

---

**Nucleotide sequence of a Sendai virus genome region covering the entire M gene and the 3' proximal 1013 nucleotides of the F gene**

---

Yuji Hidaka, Tadahito Kanda, Kentaro Iwasaki\*, Akio Nomoto<sup>1</sup>, Tatsuo Shioda<sup>2</sup> and Hiroshi Shibuta<sup>2</sup>

Dept. Physiol. Chem., Tokyo Metropolitan Inst. Med. Sci., Honkomagome, Bunkyo-ku, Tokyo 113, <sup>1</sup>Dept. Microbiol., Fac. Med., Univ. Tokyo, Hongo, Bunkyo-ku, Tokyo 113, and <sup>2</sup>Dept. Viral Infection, Inst. Med. Sci., Univ. Tokyo, Shirokanedai, Minato-ku, Tokyo 108, Japan

---

Received 24 September 1984; Accepted 8 October 1984

---

**ABSTRACT**

We determined the sequence of the 2,138 nucleotides in the Sendai virus genome just following the 3' proximal 3,686 nucleotides which we had previously reported (Nucleic Acids Res. 11, 7317-7330, 1983). This covers the entire third gene of 1,173 nucleotides and the 3' proximal 1,013 nucleotides of the fourth gene. Like the NP and P+C genes, both the third and fourth genes start from consensus sequence R1 (3'-UCCAC(or UA)UUUC) at the 3' end and the third gene terminates with consensus sequence R2 (3'-AUUCUUUUU) at the 5' end. The third gene was identified as M, and the deduced 348 amino acids indicated that the M protein is rich in basic residues and has hydrophobic domains near the C-terminal. The fourth gene, although sequencing is not complete yet, was identified as F, since a large open reading frame found in the gene contains the characteristic sequence of 20 amino acids located at the N-terminal of the F<sub>1</sub> protein. Analyses of the amino acid sequence suggested that the structure of the F gene product is NH<sub>2</sub>-signal peptide-F<sub>2</sub>-F<sub>1</sub>-COOH.

**INTRODUCTION**

Recently, research on the genome and gene structures of paramyxoviruses has greatly increased with the use of complementary DNA (cDNA) to viral RNA (1,2,3,4,5,6,7,8). Information that has thus accumulated is of importance for the understanding of both the replication and transcription mechanisms as well as the nature of viral proteins of paramyxoviruses, and hence, for the elucidation of biological events caused by these viruses.

Among the viruses, Sendai virus has been the most extensively investigated as a prototype. The Sendai virus genome encodes the leader RNA and at least seven proteins; nucleocapsid (NP), RNA polymerase (P), membrane (M), fusion (F), hemagglutinin-neuraminidase (HN), large (L) and non-structural C proteins. So far, the gene order in the genome has been established to be 3'-Leader-NP-P+C- (1,2) and the nucleotide sequences of these genes were completely determined by us and others (2,3,4). In addition, the next gene following the P+C gene was almost unequivocally established to be the M gene (1) and the gene order subsequent to the M gene

has been proposed to be -F-HN-L-5' (1,9). As a continuation of our work to determine the whole nucleotide sequence of the genome of Sendai virus strain Z from the 3' end toward the 5' end, we present here the complete sequence of the third gene and the 3' proximal 1,013 nucleotide sequence of the fourth gene. From the results of analyses of the sequence, we identified the former as the gene for the M protein which seems to play a crucial role in the maturation of the enveloped virus and the latter as that for the F protein which characterizes paramyxoviruses.

### **MATERIALS AND METHODS**

#### **Preparation of viral genome RNA**

The 50S viral genome RNA was prepared as described previously (2) from Sendai virus strain Z which had been plaque-purified and grown in chicken eggs (2).

#### **Preparation of cDNA clones**

The cDNAs were synthesized by the primer extension method (10), using the viral genome RNA as the template. The primers used were DNA fragments prepared by digesting the B36 cDNA clone (2) or the newly prepared R54 cDNA clone with restriction endonucleases as indicated in the text. The cDNAs synthesized were ligated with Hind III linker nucleotides, digested with Hind III and then inserted into the Hind III site of the pBR322 plasmid. *E. coli* K12, strain HB101, was transformed with these recombinant plasmids by the standard technique (11). The colonies that hybridized to the viral genome RNA but not to the cDNA that had provided the primer fragment were selected (12) (Fig. 1).

#### **DNA sequencing**

The cDNAs were cleaved into fragments with appropriate restriction endonucleases and the fragments obtained were sequenced either by the method of Maxam-Gilbert (13) or by the dideoxy method of Sanger *et al.* (14) after subcloning of the fragments into the M13 phage (15).

#### **In vitro translation of hybrid-arrested mRNA**

Mouse SRCDF<sub>1</sub>-DBT cells were infected with Sendai virus as described previously (2). The RNA fraction was prepared by the guanidium thiocyanate method (16) from the infected cells, and the mRNAs were selected by oligo(dT)-cellulose column chromatography (17). The viral mRNAs were hybrid-selected with the use of viral cDNA fixed on nitrocellulose filters (18) and translated *in vitro* with rabbit reticulocyte lysates (19).

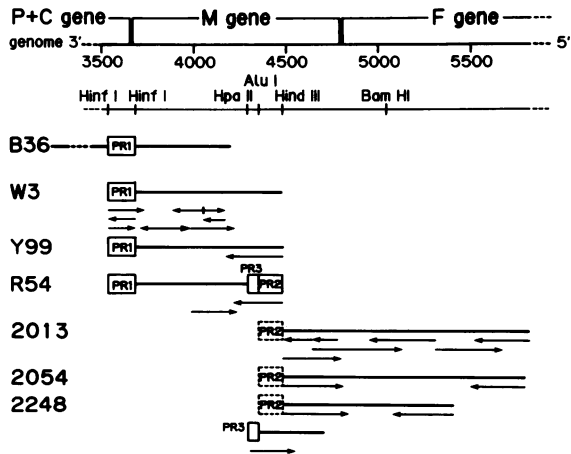


Fig. 1 Diagram showing the locations of cDNA clones (B36, W3, Y99, R54, 2013, 2054 and 2248) in the genome RNA and the sequencing strategy. The numbers under the genome line indicate the nucleotide positions. Arrows indicate the cDNA fragments sequenced and the direction of sequencing. PR1, PR2 and PR3 represent the fragments used as the primers for the reverse transcription of viral genome RNA.

### Enzymes and Other Materials

M13 cloning and sequencing kits, the rabbit reticulocyte lysate kit and all radioactive compounds were purchased from Amersham International plc., Amersham, England; HAWP nitrocellulose membrane filters from Millipore, Bedford, U.S.A.; avian myeloblastosis virus reverse transcriptase from Seikagaku Kogyo, Tokyo, Japan; and Hind III linker nucleotides, T4 DNA ligase, T4 polynucleotide kinase and bovine alkaline phosphatase, and all other restriction endonucleases from Takara Shuzo, Kyoto, Japan.

### RESULTS AND DISCUSSION

Figure 1 shows our sequencing strategy. The primer used in the first series of primer extension was the cDNA fragment from nucleotide positions 3,533 to 3,684 in the genome (PR1), which was prepared by digesting the B36 cDNA clone (2) with Hinf I. The cDNAs synthesized by this primer extension were ligated with the Hind III linker nucleotides, digested with Hind III and then inserted into the Hind III site of pBR322. Thus, we isolated cDNA clones R54, W3 and Y99. However, these clones were equal to one another in size, being smaller than expected. It was later found that this was due to the presence of a Hind III site at position 4,477 in the genome. The second series of primer extension was carried out in the same way as above,

Nucleic Acids Research

10	20	30	40	50	60	70	80	
3'---CGUAG	UCGGUCAUUU	CUAAUUCUUU	UDGAAUCCCA	CUUUCUUUAA	AGUGGAUUGU	GCCGCGUAC	CGUCUAUAGA	3680
		R2	R1					
UAUCUAAGG	AUCCAAGAGU	AUACUCCUAA	UGCCAUGACA	CCUCGGGGAC	GGAGACUCUU	GACCAGGCCU	AUUCUUUCGG	3760
▲								
UAGGGGGUGU	AGUCCUAACA	GUUCCAUCUU	CUGGGAGGAU	UUGUACCUCA	CUCUAUGGAU	CUAAAUACG	AGAACCCAAA	3840
GAACUCUGU	GGCUUUUUU	GUUGUUUAGA	UCCUCGCAU	AGACUGAACU	GUCUCGGCUG	GUCGAUGAGU	UAUACGCCGA	3920
GGCCAGCAA	UGGGUAUCCA	CACCGUUUA	UGAUGCCUG	ACUAGUCCUU	GAGAAUUUCC	GGACGUGGCU	AGAGUCUUAA	4000
UGCCACUCCU	CCUGACAAGC	UCGUCCUCUC	UACUAGCAUA	UGUACCACCU	AAGCUAACCA	CGAGGUGAGG	AUGGUACCAG	4080
UCCGUCCGAC	UCUGUCCUU	ACUAAAAUU	ACGUUUUUUC	CAGCGUGAUC	GAGGGUUUAC	GGAGGGACAC	CUGUCCUGU	4160
AUUCUGAGUC	UCACCACAAA	CAGUUACCCU	GUAGAGAUCC	CCGUUAGUGG	UAUCGGUUCU	AGGGUUUCUG	GGAACGUCUG	4240
GAACGUAAAG	GGUUGAGUAU	UAGGCAUUUA	AAUGACCACU	GUGAGUUCUG	GCCCUAGAGG	UGUCUUUUU	UCCCCAUGA	4320
GGGUCAUGAA	CUACUAGUUC	CCUCUUUUU	CGAGUUAAAA	UACCACGUGG	AGCCCAACUA	GUCCUCUUUC	CAGCCUUUCU	4400
AUAUGAGACA	ACUCAUGAGC	UUCUCGUUCU	AACUCUCUUA	CGCCGACUAA	AAGAGUGAAC	CCAAUUAGCC	GCCAAUUUCG	4480
AAGGUACAAG	UCCAAUUACC	CUGUGAUAGA	UUCUGUAAGU	ACUCAGUCGA	GCGUACCUUC	UCCCGUCAGA	CGAAGGGUAA	4560
UUACCUACAC	UUAGGGGUAU	ACUUGUACCA	CUAAACCCGC	CGUAGACAUC	UUUAGUGUCC	GCAGCUACGC	CACAAGGUUG	4640
GCCGSUAGGG	AGCACUAAAG	GCGAUGAUGG	GAUUACAACA	CCGAUUUCUUG	UAGCCUUCCU	AGUCUUUGCA	CAUUUACACG	4720
UGGGUAGUCU	CUGGACGCUU	UUACGGGGUU	CGUCUGUGGU	GGACCGUCAG	CCUCGGUGGC	CCAGUGAGGA	ACAGAAUUUA	4800
UUUUUUUUGA	AUCCUUUUUU	CAGGGAACAC	UCACGAACCA	ACGUUUUUGAG	AGGGGAACCC	UUUUUACUGU	CGUAUUAUAGG	4880
R2	R1							
UCUCUAGUGU	CAGUAGAGU	UGUAGUGAUG	ACCAACAAGA	GUGGUGUAAC	CAGAGCACAG	UCUAAGGGUC	CCUAUCCGAG	4960
AGAUGUUAUC	CCCAGUAUCA	GCUAUCUCCC	UUUAGUGACU	UCUAUCGACC	UAGGGUGCUU	AGCUCCAUGU	AUCAUGACUC	5040
AGAUCUAAAGC	CCCCAUUCGA	AACUCUUUAC	CACGCCUUUG	CGGGUCCAAU	AGGUCAUUGU	CUCGGAUGAC	UUGUCCGACA	5120
AUUAGGGUAA	CUCCUACGG	AAUCUAGAAG	UCCUCCGAGA	CUAUUGACAG	UGGUUACUUA	GCUGUGUUUU	ACGGCCACGA	5200
GGGUGACGCU	CUAAGAAGCC	ACGACACUAA	CCAUGAUAGC	GUGAACCUCA	CCGUCUGUAGU	CGUUGUUUAGU	GGCGUCCCUA	5280
ACGUGAUCGG	CUUCGCUCCC	UCCGGUUUUU	UCUGUAUCGC	GAGUAGUUUC	UUAGCUACUG	UUUUUGUGUG	UUCAGAUUUC	5360
UUGACGACGU	UUUUGGCACAC	CCCCUUUUUU	AAGAACGAGA	UUUCUGUGAG	GUCCUAAAGC	ACUUACUACU	CUAGUUUGGG	5440
CGUUUUUUCG	UUUUUCCGAC	ACUCUGACGA	CGAAAUUCUG	ACCAUAUUUU	UAACUGUGUC	GUAUUGAGGC	UCGACAAUUG	5520
ACGCAAGCCG	AGCUUUAAAGC	CUUGGUAGCC	UCUCUUCUGC	GAGUGCGACG	UCCGCGACAG	AAGUGAAUUG	AGACGAUUGU	5600
AAUGACUCUA	AUACUGGUGU	UAGUCCUGUC	CCGUCAGAUU	GUAGAGACUA	CAGUAAAUAU	GUCUUGUCUA	GUUUCCUUGC	5680
CACUAUCUAC	ACCUAGAUCU	CUCUAUGUAC	CAGUGGGACA	GACACUUCUA	GGGAUAAGAA	AGACUUCAGG	GUCCACACGA	5760
GUUAGUGUUC	CGUAGUAGAU	AAAGAUGUUU	GUUUCUGCCC	CUCCUUACCA	UACACUGACG	GGGU		5824

Fig. 2 The RNA sequence of the Sendai virus genome (Z strain) following the 3'-terminal sequence of 3,686 nucleotides previously reported (2). R1 and R2 are the repeating consensus sequences. Enclosed sequences correspond to the large open reading frame. ▲ ; As pointed out by Morgan, Re and Kingsbury (4), the nucleotide sequence of the NP gene reported in our previous paper (2) had an error in that the dinucleotides 3'-AG at positions 1,581 and 1,582 should be a mononucleotide -C. Thus, the new numbering is used in this paper. (This correction resulted in shortening by one nucleotide in the NP gene and both 7 amino acid shortening and 30 amino acid replacement at the C-terminal region of the NP protein. This has, however, little effect on our conclusions.)

employing a cDNA fragment from positions 4,352 to 4,477 as primer (PR2), which was prepared by digesting clone R54 with Alu I, and we obtained cDNA clones 2013, 2054 and 2248. Because the procedure included the digestion

## M protein

```

      10      20      30      40      50      60
MADIYRPFKF SYEDNGTVEP LPLRTGPKK AIPHIRIVKV GDPPKHGVRY LDLLLLGFFE 60
TPKQTTNLGS VSDLTEPTSY SICGSGSLPI GVAKYYGTDQ ELLKACTDLR ITVRRTVRAG 120
EMIVYMVDSI GAPLLPWSGR LRQGMIFNAN KVALAPQCLP VDKDIRLRVY FVNGTSLGAI 180
TIAKIPKTLA DLALPNSISV NLLVTLKTGI STEQKGVLPV LDDQGEKKLN FMVHLGLIRR 240
KVGKIYSVEY CKSKIERMRL IFSGLIGGI SFHVQVNGTL SKTFMSQLAW KRAVCFPLMD 300
VNPMMNVIW AASVEITGVD AVFQPAIPRD FRYYPNVVAK NIGRIRKL (348)

```

```

ALA= 20 5.75%  ASN= 13 3.74%  ASP= 17 4.89%  ARG= 21 6.03%
CYS= 5  1.44%  GLN= 9  2.59%  GLU= 11 3.16%  GLY= 27 7.76%
HIS= 5  1.44%  ILE= 27 7.76%  LEU= 38 10.92% LYS= 24 6.90%
MET= 10 2.87%  PHE= 13 3.74%  PRO= 22 6.32%  SER= 20 5.75%
THR= 20 5.75%  TRP= 3  .86%   TYR= 11 3.16%  VAL= 32 9.20%

```

M.W. = 38,557 dalton

## F protein

```

      10      20      30      40      50      60
MTAYIORSQC ISTSLLVLT TLVSCOIPRD RLSNIGVIVD EGKSLKIAGS HESRYIVLSL 60
+++  +  +++++  ++AAA
VPGVDFENG  GTAQVIQYKS LLNRLLIPLR DALDLQEALI TVTNDTTQNA GAPOSRRFFGA 120
VIGTIALGVA TSAQITAGIA LAEAREAKRD IALIKESMTK THKSIELLQN AVGEQILALK 180
TLQDFVNDIE KPAISELGCE TAALRLGIKL TQHYSELLTA FGSNFGTIGE KSLTLQALSS 240
LYSANITEIM TTIRTGQSN I SDVIYTEQIK GTVIDVDLER YMVTLSVKIP ILSEVPGVLI 300
HKASSISYNI DGEEWYVTAP ----

```

Fig. 3 The predicted amino acid sequence, amino acid composition and molecular weight of the M protein as well as the predicted NH<sub>2</sub>-terminal sequence of 320 amino acids of the F protein. Each small arrow indicates a putative cleavage site of the signal peptide. A plus sign indicates a hydrophobic residue within the signal peptide. A large arrow indicates the proposed tryptic cleavage site of the F protein into the F<sub>1</sub> and F<sub>2</sub> proteins. As to the underlining in the F protein, see the text.

with Hind III, the fragment PR2 was removed from these clones. It was, therefore, necessary to confirm the boundary sequence between R54 cDNA and 2013 cDNA and this was carried out by directly sequencing the single stranded cDNA synthesized by another primer extension using a 5'-labeled primer (PR3)

The nucleotide sequence from position 3,606, the 5' end portion of the P+C gene, to 5,824 is shown in Fig. 2. Within this sequence, we found the sequences of 3'-AUUCUUUUU-GAA-UCCACUUUC-5' and 3'-AUUCUUUUU-GAA-UCCCUUUUC-5' at positions 3,624 and 4,800, respectively. These sequences

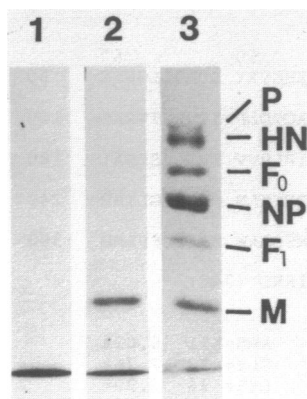


Fig. 4 In vitro translation of the mRNA hybrid-arrested by R54 cDNA. The mRNAs were prepared from uninfected SRCDF<sub>1</sub>-DBT cells (lane 1) or Sendai virus-infected cells (lane 2), and the hybrid-arrested mRNAs were translated with rabbit reticulocyte lysates in the presence of <sup>35</sup>S-methionine. The translation products were immunoprecipitated with a rabbit hyperimmune serum against Sendai virus. <sup>3</sup>H-leucine-labeled viral proteins (lane 3) were used as markers which were prepared from Sendai virus-infected MDBK cells by means of the same immunoprecipitation as above. These samples were analysed by polyacrylamide gel electrophoresis in the presence of SDS using a 7.5% gel.

are identical with the consensus sequence of R2-GAA-R1 reported in the previous paper, in which R1 was considered as the recognition site of RNA polymerase, R2 as the polyadenylation signal, and GAA as the intergenic sequence (2,3,4,20,21). Since we defined the NP and P+C genes as the sequences flanked by R1 and R2 at their 3' and 5' ends, respectively, we assigned, in a similar manner, the sequence from R1 starting at 3,636 to R2 ending at 4,808, as the third gene, and that beginning from 4,812, the start of another R1, as the 3' proximal part of the fourth gene.

In the third gene only one large open reading frame consisting of 1,044 nucleotides (348 amino acids) was detected (Fig. 2). The amino acid sequence (Fig. 3) deduced from this reading frame indicated that the molecular weight of the gene product is 38,557 dalton which is very close to the estimated molecular weight of the Sendai virus M protein, 34,000 (22). The amino acid composition of this sequence mostly resembles the estimated amino acid composition of the M protein which we previously reported (2) and is very different from the estimated amino acid compositions of other viral proteins. Four additional small open reading frames consisting of 27 to 58 amino acids could be assigned in this gene, although their physiological significance remains to be elucidated.

In order to confirm that this gene really encodes the M protein, in vitro translation of hybrid-arrested mRNA was performed. The results (Fig. 4) revealed that mRNA specifically selected by R54 cDNA was translated into the M protein. Furthermore, the R54 cDNA hybridized to the smallest viral monocistronic mRNA, i.e., M-mRNA, among the monocistronic viral mRNAs transferred by the Northern method (23) to nitrocellulose paper from an agarose gel (2) (data not shown). From these observations we concluded that

the third gene is the M gene. This conclusion is in good agreement with the report by Dowling *et al.* (1).

Although it has not been determined yet what amino acid residue is at the very end of the N-terminal of the M protein, the deduced amino acid sequence indicated that 1) the Sendai virus M protein is very rich in basic amino acid residues compared with acidic residues, 2) it has no signal polypeptide, supporting the previous suggestion (24,25) that the M protein is synthesized on free polysomes and not on membrane-bound polysomes, and 3) it has hydrophobic domains near the C-terminal.

The M proteins of paramyxoviruses, orthomyxoviruses and rhabdoviruses are supposed to play a crucial role in virus maturation; they comprise the lining of the inner surface of the viral envelope, on one hand, and show affinity to both nucleocapsid and viral glycoproteins, on the other hand (26,27,28,29). Therefore, we expected to find a constant distribution pattern of hydrophobic domains as well as that of basic domains among the structures of the M protein of Sendai virus, respiratory syncytial virus (RSV) (6), vesicular stomatitis virus (VSV) (30), influenza A (31,32) and B viruses (33). And we also expected the appearance of some conserved amino acid sequences among them. However, we could detect any homology neither in their primary structures nor in their secondary structures (34), although these proteins are rich in basic amino acid residues in common. Thus, the analysis of tertiary structures of these proteins is necessary to obtain molecular insight into the functions of the M protein.

In the fourth gene beginning at position 4,812, we found a large open reading frame, which starts at 4,865 and is still in the middle. When this sequence was converted into the amino acid sequence, we could detect a sequence of 20 amino acids (Fig. 3, underlined) starting from position 5,213 which is identical with that of the reported N-terminal 20 amino acids of the Sendai virus F<sub>1</sub> protein (35,36,37), except that the 16th residue of this sequence is Ser instead of Ala. Similar sequences to this were also reported for the N-terminal of F<sub>1</sub> proteins of Newcastle disease virus and SV5 (37). On the other hand, the sequence of about 30 amino acids in the N-terminal region of the open reading frame may be assigned as a signal peptide, since it is hydrophobic (plus signs in Fig. 3) and fulfils the format proposed by Watson (38) for a signal peptide, and each peptide bond in the sequence of Ser-Cys-Gln-Ile at amino acid positions 24 to 27 (small arrows in Fig. 3) can be a cleavage site for signal peptidase according to von Heijne (39). The presence of a signal peptide is expected to be a

characteristic of the F protein, it being synthesized on membrane-bound polysomes (24,25). Thus, we concluded that this large open reading frame encodes the F protein, that is, the fourth gene is the F gene.

It is very plausible that the arginine residue (a large arrow in Fig. 3) just N proximal to the sequence corresponding to F<sub>1</sub> is the site of the tryptic cleavage of the F protein into F<sub>1</sub> and F<sub>2</sub> proteins (40,41), and the resulting N-terminal portion without the signal peptide, consisting of 90 to 92 amino acids, corresponds to the F<sub>2</sub> protein. This assumption is supported by the facts that the N-terminals of both the F and F<sub>2</sub> proteins are blocked (35,36), the reported amino acid composition of the F<sub>2</sub> protein (36) is quite similar to that calculated from the deduced amino acid sequence, and the reported chain length of the F<sub>2</sub> protein is about 90 residues (36). From these facts, it seems reasonable to assign the sequence of the peptide within the F protein as NH<sub>2</sub>-signal peptide-F<sub>2</sub>-F<sub>1</sub>-COOH, supporting the previous proposals by Scheid and Choppin (35) and Gething *et al.* (36). From the amino acid sequence of F<sub>2</sub> protein, it is suggested that Cys at position 70 is the site of the disulfide bond between F<sub>1</sub> and F<sub>2</sub> proteins (35) and that Asn at position 104 is the potential glycosylation site, because 1) the F<sub>2</sub> protein is glycosylated (35), 2) all the glycoproteins of Sendai virus carry exclusively Asn-linked sugar chain (42) and 3) Asn-X-Thr or Asn-X-Ser has been known as the most frequent glycosylation sequence for Asn-linked sugar chain (43).

Further sequencing of the Sendai virus genome toward its 5' end is in progress.

### **ACKNOWLEDGEMENT**

This work was supported by grants from the Ministry of Education, Science and Culture of Japan.

\*To whom correspondence should be addressed

### **REFERENCES**

1. 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.
2. Shioda, T., Hidaka, Y., Kanda, T., Shibuta, H., Nomoto, A. and Iwasaki, K. (1983) Nucleic Acids Res. 11, 7317-7330.
3. Giorgi, C., Blumberg, B.M. and Kolakofsky, D. (1983) Cell 35, 829-836.
4. Morgan, E.M., Re, G.G. and Kingsbury, D.W. (1984) Virology 135, 279-287.
5. Venkatesan, S., Elango, N. and Chanock, R.M. (1983) Proc. Natl. Acad. Sci. USA. 80, 1280-1284.
6. Satake, M. and Venkatesan, S. (1984) J. Virol. 50, 92-99.



7. Billeter, M.A., Baczkowski, K., Schmid, A. and ter Meulen, V. (1984) *Virology* 132, 147-159.
8. Bellini, W.J., Englund, G., Richardson, C.D. and Rozenblatt, S. (1984) *J. Virol.* 50, 939-942.
9. Glazier, K., Raghoebar, R. and Kingsbury, D.W. (1977) *J. Virol.* 21, 863-871.
10. Lamb, R.A. and Lai, C.-J. (1980) *Cell* 21, 475-485.
11. Bolivar, F. and Backman, K. (1979) *Methods Enzymol.* 68, 245-267.
12. Grunstein, M. and Hogness, D.S. (1975) *Proc. Natl. Acad. Sci. USA.* 72, 3961-3965.
13. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA.* 74, 5463-5467.
15. Messing, J. (1983) *Methods Enzymol.* 101, 20-78.
16. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* 18, 5294-5299.
17. Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA.* 69, 1408-1412.
18. Parnes, J.R., Velan, B., Felsenfeld, A., Ramanathan, L., Ferrini, U., Appella, E. and Seidman, J.G. (1981) *Proc. Natl. Acad. Sci. USA.* 78, 2253-2257.
19. Pelham, H.R.B. and Jackson, R.J. (1976) *Eur. J. Biochem.* 67, 247-256.
20. Gupta, K.C. and Kingsbury, D.W. (1982) *Virology* 120, 518-523.
21. Gupta, K.C. and Kingsbury, D.W. (1984) *Nucleic Acids Res.* 12, 3829-3841.
22. Homma, M., Tozawa, H., Shimizu, K. and Ishida, N. (1975) *Japan. J. Microbiol.* 19, 467-470.
23. Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. USA.* 77, 5201-5205.
24. Lamb, R.A. and Choppin, P.W. (1977) *Virology* 81, 371-381.
25. Nagai, Y., Yoshida, T., Hamaguchi, M., Nagura, H., Hasegawa, H., Yoshimura, S. and Watanabe, K. (1983) *Microbiol. Immunol.* 27, 531-545.
26. McSharry, J.J., Compans, R.W. and Choppin, P.W. (1975) *Virology* 67, 365-374.
27. Shimizu, K. and Ishida, N. (1975) *Virology* 67, 427-437.
28. Yoshida, T., Nagai, Y., Yoshii, S., Maeno, K., Matsumoto, T. and Hoshino, M. (1976) *Virology* 71, 143-161.
29. Yoshida, T., Nagai, Y., Maeno, K., Iinuma, M., Hamaguchi, M., Matsumoto, T., Nagayoshi, S. and Hoshino, M. (1979) *Virology* 92, 139-154.
30. Rose, J.K. and Gallione, C.J. (1981) *J. Virol.* 39, 519-528.
31. Winter, G. and Fields, S. (1980) *Nucleic Acids Res.* 8, 1965-1974.
32. Lamb, R.A. and Lai, C.-J. (1981) *Virology* 112, 746-751.
33. Briedis, D.J., Lamb, R.A. and Choppin, P.W. (1982) *Virology* 116, 581-588.
34. Chou, P.Y. and Fasman, G.D. (1978) *Adv. Enzymology* 47, 45-148.
35. Scheid, A. and Choppin, P.W. (1977) *Virology* 80, 54-66.
36. Gething, M.J., White, J.M. and Waterfield, M.D. (1978) *Proc. Natl. Acad. Sci. USA.* 75, 2737-2740.
37. Richardson, C.D., Scheid, A. and Choppin, P.W. (1980) *Virology* 105, 205-222.
38. Watson, M.E.E. (1984) *Nucleic Acids Res.* 12, 5145-5164.
39. von Heijne, G. (1983) *Eur. J. Biochem.* 133, 17-21.
40. Homma, M. and Ohuchi, M. (1973) *J. Virol.* 12, 1457-1465.
41. Scheid, A. and Choppin, P.W. (1974) *Virology* 57, 475-490.
42. Yoshima, H., Nakanishi, M., Okada, Y. and Kobata, A. (1981) *J. Biol. Chem.* 256, 5355-5361.
43. 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.