Molecular Cloning of Complementary DNA to Newcastle Disease Virus, and Nucleotide Sequence Analysis of the Junction between the Genes Encoding the Haemagglutinin–Neuraminidase and the Large Protein

By PHILIP CHAMBERS,* NEIL S. MILLAR, RICHARD W. BINGHAM† AND PETER T. EMMERSON

Department of Biochemistry, University of Newcastle upon Tyne, Newcastle upon Tyne and
†Department of Veterinary Pathology, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, U.K.

(Accepted 20 November 1985)

SUMMARY

Complementary DNA clones to 90% of the Newcastle disease virus (NDV) genome have been produced and mapped. These clones cover the entire HN, F and M genes, most if not all of the L gene and parts of the NP and P genes. The map of overlapping clones gives the gene order 3′-NP-P-M-F-HN-L-5′ for NDV, identical to the gene order of Sendai virus, on the assumption that the NP gene of NDV is at the 3′ end of the genome as previously suggested by inactivation of NDV transcription by u.v. light. The nucleotide sequence of 453 bases covering the junction between the HN and L genes has been determined. There is nucleotide sequence homology to the consensus polyadenylation and mRNA start sites of Sendai virus and vesicular stomatitis virus. The deduced amino acid sequence of the C terminus of the HN protein of NDV shows homology to the C-terminal amino acid sequences of the HN proteins of simian virus 5 and Sendai virus. An explanation for the presence of HN0, the precursor to HN in some strains of NDV, is suggested by the presence of a long non-coding region at the 3′ terminus of the mRNA encoding the HN protein of NDV that could, by mutation, allow synthesis of a larger polypeptide.

INTRODUCTION

Newcastle disease virus (NDV) is a typical paramyxovirus and causes a severe respiratory infection in poultry. This disease is of great economic importance, requiring control by vaccination or quarantine with slaughter of all birds in confirmed outbreaks. The genome of NDV is a single strand of RNA of negative polarity, molecular weight 5·2 × 10⁶ to 5·6 × 10⁶, or approximately 15000 bases (Kolakofsky et al., 1974b), bound to three proteins in the viral nucleocapsid. These proteins are the nucleocapsid protein NP (Mountcastle et al., 1970), the phosphoprotein P (Chambers & Samson, 1980; Smith & Hightower, 1981) and the large protein L (Hightower & Bratt, 1974). The nucleocapsid is contained within a lipid envelope derived from host cell plasma membranes (Klenk & Choppin, 1969, 1970), on the inner surface of which is a shell of membrane or matrix protein M (Shimizu & Ishida, 1975; Rott & Klenk, 1977). On the outer surface of the viral envelope are the two viral glycoproteins, the haemagglutinin–neuraminidase HN, involved in virus binding to host cells (Choppin & Compans, 1975) and the fusion protein F, involved in fusion with and penetration through the membrane (Homma & Ohuchi, 1973; Nagai et al., 1976). Antibodies raised against either glycoprotein of the paramyxovirus simian virus 5 (SV5) neutralize viral infectivity (Merz et al., 1981). NDV-infected cells contain two virus-coded non-structural proteins 36K and 33K, probably encoded on the same mRNA as P (Chambers & Samson, 1982; Collins et al., 1982). Molecular cloning and nucleotide sequence determination of NDV was undertaken to obtain further information about the organization of the genome of this important virus, which is the type species of the genus Paramyxovirus of the family Paramyxoviridae.
Materials. NDV strain Beaudette C was obtained from A. C. R. Samson, University of Newcastle upon Tyne, U.K. NDV strain Hitchner B1 was obtained from J. B. McFerran, Veterinary Research Laboratories, Stormont, U.K. Madin and Darby bovine kidney (MDBK) cells were obtained from Flow Laboratories. Escherichia coli strain DH1 was obtained from P. Meacock, University of Leicester, Leicester, U.K. Pall Biodyne A membranes were obtained from Pall Process Filtration, Portsmouth, U.K. NDV strain Hitchner B1 was obtained from J. B. McFerran, Veterinary Research Laboratories, Stormont, U.K. NDV strain Beaudette C was obtained in eggs and purified by centrifugation (Chambers & Samson, 1980). DNA was extracted by digestion of NDV (approx. 5 mg protein) with 2 mg protease K in a buffer containing 10 mM-vanadyl ribonucleoside complexes, 0-15 mM-NaCl, 1% SDS, 12 mM-EDTA, 0.1 mM-Tris-HCl pH 7.5, at 37 °C for 1 h, followed by three extractions at 56 °C with phenol/0-1% 8-hydroxyquinoline. Virion DNA was concentrated into a final volume of 100 µl H2O by ethanol precipitation. Complementary DNA (cDNA) synthesis was performed on 1 to 2 µg virion RNA with 2 µg hexanucleotide primer and 20 units reverse transcriptase in a buffer containing 100 mM-NaCl, 8 mM-MgCl2, 20 mM-2-mercaptoethanol, 1 µCi [3H]dCTP, 50 µM-deoxynucleoside triphosphates, 50 mM-Tris-HCl pH 8.3 for 30 min at 37 °C, then 90 min at 42 °C.

DNA:DNA hybrids were desalted in a 2 ml Sephadex G-100 column, ethanol-precipitated, and resuspended in a final volume of 10 µl H2O. RNA:DNA hybrids were tailed with oligo(dC) by incubation for 20 min at 37 °C with 10 units terminal transferase in a buffer containing 140 mM-potassium cacodylate pH 6-9, 1 mM-CoCl2, 25 µM-dCTP and 1 µCi [3H]dCTP (a far higher proportion of radionucleotide was incorporated in this second synthetic reaction): Oligo(dC)-tailed RNA:DNA hybrids were run through a 2 ml Sephadex column in annealing buffer (100 mM-NaCl, 0.2 mM-EDTA, 10 mM-Tris-HCl pH 7.5). One-hundred µl oligo(dC)-tailed hybrids in annealing buffer was annealed to 1 µl (0.25 µg) oligo(dG)-tailed PstI-cut pBR322 at 65 °C for 5 min, 45 °C for 2 h, cooled to room temperature overnight, then kept on ice for a further 24 h (the optimal proportions of hybrid: pBR322 were determined by small scale annealings and transformations. The annealed mixture was transformed into competent E. coli strain DH1 which were plated onto agar containing 10 µg/ml tetracycline.

More than 75% of the transformants were ampicillin-sensitive and presumably contained inserts at the PBR322 PstI site. Transformants containing NDV-specific inserts were detected by colony hybridization on nitrocellulose filters (Grumentin & Hogenesis, 1975). The probe was made by labelling cDNA with [32P]-dATP in a reverse transcription similar to that described above, except that non-radioactive dATP and [3H]dCTP were omitted and 20 µCi [32P]-dATP was included. Twenty to 30% of the radiisotope was incorporated into probe in a typical reaction. Filters were prehybridized for 3 h at 65 °C in a solution containing 5 × Denhardt’s solution (Denhardt, 1966), 6 × SSC (i.e. 0.9 M-NaCl, 90 mM-sodium citrate), 50 µg/ml boiled sheared calf thymus DNA, then hybridized to boiled probe overnight at 65 °C in a similar solution. Filters were washed three times for 1 h at 65 °C in 3 × SSC/0.1% SDS, dried, then exposed to X-ray film (Kodak NS 59T) at −70 °C.

Construction of a map of cloned inserts Four techniques were used to map the cloned inserts with respect to each other and to their positions in the NDV genome.

1. Dot blot hybridization. Twenty-seven clones were selected for study based on the intensity of hybridization to [35S]-labelled cDNA and insert sizes which were estimated from small scale plasmid preparations (Birnboim & Doly, 1981). Larger scale plasmid isolations were performed and plasmid DNAs were linearized with HindIII, boiled to separate DNA strands, made up to 2 × SSPE (i.e. 0.36 M-NaCl, 2 mM-EDTA, 20 mM-NaH2PO4 pH 8.3) and dotted onto nitrocellulose filters in a known pattern, then baked under vacuum at 80 °C for 2 h to bind DNA to nitrocellulose. The pattern of DNA dots was then probed with a preparation of insert cDNA which had been cut from a particular plasmid with PstI, purified on an agarose gel followed by electroelution and phenol extractions (McDonnell et al., 1977), then labelled by nick translation with [32P]-dATP. Hybridization conditions, washing and exposure to X-ray film were identical to those described above for screening transformants.

2. Northern blot hybridization. Selected plasmids were mapped to their corresponding mRNAs by Northern blot hybridization. MDBK cells were infected with NDV strain Hitchner B1 at 1000 EID50/90 mm Petri dish. Twenty-four h post-infection, cell monolayers were rinsed with PBS (137 mM-NaCl, 8.1 mM-Na2HPO4, 2.7 mM-KCl, 1.5 mM-KH2PO4), then lysed with 4-2 M-guanidinium thiocyanate, 0.5% Sarkosyl L, 25 mM-sodium citrate, 0.3% Antifoam A (Sigma) pH 7.0. The lysate was layered over a cushion of 3.7 M-caesium chloride in 25 mM-sodium acetate pH 5.0 and centrifuged overnight to pellet total cellular RNA at 35000 r.p.m. in a MSE 6 × 15 ml rotor (Chirgwin et al., 1979). The pelleted RNA was resuspended in 5 mM-EDTA, 1% SDS, 10 mM-Tris–HCl pH 7.4,
Molecular cloning of NDV

extracted with 4:1 chloroform:butanol and precipitated at −20 °C after the addition of 0.1 vol. 3 M-sodium acetate pH 5.5 and 2.2 vol. ethanol (Maniatis et al., 1982). A control preparation was made from uninfected cells. RNA was then denatured with 50% formamide, 10% formaldehyde at 60 °C for 5 min and electrophoresed on a 1.8% agarose gel containing 6.5% formaldehyde and 20 mM-sodium phosphate buffer pH 7.0 (Rave et al., 1979). Alternate lanes were loaded with infected and uninfected cellular RNA (10 µg/lane) arranged so that, after blotting, filters could be cut into strips with one lane of each of infected and uninfected material per strip. RNA was transferred onto Pall Biodyne A membranes without any staining or pretreatment of the gel, in the manner described by Thomas (1980). From each batch of strips, one was probed with viral genomic RNA, hydrolysed with alkali and end-labelled with [α-32P]ATP by polynucleotide kinase (Clark et al., 1982). This was the control to show all the mRNA bands. The other strips were individually probed with the appropriate plasmid DNA labelled with [α-32P]dCTP by nick translation. Conditions for hybridization were based on the method of Thomas (1980). Filters were prehybridized by incubation for 4 to 6 h at 42 °C in 50% deionized formamide, 2.5 × Denhardt’s solution, 2 × SSPE, 0.375% SDS, 250 µg/ml boiled salmon sperm DNA. Hybridization was carried out overnight (18 to 20 h) at the same temperature in a similar solution containing the probe. Filters were washed four times for 15 min at room temperature in 2 × SSC, 0.1% SDS, and once for 15 min at room temperature in 1 × SET (150 mM-NaCl, 2 mM-EDTA, 30 mM-Tris-HCl pH 8.0) and then twice for 20 min at 68 °C in 0.5 × SET. The damp filters were sealed in thin polyethylene bags and exposed to X-ray film (Kodak X-Omat S) at room temperature.

3. Restriction enzyme mapping. Plasmids that overlapped, as determined by dot blot hybridization, were mapped by digestion with a panel of restriction enzymes. In the first instance, PstI, EcoRI, HindIII, BamHI, PvuII and AarI sites were mapped in cloned inserts, and in many cases this provided sufficient data to align inserts with respect to each other. Where necessary, additional restriction enzyme sites were mapped to confirm overlaps suggested by dot blot hybridization.

4. Colony hybridization. The bank of NDV-specific clones was probed with nick-translated α-[32P]-labelled insert cDNA to find clones that extended the regions mapped in the initial dot blot hybridizations. Clones were streaked onto nitrocellulose, grown up on agar overnight at 37 °C, lysed and baked (Grunstein & Hogness, 1975), then hybridized to the probe, washed and dried as for dot blot hybridization described above. Any clones that hybridized to the probe were analysed to determine the size of the insert in the plasmid (Birnboim & Doly, 1981). In promising cases, larger scale plasmid preparations were performed and restriction enzyme sites were mapped to determine whether the newly selected plasmids extended the regions present in the probe insert.

DNA sequencing. Clones 1.13 and 3.73 overlapped each other and both contained the junction between the HN and L genes of NDV. These clones were selected for DNA sequence analysis, with emphasis on the region of overlap. DNA was subcloned into the vectors M13 mp8 and mp9 (Messing & Vieira, 1982) using restriction enzyme sites mapped in cloned inserts, and in many cases this provided sufficient data to align inserts with respect to each other. Where necessary, additional restriction enzyme sites were mapped to confirm overlaps suggested by dot blot hybridization.

RESULTS

Construction of the colony bank
The selection of transformants with NDV-specific inserts is shown in Fig. 1. After transformation of E. coli strain DH1 with the annealed mixture of oligo(dC)-tailed RNA-DNA hybrids and oligo(dG)-tailed PstI-cut pBR322, cells were plated onto agar containing tetracycline. Transformants that grew up in 2 days at 37 °C were streaked onto nitrocellulose filters on agar plates, grown up, lysed and baked, then probed with α-[32P]-labelled NDV-specific cDNA. The pattern of radioactivity detected on one such filter is shown in Fig. 1. Dark streaks were presumed to contain NDV-specific inserts, and the corresponding clones were restreaked onto fresh master plates.

A bank of 700 NDV-specific clones was constructed in two stages. The first 300 clones were analysed before the remaining clones were produced. Most of the inserts were in the range 500 to 1000 base pairs, although a few inserts were larger than 2000 base pairs.

Mapping of NDV-specific clones
Twenty-seven plasmids were selected from the first 300 NDV-specific clones on the basis that they hybridized strongly to the cDNA probe and had relatively large inserts, usually 1000 to 2000 base pairs. These plasmids were analysed by dot blot hybridization, and results for the plasmid
Fig. 1. Use of colony hybridization to select transformants with NDV-specific inserts. One-hundred transformants were streaked onto a nitrocellulose filter laid on an agar plate containing tetracycline. After growth, lysis, baking and hybridization to α-35S-labelled NDV cDNA probe, the filter was exposed to X-ray film.

Fig. 2. Use of dot blot hybridization to map NDV-specific clones. DNAs from 28 plasmids, linearized with HindIII and boiled, were dotted onto a nitrocellulose filter in a known pattern, baked, hybridized to α-35S-labelled insert DNA from plasmid 2.87, washed, dried, then exposed to X-ray film. Plasmids on the filter were (1) 1.07, (2) 1.09, (3) 1.13, (4) 1.17, (5) 1.20, (6) 1.34, (7) 1.43, (8) 1.44, (9) 2.25, (10) 2.11, (11) 2.34, (12) 2.44, (13) 2.66, (14) 2.73, (15) 2.86, (16) 2.87, (17) 3.01, (18) 3.23, (19) 3.51, (20) 3.54, (21) 3.73, (22) 3.78, (23) 3.79, (24) 3.82, (25) 3.93, (26) 4.08, (27) 4.28, (28) pBR322. Plasmids 1.07 (1), 1.09 (2), 1.17 (4), 1.34 (6), 2.25 (9), 2.44 (12), 2.86 (15), 2.87 itself = 16), 3.51 (19) and 3.54 (20) gave a positive reaction to the probe and were judged to overlap plasmid 2.87. Faint hybridizations to other plasmids were not reproducible.

2.87 are shown in Fig. 2. This plasmid hybridizes to itself, and also to plasmids 1.07, 1.09, 1.17, 1.34, 2.25, 2.44, 2.86, 3.51 and 3.54. Twenty-two out of 27 plasmids were mapped into two non-overlapping groups by means of dot blot hybridizations. The remaining five did not hybridize to each other or to these two groups of plasmids. Of these five, two mapped to other regions of NDV (3.01 and 3.93), one was a clone of ribosomal RNA (2.66), and the other two (1.43 and 4.28) remain unidentified.

Northern blot hybridization was used to determine the coding assignments of NDV-specific clones. Individual NDV mRNAs can be resolved by gel electrophoresis on the basis of their mol. wt. and the protein-coding assignments of these mRNAs have been determined by in vitro translation (Collins et al., 1982). Thus, it is possible to deduce which genes are represented in a particular plasmid by the mRNAs to which it hybridizes, with the proviso that the NP and F mRNAs are not well resolved and a plasmid that hybridized to those mRNAs could not be assigned to the NP or F genes on these data alone. Results from plasmids 2.87, 3.73, 2.73, 3.01, 4.77, 4.26, 3.48 and 3.93 are shown in Fig. 3. Genomic (G), uncharacterized (U) and the various mRNA species are indicated at the left-hand side of the figure, and the two bars on the right indicate the relative positions of 28S (upper) and 18S (lower) ribosomal RNAs. Lanes (a) and (b) show NDV-infected and uninfected MDBK cell RNA respectively, probed with end-labelled alkaline-digested virion RNA. Plasmid 2.87 (lane c) represents one of the original two non-overlapping groups of inserts located entirely within the L gene. Plasmid 3.73 (lane d) represents the other group of plasmids, containing the junction of the HN and L genes. Other genes represented are HN (plasmid 2.73, lane e), F (plasmid 3.01, lane f), NP (plasmid 4.77,
Fig. 3. Use of Northern blot hybridization to map plasmid DNAs to specific mRNAs from NDV-infected MDBK cells. Total cytoplasmic RNA extracts from MDBK cells, either uninfected or infected with NDV strain Hitchner B1 24 h previously, were electrophoresed on agarose under denaturing conditions and transferred to Pall Biodyne A membranes. The filters were cut into strips each bearing two lanes representing the pair of infected and uninfected cells, and probed with either 32P-labelled viral genomic RNA or 32P-labelled plasmid DNA. The photograph is a composite of a representative selection of such strips. For clarity, the lanes from uninfected cells have been omitted, except for (b) which they all resembled. The letters on the left identify the RNA bands: G, viral genomic RNA; U, two uncharacterized bands, possibly polycistronic transcripts; L, HN, F, NP, P and M, the six mRNA bands. The two bars on the right indicate the relative positions of 28S (upper) and 18S (lower) ribosomal RNAs. (a, b) RNA from infected and uninfected cells respectively, probed with viral genomic RNA. (c to j) RNA from infected cells only, probed with the following plasmids: (c) 2.87; (d) 3.73; (e) 2.73; (f) 3.01; (g) 4.77; (h) 4.26; (i) 3.48; (j) 3.93.

lane g), P (plasmid 4.26, lane h), the P-M junction (plasmid 3.48, lane i) and M (plasmid 3.93, lane j). No hybridization to adjacent lanes containing RNA from uninfected cells was apparent (not shown). Clones 3.01 and 4.77 were mapped to F and NP respectively only when gaps in the continuous series of cloned inserts between L and M were filled, and clone 3.01 was found to lie between HN and M, that is in the F gene by analogy to Sendai virus (Hidaka et al., 1984; Blumberg et al., 1985a). The insert in plasmid 4.77 lies outside this region, as expected since all u.v. transcriptional mapping studies place the NP gene at the 3′ end of the NDV genome (Ball et al., 1978; Collins et al., 1978, 1980). Clones to fill gaps between the regions mapped in dot blot and Northern blot hybridizations were selected by colony hybridization.

The two RNA bands designated U (for uncharacterized) contain polycistronic transcripts (Thomas et al., 1978; Wilde & Morrison, 1984) and also appear to contain partial transcripts or degradation products of the L gene, as some L-specific plasmids hybridize to this region of the gels (clone 2.87, lane c). Leaving aside this L-specific hybridization, these two RNA bands can be tentatively assigned to the NP, P, M and F genes because HN-specific plasmids do not hybridize to this region of the gels (lane e). NP- and F-specific sequences appear only in the larger U band (lanes f and g), whereas P and M sequences appear in both bands (lanes h, i and j). Since the F mRNA is slightly larger than that for NP (lanes f and g), and the P mRNA is larger than that for M (lanes h and j), a NP-P bicistronic transcript would probably be similar in size to
a M–F bicistronic transcript, but a P–M bicistronic transcript would be smaller. Thus, the larger U band probably contains both NP–P and M–F bicistronic transcripts, and the smaller U band probably contains P–M bicistronic transcripts. This explanation is consistent with the gene order (Fig. 4) and the observation that plasmids specific for NP and F hybridize only to the larger U band, whereas plasmids specific for P or M hybridize to both.

By means of these types of experiment, a map of NDV-specific cloned inserts was drawn up (Fig. 4), and the order of the P, M, F, HN and L genes was deduced. It has been assumed that the NP gene is at the 3’ end of the viral genome, as suggested by transcriptional mapping (Collins et al., 1980) and Northern blot hybridization as discussed above. The entire HN, F and M genes have been cloned, together with most if not all of the L gene. Plasmids specific for the NP and P genes were identified, but do not complete the cloning of the 3’ end of the viral genome. Cloned regions identified cover about 90% of the NDV genome.

**DNA sequencing of the HN–L gene junction**

The HN–L gene junction was identified in plasmids 1.13 and 3.73 relatively early in this work. DNA sequence analysis was started in order to compare the sequence at this gene junction to those determined for Sendai virus (Gupta & Kingsbury, 1984), and vesicular stomatitis virus (VSV) (McGeoch, 1979; Rose, 1980). A partial restriction map of the cloned HN–L junction showing the sites used for DNA sequence analysis and the location of the sequences determined.
Fig. 5. DNA sequence analysis of plasmids 1.13 and 3.73. (a) Partial restriction map of cDNA inserts from clones 1.13 and 3.73, showing the restriction enzyme sites used in force-cloning to M13, and the lengths and orientations of DNA sequences obtained (horizontal arrows). A scale in kilobases is shown above. The bold vertical arrow indicates the position of the junction of the HN (to the left) and L (to the right) genes deduced from the sequence. Plasmid 1.13 extends through the region shown, and the asterisk indicates the left-hand end of the cDNA insert from plasmid 3.73. (b) Portion of a DNA sequencing gel showing the region presumed to be the junction between the HN and L genes. The DNA sequence is indicated at the right-hand side of the figure.

is shown in Fig. 5(a), with a portion of a DNA sequencing gel showing the proposed polyadenylation site at the 5' end of the HN gene and the proposed start of the L gene shown in Fig. 5(b). The nucleotide sequence of 453 bases around the gene junction, together with the deduced C-terminal 74 amino acids of the NDV HN protein and a comparison with the SV5 and Sendai virus HN coding sequences in the corresponding region are shown in Fig. 6. There is a substantial conservation of amino acid sequence with the C termini of the HN proteins of SV5 (Hiebert et al., 1985) and Sendai virus (Blumberg et al., 1985b) (20 amino acids conserved for SV5; 16 for Sendai virus). The longest stretch of uncharged amino acids in the region of the NDV HN protein shown is only nine amino acids, of which four have polar side chains. This is neither long enough nor sufficiently hydrophobic to span the membrane (Kyté & Doolittle, 1982), which suggests that the HN protein of NDV, like those of Sendai virus and SV5, is not anchored to the membrane by its C terminus.

The HN proteins of NDV strains Ulster and Queensland are synthesized in a precursor form (HN0) which is cleaved to active HN by the removal of a C-terminal glycopeptide (Nagai et al., 1976; Nagai & Klenk, 1977; Garten et al., 1980; Schuy et al., 1984). The HN mRNA of NDV strain Beaudette C has a longer non-coding region at the 3' end [167 bases between the stop codon and the proposed polyadenylation site (see below)] than that of either Sendai virus or SV5 (95 and 101 bases respectively). There is, therefore, a sufficient length of RNA in the non-coding region to increase the length of the HN protein of NDV by up to 55 amino acids if a new termination codon nearer the 3' end of the mRNA is used. A possible asparagine-linked glycosylation site Asn→Gln→Thr is present in a potential open reading frame in the non-coding region that terminates with the amino acids Val→Ala→Ser→(stop). These residues are indicated in Fig. 6. The putative glycosylation site and terminal amino acids are consistent with glycosylation of the C-terminal peptide and with some of the C-terminal amino acids of HN0 of NDV strain Ulster (Schuy et al., 1984).

A nucleotide sequence 5'-TAAGAAAAAA-3' has been found 3' to the open reading frame which encodes the HN protein and probably represents the polyadenylation site for the mRNA. This sequence is underlined in Fig. 6 and is compared to the similar conserved polyadenylation sites at Sendai virus and VSV gene junctions in Fig. 7(a). The proposed start of the L gene, 5'-AAGTGGCAATG-3', is also underlined in Fig. 6, and is compared to the L mRNA start regions of Sendai virus and VSV in Fig. 7(b). The assignment of the start of the L gene is based on (i) sequence homologies to mRNA starts in Sendai virus and VSV, (ii) the assumption that there is an untranscribed intergenic region of three bases in NDV, as is the case for Sendai virus (Giorgi et al., 1983; Gupta & Kingsbury, 1984), and (iii) a preliminary S1 nuclease mapping experiment (not shown). The L gene mRNA would, therefore, start with an adenine.
Fig. 6. DNA sequence at the junction of the HN and L genes of NDV. The DNA sequence of 453 bases from a PstI site in the HN gene to a RsaI site in the L gene is shown, beneath which is shown the deduced amino acid sequence. Underneath the NDV amino acid sequence, the corresponding regions of the HN amino acid sequences of SV5 and Sendai virus are shown. To maximize homology, a one amino acid insertion (of isoleucine, I), followed after three amino acid residues by a two amino acid deletion, has been proposed for the Sendai virus sequence relative to NDV and SV5. Residues in NDV, SV5 and Sendai virus that are conserved are boxed. Asterisks represent stop codons. Two tripeptides are indicated in a potential open reading frame in the non-coding region of HN of NDV: N-Q-T (AsnGlnThr), a glycosylation site, and V-A-S * [ValAlaSer(stop)], a new polypeptide terminus.

DISCUSSION

The cloning protocol gave a preponderance of clones at the 5' end of the viral genome. Clones of insert size 700 base pairs in the L gene cross-hybridized to about 50 other clones in the colony bank, whereas similarly sized clones in the HN or F genes cross-hybridized to about 20 other clones, and a similarly sized clone in NP cross-hybridized to three. Even with a positive strand RNA virus that has a poly(A) tract allowing priming of cDNA synthesis at the very 3' terminus complementary to the first genomic uracil after the polyadenylation site. Maximum homology is seen if 11 bases at the start of the NDV L mRNA are compared to 10 bases at the starts of Sendai virus and VSV L mRNAs. The NDV sequence shows similarity to both polyadenylation and mRNA start sequences of VSV and Sendai virus. This is of particular interest because no homology between mRNA start sequences of Sendai virus and VSV was detectable, although homology of polyadenylation signals has been found (Gupta & Kingsbury, 1984). The proposed start of the NDV L mRNA seems intermediate between the start sequences of Sendai virus and VSV L mRNAs. The apparent similarity in the L gene start regions reinforces the evidence for an evolutionary relationship between paramyxoviruses and rhabdoviruses deduced from the extensive sequence homology between the leader RNAs of NDV and VSV (Kurilla et al., 1985).
Molecular cloning of NDV

(a) Sendai consensus T A A G A A A A A A 
NDV HN gene T A A G A A A A A A A A A A A A 
VSV consensus T A T G A A A A A A A A A A A A

(b) Sendai L gene [ AA~C G ~T [G [A A T G~
NDV L gene [ A A G T G C A A T
VSV L gene A A C A A T I C

Fig. 7. Comparison of the nucleotide sequences of the polyadenylation and mRNA start sites at the junction of the HN and L genes of NDV to analogous sequences from Sendai virus and VSV. All sequences are written as cDNA in the positive sense. In (a), the sequence of the proposed polyadenylation site at the 5' terminus of the HN gene of NDV is compared to the consensus polyadenylation sequences from Sendai virus and VSV. In (b), the sequence of the proposed mRNA start of the NDV L gene is compared to the L mRNA start sequences from Sendai virus and VSV, with sequences aligned to maximize homology. Regions of conserved sequence in (b) are boxed.

of the viral genome, cloning of RNA :DNA hybrids results in an under-representation of the 3' terminus (Cann et al., 1983). With a random primer, this problem may be exaggerated if priming of cDNA synthesis is normally at an internal position in the genome. It may, therefore, be necessary to clone the 3' terminus of the NDV genome from the antigenome. To do so, it will be necessary to use the mixture of plus and minus sense RNA extracted from nucleocapsids from virus-infected cells, as virions of the Beaudette C strain used in this study do not contain antigenomes (Kurilla et al., 1985), although virions from some strains of NDV do contain a proportion of antigenomes (Robinson, 1970; Kolakofsky et al., 1974a; Kolakofsky & Bruschi, 1975). Despite this problem, 90% of the NDV genome has already been cloned.

Molecular cloning of the viral genome provided the gene order 3'-NP-P-M-F-HN-L-5' for Sendai virus (Dowling et al., 1983; Shioda et al., 1983; Hidaka et al., 1984; Blumberg et al., 1985a) which differed from that obtained from u.v. transcriptional mapping, 3'-NP-F-M-P-HN-L-5', because the coding assignments of mRNAs had to be guessed for the transcriptional mapping (Glazier et al., 1977). The gene order of NDV now appears to be 3'-NP-P-M-F-HN-L-5', on the basis of the cloning results discussed in this paper, which is identical to the gene order of Sendai virus, whereas the gene order for NDV suggested on the basis of the most recent u.v. transcriptional mapping was 3'-NP-P-M-HN-L-5' (Collins et al., 1980).

Nucleotide sequence analysis of the HN genes of Sendai virus and SV5 suggested that the products of these genes were bound to the viral membrane by a transmembrane region of hydrophobic amino acids at their N-termini (Blumberg et al., 1985b; Hiebert et al., 1985), like the influenza virus neuraminidase (Fields et al., 1981) but unlike the paramyxovirus F (Paterson et al., 1984; Blumberg et al., 1985a) and influenza virus haemagglutinin (Porter et al., 1979), which are anchored to the membrane at their C termini. Sequence determination of the NDV HN protein has not yet been completed, but the results presented in this paper show that there is amino acid sequence homology between the HN proteins of NDV, Sendai virus and SV5 and that the C terminus of the NDV HN protein lacks a long hydrophobic amino acid sequence. These results suggest that the structure of the HN protein of NDV is similar to those of Sendai virus and SV5 and, therefore, that the HN protein of NDV is anchored to the membrane by the N terminus, as suggested previously (Schuy et al., 1984). The presence of a long non-coding region at the 3' end of the mRNA for the HN protein of NDV strain Beaudette C suggests that the gene encoding the HN0 precursor for the HN protein of avirulent NDV strain Ulster may differ from the genes encoding the HN proteins of more virulent strains of NDV by mutations generating a longer open reading frame and the synthesis of a larger polypeptide.

NDV has long been considered a good model with which to study variations in viral pathogenicity because of the wide range of virulence of NDV strains (Waterson et al., 1967). When nucleotide sequence information from several strains or mutants of NDV becomes available, it may be possible to correlate features of the sequence with virulence as has been possible for poliovirus and rabies virus (Evans et al., 1983; Dietzschold et al., 1983).
It may be possible to insert cloned copies of the genes for the HN and F glycoproteins of NDV into a DNA vector such as a poxvirus, as has been achieved for the surface antigens of hepatitis B and herpes simplex viruses (Smith et al., 1983; Paoletti et al., 1984). This could provide an alternative vaccine to live attenuated strains of NDV. Such a vaccine would be particularly useful if the coding sequences of antigenic proteins from other avian pathogens were cloned into the same vector to create a polyvalent vaccine (Mackett et al., 1985).

R.W.B. is grateful to Dr J. O. Bishop, Department of Genetics, University of Edinburgh, for provision of laboratory facilities and much useful advice. P.C. thanks Dr A. C. R. Samson, Department of Genetics, University of Newcastle upon Tyne, for the use of laboratory facilities and for a critical reading of the manuscript. Kate Brown and Tony Pickard (Newcastle), Morag Robertson, Melville Richardson and Ian Bennet (Edinburgh) are thanked for technical assistance. This work was supported by grants from the Wellcome Trust (P.C. and P.T.E., Newcastle; R.W.B., Edinburgh) and Northern Cancer Research Campaign (P.C. and P.T.E., Newcastle). N.S.M. was supported by a SERC studentship.

REFERENCES


(Received 18 October 1985)