
Nucleotide sequence of the bovine parainfluenza 3 virus genome: the genes of the F and HN glycoproteins

Shinya Suzu, Yuko Sakai, 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. Y00115

ABSTRACT

By analysing complementary DNA clones constructed from genomic RNA of bovine parainfluenza 3 virus (BPIV3), we determined the nucleotide sequence of the region containing the entire F and HN genes. Their deduced amino acid sequences showed about 80 % homologies with those of human parainfluenza 3 virus (HPIV3), about 45 % with those of Sendai virus, and about 20 % with those of SV5 and Newcastle disease virus (NDV), indicating, together with the results described in the preceding paper on the NP, P, C and M proteins of BPIV3, that BPIV3, HPIV3 and Sendai virus constitute a paramyxovirus subgroup, and that BPIV3 and HPIV3 are very closely related. The F and HN proteins of all these viruses, including SV5 and NDV, however, were shown to have protein-specific structures as well as short but well-conserved amino acid sequences, suggesting that these structures and sequences are related to the activities of these glycoproteins.

INTRODUCTION

The paramyxovirus hemagglutinin-neuraminidase (HN) and fusion (F) glycoproteins, which project as spikes from the viral envelope, are deeply concerned with the biological activities and serological characteristics of the viruses. The virions attach to receptors on the cell surface via HN protein, and penetrate into the cells by envelope-cell membrane fusion, a process which requires F protein. The release of mature virions from the infected cell surface is mediated by HN protein's neuraminidase activity (for example, 1, 2). The paramyxoviruses are classified into subgroups according to their antigenicity, which is predominantly determined by the glycoproteins. Thus there have been many studies on paramyxovirus antigenicity using monoclonal antibodies against the glycoproteins (for example, 3, 4, 5). However, the structures of the functional domains and antigenic epitopes of these glycoproteins still remain obscure, and this situation has prompted many research groups to perform sequence studies of the F and HN genes. We have been carrying out such studies on Sendai virus and bovine parainfluenza 3 virus (BPIV3), and have already reported our results on the former (6).

BPIV3, which is the causative agent of "shipping fever" of calves, shows considerable strain differences in its glycoprotein activities (7, 8, 9), and is known to have a close antigenic relationship with human parainfluenza 3 virus (HPIV3) which causes a respiratory infection of children, although these two viruses can be distinguished from each other by detailed serological examinations (4, 10, 11).

We previously showed by use of monoclonal antibodies against the viral glycoproteins that F protein of BPIV3, like other paramyxoviruses (for example, 12), consists of F1 and F2 subunits linked to each other by disulfide linkages, that the cleavage of the precursor F protein into these subunits is achieved by some cellular endoprotease, and that the HN protein of BPIV3 really does carry both hemagglutinin and neuraminidase activities (13).

In the preceding paper, we presented the nucleotide sequence of the BPIV3 genome from its 3' end to the terminal point of the membrane (M) protein gene, through the nucleocapsid (NP) protein and polymerase-associated (P) protein, as well as the non-structural "C" protein genes. Here we report the nucleotide sequence of the subsequent genome region in which the F and HN genes are located, and compare their deduced amino acid sequences with those reported for Sendai virus (6, 14, 15, 16, 17), HPIV3 (18, 19), SV5 (20, 21), Newcastle disease virus (NDV) (22, 23) and measles virus (24).

MATERIALS AND METHODS

Viruses, preparation of viral RNA, synthesis and cloning of complementary DNA (cDNA)

CDNAs of viral genome RNA were constructed and cloned from the 910N strain of BPIV3, and cDNAs of viral mRNAs from the 910N, M and SC strains of the virus (8, 9) according to the method of Okayama and Berg (25, 26) as described in the preceding paper. Bacterial colonies carrying plasmids were screened for the presence of viral cDNA inserts with the viral 50S RNA probe, and for overlapping cDNA inserts with appropriate cDNA probes, as described previously (6).

Sequence determination of cDNA

CDNAs were cleaved with appropriate restriction endonucleases into fragments and which were sequenced mainly by the dideoxy method of Sanger et al. (27) using the M13 phage system (28). In some experiments the chemical cleavage method of Maxam-Gilbert (29) was also employed.

Gene expression from cDNA

SV40 transformed CV-1 (COS-1) cells were transfected according to Wong and Hirano (30) with plasmids containing cDNA inserts, which were primed and cloned with a pcDV1 expression vector from the viral mRNAs. The cells were stained by indirect immunofluorescent techniques using rabbit antiserum against 910N virus (31) and mouse monoclonal antibodies directed to the F and HN glycoproteins (13) of the same virus.

Enzymes and other materials

Fluorescein isothiocyanate conjugated (FITC) goat antibody against rabbit IgG was purchased from Behringwerke AG, Marburg, West Germany; FITC rabbit antibody against mouse IgG from Miles-Yeda, Israel; and others were the same as described in the preceding paper.

RESULTS

Nucleotide sequences of the F and HN genes of BPIV3 genome RNA

The nucleotide sequence of a viral genome region, following the M gene region presented in the preceding paper, was determined from cDNA clones

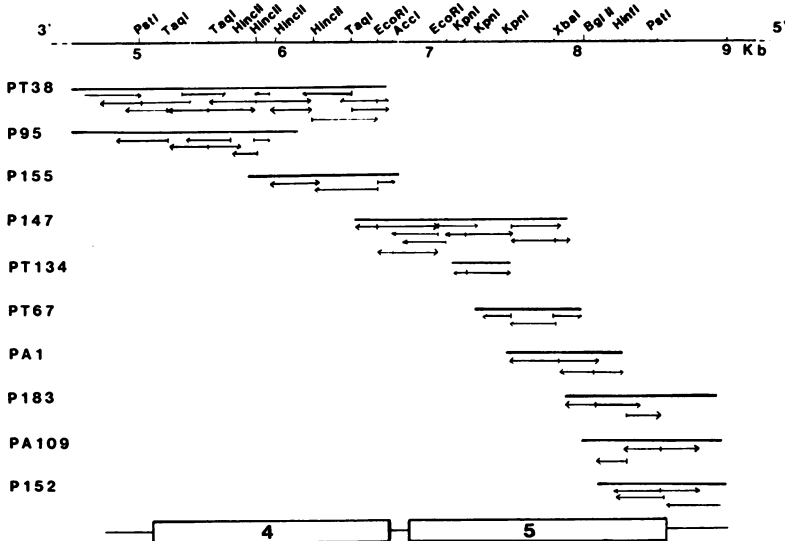


Fig. 1. Locations of cDNA clones (PT38 to P152) in the genome RNA and the sequencing strategy. The numbers in kilobases (Kb) under the top line show nucleotide positions from the genome 3' end. Restriction map is also presented on the top line. Arrows indicate the cDNA fragments sequenced and the direction of sequencing. Positions of the open reading frames of the 4th and 5th genes are denoted in the bottom.

```

      10      20      30      40      50      60      70      80      90      100
3'--- GCGUU AAAAGGUCGG AAGUAAUGGA CCACUUAAAU CUAUGAUGGG UUUUAUGUAU CGUUUUCCUC AGCCUUUUUA GUCUGUCAUU UUGGUUGUUA 4800
GGACUGUGGG CUGUGGGGUC UGGUCGCAAG GGUGUCUCUU UUUUAGUUUU UGAAUCCUAG UUUUCCUUAU GGUGUUUGAG GCUCUGUUGC AAUUUGGUUG 4900
GUUUAUGUAU UAGUGUCUCU UUUUUUUUUU CUUUUUUUUU UUCUUUUUCU UUUUUUUGUG CGUUUUCCUG UCUAGUUAGU CUCUGUGGGU UUGGUUUCGUC 5000
UUCUUUCGGU UCCGGGCGGU GGGGUUUUUU GUGGUUUUUG GGACGUCUGU UCGUGGCCUG ACUUUCGGUU UUCUUUUUUU CGUAUGUCUU ACUAGUAUGU 5100
UUUGUGUUAG UUUUAUGAUG AUUUAUUUAG GGGUAGUAAA ACAGUUUUUC UGUUUUUUUU UGAAGUUUGCA CAUCCACUAU AUCAGUUUUU GGGGUUUUCA 5200
UACUUCUAAA GUGUUUUAAA GCUUUUCUCU AUGAUAUUUA AUUCAAAUAU UGGUUUUUUU CUCUUGAGUG UGAGUACACC CCUAGUUUGU UUUUUUGGUA 5300
UGUUUUUGUA UAACCGUGUC AACUAGUAAG GAGAUUAUAC ACCUGAUUUU AAUGUUUUUU UACACUAUCA UCAUUCAGUA CUCUGAUUGU UAUUGUGGUU 5400
AAGAUCGUGU UUCGCUAAAA AGCCUCUUUA UUAACCCUGU UAACGCUAUC CCUAUCGGUG GAGUCUGUGC UAAUGGCGUC GUCAGCGAGA ACAGCUCGCU 5500
UUUGUCCGUA UCAUGUCUGA UCUCUUUGAG UUUUCUUGAU AUUCUCUGUG UUUUUUCCGU CAUGUUAGUU AAGUUUCAAG ACAUCCAUUA GACUAACAAC 5600
GUCAAUUUUG UCAAGUUCUA AUACAGUUGU UACUUUAAUA UGGAAGUUAG UGUUCUAAUC CAACACUUCG UCGACCAGAU UUUGAUCCUU AACGUAAUUG 5700
UGUUGUAAUG AGUCUUAAUU GUUUUAUUAA ACCACUAUUA UAUCUUUGUG ACUUUUUUUU CCCCUAUUUU AAUGUCCCUU AUCGUAGCGA UAUGUUGGUG 5800
UUUUUAAUG UUUUAUAAUG AUGAAGUUUU CAACUGUUUA UACUGUAAAU ACUAGAUAAU AAGUGACUUA AAGGUUUUUU CUCUAGUUUA CUACACUAAA 5900
ACUCACUAUU GAGUUAGUGA GAAGUUCAAU CUGAAGGAAA CAAUUGAUUU UAUAGUUUUU GAGUUUUAAA AUUCAUCUAU AGAUUAUGUA UGUUGUUUGU 6000
UCCGUUUUCU ACCAUUAUAG GCGAAGGGUU AGUGUAUUAC UGUUUUCCUC GAAAGGAUCC ACCACGACUA UAAUUUCUUA CGUAUCUUCG UAAUGCGUCA 6100
AUUAUACAG GAAGACUGGG UCCGAAAUUG GAUUUAGUGC UCUAUUCUUU AACAAAUAGU CCCUUUAUUU GUGUCACAGG AUUCUGAUAU CAAUGUAGUC 6200
UACAUCUUGG UGCUUAUCGU AAACACUUC ACCTUAUUUA ACGUUUGACA UAUUGUUUGU GUACAUAGUC AUUACCUCAA CUGUUUUCUU AUUUUUGUAG 6300
UGGACUAGUU CCUUUUUUUU AGUUUUUGUU AUUUUUUUUA GUUCUUUUC CAUUAUUUUC CAUUAUUUUC UAACAAUAGU UCAUGGUUUU CUCUUUCCUG UAAUCGUUGA 6400
AUUAUGAAAC UACUGUAGUA UAAUUUUUUG AGACAACGGG AAUUUAGCGUA ACUUAUAGUA UAUCUUUGAGU UGUUCGUGUU UAACCUUGAC CUUCUUAGCU 6500
UUCUUACCUA UUUUUUUUAG UUAUUUUUUA AUCUAAGACA ACCUUAACCC AUAGUUUAGU CACGUUUUUA AUGGUUUUAG UAUCACUACU AUCAGCAGGA 6600
UAAGUUUAG UUAUUUUUGU AAUAACAUA GUUUUAGUCU GUGGUUUCUU AAGUCCUUU UUUUUUUUUU CUGUUUUUUU CACUUUGCAU ACAUUUGU 6700
UCAUCUGUUA CUCUUAUUG CACUAAUUUA UAUUUUUUAU GUUUUUUGAA UCCUUUUUUC R1
R2
GUCUCUCGUC UGGUUUAAGU UUUUACCUUA UAACCUUUUG UGUUUUGUCG UGUUUCCUGU GAUUUUUUCU UGAACCCUGU UGGUCUCUGU CUGUGUCGUC 6900
GUUUCGGUGU UUAUAGUAUU ACAUGAGAA CAGCCUUGUU UGUUUUUUUU AUAAUAGUA GUAAAAUUU UACUUAUAC AUUUAAAUAU AGUUUCUUUG 7000
UUAGUGUUGU UUAUUUACAA CGUCUUUUUU UCCUUUUUUA UCUAUGGUUU UAUUUUUUUU UAGUUUUUUU UGAACCAUUG AGUUUUUGUA 7100
GUCCUUUUUU GUGUUCGAA GAUUUUUAG UCUCAGUACA AGUUUUUUAU UAAGGUAUUA GUAAUUUGU UGUUUUACAG CUAGAGUCUU UUAUUUAGUU 7200
GCUAGAGUCG UGAUUUUUCU UUGUAUUUCU CCAUGGUUUU GUCUCUUACU GAGUACUUAU ACCAUUUUCU GGAGUUUUG GCCUUUUUUA GACCUCUACA 7300
UGUAGACCAU UAGGUAGAAA UGUUUAUCA GGUUUCUAUU CUGAUUAUG UCCCGGUCCA AUUAUUUGU GUAGGUUGAU ACAUUUACCG ACGUUAUCUG 7400
AGGGUAGAAA UCGUUAGUUA UUAAGUUAGA UGCGAAUUGG GAGAUUAGAA UAAUUGGUCC CGACAGUUCU AUUUCUUUUU AGAAUGGUUC AUGAUGUUUA 7500
UCCUUUUAAA UGAUUUUUUA GUCUAACCA UGGACUAAAC UGGGUUUUCU AGUGUGUAUG UAAUUUAUUA CUACUGUUUU CCUUUAGGAC GAGGUAACGU 7600
GAUAACUUGU GUCUACAUAU AGUUUAUACA AGUUGGGUUU UUCAACUACU CUCUAGUCUA AUGCGUAGUU GUCCAUAACU CCUGUAACAU GAACGUAAC 7700
AGUGAUCAUU ACCUAAUUUA UAUUUUUUGU CUAAGUGCUU AUUAUUUAUU UGAAAACUUA UGGAAUACG UCGUAAUUAU GGUAGUCUAC CUGGUCCUUA 7800
UUAUUUUUUU CUGUUUCAUU AGAAGAGGCC UAUACCUCCA GAUCUGUAUC UUCUUUUUCC CCUGUQUAUU ACGUUUGUUU GACCAACAGG ACCGUUUUGU 7900
GUCUCUCUGA CGUUAGUCGG GAGAUAUGCG GGUACCAAGA GUUUUUCCUC UUAACUUAUA AGGUAAUAGC AACACUUAUU UCCAUAUCUA UGUUUAGAAU 8000
CGAAUUCUCA CACCGUAUA GGUUAUCUGU UUUUGAUUAC CCCCAGUCUU CCUUCUGACG AAAAUUAUCC ACUGUCUUUU AUGUAUUUAU GAUCUAGAUG 8100
CUCAACCGUG UCAUUUUUAG UCAAUCCCA UUAACUGUUA AGACUAAAUA UAUUUUUUUC UAUUUUAUCC UGAACCGUAU UACAUGAUAG UGCGCGUCC 8200
UUAUCUUAU CAGGUACCCC AGUAAGUAGC GGUCUGCCUA CGUAUUGUCC UCAAAUGUGA CUACGUUAG GGAUUUUUGG UAGUCCCUCA UAACAUAGGA 8300
GUCAGUAAGA ACUAAGUUUU UUUUGGUCU UUUUUUUUAU GUAAUGGAU AGUUGUCGAU GUUUUUCCCA UUUACUUAAC CGAUUAUUGU UGUCUUUGUA 8400
AGGUCGACGU AUGUGUGUU GUUUACAUA GUGUGUGUA CUAUUUCCCA UAACAAAAGU AUUAUCAUUU UAUUUAGUGU CUUCAACUUU GUGCAAAGUU 8500
CGAUACAUAU AAUUUUUGUC UCAAGGCUUU UGACCGUCG UUUUAGUAGG GGCGUGUAGC UUGUAAUCUA CUGUGUGUAU CUUUGGGUUU UUGUCUGUUG 8600
UGUCCUUAU UACGUUCUGU GUUUUUUUA UGUUUUUUGA AUCCUUUUU CACACAUGUU CUCUAACCGU GGCUUUAGGU GUCGCCAUGU UGUAGACUGU 8700
R2
R1
    
```

Fig. 2. The RNA sequence of the BPIV3 genome (910N strain) following the 3' proximal 4705 nucleotides presented in the preceding paper, expressed in negative sense. Enclosed sequences correspond to the open reading frames. R1 and R2 are the repeating consensus sequences.

according to the sequencing strategy illustrated in Fig. 1. The cDNA clones were constructed from genome RNA by the method of Okayama and Berg (25, 26) after *in vitro* polyadenylation of the RNA. The cDNA clones, PT38 and P95,

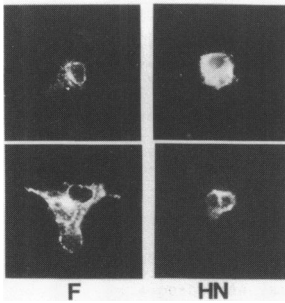


Fig. 3. Immunofluorescent monoclonal antibody staining of cDNA-transfected COS-1 cells. Left; transfected with cDNA clone M420 and positively stained with monoclonal antibody #31 against the F protein. Right; transfected with cDNA clone M176 and positively stained with monoclonal antibody H-1 against the HN protein. Two examples are presented for each.

were the same clones as those used in the preceding paper to determine the NP, P+C and M genes. Other overlapping clones were selected by colony hybridization tests using appropriate cDNA probes. The nucleotide sequence from position 4,706 to 8,700 of the genome RNA thus deduced is presented in Fig. 2.

The regions between nucleotide positions 4,855 and 6,747 and between 6,751 and 8,638 were considered to be the 4th and 5th genes, respectively, since each contained a large open reading frame (boxed in Fig. 2) flanked upstream by the starting consensus sequence R1 and downstream by the terminating consensus sequence R2.

The 4th and 5th genes were identified as the F and HN genes, respectively, by immunofluorescent monoclonal antibody staining of the gene products expressed in COS-1 cells. These cells transfected with clone M420, which was constructed from mRNAs of the M virus, and colony-hybridized with a fragment of the 4th gene cDNA of the 910N virus genome, were positively stained with monoclonal antibody #31 directed to the F protein of 910N virus (13) (Fig. 3). CDNA clone M176, constructed from M virus mRNAs, and corresponding to the 5th gene of the 910N virus, produced a protein which, upon transfection of COS-1 cells, could be stained by the monoclonal antibody H-1 directed to the HN protein of the 910N virus (31) (Fig. 3).

When compared with Sendai virus, SV5 and NDV, it was very characteristic of the BPIV3 genome for the 3' non-coding region following the R1 of the F gene to be extremely long (235 nucleotides) and highly U-rich. Although a similar feature was also observed in the F gene of HPIV3, the non-coding region of BPIV3 was 42 nucleotides longer (18).

Amino acid sequence of the F protein and its comparison with those of other paramyxoviruses

The F protein of BPIV3 910N strain, deduced from the nucleotide sequence, consisted of 540 amino acids, calculated molecular weight of the

```

EPIV3             MIITNTIIIIIIISPSFCQIDITNLQRVGLVNNPKGMKISQNFETRYLILS
HPIV3             .PT*IL*.TTM.MA.....K..H....[LSS].....
Sendai           MTAYIQRSQC.*T*****.TTL.VS...PRDR.SN*.*.DEC.SL.AGSH.*.....
SV5              .GTTIQF***SC*LAGAGS*.PAA.MQ*.*PT.V*QLMYYTEAS*AF***K
NDV ---SSTRIPILM*.IR.A*A.SCVHLASS*.GRP.AAA.*.TGD.AVN.VTSSQ.GS*.*K

BPIV3            LIPKIENSH SCGDQIQYKLLDRLIIPLYDGLKQKDVIVVSHETNNNSRTRKRF
HPIV3            .....D.N.....K.....*.....SNO*.*E.DP.....
Sendai           *.G**FEN G..TA.*I...S..N.....R.A.D..EA*.*[NT].TQ.AGAPQS...
SV5              .M.T.*SPISG[NIT].SS.[NAT]*T*.Q.*G*N.ETI*NO**P*.....R.R*..A
NDV              *.NMPKDKEA.AKAP*EA.[N*]T.TT.*T..G.S**RIQ*S*TT.GG.....R.Q...I

BPIV3            GEIIGTIAIGIATSAQITAVALVEAKQAKSDIEKLKEAIRDTNKAVQSIQSSVGNLIVA
HPIV3            .G*.....*.....*.....*.....*.....*.....*.....*.....*.....*.....
Sendai           .A*.....*.....G*..A.*E..R..AL*..SMTK.H.S*EL*.NA..EQ*.*
SV5              .V*..LA*.*..A*.....*.....K.NENAAA.LN..N..QK..A..AD*VQATQS.GT.
NDV              .A.....*.....*.....S..*Q.N.NAAN.L*...S.TA.IE..HE*TDG*SQ.A..

BPIV3            VKSVQDYVNEIVPSITRLGCEAAGLKGIALTQHYSELTNIFGDNIGTLKEKGIKLOGI
HPIV3            .....K.....A.....*.....Q.....*.....O.....*.....*.....*.....
Sendai           *.**..F..D..K.A.*E.....T.A.*.....K.....*.....LTA..S.F..*G..S*..T..A*
SV5              .QA...H*..S**S.A..AAN.K.QDAI*.SI.NLYL*..T..HNQ.TNPALSP.T*.A*
NDV              .GKM.QF...DQF[N*]AQE.D.IKITQQ*.*E.NLYL*...T*..PQ.T*PALTQT*.*A*

BPIV3            ASLYHNTITTEIFTTSTVDQYDIYDLLFTESIKMRVIDVDLSDYSITLQVRLPLTKISNT
HPIV3            .....R.....K.....*.....V.....N.....A.....*.....*.....*.....L..
Sendai           S...SA...[L].M..IRTG.S[N].S**Y..O..GT.....ER.M*..S**..*E*PGV
SV5              RI.LG*.*PT*VEK.FNT.ISAA*..S*GL*TGQ**G*..*YMQMV*K.E..T..VQPA.
NDV              YN.AGG.MDY*L.KLG.GNNQ*SS.*G*GL.TGNP.LY.SQTQL*G*..T..S*GN*N.M

BPIV3            QIYKVDISISYIQGKEWYIPLNHHMTKGAFLGGADIKECIEAFSSSYICPSDPGPTLNHE
HPIV3            .....*.....N*.....*.....S.....*.....V.....
Sendai           L.H.AS.....D.E.....*T*.S..L**AS.....*.....*.....SRL*...R..AQL*PDS
SV5              .ID**.*.AF.NNQ.VMAQ*.TR*.VT.SL*QAYPASQ.TITPN*VY.RYNDAGV.Q*
NDV              RATY*****.VSTTKGFASAL*.KV*TVQ.SV*EEL.TYS...TOLDLY.TRIVT.PMSPG

BPIV3            IENCLSCNITQCPKTIVTSVDPVRYAFVNGGLIANCITTTCTCNGVDNR[NQS]PQQGIKI
HPIV3            M.S.....*.....*.....K.....*.....*.....*.....*.....*.....*.....*.....
Sendai           QQK.*L.DT.R..V.K.VDS**.*F.....*.....*.....A.....GTGRRP.S.DRSK.VVF
SV5              TMA..Q..[L].R.TF*P.VGSF.T.FVLPD.I*Y...R*ML.K.MQPAAV.L.PSSSP*
NDV              .YS..N..T*A.MY*KTEGA*TP.MT*K.S.....KM...R.ADPPIG.S.NYGEA*

BPIV3            ITHKECQVIGINGMLFSTNREGTLATYTFDDIILNNSVALNPIDISMELNKAKLEESK
HPIV3            .....NT.....N.....*.....F..PN..T..[L]..D.....I.....*.....S*.....
Sendai           *.DN.G*.*..VELYA..R.HD..WGVG[N*]GPA*.*R*.....LN.AD.TNF.Q*..
SV5              .DMYK.VS*Q*DNLR.*ITQLAN[N*]IK*ESSQI*S*D*..*QN.AAV[NKS].S*AL
NDV              .DRHS.N.*S*D.ITLRLSG.FDATYQ[N*]S.*DSQV*VTGN*...T..GNV[NNS]SNAL

BPIV3            EWIKKSNQKLDVSGSWYQSSATITIIIVMIVVLFIIINITIIVVIIRHHRIQKGNQNDKNS
HPIV3            .....*.....*.....N.H.....T..I..**.*.....*.....*.....*.....*.....*.....*.....
Sendai           AE*E.ARKI.SE..R..N.RE.*IT..VM..*L**I.V..*..*Y.LR.SMLMGNP.DRI
SV5              QH*AQ.DTY.SA*T.ATT*.VLSI.A.CLGS*GL.*I.L*S..*W*LLT.VVA.R.RMEN
NDV              *K*E*E..S...K.NVKLT.*SA*ITY.ALTA*SL*CG.LS*.*ACYLMYK.KAQ.KTLLW

BPIV3            EPYVLTSRQ (540)
HPIV3            K.....N* (539)
Sendai           PRDTY.LEPKIRHMYTNGGFDAMAERK (565)
SV5              FV.HK (529)
NDV              LGNNTLGQMRATTKM (553)
    
```

BPIV3 F PROTEIN MOLECULAR WEIGHT = 60006

ALA=24	4.44%	ASN=39	7.22%	ASP=26	4.81%	ARG=15	2.78%
CYS=11	2.04%	GLN=27	5.00%	GLU=26	4.81%	GLY=30	5.56%
HIS= 9	1.67%	ILE=79	14.63%	LEU=45	8.33%	LYS=36	6.67%
MET= 7	1.30%	PHE=14	2.59%	PRO=15	2.78%	SER=41	7.59%
THR=42	7.78%	TRP= 3	.56%	TYR=17	3.15%	VAL=34	6.30%

Fig. 4. Comparison of the deduced amino acid sequence of the F protein of BPIV3 with those of HPIV3, Sendai virus, SV5 and NDV, and the deduced amino acid composition of BPIV3 F protein. Underlines and dashed underlines indicate hydrophobic regions and regions well conserved by all viruses, respectively. Enclosed squares are the putative carbohydrate attachment sites. Arrow indicates the putative cleavage site. (.) ; amino acid identical with that of BPIV3. (*) ; amino acid of similar nature to that of BPIV3 (R and K, S and T, D and E, and, I, L and V).

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

Combination	Protein	
	F	HN
BPIV3 - HPIV3	80	77
BPIV3 - Sendai	44	47
BPIV3 - SV5	23	21
BPIV3 - NDV	22	21
BPIV3 - Measles		<10
Sendai - SV5	24	22
Sendai - NDV	22	23
Sendai - Measles		<10
SV5 - NDV	32	32
SV5 - Measles		<10
NDV - Measles		<10

non-glycosylated form and estimated one of the glycosylated form being 60K and 75K (31), respectively. Its amino acid sequence, together with those for the F proteins of Sendai virus (6, 14, 15), HPIV3 (18), SV5 (20) and NDV (23) are presented in Fig. 4, where the amino acid sequences are aligned by insertion of minimal gaps to illustrate the maximum homology between the viruses. The deduced amino acid composition of the F gene product of BPIV3 is also presented in Fig. 4.

Comparison of the sequences showed that 80 % of amino acids were conserved between BPIV3 and HPIV3, and that the former was only one amino acid residue longer than the latter. The most different region was located near the carboxyl terminal. In contrast, the overall homologies between BPIV3 and Sendai virus, between BPIV3 and SV5, and between BPIV3 and NDV are 44 %, 23 % and 22 %, respectively (Table 1). However, all these F proteins shared common structural characteristics. In all the most hydrophobic region was near the carboxyl terminal (underlined in Fig. 4), which is thought to act as the anchor in the viral envelope (6, 14, 15, 18, 20, 23). Another hydrophobic sequence, which was detected at the amino terminal of the F1 proteins of Sendai virus, SV5 and other paramyxoviruses, and is thought to act as a functional domain during membrane fusion (32, 33), was also well conserved in BPIV3 and HPIV3 (underlined in Fig. 4), strongly suggesting that the Arg residue, which was immediately followed by this sequence and preceded by Lys, was also the cleavage site (indicated by arrow in Fig. 4)

for the BPIV3 F protein. This basic amino acid layout was identical to that of the cleavage site of the HPIV3 F protein (18). Another common hydrophobic region, located right at the amino terminal end, and ending in Cys followed by Gln (also underlined in Fig. 4), was most probably, based on evidence from Sendai virus (14), the signal peptide.

Moreover, not only 10 Cys residues but also 7 Gly and 4 Pro residues were completely conserved at corresponding positions in these five viruses. These amino acids are important in determination of tertiary structure of proteins, since Cys forms intramolecular disulfide linkages, and both Gly and Pro promote turns in the secondary structure (34).

In addition, four short but well-conserved regions were noted in all the F proteins (dashed underlines in Fig. 4).

As to the presumed Asn-linked carbohydrate attachment sites, which are also indicated in Fig. 4 (the carbohydrate signal sequences in the anchoring domain were neglected), there was considerable variation, and none was conserved among these five viruses, although one site was shared by BPIV3, HPIV3 and Sendai virus, and another by BPIV3, HPIV3, SV5 and NDV. Even between the closely related BPIV3 and HPIV3, each had two and one unique sites, respectively, although they shared three common sites.

Amino acid sequence of the HN protein and its comparison with those of other paramyxoviruses

The HN protein of BPIV3 consisted of 572 amino acid residues. Its molecular weight deduced for the non-glycosylated form was 65K while that estimated for the glycosylated form was 76K (13). The deduced amino acid sequence of the HN protein of BPIV3, as well as those reported for Sendai virus Z strain (6), HPIV3 (19), SV5 (21) and NDV (22) are presented in Fig. 5 in the same manner as employed for alignment of the F proteins. The deduced amino acid composition of the BPIV3 HN protein is also presented in Fig. 5. The HN proteins of BPIV3 and HPIV3 had identical numbers of amino acid residues, and BPIV3 showed 77 %, 47 %, 21 % and 21 % homology with HPIV3, Sendai virus, SV5 and NDV, respectively (Table 1). A region of about 100 amino acids proximal to the amino terminal showed the most difference between BPIV3 and HPIV3, a tendency which was much more obvious between BPIV3 and Sendai virus, SV5 or NDV.

Similarly, as with the F proteins, the HN proteins of these five viruses shared common structural features, despite the variation in amino acid homology among the viruses. The most hydrophobic region in all the viruses was located near the amino terminal (underlined in Fig. 5), which is

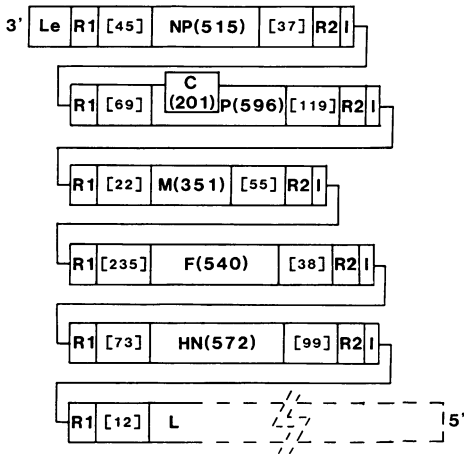


Fig. 6. Schematic illustration of the BPIV3 genome RNA. The region having been elucidated to date is written in solid line. Numbers in [] and () denote nucleotide numbers in non-coding regions and amino acid numbers of the gene products, respectively. It was determined that in the L gene a large reading frame opened after the 3' proximal non-coding region consisting of 12 nucleotides excluding R1 (data not shown). Le; leader region, R1 and R2; repeating consensus sequences, and I; intergenic trinucleotide (GAA).

thought to act both as an anchor in the viral envelope and as a signal peptide for membrane translocation (6, 16, 17, 19, 21, 22). Furthermore, 9 Cys, 12 Gly, 6 Pro and 6 Tyr residues were aligned in corresponding positions. Moreover, 9 corresponding positions were occupied by basic residues, Arg or Lys, and 12 by Ser or Thr. In addition, as marked by dashed underlines in Fig. 5, seven small regions were fairly well conserved in these five viruses, and were therefore thought to have roles in the hemagglutinating and neuraminidase activities.

On the other hand, no Asn-linked carbohydrate attachment signal was conserved in the viruses. All three presumed carbohydrate attachment sites (Asn-Pro-Thr or Ser was excluded, and the signal sequence in the region which seemed to be on the inside of the viral envelope was neglected) of HPIV3 were shared by BPIV3, but the latter had two additional sites, one of which was shared by Sendai virus.

It is very interesting to note that the H protein of measles virus (24) showed only chance homologies (less than 10%) with the HN proteins of these viruses.

DISCUSSION

In the present study, we determined the genome structure of the F and HN gene regions of BPIV3 910N strain by analysis of cDNA copies constructed from genomic RNA. Thus, together with the results in our preceding paper, the genome has been elucidated, without gaps, from its 3' end to the terminal point of the HN gene, covering 8,638 nucleotides. This region is

schematically illustrated in Fig. 6. The corresponding region of HPIV3, apart from the M gene, has already been determined (18, 19, 35, 36, 37, 38, 39).

The F and HN genes of BPIV3 share the universal feature of genes of paramyxoviruses in that they are flanked at the 3' end by the starting consensus sequence, here named R1, and at the 5' end by the terminating consensus sequence, R2, which is regarded as the signal for polyadenylation. A characteristic feature common to BPIV3 and HPIV3 (18) is that the 3' non-coding region of the F gene, from the R1 to the opening of the coding frame, is extremely long and U-rich, though it is more marked in BPIV3. This could cause the viral RNA polymerase to incorrectly recognize the polyadenylation signal of the M gene, as described in the preceding paper.

The high F and HN amino acid homologies of about 80% observed between BPIV3 and HPIV3 indicate that the viruses are closely related. However, the homologies are not as high as those observed between virus strains of Sendai virus, which are more than 95 % (6). Furthermore, differences in the presumed carbohydrate attachment sites and the nucleotide sequences of non-coding regions of these genes indicate that they are not substrains of the same virus. These findings are consistent with the previous findings of Abinanti et al. (10) that guinea pig antisera against BPIV3 and HPIV3 showed much higher neutralization and hemagglutination inhibition titers against the homologous than against the heterologous virus, and also that in similar tests sera from HPIV3-infected children exhibited much higher titers against HPIV3. In this context, alteration of glycoprotein carbohydrate attachment sites may be important in antigenic differentiation between BPIV3 and HPIV3, as suggested by Skehel et al. for influenza virus (40) and Vandepol et al. for vesicular stomatitis virus (41).

In comparison of amino acid sequences of the F and HN proteins, the BPIV3-HPIV3 group showed considerably higher homologies with Sendai virus than with SV5 and NDV. As presented in the preceding paper, the NP and M proteins, both of which constitute the basic structure of virion, show most conservation between BPIV3 and Sendai virus, while the P and C proteins, which seem to be functionally important within infected cells, are not so well conserved. The degree of conservation of the F and HN glycoproteins, which mediate attachment and penetration of virus into the cells, is lower than that of the NP and M proteins, but higher than that of the P and C proteins, suggesting that the evolutionary shifts between viruses vary from

one group of proteins to another. This tendency is also detectable even between BPIV3 and HPIV3.

On the other hand, although the amino acid homology varies greatly depending on the combinations of BPIV3, HPIV3, Sendai virus, SV5 and NDV, the F as well as HN proteins show basic structures common to all these viruses. The distribution of important hydrophobic domains in F protein (signal peptide, amino terminal of F1 subunit and anchoring region) is very similar in all viruses. Moreover, 10 Cys, 7 Gly and 4 Pro residues, which seem to be important in determination of protein structure, are conserved at corresponding positions. Four well-conserved short sequences (dashed underlines in Fig. 4) could have some role either in maintaining the F protein structure or in membrane fusion in concert with the amino terminal hydrophobic domain of the F1 subunit.

As for the HN proteins of these viruses, they all have their most hydrophobic domain near the amino terminal, which is thought to act as both signal peptide and anchor. Here again, 9 Cys, 12 Gly and 6 Pro residues are conserved at the same corresponding positions. Moreover, 9 basic residues (Arg or Lys), 12 hydroxy residues (Ser or Thr) and 6 Tyr residues are also conserved at corresponding positions, although their significance remains unclear. Furthermore, seven short regions (dashed underlines in Fig. 5) are locally well conserved in the HN proteins of these viruses, and are thought to act as functional domains in hemagglutination and hydrolysis of sialic acids. It is surprising that the H protein of measles (24) virus shows only chance homologies with any of the HN proteins of BPIV3, HPIV3, Sendai virus, SV5 and NDV. This may be a reflection of the fact that measles virus lacks neuraminidase activity and agglutinates red blood cells from only a limited number of animal species, in sharp contrast to parainfluenza viruses.

Thus, accumulating information is delineating the common basic structures as well as the evolutionary changes of the paramyxovirus glycoproteins, and it now seems essential to compare the genes of virus mutants carrying defects in activities mediated by the glycoproteins, in order to obtain more insight into functional domain structures.

During preparation of this manuscript, an amino acid sequence corresponding to the region between the 65th and 549th residue of the HN protein of the BPIV3 910N strain was reported for BPIV3 SF-4 strain (42). The main bulk of the sequence was virtually identical in the two strains, but the last region from the 523rd to 549th residue was completely

different. Our own analysis demonstrated that there was no such difference between the 910N, SC and M strains of BPIV3 (data not shown).

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. Hsu, M.-C., Scheid, A. and Choppin, P.W. (1979) *Virology* 95, 476-491.
2. Merz, D.C., Scheid, A. and Choppin, P.W. (1980) *J. Exp. Med.* 151, 275-288.
3. Long, L., Portetelle, D., Ghysdael, J., Gonze, M., Burny, A. and Meulemans, G. (1986) *J. Virol.* 57, 1198-1202.
4. Ray, R. and Compans, R.W. (1986) *Virology* 148, 232-236.
5. Rybeck, R., Oervell, C., Loeve, A. and Norrby, E. (1986) *J. gen. Virol.* 67, 1531-1542.
6. Shioda, T., Iwasaki, K. and Shibuta, H. (1986) *Nucleic Acids Res.* 14, 1545-1563.
7. Drzeniek, R., Boegel, K. and Rott, R. (1967) *Virology* 31, 725-727.
8. Shibuta, H., Kanda, T., Hazama, A., Adachi, A. and Matumoto, M. (1981) *Infect. Immun.* 34, 262-267.
9. Shibuta, H., Nozawa, A., Shioda, T. and Kanda, T. (1983) *Infect. Immun.* 41, 780-788.
10. Abinanti, F.R., Chanock, R.M., Cook, M.K., Wong, D. and Warfield, M. (1961) *P.S.E.B.M.* 106, 466-469.
11. Fischman, H.R. and Bang, F.B. (1966) *P.S.E.B.M.* 121, 966-970.
12. Scheid, A. and Choppin, P.W. (1977) *Virology* 80, 54-66.
13. Shibuta, H., Suzu, S. and Shioda, T. (1986) *Virology* 155, in press.
14. Blumberg, B.M., Giorgi, C., Rose, K. and Kolakofsky, D. (1985) *J. gen. Virol.* 66, 317-331.
15. Hsu, M.-C. and Choppin, P.W. (1984) *Proc. Natl. Acad. Sci. USA.* 81, 7732-7736.
16. Blumberg, B.M., Giorgi, C., Roux, L., Raju, R., Dowling, P., Chollet, A. and Kolakofsky, D. (1985) *Cell* 41, 269-278.
17. Miura, N., Nakatani, Y., Ishiura, M., Uchida, T. and Okada, Y. (1985) *FEBS lett.* 188, 112-116.
18. Spriggs, M.K., Olmsted, R.A., Venkatesan, S., Coligan, J.E. and Collins, P.L. (1986) *Virology* 152, 241-251.
19. Elango, N., Coligan, J.E., Jambou, R.C and Venkatesan, S. (1986) *J. Virol.* 57, 481-489.
20. Hiebert, S.W., Paterson, R.G. and Lamb, R.A. (1985) *J. Virol.* 54, 1-6.
21. Paterson, R.G., Harris, T.J.R. and Lamb, R.A. (1984) *Proc. Natl. Acad. Sci. USA.* 81, 6706-6710.
22. Millar, N.S., Chambers, P. and Emmerson, P.T. (1986) *J. gen. Virol.* 67, 1917-1927.
23. McGinnes, L.W. and Morrison, T.G. (1986) *Virus Res.* 5, 343-356.
24. Alkhatib, G. and Briedis, D. (1986) *Virology* 150, 479-490.
25. Okayama, H. and Berg, P. (1982) *Mol. Cell. Biol.* 2, 161-170.
26. Okayama, H. and Berg, P. (1983) *Mol. Cell. Biol.* 3, 280-289.

27. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA.* 74, 5463-5467.
28. Messing, J. (1983) *Methods Enzymol.* 101, 20-78.
29. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
30. Wong, T.C. and Hirano, A. (1986) *J. Virol.* 57, 343-348.
31. Shibuta, H., Kanda, T., Nozawa, A., Sato, S. and Kumanishi, T. (1985) *Arch. Virol.* 83, 43-52.
32. Gething, M.J., White, J.M. and Waterfield, M.D. (1978) *Proc. Natl. Acad. Sci. USA.* 75, 2737-2740.
33. Richardson, C.D., Scheid, A. and Choppin, P.W. (1980) *Virology* 105, 205-222.
34. Chou, P.Y. and Fasman, G.D. (1974) *Biochemistry* 13, 211-222.
35. Dimock, K., Rud, E.W. and Kang, Y. (1986) *Nucleic Acids Res.* 14, 4694.
36. Galinski, M.S., Mink, M.A., Lambert, D.M., Wechsler, S.L. and Pons, M.W. (1986) *Virology* 149, 139-151.
37. Sánchez, Á., Banerjee, A.K., Furuichi, Y. and Richardson, M.A. (1986) *Virology* 152, 171-180.
38. Luk, D., Sánchez, Á. and Banerjee, A.K. (1986) *Virology* 153, 318-325.
39. Spriggs, M.K. and Collins, P.L. (1986) *J. Virol.* 59, 646-654.
40. Skehel, J.J., Stevens, D.J., Daniels, R.S., Douglas, A.R., Knossow, M., Wilson, I.A. and Wiley, D.C. (1984) *Proc. Natl. Acad. Sci. USA.* 81, 1779-1783.
41. Vandepol, S.B., Lefrancois, L. and Holland, J.J. (1986) *Virology* 148, 312-325.
42. Coelingh, K.J., Winter, C.C., Murphy, B.R., Rice, J.M., Kimball, P.C., Olmsted, R.A. and Collins, P.L. (1986) *J. Virol.* 60, 90-96.