
Sequence of 2,617 nucleotides from the 3' end of Newcastle disease virus genome RNA and the predicted amino acid sequence of viral NP protein

Nobuhiro Ishida, Hideharu Taira¹, Toshiko Omata¹, Kiyohisa Mizumoto², Seisuke Hattori, Kentaro Iwasaki¹ and Masao Kawakita

Department of Pure and Applied Sciences, College of Arts and Sciences, University of Tokyo, Meguro-ku, Tokyo 153, ¹Department of Physiological Chemistry, Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113 and ²Department of Chemistry, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108, Japan

Received 11 June 1986; Revised and Accepted 29 July 1986

ABSTRACT

DNA fragments complementary to the Newcastle disease virus genome (strain D26) were cloned and sequenced. The sequence of 2,617 nucleotides from the 3' end of the genome was determined and an open reading frame (OP-1) consisting of 1,467 nucleotides, most likely encoding NP protein, was found in this region. This was followed by a second unfinished open reading frame (OP-2) of at least 729 nucleotides which continued beyond the 2,617th nucleotide. Another relatively short (312 nucleotides long) open reading frame (OP-2') was found overlapping with OP-2, but its significance is still unclear. The amino acid sequence deduced from the nucleotide sequence of OP-1 showed a moderate homology to that of the NP protein of Sendai virus in the central portion of the peptide. The leader sequence of 53 nucleotides was also identified.

The 5' end of mRNAs synthesized in the infected cells was analyzed and found to be m⁷GpppA, suggesting that the transcription of viral mRNAs starts with A, but not with G residue.

INTRODUCTION

In recent years understanding on the structure of paramyxovirus genome has been extended by means of the analysis of DNA clones complementary to viral RNAs. The genome of many paramyxoviruses including Newcastle disease virus (NDV), Sendai virus (HVJ), and measles virus is a single continuous RNA molecule of approximately 15 kilobases long with negative polarity (1). The most well characterized among these is HVJ RNA (2-11), whose complete nucleotide sequence has recently been determined (3,7,11). Nucleotide sequences of several individual genes of other viruses were also reported based mainly on the structure of cDNA clones synthesized by using viral mRNAs as templates for reverse transcriptase (12-20).

Structure of NDV genome has been less characterized as compared with the viruses cited above, so that there has not yet appeared any report describing the complete structure of any one of the genes, although the sequence of 3'-proximal 247 nucleotides (21) and a partial structure at the junction of the HN and L genes (22) have been quite recently reported.

In order to understand the mechanisms of the gene expression and replication of NDV and to analyze the effect of NDV infection on the physiology of host cells, we have started the analysis of the structure of the NDV genome. In this report we present the sequence of 2,617 nucleotides from the 3' end of the NDV strain D26 genome, in which the NP gene and a part of the P gene are found to be located. We also suggest the presence of a possible nonstructural protein gene which is overlappingly encoded within the P gene region and seems to be the counterpart of C gene in HVJ and measles virus.

MATERIALS AND METHODS

Virus

Strain D26 of NDV, an avirulent strain isolated from duck (23,24), was kindly supplied by Dr. Y. Nagai of Nagoya University. The virus was propagated in the allantoic cavity of chicken eggs, and purified by discontinuous sucrose density gradient centrifugation. Virions concentrated between 40 and 48 % sucrose layers were recovered.

Preparation of viral RNA

Virion RNA was extracted from purified virus by phenol in the presence of 1 % SDS. Aqueous phase containing RNA was then extracted with phenol-chloroform-isoamyl alcohol (25:24:1), and RNA was precipitated with ethanol. RNA was further purified by 15-30 % sucrose density gradient centrifugation.

Viral mRNAs were prepared from infected mouse SRCDF₁-DBT cells (25,26). The cells were infected with NDV strain D26 at a multiplicity of 1 pfu/cell and incubated at 37° in Eagle's minimum essential medium containing 5 % calf serum. At 9 h postinfection, they were harvested and RNA was extracted according to the method described in ref. 27. To prepare ³²P-labeled viral mRNAs, chick embryo primary culture cells were employed. The cells in a plate (9 cm in diameter) were infected with NDV strain D26 at a multiplicity of 75 pfu/cell, and incubated for 24 h at 37° in phosphate-free Eagle's minimum essential medium containing 3 % dialyzed calf serum, 2 µg/ml actinomycin D, and 10 mCi [³²P]orthophosphate. RNA was extracted as above and fractionated by oligo(dT)-cellulose chromatography into components with and without poly(A) tail.

Synthesis and molecular cloning of complementary DNA (cDNA)

Virion RNA was polyadenylated with ATP : RNA adenylyltransferase essentially according to Sippel (28). Double-stranded cDNAs were

synthesized essentially according to the procedures described in ref. 27. Polyadenylated RNA was isolated by using oligo(dT)-cellulose column and recovered by precipitation with ethanol. This RNA was used as the template for cDNA synthesis with reverse transcriptase, in which oligo(dT) was served as a primer. Resulting DNA-RNA hybrids were alkali-treated and double stranded DNAs were synthesized with the Klenow fragment of DNA polymerase I of Escherichia coli. The double stranded cDNAs were purified by Sephadex G-50 chromatography, treated with S1 nuclease, and purified again by Sephadex G-50 chromatography. They were then fractionated by sucrose density gradient centrifugation and fractions containing chains of 2-5 kilobase pairs long were pooled. Oligo(dC) tails were attached to the cDNAs with terminal deoxynucleotidyl transferase according to Hoeijmakers et al. (29), and the tailed cDNAs were inserted into the Pst I cleavage site of pBR322 utilizing the Pst I-cut and oligo(dG)-tailed pBR322 as a vector. E. coli HB101 was transformed with the recombinant DNAs, and tetracycline-resistant colonies were selected as transformants. Plasmid DNAs were isolated according to Clewell and Helinski (30) using Triton X-100, precipitated with polyethyleneglycol, and purified by CsCl/ethidium bromide density gradient centrifugation.

Sequencing of cDNA

The cDNAs were cleaved into fragments with appropriate restriction endonucleases and each fragment was sequenced by the dideoxy chain termination method of Sanger et al. (31) after subcloning of the fragment into M13 phage (32). A part of cDNA was also sequenced by the method of Maxam and Gilbert (33).

Identification of mRNA cap structure

Poly (A)⁺ viral RNAs labeled with [³²Pi] were prepared from NDV-infected cells as described above and digested with RNase mixture containing 2 µg/ml RNase A, 25 u/ml RNase T1 and 20 u/ml RNase T2 for 3 h at 37°. The hydrolysate was loaded on a column of DEAE-Sephadex A-25 and the cap-containing oligonucleotide fraction was separated from mononucleotides by a step elution with 0.2 and 1 M triethylamine-HCO₃⁻ buffer (pH 7.8). The 1 M triethylamine-HCO₃⁻ eluate containing highly negative-charged (-3 to -6) materials (m⁷GpppNp, m⁷GpppNmpNp, m⁷GpppNmpNmpNp, or pppNp) was lyophilized and further treated with nuclease P1 and alkaline phosphatase to liberate cap cores, i.e. m⁷GpppN or m⁷GpppNm, and was analyzed for cap structure by paper electrophoresis on Whatman DE81 at pH 3.4 and thin-layer chromatography on polyethyleneimine-cellulose plate as described(34).

Enzymes and other materials

Following enzymes were purchased: ATP : RNA adenylyltransferase from Pharmacia P-L Biochemicals, Uppsala; avian myeloblastosis virus reverse transcriptase from Wako Pure Chemicals Industry, Osaka; S1 nuclease from Seikagaku Kogyo, Tokyo; Klenow fragment of *E. coli* DNA polymerase I, terminal deoxynucleotidyl transferase and restriction endonucleases from Takara Shuzo Co., Kyoto. Pst I-cut and oligo(dG)-tailed pBR322 was purchased from Bethesda Research Laboratories, Inc., M13 mp10 and mp11 and M13 sequencing kit were from Amersham. [³²P]orthophosphate was obtained from Japan Atomic Energy Research Inst., Tokyo, [α -³²P]dCTP from New England Nuclear, and [α -³⁵S]dATP from Amersham.

RESULTS

Characterization of cDNA clones

Double stranded cDNAs to NDV genomic RNA were synthesized as described under Materials and Methods, and then inserted into the Pst I-cleavage site of plasmid pBR322. After transforming *E. coli* HB101 with these plasmids, 2,824 tetracycline-resistant transformants were obtained. Of these, 8 clones were found to have inserts which were complementary to NDV RNA when tested by Southern blot hybridization (35) with ³²P-labeled viral RNA fragments as probes (data not shown), and at the same time gave relatively large (>2,000 base pairs) fragments on digestion with Pst I.

These cDNA clones were classified into two groups of closely related ones on the bases of their restriction maps and of cross hybridization analyses by the Southern technique. One of these groups, which is composed of 3 cDNA clones was characterized in detail, as described below, and turned out to represent the structure of the 3'-terminal portion of the viral genome. The other, which includes five members, collectively covers a continuous stretch of about 4.5 kilobases long which corresponds most likely to a part of L protein gene. This is suggested by the fact that a significant homology is found between the amino acid sequence of HVJ-L protein and that of the putative product of this gene, based on the partial nucleotide sequence of the cDNAs so far determined (Ishida, N., unpublished result).

The nucleotide sequence of cDNA clone XXIV12A, a representative of the former group, was analyzed. The sequencing strategy is shown in Fig. 1. To assure the accuracy of the nucleotide sequence, both strands of the cDNA was analyzed except for positions 1-100 and 2,610-2,617. For these

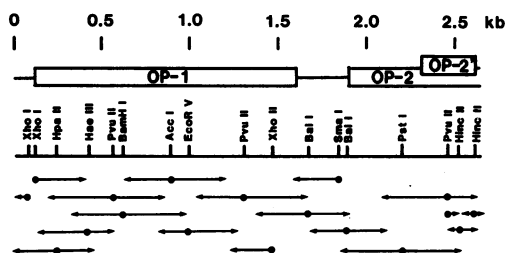


Fig. 1. The sequencing strategy for the insert of cDNA clone XXIV12A. The restriction sites used in obtaining cDNA fragments for subcloning in M13 vector are shown. The arrows indicate the cDNA fragments sequenced and the direction of sequencing. Approximate positions of large open reading frames (see also Fig. 2) are also shown.

portions, only one strand was analyzed, but the accuracy was assured by determining the sequences of two independent M13 clones having different cDNA inserts prepared by using different restriction endonucleases. The nucleotide sequence of a Pvu II cleavage product covering position 1,302-2,466 was further analyzed according to Maxam and Gilbert, because stretches of repeated G residues were frequently found between positions 1,500 and 1,750, and the M13 sequencing method alone might lead to an erroneous result for a region having such a structure. The nucleotide sequence of the 3'-proximal region of NDV genome, deduced from the sequence of cDNA clone XXIV12A, is shown in Fig. 2.

Structure of NDV RNA as inferred from the nucleotide sequence of cDNA clone XXIV12A

The unique sequence of the cloned cDNA XXIV12A was flanked by poly(dG) and poly(dC) tails, but poly(dA) tail was not found for unknown reasons. However, the nucleotide sequence adjacent to the poly(dC) tail of one strand was closely similar to the one recently reported for the sequence of the 3'-terminal portion of the genomic RNA of Beaudette C strain (21). Especially, the sequence from position 1 to 15 was exactly identical. Furthermore, the sequence from position 1 to 8 coincided with the 3'-terminal sequence of HVJ (2,3,6) and measles virus (12). These results strongly suggest that the nucleotide sequence shown in Fig. 2 represents the structure of the 3'-terminal region of the NDV strain D26 genome.

A large open reading frame composed of 1,467 nucleotides was found in the sequence, extending from position 122 to 1,588 (OP-1). This encodes a peptide of 489 amino acid residues with a calculated molecular weight of

	10	20	30	40	50	60	70	
3'	UGG	UUG	UUC	UUC	UUC	UUC	UUC	70
	CUU	AGACA	UUU	UUU	UUU	UUU	UUU	
	CUU	AGAGCUC	ACGCUCGGCC	UUUGAGUUUG	AGCUCUCUUG	GAAGAUGUUU	UUCAGAUCC	140
	UCA	UGUCUG	CGAGGAGCGC	CGAGUCUCGAG	CGGGUUUACC	UCGAGUACGC	CUCUCUCUCU	210
	GAA	UUUCAU	CUUCAGGGCC	AUAAGUGAGA	GUUGUCACUA	CUGGGUCUCU	UAUCUACCUU	280
	AAA	ACAGAAG	CCUAAGCACA	AUGCCUCCUA	CGGUUGUUUG	GUGAUCCGU	UCCACGAGAG	350
	AUA	CGAGAGU	GAGAGUCCAC	UACUCCUUGG	UACAACGGGA	ACGCCCCUUU	GUCUUAUCC	420
	ACA	AGAACUC	UAGCUACCA	AAUGGUUGCC	GCAUGGGGUC	AAGUUGUUGU	CUCUACCUCA	490
	UCU	CGUGUCU	CUAAGUACUA	CUAUGGUCCC	AGAGAAGGAG	CCCGUACGUC	GUUGCCUAGG	560
	GU	CGACCCCA	ACUUCUACUA	CGUGGUUCUC	UGUAUUGACU	AUGGGACUCC	UCCUAGGAGA	630
	AGU	UCAUACC	CAGUCCACAC	GUUUCGGGUA	CUGAGGUUA	CUCUGUGUC	UACUGAGUCU	700
	UAG	UUUAUCA	UGUACGGUGU	UCUGUCCGAG	GUCUUCUUCA	UGUAGGAGGU	GGGGUCUUG	770
	AGG	UUGAGUG	UUAGUCUGUC	AGAGACCGCC	AGGGUUGAA	GAACCAUUCG	CUGAAUUCU	840
	GUG	CCGUCCA	CCUUGGAGGU	GGAAUUGUU	GAACCAUCC	CUACAUUGA	GUUAGUAGC	910
	GAU	UGACGUA	AGAAGGACUG	UGAAUUUAUA	CCUUAUUGU	GGUUCUGUAG	UCCGGAACGU	980
	AGAC	UGCCGU	AUAGUUUUU	UACUUCGUG	AGUACGGGAA	CAUAGCCUAC	UUUCCUCUUA	1050
	GUAC	UCUGAUC	GAACCAUGU	CACUGGUCUA	CUCGAAACGU	GGACGGCUCA	UACGGUUUGA	1120
	CGG	UACCCAU	ACCGUAGUCA	GAUCUUAUU	CCUUGACCGU	UUUUGGUUAA	ACGGUCCUCU	1190
	GUAG	UAAGAC	CUCUGAACCU	CAUCUCAUGC	GAGUCCGAGU	UCCUUAUCCG	UAGUUAUCC	1260
	GCUC	GAUUUC	GAUUGGGUC	GUCGUUCCUC	UCCGACCGU	CGACGACGGG	UUUCUCACAG	1330
	UCG	UCGUUUC	UGUACGGGUG	GGUUGUUCGG	CCCCAGGAGU	GACUAGAGUC	GUCUCCUCCA	1400
	GGG	UCCACG	UGACUUGUCU	AGUGUUCGCC	UUGGCCUGUG	GCUCCGUCU	CUCGGGUA	1470
	CUAC	UCUUGC	CACCGUUUAU	CGUACUCUCU	UCCGGGUUUG	AGACGGUCC	CGUUGGGAGU	1540
	GGGG	UUGAG	GACCCGGGAG	AGUUCUGUUA	CUGUGGCUGA	CCUCCAUAC	UCCGUGUUGU	1610
	AAG	GUACUUU	AGUAGGGUUG	AGGAGACGGG	CGUUGGGUGG	GGAGUUAGGC	GUUAGGGCCU	1680
	GUG	UUUGCUU	GGGGAAACAG	AGGGAGGAGA	GGGAGUCGGG	GUGUUUGGGU	GGACGGUCC	1750
	GUG	UACCGU	GGGUGAUUAU	UAGUUUAGUC	CCGGUUUCUU	UAUCUUUUUU	UCAUCCCAU	1820
	UA	AGUCUCUA	GUCCCGCUCA	GUGGGCCAG	AGAAGAGAGG	GAAGAUGGAU	CAUUGGUCC	1890
	GGG	UAAAUG	UCUACGCCUC	UAGCUCUUG	AUAAACUCUG	GUCACCUUGA	CAGUAAUCU	1960
	CCG	GUUCCU	UUUGGUCAUC	UCUGGCAACC	UUUCCUACGU	UAGGGUUGUC	CGUUUUGAUU	2030
	CGU	CGUACCC	UCUUCGUGCC	CUCGAGGUC	AGUGGUCCGU	CGGUCCUGUG	GGACUCCGCC	2100
	GUC	UUUUUG	UGACAGGUGU	GGCCUCGUUC	GUCAGGUUUU	GUCUGCCGCC	GGUGACUGUU	2170
	CGG	GGGAUGA	GUCGACGUC	UACUUCGGCC	GUCUGUGUGC	GAGUCCUGGC	CUCGUUCGUU	2240
	AGC	UACGAAC	UAUUUGAGUC	GUUAUUCAGU	AGAUUACGAU	UUUUCCGGG	UACCGACUCC	2310
	CCG	UAGUAGU	UGCAGACUGA	GUUGUUGUCC	CCUCAGUUGG	UUCCGCCUCC	UUUCAGUUC	2380
	CUU	GGUCCGG	UUCCAGUAGG	GACUUUUGGA	CCAGUGUCUC	CGCUUGUGUC	GUUAGUACC	2450
	CCG	AGUGUUG	AUAGUCGACC	ACGUUGGGGA	GUACGAGAGG	CUAGUCUUGU	CUCGGUUCUC	2520
	GUG	GACACCU	AGUACAGGUC	GAUGGACAGC	UGAAACACGU	CCCGUACUAC	AGAUACUACC	2590
	UGUC	UCCCAU	UCAUUUCAAC	UGAUUGU	(2,617)			

Fig. 2. The nucleotide sequence of 2,617 nucleotides from the 3' end of the NDV (strain D26) genome deduced from that of cDNA clone XXIV12A. OP-1, OP-2 and OP-2' indicate large open reading frames. See text for expositions of N1 and N2.

53,161, whose size is consistent with that of NP protein of NDV strain D26 (Mr. ca. 52k as estimated by SDS-polyacrylamide gel electrophoresis). The second open reading frame, beginning at position 1,887 and continuing beyond the 2,617th nucleotide (OP-2) was also detected. It is noteworthy that another open reading frame (OP-2') of 312 nucleotides, which is overlapped

```

OP-1
MSSVFDEYEQ LLAAQTRPNG AHGGGEGKST LKVEVPVFTL NSDDPEDRWN FAVFCLRIAV 60
SEDANKPLRQ GALISLLCSH SQVMRNHVAL AGKQNEATLA VLEIDGFTNG VPQFNRRSGV 120
SEERAQRFMF IAGSLPRACS NGTPFVVTAGV EDDAPEDITD TLERILSIQA QVWVTAKAM 180
TAYETADESE TRRINKYMQQ GRVQKKYILH PVCRSAILT IRQSLAVRIF LVSELKRGRN 240
TAGGTSTYYN LVGDVDSYIR NTGLTAFFLT LKYGINTKTS ALALSSLSGD IQKMKQLMRL 300
YRMKGDNAPY MTLGLDSDQM SFAPAAYAQL YSFAMGMASV LDKGTGKYQF ARDFMSTSFV 360
RLGVEYAQAQ GSSINEDMAA ELKLTAAARR GLAAAAQRVS EETSSIDMPT QQAGVLTGLS 420
DGGSQAPQGA LNRSQQQPD TGDGETQFLDL MRAVANSMRE APNSAQGTPQ PGPPPTPGPS 480
QDNDTDWGY (489)

OP-2
MATPTDAEID ELFETSGTVI DSIITAQGKP VETVGRSAIP QGKTKALSAA WEKHGSVQSP 60
ASQDTPDRQD RSDKQLSTPE QASPNDSPPA TSTDQPPTQA ADEAGDTQLR TGASNSLLSM 120
LDKLSNKSSN AKKGPWSSPQ EGHHQRLTQQ QGSQPSRGN QERPQNAKV IPGNLVTAN 180
TAYHGQWEGS QLSAGATPHA LRSEQSQDNT PAPVDHVQLP VDFVQAMMSM MEAISQRVSK 240
VDY-----

OP-2'
MVEPSRRASS TSDSTTGEST KPRKQSRETA EPGQGHWPKP GHRREHSISW TMGGVTTISW 60
CNPSCSPIRA EPRQYSCTCG SCPATCRLCA GDDVYDGGDI TEGK (104)

```

Fig. 3. Predicted amino acid sequences of peptides encoded by the open reading frames presented in Fig. 2.

with OP-2, was found at position 2,291 to 2,602. The size of the peptide encoded by this gene (104 amino acids), however, does not correspond to any of virion components. It is possible that this open reading frame codes for a nonstructural viral protein.

Both OP-1 and OP-2 are preceded by a common sequence of UGCCCCAUUUCC (N1) which may be the NDV-counterpart of the consensus sequences which are found in HVJ as well as VSV genome (36) and form a part of the transcriptional initiation site. The sequence CA at positions 54 and 1802 may correspond to the intergenic sequence of HVJ and VSV genomes. The sequence N2 (AAUCUUUUUU) which precedes the second N1 sequence closely resembles the transcriptional termination signal of the HN gene of NDV strain Beaudette C (AUUCUUUUUU)(22), the HVJ genes (AUUCUUUUUU)(3), and the VSV genes (AUACUUUUUU)(36).

The predicted amino acid sequences of peptides potentially encoded by OP-1, OP-2 and OP-2' are presented in Fig. 3. A part of amino acid sequence of the OP-1 shows a moderate homology to that of NP proteins of HVJ (3,4) and measles virus (13) in the middle of the polypeptide chain, as shown in Fig. 4. When position 170 to 338 of NDV NP protein is compared with position 172 to 340 of HVJ and measles virus NP proteins the homology between them are 32 % and 39 %, respectively. Thus it is likely that OP-1 codes for NP protein.

```

1         10        20        30        40        50        60
HVJ  MA  GLLSTFDTFSS  SRSESINKS  GRGAVIPQQR  STVSVFVLGL  SVTDDADKLF  IATTFLAHSL
NDV   MSSVFDEYEQ  LLAQAQTRPAG  AHGGGEGKST  LKVEVPVFTL  NSSDDPEDRW  FAVFCLRIAV
measles MA  TLLRSLALFK  RNKDKPPIYS  GSGGAIINGIK  HIIIVIPPGD  SSITTRSRLL  DRLVRLIGNP

70         80        90        100       110       120       130
DTRKRSQSG  GFLVSLAMA  YSSPELAYLT  NGVNADVQYV  IYNIKDPKR  TKTDGFIVKT  RMEYERTTE
* * * * *
SEDANKPLRQ  GALISLLCSH  SQVMRNHVAL  AGKQNEATLA  VLEIDGFPTG  VPQFNNRSGV  SEERAQRFMM
* * * * *
DVSGPKLTGA  LIGILSLFVQ  SPQQLIQRIT  DDPDVSIRLL  EVVQSDQSQS  GLTFASRGIN  MEDEAQYFSS

140        150       160       170       180       190       200
WLFQPMVWKS  PLFQQRDAA  DPTLLIQIYG  YPACLGAIIV  QVWIVLVKAI  TSSAGLRKGF  FNRLEAFRQD
* * * * *
IAGSLPRACS  NGTFFPVTAGV  EDDAPEITD  TLERILSIQA  QVWIVAKAM  TAYETADESE  TRRINKYMQQ
* * * * *
HDDPISSDQS  RFGWFENKEI  SDIEVQDPBG  FNMILGTILA  QIWWLLAKAV  TAPDTAADSE  LRRWIKYQQQ

210        220       230       240       250       260       270
GTVKGALVFT  GETVEGIGSV  MRSQQSLVSL  MVETLVIMNT  ARSDLITLLEK  NIQIVGNYIR  DAGLASFMNT
* * * * *
GRVQKYLILH  PVCRSAIQLF  IRQSLAVRIF  LVSELRGRFN  TAGGTSTIYIN  LVGDVDSYLR  NIGLTAFFIL
* * * * *
RRVVGEPRLR  RKWLDVVRNI  IAEDLSLRRF  MVALILDIKR  TPGNKPRIAE  MICDIDTYIV  EAGLASFILT

280        290       300       310       320       330       340
IKYGVETKMA  ALITLSNLRP  INKLRSLDIT  YLSKGPRAFF  ICILKDFVHG  EFAPGNYPAL  WSYAMGVAVV
* * * * *
LKYGINTKTS  ALALSSLSGD  IQKMKQLMRL  YRMKGNAPY  MTLGDSQDM  SFAPAETAGL  YSFAMGMSV
* * * * *
IKPGIETMYP  ALGLHEPAGE  LSTLESIMDL  YQQMGKPAPY  MVNLENSIQN  KFSAGSYPLL  WAYAMGVGVE

350        360       370       380       390       400       410
QNKAMQYVIT  GRTYLDMEMF  LLGQAVAKDA  ESKISSALED  ELGVTEAARG  RLRHHLASLS  GQNGAYRKPT
* * * * *
LDKGTGKYQF  ARQFMSTSFV  RLGVEYAAQ  GSSINEDMAA  ELKLPAPARR  GLAAAQRVS  EETSSIDMET
* * * * *
LENSMGGLNF  GRSYPDPAYP  RLQDEMVRRS  AGKVSSTLAS  ELGITAEADR  LVSEIAMHIT  EDKISRACVP

420        430       440       450       460       470       480
GGGAIEVALD  NADIDLETKA  HADQDARGWG  GDGGERWARQ  VSGGHFVTLH  GAERLEEETN  DEDVSDIERR
* * * * *
QQAGVLTGLS  DGGSQAPQGA  INRSQQPDT  GDGETQFLDL  MRVANSMRE  APNSAQTPQ  PGPPPTPGPS
* * * * *
RQAGVSPLOQ  DQSENELPRL  GKKEDRRVKQ  SRGEARESYR  ETCPSRASDA  RAAHLPTGTP  LDIDTASESS

489
IAMRLAERR
QNDITDWGY
**
QQPQDSRRS

```

Fig. 4. Comparison of amino acid sequences of NP protein between NDV, HVJ and measles virus. Conserved amino acid residues are indicated by asterisks.

Hybridization of cDNA fragments with viral mRNAs

As shown in Fig. 5A, two cDNA fragments were prepared. The one having ca. 700 base pairs is produced by Pvu II digestion of XXIV12A DNA and covers the middle portion of OP-1. The other, having ca. 1,000 base pairs, is the (Sma I + Sca I) cleavage product of XXIV12A DNA and contains a stretch covering OP-2 together with a short segment of the vector DNA, but is devoid of any of the OP-1 portion. Each of these fragments was labeled with ³²P and used separately to detect mRNA species complementary to it. RNAs were

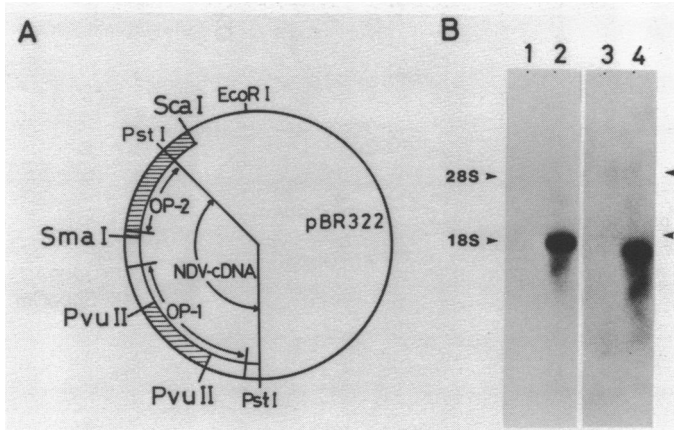


Fig. 5. (A). Restriction fragments used as probes (hatched) in the hybridization. (B). Hybridization of ^{32}P -labeled restriction fragments derived from OP-1 (1,2) and OP-2 (3,4) portion with RNAs from NDV-infected (2,4) and uninfected (1,3) SRCDF₁-DBT cells.

isolated from infected SRCDF₁-DBT cells, and resolved by neutral agarose gel electrophoresis after glyoxalation (37). Resolved RNAs were then tested for hybridization with each ^{32}P -labeled cDNA fragment by the Northern technique (38-40).

As shown in Fig. 5B, the fragment covering the OP-1 portion hybridized with 18S RNA species (lane 2), while the one containing the OP-2 portion with RNA which is significantly smaller than 18S (lane 4). These cDNA fragments, however, did not hybridize with mRNAs obtained from uninfected cells (lanes 1 and 3). According to Collins *et al.* (41), the 18S species and the one a little smaller than that are assumed to be mRNAs for NP and P proteins, respectively. These results therefore indicate that OP-1 and OP-2 of the NDV genome are transcribed in infected cells to form discrete species of mRNAs, and suggest that OP-1 and OP-2 represent the coding regions of NP and P proteins, respectively. This assignment is in agreement with the assumption deduced from the partial homology of the amino acid sequence of the putative OP-1 product to that of HVJ NP protein as described above. Furthermore, this is consistent with the gene order reported recently (22). Analysis of the 5'-terminal nucleotide of NDV mRNAs synthesized in infected cells

In a recent publication, Kurilla *et al.* (21) analyzed 5'-terminal nucleotide sequence of *in vitro*-synthesized NDV mRNAs and claimed that the mRNA species transcribed from the 3'-proximal gene (NP) starts with G. In

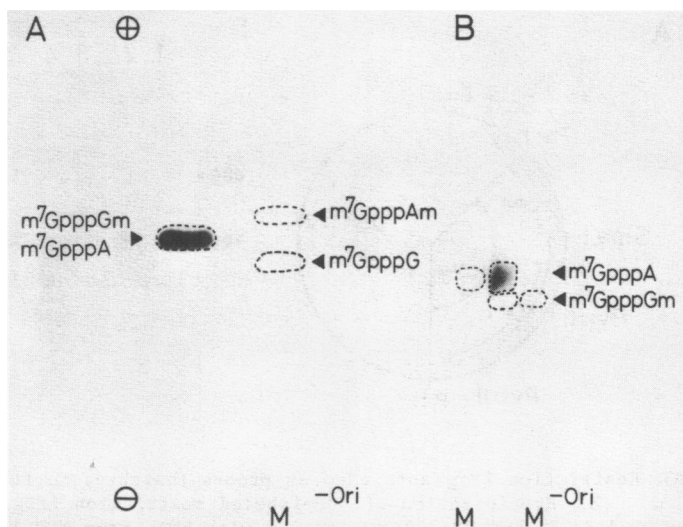


Fig. 6. Identification of the cap structure of NDV mRNA. (A). The cap nucleotides derived from ^{32}P -labeled poly(A)⁺ RNA from NDV-infected chick embryo cells were analyzed by paper electrophoresis on DEAE-cellulose at pH 3.4. (B). The radioactive spot in A was eluted and reanalyzed by polyethyleneimine-cellulose TLC using 0.6 M LiCl. The dotted circles indicate the position of authentic cap markers. M denotes cap markers alone.

contrast, mRNAs of other paramyxoviruses such as HVJ and measles virus were reported to start with A (5,42) and have m^7GpppAm at their 5'-termini (Mizumoto, K., unpublished observation). We therefore analyzed the cap structure of NDV strain D26 mRNAs synthesized in infected cells. The cap cores, *i.e.* m^7GpppN or m^7GpppNm , were isolated from uniformly ^{32}P -labeled mRNA and were analyzed as described in Materials and Methods. Fig. 6A shows the results of paper electrophoresis of a nuclease P1 and alkaline phosphatase-treated sample on a DEAE-cellulose sheet. Almost all of the radioactivity co-electrophoresed with authentic m^7GpppGm and m^7GpppA which could not be separated each other under these conditions, but was clearly separated from the m^7GpppG marker. The radioactive material was eluted from the electropherogram and was further analyzed by thin-layer chromatography. As shown in Fig. 6B, the ^{32}P label comigrated with the m^7GpppA marker. The results presented in Fig. 6 substantiate that the cap structure of strain D26 mRNA is m^7GpppA but neither m^7GpppG nor m^7GpppGm , indicating that the transcription starts with A instead of G residue. The absence of 2'-O-

methylation at 5'-penultimate residue is in accordance with the previous finding of Colonno and Stone (43).

DISCUSSION

The sequence of 2,617 nucleotides from the 3' end of the genome of NDV strain D26 was deduced by analyzing cDNA clones obtained from the viral RNA template. The sequence of the 3'-proximal 53 nucleotides shown in Fig. 2 is quite similar to that of the leader RNAs of 47 and 53 nucleotides long reported by Kurilla *et al.* (21) except for substitution of 4 bases. In agreement with their notion, 8 nucleotides from the 3' end was exactly the same with HVJ and measles virus, which suggests that they are functionally important in the viral replication cycle, but the remainder of the leader sequence was not homologous to that of either HVJ (3,6) or measles virus (12).

Assuming N1 and N2 to be the transcriptional initiation and termination signals, respectively, by analogy with the structure of HVJ and VSV genomes, we tentatively assigned the sequence at position 56 to 1,801 as the first gene (gene 1) and the sequence at 1,803 to 2,617 as part of the second gene (gene 2). The sequence CA at positions 54 and 1,802 was assumed to be the intergenic sequence, since the analysis of the 5'-terminal cap structure revealed that the transcription of NDV genes starts with A.

It is highly likely that gene 1 codes for NP protein, since (i) predicted amino acid sequence deduced from the nucleotide sequence of the OP-1 shows a moderate homology with that of NP proteins of HVJ (3,4) and measles virus (12) (Fig. 4), (ii) the size of mRNA transcribed from this gene coincides with that of NP mRNA (Fig. 5), and (iii) this assignment is consistent with the so far accepted gene order (22). To our knowledge this is the first communication that reports the complete structure of any one of NDV genes.

As judged by the size of mRNA to which OP-2 cDNA fragment hybridized, gene 2 probably codes for P protein. It should be noted here that a cDNA clone to the mRNA of M protein, which was isolated and kindly supplied by Dr. T. Toyoda of Nagoya University did not hybridize with the cDNA insert of clone XXIV12A. Recently we have obtained a cDNA clone which has a partial overlapping with XXIV12A and also hybridizes with the clone complementary to the M-mRNA (Ohira, M., unpublished result). This suggests that the M gene is located downstream with respect to the P gene.

Although only about half of the structure of the gene 2 has so far been

determined, it shows an interesting feature that an open reading frame potentially encoding a peptide of 104 amino acids (OP-2') is found overlapping with the longer open reading frame (OP-2). The phase of two frames are different, so that two quite different peptides are encoded by these open reading frames. So far there has been no evidence that a peptide of this size was synthesized in NDV-infected cells. Collins *et al.* (41) reported that peptides of Mr. 33,000 and 36,000 were produced as byproducts of *in vitro* translation of P protein mRNA. However, they were unlikely related to the OP-2' product, since they gave similar peptide maps as P protein. On the other hand, it has been reported that within the P protein gene of HVJ and measles virus, an overlapping open reading frame was found (3,5,14) and was suggested to code for the C protein of these viruses. An analogy based on the similarity of the genomic construction among these viruses leads to a suggestion that OP-2' may be the NDV-counterpart of the C protein gene. In view of possible importance of C protein in the physiology of infected cells, it is interesting to see if this overlapping gene is expressed in the NDV-infected cells. The discovery and cloning of this gene will greatly help us to answer this question.

The 3'-proximal nucleotide sequences presented in this communication and those reported recently for the strain Beaudette C genome (21) are closely similar though not identical (for comparison see Fig. 7). Thus, only 10 out of 174 3'-proximal residues are different, and 4 of those discrepancies which occurred in the putative coding region of gene 1 turned out not to affect its predicted amino acid sequence. Discrepancies gradually increased beyond this point. However, if difficulties in sequencing a long stretch of RNA

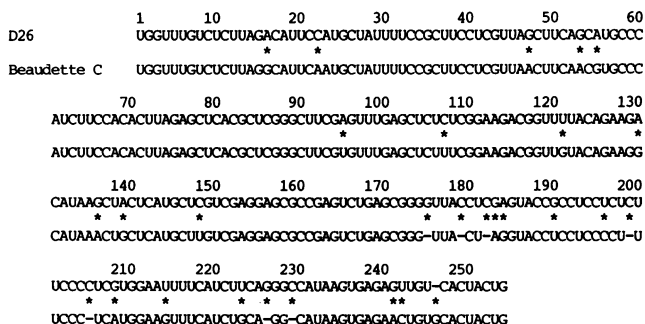


Fig. 7. Comparison of 3'-proximal RNA sequence of NDV genome between strains D26 and Beaudette C. Substitutions, deletions and insertions are indicated by asterisks.

chain directly is taken into account, we may not be able to take these differences at their face value. We should rather say that the nucleotide sequences of the genomes of two strains are quite similar at least in their 3'-proximal region. It is interesting to note that the common sequence N1 which precedes both OP-1 and OP-2 is also found at the beginning of gene 1 of the Beaudette C genome. The sequence N1 is homologous to the consensus sequence R1 found in the transcriptional initiation site of HVJ, if two insertions be admitted as follows, UGCCCCAUCUUUCC (NDV) and U-CCCA-CUUUC (HVJ), and this strongly suggests the importance of this sequence in transcription. However, it shows less homology to the sequence of the initiation site of L-mRNA of Beaudette C strain (22). This may suggest that either the initiation mechanism of the L gene is different from that of gene 1 and gene 2, or the proposed sequence of the L mRNA may not correspond to its 5' sequence. Since strain D26 is an avirulent strain while Beaudette C is virulent, and since the molecular basis for the difference in virulence must lie in the structure of their respective genomes, the comparison of the nucleotide sequences of these strains would be quite interesting. The cDNA cloning of the rest of NDV genome is currently in progress in this laboratory.

ACKNOWLEDGEMENT

We wish to express our thanks to Drs. H. Shibuta and T. Shioda for valuable discussions.

REFERENCES

1. Kolakofsky, D., Boy de la Tour, E. and Delius, H. (1974) J. Virol. **13**, 261-268
2. Re, G. G., Gupta, K. C. and Kingsbury, D. W. (1983) J. Virol. **45**, 659-664
3. Shioda, T., Hidaka, Y., Kanda, T., Shibuta, H., Nomoto, A. and Iwasaki, K. (1983) Nucleic Acids Res. **11**, 7317-7330
4. Dowling, P. C., Giorgi, C., Roux, L., Dethlefsen, L. A., Galantowicz, M. E., Blumberg, B. M. and Kolakofsky, D. (1983) Proc. Natl. Acad. Sci., U. S. A. **80**, 5213-5216
5. Giorgi, C., Blumberg, B. M. and Kolakofsky, D. (1983) Cell **35**, 829-836
6. Gupta, K. C. and Kingsbury, D. W. (1984) Nucleic Acids Res. **12**, 3829-3841
7. Hidaka, Y., Kanda, T., Iwasaki, K., Nomoto, A., Shioda, T. and Shibuta, H. (1984) Nucleic Acids Res. **12**, 7965-7973
8. Blumberg, B. M., Rose, K., Simona, M. G., Roux, L., Giorgi, C. and Kolakofsky, D. (1984) J. Virol. **52**, 656-663
9. Blumberg, B. M., Giorgi, C., Rose, K. and Kolakofsky, D. (1985) J. Gen. Virol. **66**, 317-331
10. Blumberg, B., Giorgi, C., Roux, L., Raju, R., Dowling, P., Chollet, A. and Kolakofsky, D. (1985) Cell **41**, 269-278
11. Shioda, T., Iwasaki, K. and Shibuta, H. (1986) Nucleic Acids Res. **14**, 1545-1563

12. Billeter, M. A., Baczko, K., Schmid, A. and ter Meulen, V. (1984) Virology **132**, 147-159
13. Rozenblatt, S., Eizenberg, O., Ben-Levy, R., Lavie, V. and Bellini, W. J. (1985) J. Virol. **53**, 684-690
14. Bellini, W. J., Englund, G., Rozenblatt, S., Arnheiter, H. and Richardson, C. D. (1985) J. Virol. **53**, 908-919
15. Barrett, T. and Mahy, B. W. J. (1984) J. Gen. Virol. **65** 549-557
16. Russel, S. E. H., Clarke, D. K., Hoey, E. M., Rima, B. K. and Martin, S. J. (1985) J. Gen. Virol. **66**, 433-441
17. Rozenblatt, S., Eizenberg, O., Englund, G. and Bellini, W. J. (1985) J. Virol. **53**. 691-694
18. Paterson, R. G., Harris, T. J. R. and Lamb, R. A. (1984) Virology **138**, 310-323
19. Paterson, R. G., Harris, T. J. R. and Lamb, R. A. (1984) Proc. Natl. Acad. Sci., U. S. A. **81**, 6706-6710
20. Hiebert, S. W., Paterson, R. G. and Lamb, R. A. (1985) J. Virol. **54**, 1-6
21. Kurilla, M. G., Stone, H. O. and Keene, J. D. (1985) Virology **145**, 203-212
22. Chambers, P., Millar, N. S., Bingham, R. W. and Emmerson, P. T. (1986) J. Gen. Virol. **67**, 475-486
23. Yamane, N., Odagiri, T., Arikawa, J., Morita, M., Sukeno, N. and Ishida, N. (1978) Japan. J. Med. Sci. Biol. **31**, 407-415
24. Nagai, Y., Yoshida, T., Hamaguchi, M., Naruse, H., Iinuma, M., Maeno, K. and Matsumoto, T. (1980) Microbiol. Immunol. **24**, 173-177
25. Kumanishi, T. (1967) Japan. J. Exp. Med. **37**, 461-474
26. Hirano, N., Fujiwara, K. and Matumoto, M. (1976) Japan. J. Microbiol. **20**, 219-225
27. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory
28. Sippel, A. E. (1973) Eur. J. Biochem. **37**, 31-40
29. Hoeijmakers, J. H. J., Borst, P., Van den Burg, J., Weissmann, C. and Cross, G. A. M. (1980) Gene **8**, 391-417
30. Clewell, D. B. and Helinski, D. R. (1969) Proc. Natl. Acad. Sci., U. S. A. **62**, 1159-1166
31. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci., U. S. A. **74**, 5463-5467
32. Messing, J. (1983) Methods in Enzymol. **101**, 20-78
33. Maxam, A. M. and Gilbert, W. (1980) Methods in Enzymol. **65**, 499-560
34. Mizumoto, K. and Lipmann, F. (1979) Proc. Natl. Acad. Sci., U. S. A. **76**, 4961-4965
35. Southern, E. M. (1975) J. Mol. Biol. **98**,503-517
36. Rose, J. K. (1980) Cell **19**, 415-421
37. Carmichael, G. G. and MacMaster, G. G. (1980) Methods in Enzymol. **65**, 380-391
38. Thomas, P. S. (1980) Proc. Natl. Acad. Sci., U. S. A. **77**, 5201-5205
39. Thomas, P. S. (1983) Methods in Enzymol. **100**, 255-266
40. Wahl, G. M., Stern, M. and Stark, G. R. (1979) Proc. Natl. Acad. Sci., U. S. A. **76**, 3683-3687
41. Collins, P. L., Wertz, G. W., Ball, L. A. and Hightower, L. E. (1982) J. Virol. **43**, 1024-1031
42. Alkhatib, G. and Briedis, D. J. (1986) Virology **150**, 479-490
43. Colonna, R. J. and Stone, H. O. (1976) Nature **261**, 611-614