
Nucleotide sequence of the bovine parainfluenza 3 virus genome: its 3' end and the genes of NP, P, C and M proteins

Yuko Sakai, Shinya Suzu, Tatsuo Shioda and Hiroshi Shibuta*

Department of Viral Infection, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108, Japan

Received December 31, 1986; Revised and Accepted March 12, 1987

Accession no. Y00114

ABSTRACT

We present the nucleotide sequence of bovine parainfluenza 3 virus (BPIV3) genome from its 3' end to the opening region of the F gene, through the NP, P plus C, and M genes. Comparison of the sequence with those reported for other paramyxoviruses indicated that BPIV3 was most similar to human parainfluenza 3 virus (HPIV3), and also very similar to Sendai virus in the structural make-up of its genome and the amino acid sequences of its gene products, suggesting that these three viruses constitute a paramyxovirus subgroup from which Newcastle disease and measles viruses are separable. In BPIV3 and Sendai virus, the NP and M proteins, the main structural elements, were more highly conserved than the functionally important P and C proteins. This tendency was also observed even in BPIV3 and HPIV3. Virus-specific amino acid sequences of the NP and M proteins were found at the carboxyl and amino terminal regions, respectively. BPIV3 M mRNA was found to have aberrations in its poly A attachment site.

INTRODUCTION

Paramyxoviruses have as their genome single stranded non-segmented RNA of negative polarity. Recent rapid progress in gene analysis of several paramyxoviruses will undoubtedly reveal structures specific to each virus, as well as some common to the group, and cast light on the evolutionary relationships among the paramyxoviruses, the molecular structure of functional and antigenic domains of the viral proteins, mode of transcription and replication, and molecular events that determine the host-range specificity of the viruses.

We have been investigating the genome structures of Sendai virus (murine parainfluenza 1 virus) (1, 2, 3) and bovine parainfluenza 3 virus (BPIV3), both belonging to the paramyxovirus group. Parainfluenza 3 virus is known on the basis of serological studies to consist of human- (HPIV3) and bovine-specific viruses; They are antigenically distinguishable from each other, although they strongly share common antigenic properties (4, 5). Very recently, reports were presented on the nucleotide sequences of HPIV3 genes, other than those of the membrane (M) and RNA polymerase (L) proteins, as

well as its 3' terminal nucleotide sequence (leader region) (6, 7, 8, 9, 10, 11, 12). The complete genome structure of Sendai virus has already been elucidated (1, 2, 3, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22).

To date, we have determined the complete nucleotide sequence of the 3' proximal two thirds of the BPIV3 genome where, in sequence from the 3' terminus, are located the leader region and the genes of the nucleocapsid (NP), polymerase-associated (P) and nonstructural C, M, fusion (F) and hemagglutinin-neuraminidase (HN) proteins. In this paper we present the complete nucleotide sequence from the 3' end of viral genome RNA to the opening region of the F gene, and compare the amino acid sequences of the products of the NP, P+C and M genes with those of HPIV3 (6, 7, 8), Sendai virus (1, 2, 13, 14, 15, 16), measles virus (23, 24, 25) and Newcastle disease virus (NDV) (26). In the accompanying paper, we will present the nucleotide sequence of the entire F and HN genes.

MATERIALS AND METHODS

Viruses and preparation of viral RNA

Complementary DNAs (cDNAs) of viral genome RNA were constructed from the 910N strain of BPIV3, and cDNAs of mRNA from the 910N, M and SC strains. These viral strains were described in detail previously (27, 28, 29, 30), and in this study were used at the second passage level after three successive single plaque isolations.

The 50S viral genome RNA of the 910N virus was purified from virions purified from culture fluids of infected Madin-Darby bovine kidney (MDBK) cells, as described previously (31). For preparation of the mRNAs of M and SC viruses, trypsin-monodispersed mouse SRCDF1-DBT (1) cells in a small volume were infected with the M or SC virus at 20 or more plaque-forming units (PFU) per cell in the presence of 20 µg/ml DEAE-dextran, were incubated at 37 C for 1 hr, sedimented by low speed centrifugation, suspended in Eagle minimum essential medium (MEM) containing 5 % fetal calf serum (FCS) and 5 % tryptose phosphate broth, and were dispensed into plastic dishes (9 cm in diameter) at a concentration of 2×10^7 cells per dish. After incubation at 37 C for 2 hr, the medium was replaced with that containing 3 µg/ml actinomycin D, and the cells were further incubated at 37 C for 18 hr. For preparation of 910N virus mRNAs, human embryonal lung R66 cell monolayers (29) were infected with the virus at 20 PFU/cell, and were incubated at 37 C in MEM containing 2 % FCS for 12 hr, followed by the same medium containing 20 µg/ml actinomycin D for an additional 6 hr. The

infected cells were lysed with guanidine thiocyanate for RNA extraction (32), and the viral mRNAs were selected from crude RNA by oligo-(dT) cellulose column chromatography.

Synthesis and cloning of cDNA

We previously found that treatment of the 50S Sendai virus genome RNA with poly A polymerase at about 20-fold the amount used for the standard polyadenylation reaction resulted in partial digestion of the RNA to the average size of 28S which served satisfactorily as the template for oligo(dT)-primed cDNA synthesis by the method of Okayama and Berg (33, 34), and we obtained a good number of cDNA clones starting from multiple sites of the genome RNA (3). This suggested that ribonuclease and phosphatase activities contaminated the poly A polymerase preparation. In the present study, amount of the poly A polymerase was somewhat reduced, and 16 µg of the 50S 910N virus genome RNA was incubated with 2 units of poly A polymerase at 37 C for 10 min. Synthesis and molecular cloning of cDNAs from the enzyme-treated RNA were performed by the method of Okayama and Berg using a pcDV1-derived vector primer (33). All the cDNA clones thus far examined were revealed to have poly A tail. CDNAs of mRNA were also synthesized and cloned by the method of Okayama and Berg using both pcDV1-derived (33) and pSV7186-derived (34) vector-primers. *E. coli* K12 strain HB101 was transformed with the resulting recombinant plasmids (35).

Sequence determination of cDNA

CDNAs were cleaved with appropriate restriction endonucleases into fragments, which were sequenced mainly by the dideoxy method of Sanger et al. (36) using the M13 phage system (37). In some experiments the chemical cleavage method of Maxam-Gilbert (38) was also employed.

Colony hybridization

Bacterial colonies carrying recombinant plasmids were screened for viral cDNA inserts with the viral 50S RNA probe, and for overlapping cDNA inserts with appropriate cDNA probes, as described previously (3).

In vitro translation of viral mRNA

The 910N virus mRNAs prepared as above were hybrid-arrested (39) with viral cDNA and translated in vitro with rabbit reticulocyte lysates (40). Translational products were analysed by sodium dodecylsulfate (SDS)- 7.5 % polyacrylamide gel electrophoresis (SDS-PAGE) using Maizel's system (41).

In vitro transcription

Appropriate fragments cut from cDNAs were subcloned into pGEM vectors downstream of the promoter for bacteriophage SP6 RNA polymerase and

transcribed in vitro into RNAs in the presence of m⁷GpppG (42). The RNAs were subsequently translated in vitro as above.

Enzymes and other materials

ATP:RNA adenylyltransferase, 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, the rabbit reticulocyte lysate kit, and all the radioactive compounds from Amersham International plc, Amersham, England; pGEM3 and pGEM4 vectors from Promega Biotec, Wisconsin, U.S.A.; SP6 RNA polymerase from Boehringer, West Germany; 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.

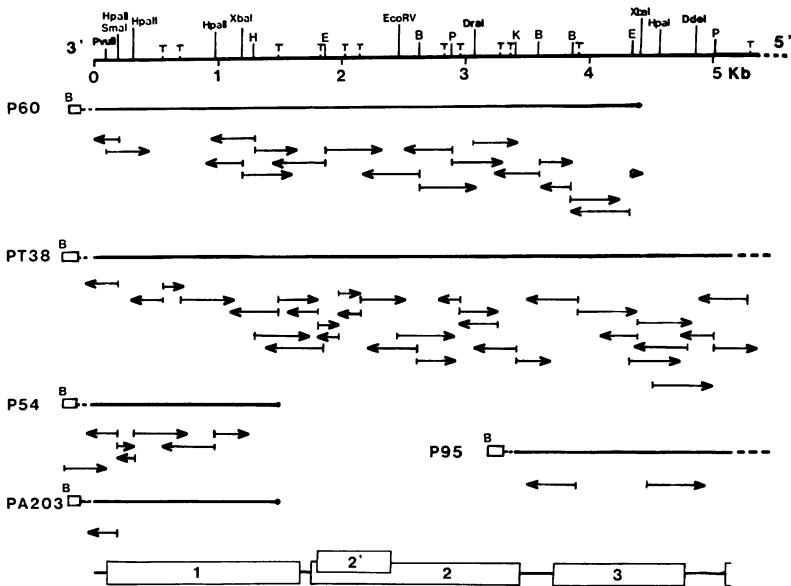


Fig. 1. Locations of cDNA clones (P60, PT38, P54, PA203 and P95) in the genome RNA and the sequencing strategy. The numbers in kilobases (Kb) under the top line show nucleotide positions from the genome 3' end. Restriction map is also presented on the top line (T; TaqI, H; HindIII, E; EcoRI, B; BamHI, P; PstI, and K; KpnI). Arrows indicate the cDNA fragments sequenced and the direction of sequencing. Dashed line and small box at the 3' end of each cDNA clone represent the poly A tail and the vector sequence from the BamHI site, respectively. Positions of the open reading frames of the 1st, 2nd and 3rd genes are denoted in the bottom.

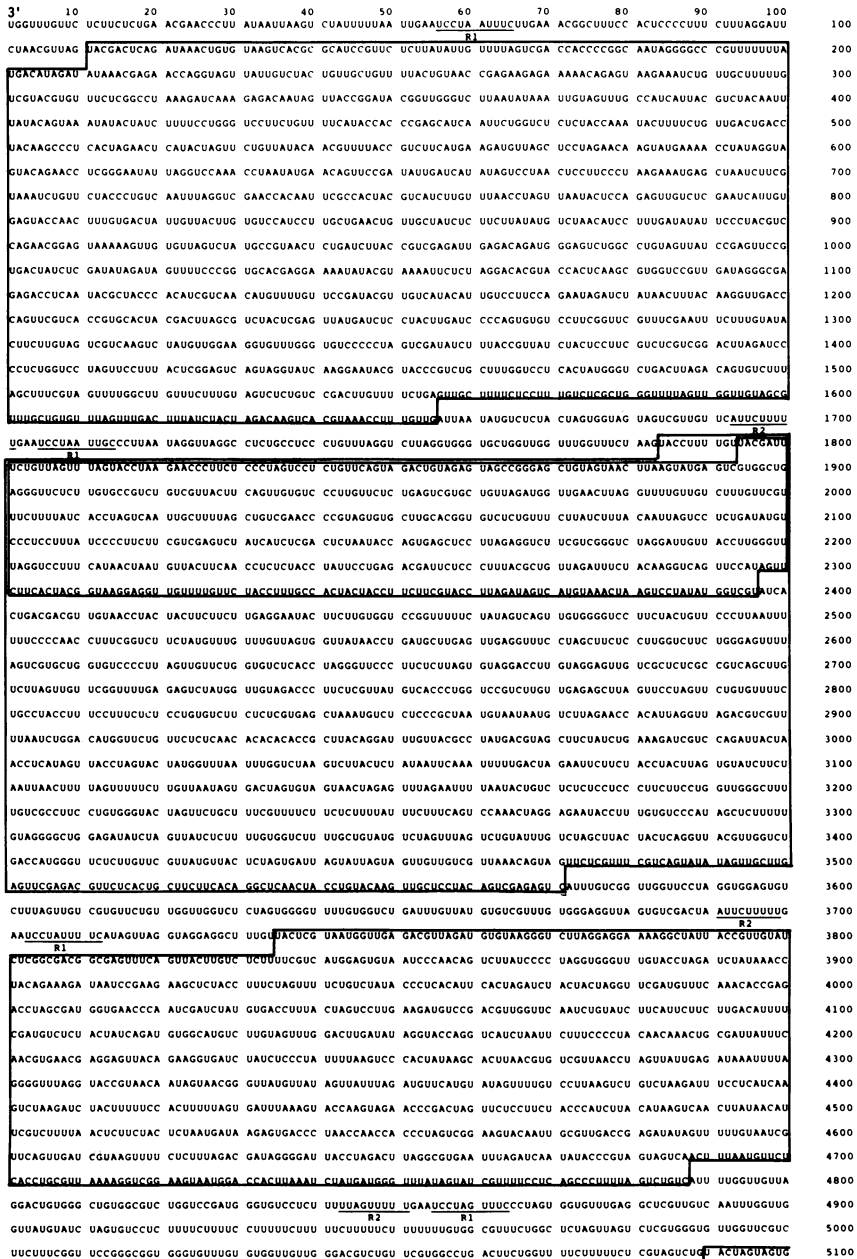


Fig. 2. The RNA sequence from the 3' end of the BPIV3 genome (910N strain), expressed in negative sense. Enclosed sequences correspond to the open reading frames. R1 and R2 are the repeating consensus sequences.

RESULTS

Sequencing of nucleotides from the 3' end of the viral genome

Of a total of 872 cDNA clones which were constructed from 8 µg of 910N virus genome RNA, 117 clones were strongly positive to the viral genome RNA probe in the colony hybridization test. Clones P60 and PT38 contained large inserts, about 4.5 and 7 kilobases long, respectively, and it was suggested by digestion patterns with several restriction endonucleases that the former was part of the latter. Determination of the nucleotide sequences according to the strategy illustrated in Fig. 1 actually revealed that both clones initiated from the same position, since the sequences immediately following the poly A tail were identical in the two clones. Analyses of two additional clones, P54 and PA203 which were selected according to their digestion patterns with appropriate restriction endonucleases, indicated that they too started from the same position as P60 and PT38; these 4 clones showed an identical nucleotide sequence next to the poly A tail. Moreover, except for 3 the sequence of the first 60 nucleotides was identical with that reported for the 3' end of HPIV3 genome by direct analysis of the RNA (11), and the first 14 were identical with the 3' end 14 nucleotides of the Sendai virus genome (1). Thus, we concluded that these 4 clones were cDNA copies that originated from the very 3' end of the 910N virus genome RNA. Another clone p95, which was selected by colony hybridization using a cDNA probe cut from the PT38 right region non-overlapping the P60, was also used as a sequencing partner.

Figure 2 shows the RNA sequence of about 5,000 nucleotides from the 3' end of the viral genome deduced from cDNA analysis. In this region, three large open reading frames composed of more than 1,000 nucleotides were detected (indicated by boxes in the figure). Each of these open reading frames was flanked on both sides by non-coding sequences, and further flanked upstream by the 3' consensus sequence of R1 (UCCUNNUUNC), and downstream by the 5' consensus sequence R2 (NUNUUUUU). This feature of the BPIV3 genome closely resembled those found in the genomes of Sendai virus (1, 2, 3, 22), HPIV3 (12, 26) and other paramyxoviruses, indicating that the region starting from the top of the R1 sequence and terminating at the end of the R2 sequence was a gene. It was also common to both BPIV3 and Sendai virus for the genes to be connected with the trinucleotide GAA, forming R2-GAA-R1 (1, 2, 3, 22). Thus, the first, second and third genes were composed of 1,646, 1,995 and 1,149 nucleotides, respectively.

As observed with Sendai virus (1,13), HPIV3 (8), NDV (26) and measles

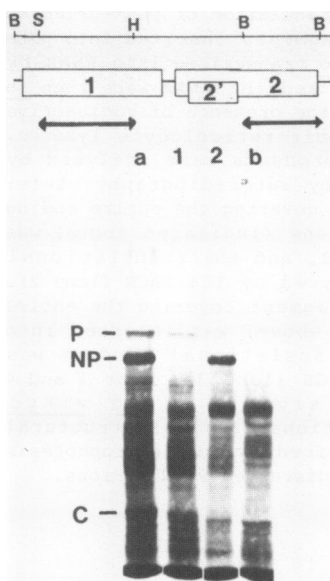


Fig. 3. *In vitro* translation of mRNAs. The mRNAs were prepared from BPIV3 910N strain-infected R66 cells (lane a) and hybrid-arrested by a cDNA fragment from the first gene (lane 1) or by that of the second gene (lane 2), or prepared from uninfected cells (lane b). Translation products were labeled with radioactive methionine and, without immunoprecipitation, analysed by SDS-PAGE. The cDNA fragments used for hybrid-arrest are shown by arrow-lines, and the restriction map (B; BamHI, S; SmaI and H; HindIII) and positions of the open reading frames are illustrated in the upper part.

virus (24), a small overlapping reading frame consisting of 603 nucleotides was detected within the large reading frame in the 2nd gene.

The sequence of the 3' proximal 55 nucleotides upstream of the first gene was U-rich, and terminated with the trinucleotide GAA. This region most probably corresponded to the leader region known in Sendai virus (43).

Identification of the genes

Products of *in vitro* translation of hybrid-arrested mRNAs were analysed by SDS-PAGE (Fig. 3). The profile for 910N viral proteins in SDS-PAGE was previously described (31, 44).

A cDNA fragment (Fig. 3) from the first gene arrested the NP mRNA, resulting in disappearance of the NP protein from the electrophoresis gels. This was clearly recognizable, regardless of whether or not immunoprecipitation of translation products by hyperimmune antiserum against the 910N virus was employed. The antiserum was raised in rabbits by multiple intravenous doses of the purified virus (27). When the mRNAs were hybrid-arrested with a cDNA fragment of the second gene (Fig.3), the P protein as well as a smaller protein of around 20K could no longer be detected. This observation was confirmed by *in vitro* transcription. The cDNA fragment between the HapII site at nucleotide position 1,729 and the PvuII site at 3,685, obtained from clone PT38, inserted into pGEM3 transcription vector cut by AccI and SmaI. This fragment covered the entire large reading frame of the second gene,

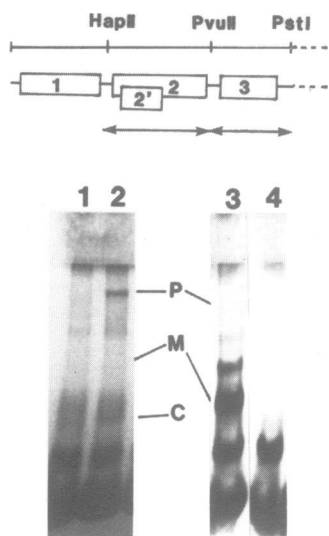


Fig. 4. *In vitro* translation of pGEM-oriented RNAs. Viral cDNA fragments inserted into pGEM vectors were *in vitro* transcribed into RNAs by SP6 RNA polymerase and the RNAs were then *in vitro* translated in the presence of radioactive methionine with rabbit reticulocyte lysates. The translational products were analysed by SDS-PAGE followed by autoradiography. Left; HapII-PvuII fragment covering the entire coding frames of the 2nd gene (indicated above) was inserted into pGEM3, and the translational products were analysed by 10% PAGE (lane 2). Right; PvuII-PstI fragment covering the entire 3rd gene (indicated above) was inserted into pGEM4, and the translational product was analysed by 7.5% PAGE (lane 3). Lane 1 and 4 show background signals of *in vitro* translation. Positions of viral structural proteins were determined by co-electrophoresis of purified and SDS-disrupted 910N virions.

although it lacked the 25 nucleotides of the 3' proximal non-coding region. The RNA transcribed *in vitro* by SP6 RNA polymerase produced the P protein and a small protein upon *in vitro* translation (Fig. 4). Since the small protein could not be detected in virions and had a molecular weight of about 20K, and mainly on the basis of analogy with Sendai virus, it seemed highly probable that this protein was the viral nonstructural C protein encoded by the small overlapping reading frame. However, this remains to be substantiated. It is also to be examined whether or not another nonstructural C' protein, starting from the second methionine codon 129 nucleotides downstream of the first codon of the presumed C protein, exists, as observed for Sendai virus (45).

The *in vitro* transcription system was also effective for identification of the third gene product. The cDNA fragment of clone PT38 between the PvuII site (3,685) and the PSTI site (5,046), which covered the entire third gene, was inserted into pGEM4 vector cut by SmaI and PstI. The RNA transcribed from the inserted cDNA fragment clearly produced the M protein in *in vitro* translation (Fig. 4).

Thus, the first and third genes were identified as the NP and M genes, respectively, while the second gene was regarded as the P+C gene, the large and small reading frames encoding the P and C proteins, respectively. As presented in the accompanying paper, we identified the genes following the M gene toward the 5' end as the F and then HN genes. Therefore, the gene order

of BPIV3 was found to be the same as those of other paramyxoviruses for which gene orders have been analysed to date (3, 12, 46, 47, 48, 49).

Comparison of genome structures of BPIV3 and other paramyxoviruses

The 3' proximal 55 nucleotides of the BPIV3 genome were compared with other paramyxoviruses. Mismatch between BPIV3 and HPIV3 (11) was only 7 nucleotides, thus showing 87 % nucleotide homology. That between BPIV3 and Sendai virus (1) was 67 %, while any combination of BPIV3, measles virus (50) and NDV (51), or of Sendai virus, NDV and measles virus showed homologies less than 40 %. However, the first eight nucleotides were completely conserved by all these viruses, suggesting that they may be an essential sequence in the paramyxovirus genome.

The structural make-up of the NP gene was identical in BPIV3 and HPIV3. The 3' non-coding region including R1, the coding frame and the 5' non-coding region including R2 consisted of 55, 1,545 and 46 nucleotides, respectively, in both viruses, while those of Sendai virus consisted of 64, 1,551 and 66, respectively.

The presence of the overlapping P+C gene is regarded as one of the most characteristic features of paramyxoviruses. However, the similarity in structural make-up of the overlaps of BPIV3, HPIV3 (8) and Sendai virus (1, 13) was striking: the opening methionine codon of the smaller reading frame was located 7 nucleotides downstream of that of the larger reading frame. This structural make-up differed considerably from that of measles virus (24), in which the initiation codon of the C protein is located 19 nucleotides downstream of that of the P protein. However, the numbers of nucleotides in the 3' non-coding region including R1, the P reading frame, C reading frame and the 5' non-coding region including R2 were different in those viruses, being 79, 1788, 603 and 128, respectively for BPIV3, 79, 1845, 597 and 95 for HPIV3 (8), and 103, 1704, 612 and 86 for Sendai virus (1, 13).

As for the M gene, the non-coding regions and reading frames of BPIV3 and Sendai virus were of different lengths.

The gene-starting consensus sequences (here named R1) and the gene-terminating consensus sequences (here named R2, and which is presumably the poly A signal) of the NP, P+C and M genes were also found in the F and HN genes, as presented in the following paper. The R1 sequence was UCCUAAUUUC for NP, UCCUAAUUGC for P+C, UCCUAAUUUC for M, UCCUAGUUUC for F and UCCUUGUUUC for HN. Thus, the R1 of BPIV3 could be expressed as 3'-UCCUNNUUNC-5' which is most similar to 3'-UCCUNNUUNC-5' of HPIV3 (12), and also very similar to 3'-UCCNNUUNC-5' of Sendai virus (1, 2, 3, 22).


```

BPIV3  MENNAKDQIMDS WEEGSGDKSSDISSALDIIEFILSTDSQENTADSN  EVNTGNK
HPIV3  ..SD..NY.....EPR..*N.....P..DLS*ND  T*...S
Sendai  .*QD..FILKE ..EV.REAPGGRES*.DV*GF**A*...P*P*IGG.RSWLHNT*..PQG
Measles  .AEQARHVKNGLERICALKAEPIG .L.*EAMA*AW.EI.DNPGQ*RATCRE.KAGSSG

BPIV3  RLSTTIYQLESKTTETSKEGSGVNNRQLGASHERATETK  NRRNVQETIQGGN
HPIV3  Q..A..C.P.I.P...EKV...TDK...S.S...CT..A.  D..*D...*...S
Sendai  PG.AHRAKS.GEG .V. *P.TQ*..SGEE.RVSGRT*.PEAEAHAG.*DKQN.HRAF
Measles  ..KPLCSAIG *..GGAPRIRGQPGESDDDAETLGI*P*NLQASSTGLQCY*YDHS

BPIV3  RGRSSSDSRAEIMVTRGISRSSPDNNGTQIQESIDYNEVGEMDKDSAKREMRQSKDVPV
HPIV3  GR.....TV.*G...G.IT.SK...NT.N...L.*RK.....IE.K...A...S
Sendai  G..*G*N.VSQDLGDG.D.GILEN.P.ERGYPR.GIED.NR..AAHPD..GED.AEG*.E
Measles  GEAVKGIQD.*SIMVQSGLDGDSLTS.G.DNES.NS.V **...P.T*GYAITD.G.AP*SM

BPIV3  KVSRSDAIPPTKQDNGDDGRSMESISTFDSGYTSIVTAATLDDEEELMKNTRPKRYQS
HPIV3  E*.G..V.FT.E.SR.S.H...L.P...P.*RSM.*...P*...*...*...M.*SS.
Sendai  E.R.GT.S*DEGEG.ASN...PG.*HSARV.G**VIP*P*L..A*.R*.K.RPTNSG
Measles  GFRA..VE.TAEGGEIH*LL. LQ.RGNFPKLGK*TVNPPPP.PGRASTSG.PI.*GTE

BPIV3  TPQEDDKGIKKGVG KPEDTNKQSPILDYELNSKSGSKRNQKTLK IS  TTTGES TRPQ
HPIV3  .H....R....G.G.GK.WF.K.RDT.NQTS*SDH.PTS.GQ.K..KT...NTD.*G.
Sendai  *KPLTPAT*PGTRSPLNRY.STGSPPGKPPS*QDEHI.SGDTPA*RVKDRKPPIG..
Measles  RRLASFGTEIAS*LTGGATQCARKSPS*PSGPGAPAGNVEPC*SNAALIQEWTPESGTTI

BPIV3  SGSQ  GKRTISWNILNSESGSRTESTSQNSQIPTSGKSNVTGPGRTTLESRIKTKQT
HPIV3  *E*.TESSETQSP...P.IDNNTD...Q..TTPPTT.PKS*R.KESI..NS..*P....
Sendai  ..V*DC.ANGRSIHPGLETD*
Measles  .PRSQNNEEGDYDDE.F.*VQDIKTALAKIHEDNQKII.KLESLLLK.G.VESIKKQI

BPIV3  DGKEREDTEESTRFTERAITLLQNLGVIQSAAKLDLYQDKRVVVCVANLNNADTASKIDF
HPIV3  I....K.....N.....TS.....V.....*.....*.....*.....
Sendai  TK.GIG*NTS. MK.M. ...TS.....QEF*SSR.ASY.FARRA.KS.NY.EMTFN
Measles  NRQNISI*TLEGLH*SIM. A*PG..KDPNDPTA.*EINPD*KP*IGRDSGRAL.EV*KK

BPIV3  LAGLMIG  VSMHDHTKLNQIQNEILSLKTDLKKMDESHRRIENQKEQLSLITS
HPIV3  ....V... ..N.....I.....M.N..A...*.....*.....*.....
Sendai  *C..I*SAEKSSARKVD*NKQL.K...ES*E.F*DTY.* FS.Y...N..*M.
Measles  PVASRQLQGMNTNGRT.SRG QL.KEF.LKP*GK.MSSAVGFVPDTGPAASRVIR .*K.

BPIV3  LISNLKIMTERGGKKDQENSGRTPMIKTKAKEEKIKKVRPDLMETQGIKKNIPDLYRS
HPIV3  .....N.N.SNE.VS.....L.....T.....A...*.....*.....H
Sendai  N*.T.H.I.*...T.NT*SLT.*.S*FA.S..N.T.AT...S...LEDM.YK...I.E
Measles  .R.EEDRK.YLMTLLD*IK.ANDLA.FHQMLM..IM (507)

BPIV3  IEKTPENDIQIKSDINRSNDESNA*RLV*PKRTSNTMRS*LI*II*NN*SL*SS*RAKQSYINEL
HPIV3  AGN.L...*.*.*.LS.YN.....*.*V.S.....*A*.....PQST.....
Sendai  D.FR.D.IRNP*YQ*RDTEPRA...*.*.*S*EKP..H..R**.*ES.P..RAE.AA.*KS.

BPIV3  KLCKSDDEEVSELMDFNEDVSSQ (596)
HPIV3  .H.....NNAKDQIKKTPNK (615)
Sendai  SK...*Q..KA*.*LVE..*E.LTN (568)

```

BPIV3 P PROTEIN		MOLECULAR WEIGHT = 66409							
ALA= 20	3.36%	ASN= 46	7.72%	ASP= 43	7.21%	ARG= 36	6.04%		
CYS= 2	.34%	GLN= 31	5.20%	GLU= 52	8.72%	GLY= 34	5.70%		
HIS= 3	.50%	ILE= 44	7.38%	LEU= 38	6.38%	LYS= 51	8.56%		
MET= 16	2.68%	PHE= 6	1.01%	PRO= 18	3.02%	SER= 74	12.42%		
THR= 51	8.56%	TRP= 2	.34%	TYR= 8	1.34%	VAL= 21	3.52%		

Fig. 6. Comparison of the deduced amino acid sequence of the P protein of BPIV3 with those of HPIV3, Sendai virus and measles virus, and the deduced amino acid composition of BPIV3 P protein. See the legend for Fig. 4.

some clones conserved the complete complementary R1 sequence with no extra nucleotides at their 5' ends (data not shown). An unexpected finding was that two cDNA clones, one derived from the M mRNA of SC virus and the other

from the M mRNA of M virus, carried the poly A tail beginning 67 and 69 nucleotides, respectively, downstream of the R2 end of the M gene, i.e., within the 3' non-coding region of the F gene, although these clones showed exactly the same nucleotide sequence as the 910N virus around the M-F gene junction. On the other hand, however, other two cDNA clones derived from the M mRNA of M virus were demonstrated to have the poly A tail immediately following the R2 sequence (data not shown).

Comparison of deduced amino acid sequences of the NP, P, C and M proteins of BPIV3 and other paramyxoviruses

The deduced amino acid sequence of the NP protein of BPIV3 is shown in Fig. 5 along with those of HPIV3 and Sendai virus. Although we determined each part of the sequence by analyzing two or more independent cDNA clones, the calculated molecular weight of the protein (57K) was much less than that estimated (68K) by SDS-PAGE (44). The same situation has been reported for HPIV3 (6, 7). Moreover, although the calculated molecular weight of the NP protein of BPIV3 was very similar to that of Sendai virus (57K), the latter moved significantly faster than the former in SDS-PAGE (31). Many recent papers have reported such discrepancies between the calculated and SDS-PAGE-estimated molecular weights of viral proteins.

The amino acid sequence of the BPIV3 NP protein was very similar to that of the HPIV3 NP protein (6, 7); 76 of the total 515 amino acid residues were different, thus showing 86 % homology. Of the 76 amino acid substitutions, 41 were located within the carboxyl terminal quarter where the homology was 67 %. The NP protein of BPIV3 also showed a fairly high homology with that of Sendai virus (1, 14, 15), being 59 % in total. However, the carboxyl terminal quarter again showed less homology, only 15 %, while the middle half showed about 80 % homology.

As with HPIV3 (6, 7), the middle part of the NP protein of BPIV3 shared fairly well-conserved sequences with those of measles virus (23) and NDV (26) (underlined sequences in Fig. 5), but the total homologies between them were only about 20 %.

Compared with the NP protein, the P protein (Fig. 6) showed slightly less amino acid homology, 62 %, between BPIV3 and HPIV3 (8). Furthermore, the P protein of BPIV3 was quite different from that of Sendai virus (1, 13), the amino acid homology of the total protein being only about 20 %, though the carboxyl terminal half showed a significant homology, 31 %. Comparison of BPIV3 and measles virus (24) indicated that there was only a coincidental and therefore insignificant homology (less than 10 %) between

```

BPIV3      MLKTIKSWILGKRQDQETSHLTSHRP STSLN SYSAPTPKRT RQTAMKSTQGTDRS
HPIV3      .....N..INQ.I.P. ....*..Y.K TTQ...EPSN
Sendai     MPSF..K. LK*R.*. .DESR*RMLSD.*M.SCRVNQL.SEG.EAG*TP..LPK .Q
Measles    .S..DWNAS .LSRPS.P.AHWPS.KLWQHGGQKY*QTDRSEPPAGKRRQAVRV*ANH

BPIV3      ARQSTNLNP      KQQKQAKKIVDQLTKIDSLGHHTNVPQRQRIEMLIRRLRYREEIGEE
HPIV3      .PP.V.QKS      N...V...*.....Q.*..I...*.....
Sendai     .LLIEPKVRAKE .S.HRRP.*...*R***...EQASQR.*HM*.T.N**..TGP*...
Measles    .S.QLDQLKAVHLASAV*DLE*AMTT.KLW*.PQEI*RHQALGYSVIMFMITAVKR*R.S

BPIV3      AAQIVELRLWSLEESPEAAQILTMPEKSRKVLITMKLERWIRTLLRGKCDNLKMFQSRYO
HPIV3      .....L..S...K....*...*.....Q...A...
Sendai     LV.T*Y...*..AM...*..SLK..Q.REDI.DQ**K..T...*...*..EKT.K..D..K..E
Measles    KMLT*SWFNQA.MVIAPSQEET .NL.*AMW**ANL*P.DNL*.T . .L.PSLWGSSL

BPIV3      EVMFPFLOQNKMETVMMEAWNLSVHLIQDIPA (201)
HPIV3      ...SY....V...I.....Q (199)
Sendai     .H.Y.MKE.V.Q.I.....S.AA.*... (204)
Measles    L .LK..KEGRS.SS (186)

```

```

BPIV3 C PROTEIN          MOLECULAR WEIGHT = 23509

ALA= 10  4.98%  ASN=  7  3.48%  ASP=  6  2.99%  ARG= 17  8.46%
CYS=  1  .50%  GLN= 17  8.46%  GLU= 16  7.96%  GLY=  5  2.49%
HIS=  5  2.49%  ILE= 13  6.47%  LEU= 21 10.45%  LYS= 16  7.96%
MET= 10  4.98%  PHE=  2  1.00%  PRO=  9  4.48%  SER= 16  7.96%
THR= 16  7.96%  TRP=  4  1.99%  TYR=  3  1.49%  VAL=  7  3.48%

```

Fig. 7. Comparison of the deduced amino acid sequence of the C protein of BPIV3 with those of HPIV3, Sendai virus and measles virus, and the deduced amino acid composition of BPIV3 C protein. See the legend for Fig. 4.

the P protein amino acid sequences; this was also the case between Sendai and measles viruses. Characteristics common to the P proteins of BPIV3, HPIV3, Sendai virus and measles virus were that they were all hydrophilic and were rich in Thr and Ser residues, the latter providing sites for phosphorylation. The calculated molecular weight of the BPIV3 P protein (66K) was again different from that estimated (81K) by SDS-PAGE (44).

The amino acid homology between the C proteins of BPIV3 and HPIV3 (8) was 76 % (Fig. 7), while that between BPIV3 and Sendai virus (1, 13) was 35 % as a whole. However, the latter reached a considerably higher level, 43 %, in the carboxyl terminal two thirds. This indicated that the amino terminal one third was very different in the two viruses. The C proteins of BPIV3, HPIV3 and Sendai virus had virtually no homology with that of measles virus (24), although they were all strongly basic.

The amino acid sequence of the M protein (Fig. 8) was conserved fairly well by BPIV3 and Sendai virus (2, 16), showing 62 % homology as a whole, as was observed for the NP protein. In contrast to the NP protein, however, an area with less homology (50 %) was found in the amino terminal quarter. Only the M protein showed fairly high homologies between the BPIV3-Sendai virus group and measles virus (25), being 34 % between BPIV3 and measles virus, and 33 % between Sendai and measles viruses. The measles M protein differed

```

BPIV3      MSITNSAIYTFPESSFDNGNIEPLPLKVNQQRKAVPHIRVVRIGDPPKHGSRYLVDVFL
Sendai     MAD..R..KF.YE...T*.....*TGPD*...*.....*V...*L..
Measles    M*Q..D.DK.AWDIK.S.A.*QPTYSDG*L..Q*...*DP.LGD*KDECFMYM...

BPIV3      GFFEMERSKDRYGSVSLDDDDPSYKVCGS  GSLPLGLARYTGNDDQELLQAATKLDIEVR
Sendai     ...TP*QTNL.....T*PT..S*... ..*..*..*Y.T.....K.C.D.R.T..
Measles    .VV. .S..LGPP.I.RAF.....*G.S.AKPEK..KE..E...V..

BPIV3      RTVKATEMIVYTVQNIKPELYPWSSRLRKGMFLDANKVALAPQCLPLDRGIKFRVIFVNC
Sendai     ...*G...M.DS.GAP.L...G...Q...*N.....*D.*L...*G
Measles    ..AGLN.K*.FYNNTPLTL.T..RKV.TT.S*.N..Q.CS.VNL*...TPQ*...*YMSI

BPIV3      TAIGSITLFPKIPKSMALLSLPNTISINLQVHIKTGIQDTSKGVVQILD EKGEKSLNFMV
Sendai     .S*.A...*A...*L.D.A...*...L.T*...S*Q...*P... *Q...K....
Measles    .R*SDNGYYT*.*R.LEFRSV.A*AF..L.T**IDKAI .PGK.*.NTEQLPEAT...

BPIV3      HGLGIKRKMGRMYSVEYCKQKIEKMRLFLSLGLVGGISLHVNTATGSISKTLASQLAFKRE
Sendai     ....*..V.*I.....S...*...*.....*F..QVN.**...FM...W..A
Measles    *.NF*..KSEV..A*...M.....G.*.A.G*..T...*RS..KM...HA..G..*T

BPIV3      ICYPLMDLNLPHLNLVIVWASSVEITRVDAIFQPSLPGEFRRYPNIIAKGVGKIRQ (351)
Sendai     *.F.....*M.M...A....G...*..A*.R*.....*..N*...KL (348)
Measles    *.....*..ED..R*..R.RCK.V.*Q.*L...*Q...I.DD*.INDDQG*FKVL(335)
    
```

BPIV3 M PROTEIN		MOLECULAR WEIGHT = 39309													
ALA= 17	4.84%	ASN= 15	4.27%	ASP= 16	4.56%	ARG= 20	5.70%	CYS= 5	1.42%	GLN= 12	3.42%	GLU= 16	4.56%	GLY= 24	6.84%
HIS= 6	1.71%	ILE= 29	8.26%	LEU= 39	11.11%	LYS= 26	7.41%	MET= 10	2.85%	PHE= 14	3.99%	PRO= 19	5.41%	SER= 29	8.26%
THR= 15	4.27%	TRP= 2	.57%	TYR= 12	3.42%	VAL= 25	7.12%								

Fig. 8. Comparison of the deduced amino acid sequence of the M protein of BPIV3 with those of Sendai virus and measles virus, and the deduced Amino acid composition of BPIV3 M protein. See the legends for Fig. 4.

mostly from the other viruses in the amino terminal quarter, showing only about 10 % homology. As pointed out previously by Bellini et al. (25), it is noteworthy that 13 Gly, 9 Pro and 20 basic amino acid residues (Arg or Lys) were conserved at the same positions in the M proteins of these three viruses. Furthermore, eight corresponding positions were occupied by the hydroxyamino acid residues, Ser and Thr. BPIV3 M protein's deduced molecular weight (39K) was fairly consistent with that estimated by SDS-PAGE (36K). The amino acid sequence of HPIV3 M protein has not yet been reported.

The above amino acid homology observations are summarized in Table 1.

DISCUSSION

The present study elucidated the complete nucleotide sequence of the BPIV3 genome RNA from the 3' end to the terminal point of the M gene. The gene order and the presence of the overlapping P+C gene, as well as the consensus R1 and R2 sequences are characteristics common to other paramyxoviruses.

The nucleotide sequence homology of the leader region clearly indicates that BPIV3 and HPIV3 are most closely related to each other and also show a

Table 1. Amino acid homologies (%) of viral proteins between viruses.

Combination	Protein			
	NP	P	C	M
BPIV3 - HPIV3	86	62	76	
BPIV3 - Sendai	58	24	35	62
BPIV3 - NDV	18		<10	
BPIV3 - Measles	20	<10	11	36
Sendai - NDV	18		<10	
Sendai - Measles	18	<10	11	33
NDV - Measles	19		<10	

high resemblance to Sendai virus. It is also clear that these three viruses are quite different from measles and Newcastle disease viruses, the latter two being again different from each other. However, all five viruses conserve eight nucleotides at the very 3' end of the genome, suggesting that this sequence is an important element of the paramyxovirus genome.

Comparison of the NP gene of BPIV3 with that of HPIV3 provided a further molecular basis for their current classification. They are much more closely related to each other than to Sendai virus, since their gene nucleotide and gene product amino acid sequences show extremely high homologies compared to those between BPIV3 and Sendai virus. Moreover, the numbers of nucleotides in the non-coding regions and reading frames are identical in BPIV3 and HPIV3, but different in BPIV3 and Sendai virus. However, the level of difference between BPIV3 and HPIV3 is greater than that detectable among strains of the same virus. For example, the NP proteins of Sendai virus Z and Enders strains (1, 15) show only 15 different amino acids out of a total of 517 residues, whereas the difference between BPIV3 and HPIV3 is 76 out of 515. These findings support the conclusion that, although BPIV3 and HPIV3 belong to the parainfluenza 3 virus group, they are separable subtypes. This view is further supported by the finding that the amino acid homologies of the P as well as C proteins in BPIV3 and HPIV3 are somewhat but significantly lower than that of the NP protein, and that the two proteins in each virus contain different numbers of amino acids.

The present study also demonstrated similarity as well as dissimilarity between the genome and gene structures of BPIV3 and Sendai virus. The leader region, the nucleotide sequences of the NP and M genes and the structural make-up of the overlapping P+C gene are strikingly similar

between the two viruses. Accordingly, the NP and M proteins, which are fundamental elements in the maintenance of the virion structure, are very similar in the two viruses, both showing about 60 % homologies. However, it is interesting that the carboxyl terminus of the NP protein and the amino terminus of the M protein show reduced levels of homology, indicating that these regions harbor viral specificity. A similar observation was reported with measles and canine distemper viruses, both belonging to the paramyxovirus morbilli virus subgroup; the viruses have very similar NP protein amino acid sequences, but a considerable difference is detectable in the region proximal to the carboxyl terminus (25). Recently, lack of homology at the amino terminus of M proteins of NDV, Sendai virus and measles virus was also reported (53).

On the other hand, the homologies of the P and C proteins between BPIV3 and Sendai virus are lower than those of the NP and M proteins. This indicates that a great deal more virus specificity exists in the functionally rather than the structurally important proteins, although the functions of the P and C proteins remain obscure. In both P and C proteins, differences in amino acid homology in BPIV3 and Sendai virus are more obvious in the region proximal to the amino terminus than in the region proximal to the carboxyl terminus. Moreover, when the BPIV3-Sendai virus group is compared with measles virus, the amino acid homologies are relatively high in the structurally important NP and M proteins (about 20 % and 30 %, respectively), while there is virtually no homology in the functionally important P and C proteins.

These observations suggest that although paramyxoviruses conserve common structural elements as well as a common mode of genome transcription and replication, each virus may have virus-specific elements in their multiplication. Such virus specificity might be concerned with either host cell factors needed for transcription and replication, or with R1 and R2 signal sequences which are subtly different from virus to virus, or with both. In this context, it is important to compare the viral RNA polymerase genes, i.e., the L genes of paramyxoviruses, and for this reason we are now sequencing the BPIV3 L gene.

One of the characteristic findings of the present study was that the poly A tails of two mRNAs of BPIV3 M gene were found to begin several tens of nucleotides downstream beyond the R2 sequence whereas those of other two M mRNA were revealed to start from the R2 sequence. This seems to be correlated with the extremely long U-rich sequence in the non-coding region

of the following F gene where R2-like sequences could be detected, and also with the considerable variation in the R2 sequence from gene to gene; such a situation could cause the RNA polymerase to incorrectly recognize the polyadenylation signal. Recently, Spriggs and Collins (12) reported that read-through frequently occurred in transcription of HPIV3 M gene, in which the same situation as above was observed. It would be interesting to evaluate the biological significance of such a miss-read.

ACKNOWLEDGEMENT

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. Shioda, T., Hidaka, Y., Kanda, T., Shibuta, H., Nomoto, A. and Iwasaki, K. (1983) *Nucleic Acids Res.* 11, 7317-7330.
2. Hidaka, Y., Kanda, T., Iwasaki, K., Nomoto, A., Shioda, T. and Shibuta, H. (1984) *Nucleic Acids Res.* 12, 7965-7973.
3. Shioda, T., Iwasaki, K. and Shibuta, H. (1986) *Nucleic Acids Res.* 14, 1545-1563.
4. Abinanti, F.R., Chanock, R.M., Cook, M.K., Wong, D. and Warfield, M. (1961) *P.S.E.B.M.* 106, 466-469.
5. Fischman, H.R. and Bang, F.B. (1966) *P.S.E.B.M.* 121, 966-970.
6. Galinski, M.S., Mink, M.A., Lambert, D.M., Wechsler, S.L. and Pons, M.W. (1986) *Virology* 149, 139-151.
7. Sánchez, Á., Banerjee, A.K., Furuichi, Y. and Richardson, M.A. (1986) *Virology* 152, 171-180.
8. Luk, D., Sánchez, Á. and Banerjee, A.K. (1986) *Virology* 153, 318-325.
9. Spriggs, M.K., Olmsted, R.A., Venkatesan, S., Coligan, J.E. and Collins, P.L. (1986) *Virology* 152, 241-251.
10. Elango, N., Coligan, J.E., Jambou, R.C and Venkatesan, S. (1986) *J.Virol.* 57, 481-489.
11. Dimock, K., Rud, E.W. and Kang, Y. (1986) *Nucleic Acids Res.* 14, 4694.
12. Spriggs, M.K. and Collins, P.L. (1986) *J.Virol.* 59, 646-654.
13. Giorgi, C., Blumberg, B.M. and Kolakofsky, D. (1983) *Cell* 35, 829-836.
14. Blumberg, B.M., Giorgi, C., Rose, K. and Kolakofsky, D. (1984) *J.gen.Virol.* 65, 769-779.
15. Morgan, E.M., Re, G.G. and Kingsbury, D.W. (1984) *Virology* 135, 279-287.
16. Blumberg, B.M., Rose, K., Simona, M.G., Roux, L., Giorgi, C. and Kolakofsky, D. (1984) *J.Virol.* 52, 656-663.
17. Blumberg, B.M., Giorgi, C., Rose, K. and Kolakofsky, D. (1985) *J.gen.Virol.* 66, 317-331.
18. Hsu, M-C. and Choppin, P.W. (1984) *Proc.Natl.Acad.Sci.USA.* 81, 7732-7736.
19. Blumberg, B.M., Giorgi, C., Roux, L., Raju, R., Dowling, P., Chollet, A. and Kolakofsky, D. (1985) *Cell* 41, 269-278.
20. Miura, N., Nakatani, Y., Ishiura, M., Uchida, T. and Okada, Y. (1985) *FEBS lett.* 188, 112-116.
21. Morgan, E.M. and Rakestraw, K.M. (1986) *Virology* 154, 31-40.
22. Gupta, K.C. and Kingsbury, D.W. (1984) *Nucleic Acids Res.* 12, 3829-3841.

23. Rozenblatt,S., Eizenberg,O., Ben-Levy,R., Lavie,V. and Bellini,W.J. (1985) *J.Virol.* 53, 684-690.
24. Bellini,W.J., Englund,G., Rozenblatt,S., Arnheiter,H. and Richardson,C.D. (1985) *J.Virol.* 53, 908-919.
25. Bellini,W.J., Englund,G., Richardson,C.D., Rozenblatt,S. and Lazzarini,R.A. (1986) *J.Virol.* 58, 408-416.
26. Ishida,N., Taira,H., Omata,T., Mizumoto,K., Hattori,S., Iwasaki,K. and Kawakita,M. (1986) *Nucleic Acids Res.* 14, 6551-6564.
27. Shibuta,H., Kanda,T., Hazama,A., Adachi,A. and Matumoto,M. (1981) *Infect.Immun.* 34, 262-267.
28. Shibuta,H., Adachi,A., Kanda,T. and Matumoto,M. (1982) *Infect.Immun.* 35, 437-441.
29. Shibuta,H., Nozawa,A., Shioda,T. and Kanda,T. (1983) *Infect.Immun.* 41, 780-788.
30. Shibuta,H., Kanda,T., Nozawa,A., Sato,S. and Kumanishi,T. (1985) *Arch.Virol.* 83, 43-52.
31. Shibuta,H., Kanda,T., Adachi,A. and Yogo,Y. (1979) *Microbiol.Immunol.* 23, 617-628.
32. Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry* 18, 5294-5299.
33. Okayama,H. and Berg,P. (1983) *Mol.Cell.Biol.* 3, 280-289.
34. Okayama,H. and Berg,P. (1982) *Mol.Cell.Biol.* 2, 161-170.
35. Boliver,F. and Backman,K. (1979) *Methods Enzymol.* 68, 245-267.
36. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc.Natl.Acad.Sci.USA.* 74, 5463-5467.
37. Messing,J. (1983) *Methods Enzymol.* 101, 20-78.
38. Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.* 65, 499-560.
39. Paterson,B.M., Roberts,B.E. and Kuff,E.L. (1977) *Proc.Natl.Acad.Sci. USA.* 74, 4370-4374.
40. Pelham,H.R.B. and Jackson,R.J. (1976) *Eur.J.Biochem.* 67, 247-256.
41. Maizel,J.V., Jr., White,D.O. and Scharff,M.D. (1968) *Virology* 36, 115-125.
42. Gill,D.S., Chattopadhyay,D. and Banerjee,A.K. (1986) *Proc.Natl.Acad. Sci.USA.* 83, 8873-8877.
43. Leppert,M., Rittenhouse,L., Perrault,J., Summers,D.F. and Kolakofsky,D. (1979) *cell* 18, 735-747.
44. Shibuta,H., Suzu,S. and Shioda,T. (1986) *Virology* 155, 688-696.
45. Curran,J.A., Richardson,C. and Kolakofsky,D. (1986) *J.Virol.* 57, 684-687.
46. 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.
47. Paterson,R.G., Harris,T.J.R. and Lamb,R.A. (1984) *Virology* 138, 310-323.
48. McGinnes,L.W. and Morrison,T.G. (1986) *Virus Res.* 5, 343-356.
49. Chambers,P., Millar,N.S., Bingham,R.W. and Emmerson,P.T. (1986) *J.gen.Virol.* 67, 475-486.
50. Billeter,M.A., Baczko,K., Schmid,A. and Ter Meulen,V. (1984) *Virology* 132, 147-159.
51. Kurilla,M.G., Stone,H.O. and Keene,J.D. (1985) *Virology* 145,203-212.
52. Millar,N.S., Chambers,P. and Emmerson,P.T. (1986) *J.gen.Virol.* 67, 1917-1927.
53. Chambers,P., Millar,N.S., Platt,S.G. and Emmerson,P.T. (1986) *Nucleic Acids Res.* 14, 9051-9061.