Nucleotide Sequence of the Fusion and Haemagglutinin–Neuraminidase Glycoprotein Genes of Newcastle Disease Virus, Strain Ulster: Molecular Basis for Variations in Pathogenicity between Strains

By NEIL S. MILLAR, PHILIP CHAMBERS† AND PETER T. EMMERSON*
Department of Biochemistry, University of Newcastle upon Tyne,
Newcastle upon Tyne NE1 7RU, U.K.

(Accepted 9 November 1987)

SUMMARY

The nucleotide sequences of the fusion (F) and haemagglutinin–neuraminidase (HN) glycoprotein genes of the extremely avirulent Newcastle disease virus (NDV) strain Ulster have been determined by sequencing cDNA clones derived from viral genomic RNA. Open reading frames, assumed to encode the Fo and HNo glycoprotein precursors, were 553 and 616 amino acids long, respectively. Comparisons of the two glycoprotein sequences with those of more virulent NDV strains suggested an explanation for the molecular basis of the wide-ranging differences in virulence observed between strains of NDV. The open reading frame corresponding to the Ulster HN glycoprotein extended beyond the C terminus of more virulent strains. This C-terminal extension was assumed to be responsible for the origin of the HN precursor (HNo) found in strain Ulster and other extremely avirulent strains of NDV. There were fewer basic amino acids at the cleavage site of Fo in strain Ulster than are present in more virulent strains, which may be responsible for the absence of cleavage and activation of Fo from this strain in many host cells. In more virulent strains of NDV, as well as in other paramyxoviruses, a phenylalanine residue occurs at the N terminus of the F1 cleavage fragment. The occurrence of a leucine residue at this position in strain Ulster may be partly responsible for the lack of virulence of this strain.

INTRODUCTION

Newcastle disease virus (NDV), a member of the family Paramyxoviridae, is responsible for the avian respiratory disease fowl pest. Of particular interest is the wide range of virulence displayed by different NDV isolates, which can cause a broad spectrum of disease ranging from a fatal to an essentially asymptomatic infection.

NDV contains two glycoproteins which are associated with the viral envelope. The haemagglutinin–neuraminidase (HN) glycoprotein has both haemagglutinating and neuraminidase activities which are responsible for attachment of virus to host cell receptors and receptor-destroying activity, respectively (Scheid & Choppin, 1973; Choppin & Compans, 1975). The fusion (F) glycoprotein is responsible for fusion with, and penetration through the host cell membrane (Homma & Ohuchi, 1973; Nagai et al., 1976).

It has long been thought that the degree of virulence displayed by NDV strains is determined by its two glycoproteins. In particular, it appears that viral infectivity is dependent upon the ability of host cell proteases to cleave the F0 precursor, generating the disulphide-linked F1 and F2 fragments (Homma & Ohuchi, 1973; Samson & Fox, 1973; Scheid & Choppin, 1974, 1977; Nagai et al., 1976). Cleavage of F0 may result in a conformational change which exposes the hydrophobic region at the N terminus of F1, thus promoting virus–host cell fusion (Kohama et al., 1981). In virulent NDV strains, F0 is cleaved in a wide range of host cell types in tissue culture, whereas in avirulent strains, it is only cleaved in a restricted range of cell types (Nagai et al., 1976).

† Present address: Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, U.K.
Extremely avirulent strains of NDV are unusual in their requirement for the cleavage of a precursor of HN (Nagai et al., 1976; Nagai & Klenk, 1977). This precursor (HN0) requires the proteolytic removal of a glycopeptide for the generation and activation of HN (Garten et al., 1980). In more virulent strains of NDV, as in other paramyxoviruses, the HN glycoprotein does not require cleavage for activation.

To understand further the molecular basis for the differences in virulence shown by strains of NDV we have constructed cDNA clones from viral genomic RNA of strain Ulster and determined the complete nucleotide sequence of the F and HN genes. The sequence has been compared with the previously published glycoprotein gene sequences of several other NDV strains.

**METHODS**

Clones were constructed from genomic RNA of NDV strain Ulster 2C, either as DNA-RNA hybrids, or as double-stranded (ds) cDNA. The methods used for synthesis of DNA : RNA hybrids were essentially as described for strain Beaudette C (Chambers et al., 1986a). The ds cDNA clones were synthesized using a cDNA synthesis kit (Amersham); however, the procedure and some of the components of the kit were changed as described below.

The first strand was synthesized in a total volume of 12 μl, using 6 μl (approximately 1 μg) NDV genomic RNA [as described by Chambers et al. (1986b)], 2 μg random hexanucleotide primer and the following components of the kit (full details of which were not supplied), 4 μl 5 × first strand synthesis buffer, 1 μl sodium pyrophosphate, 25 units human placental ribonuclease inhibitor, 2 μl deoxynucleoside triphosphate mix and 20 units reverse transcriptase. The mixture was incubated at 37 °C for 30 min then 42 °C for 40 min. The second strand was synthesized by the addition of 37.5 μl second strand synthesis buffer, 0.8 units RNase H, 23 units DNA polymerase I, to a final volume of 100 μl. The mixture was incubated at 12 °C for 1 h, 22 °C for 1 h and then 70 °C for 10 min. After cooling on ice, 2 units of T4 DNA polymerase were added and the mixture was incubated at 37 °C for a further 10 min. The reaction was stopped by the addition of 10 μl 0.25 M-EDTA and 10 μl 10% SDS. The ds cDNA was phenol-extracted, ethanol-precipitated, resuspended, tailed with oligo(dC), annealed to pBR322 and transformed into Escherichia coli, as described for the Beaudette C strain (Chambers et al., 1986a).

Clones containing NDV-specific inserts were identified by colony hybridization (Grunstein & Hogness, 1975) using a radioactively labelled probe derived from Ulster genomic RNA, as described for the Beaudette C clones (Chambers et al., 1986a). Clones corresponding to the F and HN genes were identified by screening the bank of NDV-specific clones with α-35S-labelled cDNA derived from strain Beaudette C, which had been previously mapped by Northern blot hybridization and DNA sequencing (Chambers et al., 1986a, b; Millar et al., 1986). In order to maximize hybridization between cDNA and the labelled probe, which were derived from different strains, the conditions used for this colony hybridization experiment were as described for the mapping of Beaudette C clones, but with the modification that a lower temperature was used for the hybridization (12 h at 65 °C, 2 h at 55 °C) and washing (55 °C) steps.

Restriction fragments from clones covering the F and HN genes were subcloned into the phage vectors M13mp18 and -mp19 (Yanisch-Perron et al., 1985). DNA sequencing was performed by the dideoxy chain termination method (Sanger et al., 1977). Oligonucleotide sequencing primers made on an Applied Biosystems model 381A DNA synthesizer were used in addition to a 17-base universal primer (Pharmacia). Sequence data were assembled and analysed using the computer programs of Staden (1982, 1984).

**RESULTS AND DISCUSSION**

In earlier papers we have discussed many of the general features which emerged from studies of the F and HN gene sequences of the Beaudette C strain of NDV (Chambers et al., 1986a, b; Millar et al., 1986). Our aim in sequencing an avirulent strain of NDV was to identify regions of the sequence which differed from those of virulent strains and which might help to explain the molecular basis for variations in virulence. A bank of cDNA clones covering the F and HN genes of strain Ulster was constructed from purified virion genomic RNA, and the positions of clones were determined essentially as described for the Beaudette C strain (Chambers et al., 1986a) and outlined in Methods. The nucleotide sequence of the region of the genome encoding the F and HN genes of NDV strain Ulster is shown in Fig. 1 as a cDNA sequence in the positive (mRNA) sense. Also shown are the amino acid sequences of the major open reading frames which correspond to the F and HN glycoproteins.
NDV F and HN gene sequences

The fusion glycoprotein

The general features of the F glycoprotein of strain Ulster appear to be similar to those of other strains of NDV (Chambers et al., 1986b; McGinnes & Morrison, 1986; Espion et al., 1987; Sato et al., 1987; Toyoda et al., 1987). The length of the F₀ glycoprotein is identical in all strains of NDV that have been sequenced (553 amino acids), as is the length of the F gene (1792 nucleotides) in all strains for which the sequencing data appear to span the entire F gene (between the proposed mRNA start and polyadenylation sites). The locations of cysteine residues and potential glycosylation sites are well conserved. Regions of high hydrophobicity within the F glycoprotein (which have been proposed to correspond to the signal sequence, the N terminus of F₁ and the membrane attachment site) are present in similar locations. There is 94% identity in amino acid sequence between the F glycoproteins of strains Ulster and Beaudette C.

The two most striking features which emerge from the sequence of the F glycoprotein gene of strain Ulster are that there is a leucine residue at the N terminus of the F₁ cleavage fragment, and that there are fewer basic amino acid residues in the region immediately preceding this cleavage site compared to more virulent strains (Fig. 2).

In strains of avian influenza virus there is a correlation between strain virulence, cleavage of the haemagglutinin glycoprotein precursor (HA₀), and the presence of basic amino acids. In virulent strains with readily cleaved HA₀ there is a highly basic amino acid sequence at the site of cleavage to the HA₁ and HA₂ fragments, whereas in less virulent strains this region of the glycoprotein is less basic (Bosch et al., 1979, 1981; Kawaoka et al., 1987). For a recent review of the role of HA in the pathogenicity of influenza virus, see Webster & Rott (1987).

There is a direct relationship between the virulence of NDV strains and the cleavage of the F₀ precursor into the disulphide-linked F₁ and F₂ fragments (Nagai et al., 1976). In virulent strains, the F₀ glycoprotein is cleaved in a wide range of host cells, whereas in avirulent strains it is cleaved in a restricted range of cell types (chick embryo lung and chorioallantoic membrane cells). In simian virus 5 (SV5) there is a highly basic sequence at the cleavage site of F₀ (Paterson et al., 1984) and F₀ is readily cleaved in chick embryo cells (Peluso et al., 1977), whereas Sendai virus F₀ has a single basic amino acid in this region (Blumberg et al., 1985) and is not cleaved in chick embryo cells (Lamb et al., 1976). This relationship between basic amino acids and cleavage of F₀ extends to strains of NDV (Fig. 2), as described recently by Toyoda et al. (1987).

The absence of paired basic amino acids at the F₀ cleavage site of NDV strain Ulster reflects the lack of cleavage of F₀ and infectivity of progeny virions of this strain in some tissue culture systems (Nagai et al., 1976) and certain germinal layers in ovo (Nagai et al., 1979).

The sequence at the N terminus of the F₁ cleavage fragment of paramyxoviruses is highly conserved. In 1978 it was shown that there is also similarity in sequence between the N terminus of Sendai virus F₁ and the N-terminal region of influenza virus HA₂ (Gething et al., 1978). It was surprising to find a leucine residue at the extreme N terminus of F₁ in strain Ulster, since in all other paramyxoviruses (including more virulent strains of NDV), morbilliviruses and pneumoviruses which have been sequenced, a phenylalanine is always present at this position. The importance of this substitution in determining cleavage is not known. It is, however, possible that this is partly responsible for the lack of virulence shown by strain Ulster.

The recently published sequence of the F glycoprotein of the avirulent strain D26 (Sato et al., 1987) and the sequencing of a region of 26 amino acids from the F₀ cleavage site of several NDV strains (Toyoda et al., 1987), suggest that the presence of less basic cleavage sites is common among avirulent strains of NDV, but that an N-terminal leucine on F₁ may be exclusive to the relatively small number of extremely avirulent isolates.

The haemagglutinin–neuraminidase glycoprotein

The length of the HN gene of strain Ulster (between the proposed mRNA start and polyadenylation sites) is identical to that of NDV strain Beaudette C (Millar et al., 1986), assuming that the distance between the F and HN genes is 31 nucleotides, as suggested by nuclease mapping of the HN 5’ mRNA terminus (Jorgensen et al., 1987; Yusoff et al., 1987).
NDV and HN gene sequences

![Image of nucleotide sequence and amino acid sequences]

**Fig. 1.** Nucleotide sequence of the region of the NDV strain Ulster genome encoding the F and HN genes, shown as cDNA in the positive (mRNA) sense. The probable position of the F and HN mRNA start sites are underlined twice, probable mRNA polyadenylation signals are underlined once. The deduced amino acid sequences of the F and HN glycoproteins are shown above the nucleotide sequence and potential asparagine-linked glycosylation sites are boxed.

**Fig. 2.** Amino acid sequences at the cleavage-activation site of the F0 glycoprotein precursor. The sequence of NDV strain Ulster is compared with NDV strain Beaudette C (Chambers et al., 1986b), Sendai virus (Blumberg et al., 1985) and SV5 (Paterson et al., 1984). Basic amino acids are indicated by a plus sign and the N terminus of the F1 cleavage fragment by an arrow.

However, the major open reading frame which encodes the HN0 glycoprotein of strain Ulster is 616 amino acids long and extends beyond the C termini of the HN glycoproteins of more virulent strains (Fig. 3). This corresponds to a 39 amino acid extension at the C terminus of the HN glycoprotein relative to strains Beaudette C (Millar et al., 1986) and BI (Jorgensen et al., 1987), and a 45 amino acid extension relative to strain A-V (McGinnes et al., 1987). Apart from this C-terminal extension, there is a high degree of amino acid sequence conservation between different strains of NDV (94% between strains Ulster and Beaudette C). Other than differences at the C terminus of HN, the major difference in HN between NDV strains is the apparent deletion of three bases in the HN gene of strain A-V which corresponds to the tyrosine residue at amino acid position 185 in strains Ulster, Beaudette C and BI.
Fig. 3. Amino acid sequences at the C termini of the HN0 glycoprotein precursor of NDV strain Ulster and the HN glycoproteins of NDV strains Beaudette C (Millar et al., 1986), B1 (Jorgensen et al., 1987) and A-V (McGinnes et al., 1987) are shown boxed. A translation of the non-coding regions beyond the HN C termini is also shown to indicate the extent of sequence conservation in this region with the C terminus of HN0 (data from strain B1 does not extend to the end of the non-coding region). Amino acids identical to the Ulster sequence are indicated by dots and termination codons by asterisks. There appears to be a (+1) frameshift in the A-V non-coding region, relative to the other strains; this is indicated by a plus sign.

The general features of the HN glycoprotein of strain Ulster are similar to those of other strains of NDV (Millar et al., 1986; Jorgensen et al., 1987; McGinnes et al., 1987). The major hydrophobic region, corresponding to the membrane attachment site, is close to the N terminus, as found in other paramyxoviruses. One of the potential glycosylation sites (amino acid positions 500 to 502) in the HN glycoprotein sequence of strains Beaudette C is not present in strain Ulster. This site was previously suggested to be less favourable for glycosylation due to the presence of proline, despite being conserved between NDV strain Beaudette C, SV5 and Sendai virus (Millar et al., 1986).

Strain Ulster and other extremely avirulent strains of NDV (Queensland and D26) are as yet unique among paramyxoviruses in that they synthesize a precursor glycoprotein (HN0) which is larger than the normal HN and requires proteolytic cleavage for activation (Nagai et al., 1976; Nagai & Klenk, 1977). Sequencing data from the more virulent strain, Beaudette C, revealed the presence of a long non-coding region at the 3' end of the HN mRNA which contains an open reading frame that is in phase with HN (Chambers et al., 1986a; Millar et al., 1986). This suggested that HN0 in avirulent strains might be encoded by a longer open reading frame which extends beyond that of more virulent strains into this 3' non-coding region.

There is a potential asparagine-linked glycosylation site, Asn-Gln-Thr (amino acid positions 600 to 602), in the C-terminal extension region of HN0, which agrees with the finding that the proteolytic activation of HN0 requires the removal of a glycopeptide (Garten et al., 1980). Treatment of whole virions of strain Ulster with trypsin releases a glycopeptide of Mr 9000 (9K) Schuy et al. (1984). From amino acid sequencing data inferred that this 9K fragment was derived from the C terminus of HN0, thereby providing evidence for the N-terminal attachment of the glycoprotein to the viral membrane. However, the pentapeptide sequence (Leu-Gly-Pro-Gly-Val) reported as the N-terminal sequence of the 9K fragment does not occur in the sequence presented here (Fig. 1). Although changes may have accumulated on passage of this strain in the two laboratories, it is surprising that this pentapeptide sequence is absent from our sequence, in view of the high level of amino acid sequence conservation observed between different strains of NDV.

It is possible to speculate about the site at which proteolytic cleavage of the Ulster HN0 to active HN occurs. The sequence Gly-Arg-Leu (positions 577 to 579) in strain Ulster is identical to the Ulster F0 cleavage site. If cleavage occurred at this arginine followed by removal of this residue by a carboxypeptidase, as has been proposed to act after cleavage of F0 (Kohama et al., 1981), the cleaved HN of strain Ulster would end at the same position as those of strains Beaudette C and B1. In strains of Beaudette C and B1, a stop codon (TGA) replaces the arginine-578 codon (CGA) present in strain Ulster. The termination codon of HN in strain A-V (TAG) also replaces an arginine codon (AGG) in strain Ulster, in this case corresponding to arginine-572. There is considerable conservation of sequence between the longer C-terminal
region of HN0 in strain Ulster and the potential coding region beyond the true C-termini of HN in more virulent strains (Fig. 3). This, and the fact that the sequences of virulent strains A-V and Beaudette C terminate at different positions, suggest that mutations occurred within the longer open reading frame of HN0 in some ancestral form of NDV to produce the shorter reading frames found in the more virulent strains.

We thank June Ramsay for technical assistance. This work was supported by grants from the British Technology Group and North of England Cancer Research Campaign. N.S.M. was supported by an SERC studentship.

Note added in proof: The HN gene sequence of NDV strain Italien has been published (Wemers et al., 1987; Archives of Virology 97, 101-113). The HN amino acid sequence terminates at the same position as that of strain A-V (Fig. 3). The HN0 gene sequence of NDV strain D26 has also been published (Sato et al., 1987; Virus Research 8, 217-232); the HN0 amino acid sequence is of identical length to that of strain Ulster and appears to display similar general features.

REFERENCES


(Received 14 October 1987)