' terminal consensus sequences of the genes, discuss the gene position of protein C,

and compare the genome structure of Sendai virus with that of vesicular stomatitis virus (VSV).

MATERIALS AND METHODS

Virus

Strain Z (6) of Sendai virus was newly plaque-purified before use by 3 successive single-plaque isolations. The virus was propagated in the allantoic cavity of chicken eggs or in Madin-Darby bovine kidney (MDBK) cell cultures, and purified as described previously (2).

Preparation of viral RNA

Virion RNA was prepared by phenol/chloroform extraction of purified egg-grown virus in the presence of 0.5 % sodium dodecylsulfate (SDS) (7). To prepare viral mRNAs, mouse SRCDF₁-DBT cells (8,9) were infected with Sendai virus at an input of 10 plaque-forming units per cell and incubated at 37 C in Eagle minimum essential medium containing 5 % calf serum. From 14 to 18 hr postinfection (pi), the cultures were incubated in the presence of 3 µg of actinomycin D per ml. For preparation of labeled mRNAs, the cultures were pretreated with 3 µg of actinomycin D per ml in the medium with 1/10 the normal concentration of phosphate from 14 to 16 hr pi, and then incubated with phosphate-free medium containing both 3 µg of actinomycin D per ml and 0.6 to 1.0 mCi of ³²P orthophosphate per plate (9 cm in diameter) from 16 to 18 hr pi. The cells harvested with a rubber policeman were lysed with 0.5 % NP-40 and the cytoplasmic fraction of the cell lysates was deproteinized with phenol/chloroform in the presence of 0.5 % SDS and 2.5 µg of proteinase K per ml. The mRNAs were selected by oligo(dT)-cellulose column chromatography.

Synthesis of complementary DNA (cDNA)

Virion RNA was polyadenylylated with ATP:RNA adenylyltransferase according to Inokuchi et al. (10). The polyadenylylated RNA was purified by oligo(dT)-cellulose column chromatography, precipitated with ethanol, and centrifuged through a 15-30 % sucrose density gradient in 0.1 M NaCl-5 mM Tris-HCl-1 mM EDTA, pH 7.4. The 50S RNA was collected and used as the template for cDNA synthesis. The procedures for double-stranded cDNA synthesis were described previously (11). The double-stranded cDNAs were purified by Sephadex G-75 column chromatography, and treated with S1 nuclease and then with E. coli DNA polymerase I. After ligation with Hind III or BamH I linker nucleotides, the cDNAs were digested with the respective restriction endonucleases, purified by Sephadex G-150 column

chromatography and inserted into the Hind III or BamH I cleavage site of pBR 322. *Escherichia coli* K 12 strain X1766 was transformed with the recombinant DNAs (12). Ampicillin-resistant and tetracycline-sensitive colonies were selected as transformants that carry the recombinant plasmid with inserted cDNA. The plasmids chosen for sequencing (see later) were transferred to *E. coli* K 12 strain HB 101. Transformed X1776 or HB 101 was grown in X1776 broth or L broth, respectively, in the presence of ampicillin, and the plasmids were amplified by using chloramphenicol. The plasmid DNA was isolated according to Clewell and Helinski (13) using Triton X-100, sedimented with polyethyleneglycol, and purified by centrifugation in CsCl/ethidium bromide gradients.

Sequencing of cDNA

The cDNAs were cleaved into fragments with restriction endonucleases and the fragments sequenced by the method of Maxam-Gilbert (14).

Hybridization of cDNAs with viral RNAs (Northern technique)

Viral 50S RNA or mRNAs, after glyoxal treatment (15), were electrophoresed in 1 % agarose gel in 0.01 M sodium phosphate buffer, pH 7.2, and the resolved RNAs were transferred to nitrocellulose filter papers according to Thomas (16). The cDNAs were cleaved into fragments with Hinf I and the 5' end of the cDNA fragments was labeled with ^{32}P . The labeled cDNA fragments were used as the probe for hybridization between the cDNA and viral RNAs by the method of Wahl et al. (17).

Amino acid composition

Viral proteins (P, HN, Fo, NP and M) were separated by preparative SDS-8.5 % polyacrylamide gel electrophoresis (18). Protein bands visualized by soaking the gel in 4 M CH_3COONa were cut out. The sliced gels were homogenized with a Dounce homogenizer and the protein was extracted overnight with 5 % SDS in 20 mM sodium phosphate buffer, pH 7.0, containing 1 mM PMSF at room temperature. After removing the homogenized gel by centrifugation, 4 volumes of cold acetone was added to the supernatant and the mixture was kept at -20°C for a few hr to precipitate the extracted protein. The precipitates were collected by centrifugation, and the pellets washed several times with cold acetone. About 100 μg of each protein was hydrolysed with 6 N HCl for 22 hr, and its amino acid composition was determined with an amino acid analyzer, Hitachi model 835.

Enzymes and other materials

ATP:RNA adenylyltransferase was a gift from Dr. A. Hirashima, Keio University, Tokyo. Avian myeloblastosis virus reverse transcriptase was a

gift from Dr. S. Kawai, Institute of Medical Science, University of Tokyo. The following materials were purchased: S1 nuclease from Bethesda Research Laboratories; *E. coli* DNA polymerase I and T4 DNA ligase from New England Biolabs; calf intestine alkaline phosphatase and proteinase K from Boehringer Mannheim; Hind III and BamH I linker nucleotides, and all the restriction endonucleases from Takara Shuzo, Kyoto; nitrocellulose filter paper (HAWP) from Millipore; and ^{32}P -orthophosphate from Amersham.

RESULTS

Sequence of 3,687 nucleotides from the 3' end of the viral genome

Double-stranded cDNAs were synthesized from in vitro polyadenylated viral genome RNA by reverse transcription using oligo(dT) as primer. The cDNAs ligated with Hind III or BamH I linker nucleotides were inserted into the respective cleavage sites of plasmid pBR 322. Forty-nine ampicillin-resistant and tetracycline-sensitive colonies were obtained for Hind III site insertion and 116 for BamH I insertion. The maximum length of the inserted DNA, estimated by 1 % neutral agarose gel electrophoresis, was about 4 kilobases which corresponds to about 1/4 of the total length of the viral genome.

Twelve cDNA clones of 2 to 4 kilobases were chosen and examined for their ability to hybridize with viral 50S RNA by the Northern technique (data not shown). Two clones did not hybridize with the RNA, suggesting that these clones possibly contained complementary sequences to ribosomal RNA of host cells which was not separated from 50S RNA by sucrose gradient centrifugation (19). For further analysis of the remaining positive clones, their Hinf I digestion patterns were compared (data not shown). Nine out of the 10 clones showed similarity to each other in the digestion pattern and were used for further analysis in this study. One clone, B85, which showed a quite different Hinf I digestion pattern was left for subsequent studies.

The nucleotide sequences of clones B36, B99 and H33 were determined, the restriction maps of which are presented in Fig. 1. The nucleotide sequences of the other clones were partially examined. To assure the accuracy of sequences both strands of the cDNA fragments were analyzed.

The sequence analysis revealed that 6 clones including B36 and B99 had a poly(A) tail (Fig. 1) of lengths from 11 to 20 nucleotides. The upstream sequences adjacent to the poly(A) tail were identical in every clone. These results strongly indicated that the DNA sequence next to the

Fig.1. Locations of the inserted cDNAs of clones H33, B36 and B99 in the 3' terminal 3,687 nucleotide region of the Sendai virus genome, and restriction maps of these cDNAs. "●" indicates a poly(A) tail. For A and A', see text.

poly(A) reflected the sequence of the very 3' end of the viral genome. In fact, the sequence was identical with the reported sequence for the 3' terminal 25 nucleotides of the Sendai virus genome (20) except for the substitutions of C with U at positions 15 and 20 from the 3' end.

Clones H33 and B36 had different sequences at one end which are indicated in Fig. 1 as parts A and A'. This discord seems to be due to cDNA-cDNA ligation which could occur during the process of linker nucleotides ligation to the cDNAs. The sequence of the remaining part of H33 was identical to that of the counterpart of B36. The sequence of B99 was completely maintained in B36. Thus, the RNA sequence of 3,687 nucleotides from the 3' end of the Sendai virus genome was deduced to be as shown in Fig. 2.

Analysis of the 3,687 nucleotide sequence

Two large open reading frames composed of more than 1,500 nucleotides were detected in the sequence, that is, from position 120 to 1,691 (OP-1) and from 1,844 to 3,547 (OP-2). The former corresponds to 524 amino acids and the latter to 568 amino acids. Each of the open reading frames is flanked by non-coding sequences of several tens of nucleotides, and further by the consensus sequences of R1 ($\text{AUC}_3\text{A}^{\text{G}}/\text{C}^{\text{U}}_3\text{C}$) and R2 (AU_2CU_5), as indicated in Fig. 2. R2 is identical with the consensus sequence reported by Gupta et al. (21) as being complementary to the 3'

	10	20	30	40	50	60	70	80	
	3'-UGGUUUUUGUC UCUUUUUUUGU ACUAUACCCUA UACAUAUACUU CAAUAUUGCC UAAAAUCCCA GUUUCAUAGG UGGGACUCCU								80
	Leader RT								
	CGUCCAAGGU CUGGGAACG AACGACGGU UUCAAGUGU ACCGGCCCA CAACUCUGG AAGCUAUGA AUUCGAGAU								160
OP-1	CUCCUCGCU UCAUAUUUUU UAGCCCUUC UCCACGACA UAGGGGCGG UCUCUCUGU UCAGAGUAC AAGCAUGAU								240
	CGAAUUCACA CUGACUACU CGUCUGUUA AUAAGUAACG UUGAUGAAG GAUCGAGUGA GUACUCUGU UCUAUUCGCC								320
	GUGAGAGUCU CUCCCCCAA GGAGCAGAGA GACGAACGGU ACCGAUUGU AUCAGGUCUU AACAUGAACU GUUGUUUGCC								400
	UCAUUUCGG CUACAGUUUA UACCUUAGAU GUUGUAUCU UUUCUGGGA UCUCUCUGU CUUCUGCCU AAGUAACACU								480
	UCUGCUCUCU AUACCUUUAU CCUCUCUGU GUCUUAACGA CAAACUUGA UACCAUUGU UCUCUGGUGA GAAGGUCCCA								560
	GUUGCCCUAC GACGUCUGG ACUGUGUGAG GAAGUUUAGA UACCAUAGG ACGUACGGAU CCUCGUUAUU AACGAGUUA								640
	GACCUAACAC GACCAUUCG GGUAGUGUUC GUCGCGCGG AAUUCUUUC CCAAGAAGU GUCCAUCUC CGCAAGUCUG								720
	UUCUGCCGUG GCACUUAUCA CGAAUACAA AGUGACCCU CUGUCAACU CCCUACCGA GCCAAUACU UAGAGUCGUU								800
	UCGGAACAU GAGAGUACCA ACUCUGGGA CACUGAUAC UAUAGACGUU UAGACUAGAG UGGUGUAUC UCUUUCUGA								880
	GGUCUAGCA CCCUUGAUG AGGCUUACG UCCCGACCGA AGGAAGUAC UGUGUAUUU UAUACCCAC CUUUGUUUCU								960
OP-2	ACCGUCAGA UUGCAACAGU UUGGACUCCG GGUUAUUUU AUUCGAUUCU UCGGAGUAUC UGGUAGUGA CAGUUUCCG								1040
	GGGUCUCAG GGAUAUAGC AUAGGAGUUC UGGGACAAG UACCAUUAU ACGAGGUCG UUAUAGGAC GUGAUCCUC								1120
	AUAGCGGUC CCUCAGCGGG ACGUAGUCU GUUCCGUUAC GUCGUCAUG AGUUGCCUC CUGUAUGAA CUUAUCCUUU								1200
	ACAAGAUGA UCCUGUUCGG CACCGUUUC UGCAGUUCG CUUCUUAUC UACGGAACC UCUCAUCAA UCCUACUCC								1280
	CUUCGUCGU UCCCUCCGA GUCUGUAGA GACCGUCGA ACAGGCCACC UUUACACGA AUGGCUUUG GUUGUCCGCC								1360
	ACCACGUAA CUCAUCGAG AUCUGUACG GCUGUAGCUG GAUCUUUGU UUCGGUACG CCUGGUCCG CGAUCCCAA								1440
	CCCCACUCU AUACACCAU UCUAUCCGUG CAGUCCACU ACCACGGUG AAACAGUGU AUGUAUCCG ACUUGCCAAU								1520
	CCUCCUCU UGUUAUCU CCUAUAGU CUGUAUCUC CUUCUUAUG GUACUCUGA AGGUUCUCU GCCGUUCUC								1600
	AAGACUUGG GUACCUUAC UUCGGCGUU AUUGCCACAG CUAAUACUG UUCUACUGU AUGGCGUGU CAUGGACCC								1680
	AUUCUUAU GUUCCUAGA UGUCCGAAG UUCAUGAAC UAGGCAUCA UCUUUUUUA UCCACUUAU AAGUAGGUG								1760
	RT								
	CUAGCCGAG CCGUCCGUG GUGGUGUGG UGGGUGUGA GCUCUGUGG UGCCGAAGC GAUGGUAUG								1840
OP-3	GCUACUAG UUCUACGAA GUAAGAUUU CUUCUAGC UUCAACUCU CCUCCGCGU CCUCUCUGU UCAGCGAGAU								1920
	CCUACAUAG CCUAAGGAGC UACGACAGA CAGCUCACU GGUUACUGU AGCUCCCUU GUCUUCGAC GAGGUGUUGU								2000
	GGUAGUUGU AGGGGUCCU GGUCCGAGC GAGUAUCUG GUUUACUCU CCGUUCUUC UUCAGAGUUG UGGCAGCUGG								2080
	GUUCUUAUG CUAGUCCACU CCUCUACU CAGAGACCU CUUGUUGUU CGGUUCCGU CUUCUGUAC GACCUUGGA								2160
	ACUUAUUGU UUAUUGUGG CCCGAAACC CCUCUUGA CAUUGUUGA GACUAGAGU CUUAGACCG CUACUCCUC								2240
	UGAGGCCUUA GGAACUUUA GGAAGUUUAC UCUCUCCAU AGGCUCUAG CCAUAUUC UACUUUGU UCUCUACCGA								2320
	CGGUGGGAC UAUUUCUCC UCUCUGGUU CGAUUCCUG AAGGUUCUU CCAUGUCCU UCAUUGAGG AUGGACUACU								2400
	UCCUCUCCA CCUCGUUCA UUAUACCUUC UUCGUACCU GAGCCGUGA GUGUACAG UUCUUAUGA CCCGAGACC								2480
	ACUAAGGAUC GGGGCUAGAA CUUCUCCAG ACGAUCCUC CUUGUUUUC UCUGGAUGU UGUACCCAG GUUUGAGAA								2560
	UAGGUGUGU GGCACGACC GUGGCCAGG GUGGCGACU UAGCAUUGU GUCUGUCC CAGUGUGUG CUUUUGGGG								2640
OP-4	UAGAUUGUC CUACUCUGU AGUUGAGACC CCUGGGGGG CGGACGUCC AGUUCUGGC CUUUGUGGU UAUCCUGGG								2720
	CGAGACAGU UCUAACAGG CCGUUGCCG CAGUUAAGU GGGCCAGAU CCUGGUGA GUUUGUUUU CCCGUACCU								2800
	CUCUUGUGA GUAGUAUUC UCUCUACCG UGUAAACAU GCUCAGAACC ACUUAAGG AGACGAGUC UUAAGCUAG								2880
	UAGGCGUCG CGCUAAUAC ACAAAAGUUC UGCACGGAU UUCAGACGU UGAUACGUC CUACUGUAG UUAUACUAC								2960
	CGGACUAGA AAGACGGUC UUAAGAAGC GAGCAUCCA UCUACUCUG UUUUGUAGC AGUUUGUCU GGUUCUCUG								3040
	CACCUCAGU AGGCCUCUG GAUAUUCUC AAGAGACUA UAGUCUUUC UGUCUUGAG AACGACUA CUGUUGAGU								3120
	AUGGAAGUA UAGUAGUUC UAUUCCACC GUUCUGACU UUGUGUCUG GGGAAUGUUC CAGGGGAGG CAAAAACGU								3200
	UUAUUUUCU CUUGUUGA UUCGAGGU CAAACUGG UAGAUACCU UGGGAUUC UAUACUUA GUUUGCCUG								3280
	GAUAGGUC UCUCUUAU AUUCUACUC UAGGCGUUG GCACAUUGU UCUCUCCUG UGUUUGGUA CCCGAGUUU								3360
	GGUAGUGCA GAGGAGGGA GGUUCUCU CGGUGUAC GUGAGAGAU CCGAGCAGU UCUCUGUA GGGGAUUCU								3440
OP-5	CUGCAGUCU UCUGUUAU CACUUAUA AUAGGUAC GUUCUGUGU GUUCUCAA UCCGUCAGU CCUUGAGAU								3520
	CUUCUCCUG AUUCUGGUA CUGGUUUC UAGGCCAC UCCUAGGAG UGUAGGAGU AGUAUCUC CUUUGAGAU								3600
	RT								
	GGUAGUGUA GUCGUCUUA UCUAUUUUU UUGAAUCC ACUUUUUA AAGUGAUUG UGCCGCUUA CCGUCUAG								3680
	AUAUCA (3687)								

terminal of Sendai virus mRNAs where polyadenylation would occur. At position 1,729-1,750 and 3,625-3,646, the sequence of R2-GA-R1 is repeated. Assuming that the dinucleotide GA is the intergenic sequence of the Sendai virus genome by analogy with VSV (22), R1 is possibly considered as the putative promotor consensus sequence to be transcribed for the 5' terminal region of viral mRNAs. Thus, we assigned the sequence at position 55-1,737 and the sequence at position 1,740-3,633 as the first gene (gene-1) and the second gene (gene-2), respectively. Accordingly, the R1 at 3,636-3,646 seems to be the 3' terminal of the subsequent third gene. These features are schematically presented in Fig. 3. Very interestingly, the region of OP-2 in gene-2 contains another overlapping open reading frame (OP-3) of 612 nucleotides (204 amino acids). The opening methionine codon of OP-3 is located at only 7 nucleotides downstream of that of OP-2. The deduced amino acid sequences of these 3 open reading frames are shown in Fig. 4.

Before the start of gene-1, there is a 3' terminal region of the genome consisting of 54 nucleotides that has trinucleotide AAA at its 5' end. This structure resembles that of the 3' terminal region of the VSV genome. Based on the finding that Sendai virus yields leader RNA of about 50 nucleotides (19), and the observation that AAA or AAAA is the junction between the leader region and the first gene of the VSV genome (23), we consider it likely that the 51 nucleotides at the 3' terminal followed by AAA-R1 is the leader region of the Sendai virus genome. This size is very similar to the 47 nucleotides of the VSV leader region. The 3' terminal 20 nucleotides of the leader region are extremely U-rich like the leader region of the VSV genome (23,24).

A noteworthy possibility is that the intergenic sequence is not the dinucleotide GA but the trinucleotide GAA as suggested by Kolakofsky (personal communication). In this case, R1 is UC_3A^G/C_3U_3C and the junction between the leader region and gene-1 is AAAA.

Identification of gene-1 and gene-2 products

To determine the products of gene-1 and gene-2, amino acid compositions calculated from the predicted amino acid sequences of OP-1 and OP-2 were compared with those determined with purified viral

Fig. 2. The RNA sequence of 3,687 nucleotides from the 3' end of the Sendai virus genome (Z strain). OP-1, OP-2 and OP-3 indicate large open reading frames. R1 and R2 are repeating consensus sequences.

Fig. 3. A schematic illustration of the Sendai virus genome in the region of 3,687 nucleotides from the 3' end.

proteins, P, HN, NP and M, of egg-grown Sendai virus as well as Fo of MDBK cell-grown Sendai virus. The results (Fig. 5) demonstrated that the predicted amino acid compositions of OP-1 and OP-2 coincided quite well with the estimated amino acid compositions of NP and P, respectively, but not those of other proteins. On the contrary, the predicted amino acid composition of OP-3 did not coincide with any of the estimated values of viral proteins.

Hybridization of cDNA with viral mRNAs

Sendai virus mRNAs isolated from infected SRCDF₁-DBT cells, selected by oligo(dT)-cellulose column chromatography and treated with glyoxal, were resolved into 4 bands on 1 % neutral agarose gel electrophoresis (lane 4 in Fig. 6). Each of bands B and C was reported to involve 2 mRNAs of similar size (3, 21). When the inserted cDNA of clone B36 was tested for hybridization with viral mRNAs by the Northern technique, the cDNA was retained by bands B and C (lane 2 in Fig. 6). The cDNA did not hybridize with the host cell mRNAs obtained from uninfected cells (lane 3 in Fig. 6). These results indicated that clone B36 contains the cDNA of at least two mRNAs.

DISCUSSION

The sequence of 3,687 nucleotides from the 3' end of the genome of Sendai virus strain Z was deduced from the cDNA synthesized by reverse transcription. In this sequence, we assigned the leader region, gene-1 and gene-2 in this order from the 3' end.

OP-1

MAGLLSTFDT FSSRRSEIN KSGRGAIVPG QRSTVSFVL GLSVTDDADK LFIATTFLAH SLODKRHSQ RGGFLVSLLA 80
 MAYSSPELYL TTNGVNADVK YVIYNIKDP KRTKTDGFIV KTRDMEYERT TEWLFQPMVN KSPLFQQQRD AAPDTLLQI 160
 YGYPACLGAI IVQVWIVLVK AITSSAGLRK GFFNRLEAFR QDGTVKGALV FTGETVEGIG SVMRSQQSLV SLMVETLVTM 240
 NTARSDLTTL EKNIQIVGNY IRDAGLASFM NTIKYGVETK MAALTLSNLR PDINKLRSLI DTYLKSGPRA PFICILKDPV 320
 HGEFAPGNY ALWSYAMGVA VVQNKAMQQY VTGRTYLDMF MFLLGQAVAK DAESKISSAL EDELGVTEAA KGRLRHHLAS 400
 LSGGNGAYRK PTGGGAIEVA LDNADIDLET KAHADQDARG WGGDSGERWA RQVSGGHFVT LHGAERLEEE TNDEDVSDIE 480
 RRIAMRLSRE TARSDATHGD EGRNNGVDYD EDDDTAAVAG VGGI (524)

GLY 50 VAL 38 ALA 49 THR 38 SER 37 ARG 33 LEU 48 PRO 14 GLU 30 ASP 38
 LYS 24 ASN 19 MET 14 ILE 26 TRP 5 CYS 2 TYR 16 PHE 18 GLN 16 HIS 9
 M.W. 57,017

OP-2

MDQQAFILKE DSEVEREAPG GRESLSDVIG FLDAVLSSEP TDIGGDRSWL HNTINTPQGP GSAHRAKSEG EGEVSTPSTQ 80
 DNRSGEESRV SGRTSKPEAE AHAGNLKQNK IHRAFGGRTG TNSVSQDLGD GGDGSGILENP PNERGYPRSG IEDENREMAA 160
 HPDKRGEDQA EGLPEEVRGS TSLPDEGEGG ASNNGRSMEP GSSHSARVTG VLVIPSPELE EAVLRNRKRR PTNSGSKPLT 240
 PATVPGTRSP PLNRYNLTGS PPGKPPSTQD EHINSGDTPA VRVKDRKPPI GTRSVSDCPA NGRSIHPGLE TDSTKKGIGE 320
 NTSSMKEMAT LLTSLGVIQS AQEFESSRDA SYVFARRALK SANYAEMTFN VCGLILSAEK SSARKVDENK QLLKQIQESV 400
 ESFRDITYKRF SEYQKEQNSL LMSNLSTLHI ITDRGGKTDN TDSLTRSPSV FAKSKENKTK ATRFDPSMET LEDMKYKPD 480
 IREDEFRDEI RNPVYQERDT EPRASNARSL LPSKEKPTMH SLRLVIESSP LSRAEKAAYV KSLSKCKTDQ EVKAVMELVE 560
 EDIESLTN (568)

GLY 44 VAL 27 ALA 35 THR 37 SER 69 ARG 43 LEU 39 PRO 38 GLU 57 ASP 35
 LYS 34 ASN 28 MET 11 ILE 21 TRP 1 CYS 3 TYR 9 PHE 11 GLN 16 HIS 10
 M.W. 61,958

OP-3

MPSFLKKILK LRGRQREDES RSRMLSDSSM LSCRNVQLTS EGTEAGSTTP SILPKDQALL IEPKVRACEK SQHRRPKIID 80
 QVRRVESLGE QASQRQKML ETLINKIYTG PLGEELVQTL YLRIWAMEET PESLKILQMR EDIRDQVLKM KTERWLRTLI 160
 RGEKTKLKDF QKRYEEVHPY LMKKEVQVI MELAWSLAHH IVQE (204)

GLY 7 VAL 10 ALA 8 THR 12 SER 15 ARG 19 LEU 24 PRO 8 GLU 24 ASP 7
 LYS 19 ASN 2 MET 9 ILE 12 TRP 3 CYS 1 TYR 4 PHE 2 GLN 14 HIS 4
 M.W. 23,987

Fig. 4. The predicted amino acid sequences, numbers of each amino acid and molecular weights of the gene products of OP-1, OP-2 and OP-3.

Previously, a gene order of 3'-NP-Fo-M-P-HN-L-5' was proposed for Sendai virus based on kinetic analysis of mRNA synthesis after UV-irradiation of infected cells (4), although these authors indicated the possibility that the positions of Fo and P could be inverted. In the present study, the predicted amino acid compositions of gene products of OP-1 and OP-2 coincided quite well with the estimated amino acid compositions of NP protein and P protein, respectively, but not those of any other viral proteins. These observations led us to conclude that gene-1 is the NP gene and gene-2 is the P gene. The gene order of

Fig. 5. The predicted amino acid compositions of the gene products of OP-1(A), OP-2(B) and OP-3(C) as well as the estimated amino acid compositions of P(D), HN(E), Fo(F), NP(G) and M(H) proteins. Each amino acid indicated in (I) is expressed as mol %.

3'-NP-P of Sendai virus resembles the gene order of 3'-N-NS of VSV. However, the predicted NP and P proteins consist of 524 and 568 amino acids, respectively, while the predicted N and NS proteins consist of 422 and 222 amino acids, respectively, (25), indicating that these proteins of Sendai virus are much larger than their counterparts of VSV. The molecular weights of gene products of OP-1 and OP-2 calculated from the amino acid sequence are 57,017 and 61,958, respectively, and thus smaller than the estimated values of NP (60,000) and P (75,000) proteins, respectively, which were determined by SDS-polyacrylamide gel electrophoresis (26). However, such discrepancies in molecular weight between calculated and estimated values have also been reported for VSV (25) and influenza A virus (27), although the reasons remain to be clarified.

The most interesting finding in the present study is that gene-2 has two overlapping open reading frames, OP-2 and OP-3. The opening

Fig. 6. Left: Hybridization of ^{32}P -labeled cDNA of clone B36 with mRNAs from Sendai virus infected SRCDF₁-DBT cells (lane 2) and those from uninfected cells (lane 3). Right: Agarose gel electrophoresis of ^{32}P -labeled mRNAs from infected cells (lane 4) and uninfected cells (lane 6). ^{32}P -Labeled ribosomal RNA fraction from uninfected cells served as markers (lanes 1 and 5).

methionine codon of OP-3 is located at only 7 nucleotides downstream of that of OP-2, and OP-3 has coding capacity for a protein of 204 amino acids which corresponds to a molecular weight of 23,987. Therefore, OP-3 seems to really encode a virus-specific protein. The predicted amino acid composition of the protein is completely different from any viral structural protein (Fig. 5). Dethlefsen and Kolakofsky (28) showed that the mRNA of non-structural C protein of Sendai virus is larger than expected from the estimated molecular weight of the protein, 22,000, and it cosediments with the mRNA of P protein on a sucrose gradient. Very recently, Dowling et al. reported that the gene order of Sendai virus is 3'-NP-P+C-M and the genetic loci of the viral P and C proteins cannot be separated (29). These observations strongly suggest that OP-2 and OP-3 in gene-2 encode P protein and C protein, respectively. The existence of two overlapping open reading frames was reported for influenza A virus (30,31) and Bunya viruses (32,33) whose genomes are segmented RNA of negative polarity. As other approaches to identify the gene products of OP-1, OP-2 and OP-3, *in vitro* translation of the mRNAs selected by restriction fragments of the cDNA of clone B36 as well as determination of partial

OP-1 (NP)											
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	5	7	7	1
UUC		UCC		UAC		UGC		13	3	9	1
UUA	Leu	UCA		UAA	Term	UGA	Term	9	5	0	0
UUG		UCG		UAG		UGG	Trp	8	3	0	5
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	6	6	6	1
CUC		CCC		CAC		CGC		9	4	3	2
CUA		CCA		CAA	Gln	CGA		9	4	6	1
CUG		CCG		CAG		CGG		7	0	10	3
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	9	6	8	8
AUC		ACC		AAC		AGC		11	10	11	11
AUA		ACA		AAA	Lys	AGA	Arg	6	17	8	15
AUG	Met	ACG		AAG		AGG		14	5	16	11
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	7	14	20	13
GUC		GCC		GAC		GGC		11	14	18	10
GUA		GCA		GAA	Glu	GGA		8	18	16	13
GUG		GCG		GAG		GGG		12	3	14	14
GC content (%), -CG-bridge*								49.3 33/55			
5		17		6		1		0	3	1	1
6		12		3		2		2	4	3	0
1		17		0		0		2	2	0	0
3		4		0		1		5	3	0	3
10		9		4		5		0	2	1	1
9		8		6		2		4	3	3	3
8		15		10		5		0	2	7	3
8		6		6		5		13	1	7	1
3		7		8		10		1	2	0	1
13		12		20		9		4	4	2	2
5		17		20		17		7	6	4	5
11		1		14		9		9	0	15	6
4		11		19		7		0	3	4	0
9		8		16		9		0	0	3	1
7		13		24		20		5	3	9	1
7		3		33		8		5	2	15	5
49.7 22/50								50.3 4/10			
OP-2 (P)								OP-3 (C)			

Fig.7. Codon usage in OP-1, OP-2 and OP-3.

* the observed numbers of NNC-GNN/the expected numbers of NNC-GNN in the open reading frames.

amino acid sequences of NP and P proteins are under investigation.

In the present study, we could not determine which is the intergenic sequence, GA or GAA. Therefore, it is still unknown whether R1, the 3' terminal consensus sequence of the genes, is AUC_3^G/U_3C or UC_3^G/U_3C . However, it is obvious that the R1 sequence is quite different from $U_2GUCNNUAG$ which is the 3' terminal consensus sequence of the genes of VSV (22), suggesting different manners of recognition by RNA

polymerase for transcription. On the contrary, the 5' terminal consensus sequence of the genes, R2 (AU_2CU_5), is very similar to the sequence of $AUACU_7$ detected for VSV which was reported to be the polyadenylation signal for mRNAs (22). The lengths of non-coding regions next to R1 in gene-1 and gene-2 are 54 and 93 nucleotides, respectively, which are much longer than the corresponding non-coding regions of the first gene (N) and second gene (NS) of VSV, that is, 3 and 0 nucleotides, respectively (22,25). The leader sequence of VSV contains no anti-codon of methionine, UAC, which is claimed as one of the characteristics of the leader sequence (23). However, the sequence assigned as the leader region of Sendai virus does contain UAC at position 20-22, from which an open reading frame corresponding to 43 amino acids can be arranged (Fig. 2), but the existence of such a polypeptide is doubtful, since this open reading frame is not flanked by R1 and R2. Thus, the present study demonstrated some resemblance of the genome structure between Sendai virus and VSV on one hand, but pointed out differences of the genome structure between the two viruses on the other hand. Therefore, it is inappropriate here to mention to the genetical relationship between the two viruses before further analysis of the Sendai virus genome has been carried out.

A deficiency in the usage of CG dinucleotide is reported for VSV (25, 34), influenza A virus (35) and polio virus (36). This is also the case for Sendai virus, since the codon usage listed in Fig. 7 for OP-1, OP-2 and OP-3 shows a clear bias against CGN, NCG and NNC-GNN.

A computer search of the determined sequence revealed many secondary structures. Of particular interest is a kind of palindrome structure observed from nucleotide position 3,635 to 3,659, that is,

(3,635)AAU-C-CCACUUU-CUUU
 (3,659)UUA---GGUGAAA-----|

This region includes the R1 sequence of the third gene following gene-2. Therefore, it is very interesting to speculate that this structure might play a regulatory role in the transcription of the gene or translation of the mRNA.

Determination of the complete primary structure of the Sendai virus genome, which is currently in progress, should provide more information about the gene expression and RNA replication of the virus.

ACKNOWLEDGEMENT

We wish to express our thanks to Drs.S.Uchida, K.Toyoshima and Y.Yogo for their encouragement and discussions, to Dr.H.Iizuka for

performing amino acid analyses and to Mr.T.Tanaka for his skillful technical assistance. This work was supported by grants from the Ministry of Education, Science and Culture of Japan.

*To whom correspondence should be addressed

REFERENCES

1. Kolakofsky,D., DE LA Tour,E.B. and Delius,H. (1974) J. Virol. 13, 261-268.
2. Shibuta,H., Kanda,T., Adachi,A. and Yogo,Y. (1979) Microbiol. Immunol. 23, 617-628.
3. Roux,L. and Kolakofsky,D. (1975) J. Virol. 16, 1426-1434.
4. Glazier,K., Raghov,R. and Kingsbury,D.W. (1977) J. Virol. 21, 863-871.
5. Lamb,R.A. and Choppin,P.W. (1978) Virology 84, 469-478.
6. Kanda,T. and Shibuta,H. (1981) Virology 108, 318-324.
7. Dorsch-haesler,K., Yogo,Y. and Wimmer,E. (1975) J. Virol. 16, 1512-1527.
8. Kumanishi,T. (1967) Japan. J. Exp. Med. 37, 461-474.
9. Hirano,N., Fujiwara,K. and Matumoto,M. (1976) Japan. J. Microbiol. 20, 219-225.
10. Inokuchi,Y., Hirashima,A. and Watanabe,I. (1982) J. Mol. Biol. 158, 711-730.
11. Nomoto,A., Toyoda,H., Imura,N., Noguchi,S. and Sekiya,T. (1982) J. Biochem. 91, 1593-1600.
12. Enea,V., Vovis,G.F. and Zinder,N.D. (1975) J. Mol. Biol. 96, 495-509.
13. Clewell,D.B. and Helinski,D.R. (1970) Biochemistry 9, 4428-4440.
14. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol. 65, 499-560.
15. Carmichael,G.G. and McMaster,G.K. (1980) Methods enzymol. 65, 380-391.
16. Thomas,P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.
17. Wahl,G.M., Stern,M. and Stark,G.R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687.
18. Laemmli,U.K. (1970) Nature 227,680-685.
19. Leppert,M., Rittenhouse,L., Perrault,J., Summers,D.F. and Kolakofsky,D. (1979) Cell 18, 735-747.
20. Re,G.G., Gupta,K.C. and Kingsbury,D.W. (1983) J. Virol. 45, 659-664.
21. Gupta,K.C. and Kingsbury,D.W. (1982) Virology 120, 518-523.
22. Rose,J.K. (1980) Cell 19, 415-421.
23. Rowlands,D.J. (1979) Proc. Natl. Acad. Sci. USA 76, 4793-4797.
24. Colono,R.J. and Banerjee,A.K. (1978) Cell 15, 93-101.
25. Gallione,C.J., Greene,J.R., Iverson,L.E. and Rose,J.K. (1981) J. Virol. 39, 529-535.
26. Homma,M., Tozawa,H., Shimizu,K. and Ishida,N. (1975) Japan. J. Microbiol. 19, 467-470.
27. Sivasubramanian,N. and Nayak,D.P. (1982) J. Virol. 44, 321-329.
28. Dethlefsen,L.A. and Kolakofsky,D. (1983) J. Virol. 46, 321-324.
29. Dowling,P.C., Giorgi,C., Roux,L., Dethlefsen,L.A., Galantowicz,M.E., Blumberg,B.M. and Kolakofsky,D. (1983) Proc. Natl. Acad. Sci. USA 80, 5213-5216.
30. Lamb,R.A., Choppin,P.W., Chanok,R.M. and Lai,C.-J. (1980) Proc. Natl. Acad. Sci. USA 77, 1857-1861.
31. Lamb,R.A. and Lai,C.-J. (1982) Virology 123, 237-256.
32. Bishop,D.H.L., Gould,K.G., Akashi,H. and Clerx-van Haaster,C.M. (1982) Nucleic Acids Res. 10, 3703-3713.
33. Akashi,H. and Bishop,D.H.L. (1983) J. Virol. 45, 1155-1158.
34. Rose,J.K. and Gallione,C.J. (1981) J. Virol. 39, 519-528.
35. Van Rompuy,L., Min Jou,W., Huylebroeck,D., Devos,R. and Fiers,W. (1981) Eur. J. Biochem. 116, 347-353.
36. Rothberg,P.G. and Wimmer,E. (1981) Nucleic Acids Res. 9, 6221-6229.