Nucleotide Sequence of the Gene Encoding the Fusion Glycoprotein of Newcastle Disease Virus

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

(Received 1 August 1986)

SUMMARY

The nucleotide sequence of the gene encoding the fusion (F) glycoprotein of the Beaudette C strain of Newcastle disease virus (NDV) has been determined from cDNA clones obtained from virion RNA. The gene is 1792 nucleotides long, including mRNA start and polyadenylation signals typical of paramyxoviruses. The single open reading frame encodes a polypeptide of 553 amino acids, with a predicted molecular weight of 59042. The F polypeptide has three regions of high hydrophobicity: an N-terminal signal peptide, the N terminus of F₁ (known from protein sequencing) and a C-terminal membrane-spanning region by which the F glycoprotein is anchored to the membrane. The cleavage site of F₀ is located in a highly basic region of the F polypeptide. Five potential asparagine-linked glycosylation sites are present in the amino acid sequence, of which one is in F₂ and the others in F₁. Comparison of the NDV F amino acid sequence to those from other paramyxoviruses reveals homology to Sendai virus, simian virus 5 and human respiratory syncytial virus. There is also limited homology between the N terminus of F₁ of NDV and the N termini of HA₂ of influenza viruses. Post-translational modifications of the NDV F polypeptide are discussed in the light of information provided by the amino acid sequence.

INTRODUCTION

Newcastle disease virus (NDV) is a paramyxovirus that infects birds and causes the disease fowl pest in poultry. The severity of the symptoms of the disease varies from asymptomatic infection to near 100% mortality, depending on the virulence of the virus present (Waterson et al., 1967). Proteolytic cleavage of the F₀ precursor glycoprotein to generate the disulphide-linked polypeptides F₁ and F₂ is necessary for the infectivity of paramyxoviruses (Homma & Ohuchi, 1973; Samson & Fox, 1973; Scheid & Choppin, 1974; Nagai et al., 1976). The presence or absence of this cleavage in certain cell lines correlates well with the virulence of isolates of NDV (Nagai et al., 1976). The F₀ glycoprotein of virulent strains of NDV is cleaved to F₁ and F₂ in a wide range of host cells (chick embryo fibroblasts and lung cells, chorioallantoic membrane, baby hamster kidney and Madin–Darby bovine kidney cells), whereas F₀ of avirulent strains is cleaved only in a restricted range of cell types (chick embryo lung and chorioallantoic membrane cells). NDV virions of avirulent strains released from cells in which F₀ is not cleaved are not infectious unless treated with trypsin, which performs the cleavage of F₀ [and also the cleavage of the precursor glycoprotein HN₀ to HN for the extremely avirulent NDV strains Ulster and Queensland (Nagai & Klenk, 1977)]. The degree of susceptibility of F₀ to host cell proteases is probably determined by the structure of F₀, possibly by the amino acid sequence at the site where F₀ is cleaved to F₁ and F₂. In strains of avian influenza virus of differing virulence, cleavage of the haemagglutinin precursor glycoprotein HA₀ to HA₁ and HA₂ is dependent on the sequence of the peptide that connects HA₁ to HA₂ in uncleaved HA₀ (Bosch et al., 1981). Analysis of the nucleotide sequence of the F gene of NDV, and thus the deduced amino acid sequence of the F polypeptide may, therefore, help us to understand the molecular basis of the virulence of NDV.
large differences in virulence shown by various isolates of NDV. We have, therefore, determined the sequence of the F gene of the Beaudette C strain of NDV from cDNA clones of virion RNA (Chambers et al., 1986).

The F glycoproteins of paramyxoviruses undergo extensive post-translational modification in addition to the proteolytic cleavage of F₀ to the active form. Modifications to F₀ include cleavage of the signal sequence (Blumberg et al., 1985), glycosylation (Mountcastle et al., 1971), fatty acid acylation (Schmidt, 1982; Chatis & Morrison, 1982) and possibly rearrangements of intramolecular disulphide bonds (McGinnes et al., 1985). Modifications to F₂ include blockage of the N terminus (Scheid et al., 1978) and trimming of the C terminus by carboxypeptidase (Kohama et al., 1981). Determination of the amino acid sequence of F of NDV may suggest the locations of some of these processing events in the F polypeptide chain.

Monospecific antibodies to either of the NDV surface glycoproteins, HN or F, can neutralize viral infectivity (Umino et al., 1984). Since antibodies to the F glycoprotein of the paramyxovirus simian virus 5 (SV5) are important in the prevention of spread of viral infection (Merz et al., 1981), the F glycoprotein of NDV may be important in vaccines against Newcastle disease. Knowledge of the amino acid sequence of the F polypeptide is necessary for the location of the epitopes responsible for the stimulation of the immune response to NDV.

**METHODS**

Molecular cloning of the F gene of NDV. Details of the cloning of cDNA to NDV have been published previously (Chambers et al., 1986). Four plasmids containing cDNA inserts that span the NDV genome region extending from the M gene, through the F gene and into the HN gene have been used to determine the nucleotide sequence of the F gene. These plasmids were designated 3.99 (M-F), 3.01 (F), 7.58 (F) and 7.44 (F-HN).

Nucleic acid sequencing. Appropriate restriction fragments were subcloned into M13 vectors, sequenced by the dideoxy chain termination technique, and sequences were compiled and analysed as described for the NDV HN gene (Millar et al., 1986).

**RESULTS**

The nucleic acid sequence of the NDV F gene and its deduced amino acid sequence are shown in Fig. 1. It has been assumed that the NDV F gene starts with the nucleotide sequence ACGGGTAGAAG, which is similar to sequences at the starts of the NDV NP, HN and L genes (Kurilla et al., 1985; Millar et al., 1986) and terminates with a polyadenylation signal TTAGAAAAAA, which is similar to those of the NDV HN, Sendai virus and vesicular stomatitis virus (VSV) genes (Gupta & Kingsbury, 1984; McGeoch, 1979; Rose, 1980; Millar et al., 1986). The NDV M gene polyadenylation signal and the M–F intergenic nucleotide immediately precede the NDV F sequence (Millar et al., 1986). Thus, the NDV F gene is 1792 nucleotides long and encodes one long open reading frame of 553 amino acids corresponding to a polypeptide of molecular weight 59042 if translation is initiated at the first methionine in the open reading frame (nucleotides 47 to 49). Although this corresponds to the first AUG codon in the mRNA, which is normally preferred for initiation of translation, two additional methionines are present at positions 13 and 14 in the open reading frame. The nucleotide sequences flanking the codons for methionines 13 and 14 do not match the consensus eukaryotic translation initiation sequence (A-X-X-A-U-G-G: Kozak, 1981, 1986) whereas that surrounding the codon for the first methionine does. This suggests that the first ATG codon in the nucleotide sequence of the F gene is used to initiate protein synthesis.

Five potential asparagine-linked glycosylation sites are present in the amino acid sequence. A sixth such sequence, Asn–Asn–Thr, which occurs at amino acid residues 541–543 is in the presumed C-terminal cytoplasmic tail region of the polypeptide chain (see below). This sixth sequence is probably not glycosylated since it would not be translocated through the membrane during polypeptide synthesis, and glycosylation is performed in the endoplasmic reticulum and the Golgi complex (Bennet & Leblond, 1977).

The sequence of 20 amino acids at the N terminus of the F₁ polypeptide of NDV has been determined by direct protein sequencing procedures (Richardson et al., 1980) and corresponds exactly with amino acid residues 117 to 136 in the deduced amino acid sequence of F (Fig. 1). Thus, the basic amino acids on the N-terminal side of this region (residues 112 to 116) form the
Fig. 1. Nucleotide sequence of the NDV strain Beaudette C F gene and deduced amino acid sequence of the F polypeptide. The cDNA sequence is shown in the positive (mRNA) sense. The sequence is numbered from the presumed first nucleotide of the mRNA, and the consensus 11-nucleotide start and 10-nucleotide polyadenylation signals are underlined. The deduced amino acid sequence of the major open reading frame is shown above the nucleotide sequence. Potential asparagine-linked glycosylation sites in the amino acid sequence are boxed. The three major hydrophobic regions are indicated by heavy bars above the amino acid sequence. From the N terminus they are the presumed signal sequence, the known N terminus of F and the presumed membrane-spanning region.

cleavage site of the F0 precursor. The F1 and F2 polypeptides must, therefore, be the C- and N-terminal parts of F0 respectively. The orientation of F1 and F2 within F0 has also been determined by salt-shock mapping (Samson et al., 1980) and is the same as the orientation N-F2-F1-C deduced from the amino acid sequence.

Paramyxovirus F glycoproteins are integral membrane proteins anchored in the viral membrane near their C termini (Li et al., 1980; Caldwell & Lyles, 1981; Scheid et al., 1978). Deduced amino acid sequences of the Sendai virus and SV5 F polypeptides predict that both contain three highly hydrophobic regions (Blumberg et al., 1985; Paterson et al., 1984). The first is near the N terminus and has the characteristics of a signal sequence which is subsequently removed by proteolytic cleavage (von Heijne, 1982). The second is internal and is the known N terminus of F1 for NDV, SV5 and Sendai virus (Gething et al., 1978; Richardson et al., 1980).
Fig. 2. Hydropathy profile of the NDV F amino acid sequence. A window of 11 amino acids was used to calculate hydropathy, using the procedure of Kyte & Doolittle (1982). Hydrophobic regions lie above, and hydrophilic regions below the horizontal line, which is the average hydropathy of a large number of sequenced proteins. The NDV F amino acid sequence is numbered from the N terminus. The three highly hydrophobic regions indicated with arrows are (1) the presumed signal peptide, (2) the known N terminus of F₁ and (3) the presumed membrane-spanning region.

The third is near the C terminus and has the characteristics of a membrane-spanning segment (von Heijne, 1982; Kyte & Doolittle, 1982). This pattern is repeated in the amino acid sequence of the F polypeptide of NDV. Fig. 2 shows a hydropathy profile of the NDV F amino acid sequence, in which the corresponding three regions of high hydrophobicity are indicated. In contrast to the NDV HN amino acid sequence (Millar et al., 1986), the F sequence is markedly more hydrophobic than the average. On the basis of the similar high hydrophobicity of the F polypeptide of Sendai virus, it has been suggested that regions of the F glycoprotein in addition to the N terminus of F₁ may be capable of hydrophobic interactions with membranes during the fusion process (Blumberg et al., 1985).

The NDV F amino acid sequence is compared to those from SV5, Sendai virus, respiratory syncytial virus (RSV) (Collins et al., 1984) and human influenza C virus HA (Nakada et al., 1984) in Fig. 3(a) to (d) respectively. Some similarities to SV5 and Sendai virus were anticipated because of the homology detected between the HN amino acid sequences (Millar et al., 1986) and the known amino acid homologies at the N termini of F₁ for NDV, SV5 and Sendai virus (Richardson et al., 1980). Mumps virus and measles virus also have homologous F₁ N termini (Varsanyi et al., 1985; Server et al., 1985), but the complete sequences of the F polypeptides of these viruses are not yet available for comparison. The homology to the RSV F amino acid sequence is more surprising in view of the relative lack of homology detected at the N terminus of F₁ (Collins et al., 1984). Similarities in the overall structures and the locations of cysteines in the F polypeptides among Sendai virus, SV5 and RSV were, however, noted by Shioda et al. (1986). Homology of the N terminus of the Sendai virus F₁ polypeptide to the N terminus of influenza virus HA₂ has been described (Gething et al., 1978; Blumberg et al., 1985). This is also seen with NDV F₁ and influenza virus HA₂ and is shown in Fig. 3(d), in which there is a diagonal line that suggests homology between the N termini of NDV F₁ and influenza C virus HA₂. Homologies of the NDV F amino acid sequence to influenza A and B virus HA sequences (Porter et al., 1979; Krystal et al., 1982) were less strong (not shown).

An alignment of the NDV F amino acid sequence with those of SV5 (Paterson et al., 1984) and Sendai virus (Blumberg et al., 1985) is shown in Fig. 4. A few gaps were positioned in the various sequences to obtain this alignment: the most significant gaps are those adjacent to the cleavage site and the transmembrane region in F₀. Alternative alignments are possible in the transmembrane region itself, due to the high proportion of hydrophobic residues present. The
sequences show homology over the entire lengths of the F polypeptides, except at the extreme N termini of the NDV and Sendai virus sequences and at the C-terminal side of the transmembrane regions in all three viruses. All of the cytoplasmic tail regions have a net positive charge but otherwise seem to be variable in length and sequence. The high proportion of methionines present in the Sendai virus cytoplasmic tail and noted by Blumberg et al. (1985) is not seen in the other paramyxoviruses. In this alignment, the overall level of amino acid identities between the F polypeptides of NDV and SV5 is 33%, between NDV and Sendai virus 25%, and between SV5 and Sendai virus 23% (similar to the 23.5% reported by Shioda et al., 1986).

**DISCUSSION**

The F glycoprotein of NDV has an overall structure similar to its counterparts in other paramyxoviruses, most notably in SV5 and Sendai virus but also in RSV. The overall levels of amino acid identities between the NDV, SV5 and Sendai virus F amino acid sequences are similar to those between the HN sequences (Millar et al., 1986) but it was necessary to insert fewer gaps to align the F sequences than was necessary in the alignment of the HN sequences. This suggests that the structure of the paramyxovirus F glycoprotein may have been conserved more than that of HN, although a comparable degree of sequence variation has occurred in both
polypeptides. Amino acid sequence variation in both HN and F₀ between strains of NDV is suggested by the variation in the isoelectric points of these glycoproteins. For HN, the isoelectric point varies from 6.5 in strain Ulster to 7.0 in strains B1, F and N. For F₀, the isoelectric point varies from 6.0 in strain Ulster to 7.0 in strain Texas (Chambers, 1982; Chambers & Samson, 1982).

The most highly conserved region of the amino acid sequences of F of NDV, Sendai virus and SV5 is near the N terminus of F₁ (residues 126 to 141 in NDV). Presumably this peptide has a crucial role in the fusion process mediated by the F glycoprotein. Another region of the amino acid sequence of F of NDV shows particularly good homology to SV5, but only average homology to Sendai virus (residues 210 to 230 in NDV: Fig. 4). The function of this region is unknown at present.

All ten cysteines between the signal and transmembrane regions of the F polypeptides are conserved in NDV, SV5 (Paterson et al., 1984) and Sendai virus strain Z (Shioda et al., 1986). In the sequence of the F polypeptide of Sendai virus strain Harris (Blumberg et al., 1985) shown in Fig. 4, cysteine is replaced by arginine at position 394. McGinnes et al. (1985) have shown that conformational changes (detected by a shift in the mobility of F₀ on polyacrylamide gels) occur in the NDV F glycoprotein shortly after synthesis but prior to cleavage of F₀, possibly as a result of changes that occur in intramolecular disulphide bonds. Since all of the relevant cysteines are conserved in SV5 and in one of the Sendai virus strains, similar conformational changes might be detectable during the biosynthesis of the corresponding glycoproteins.

When the sequence of RSV F is aligned, few amino acid identities are seen but the general structure of RSV F (particularly F₁) is similar to those of the other paramyxoviruses (Fig. 3c).
the sequence of RSV F₁ is aligned with the other paramyxoviruses as in Fig. 4, without gaps in the RSV sequence from the start of F₁ (at phenylalanines 117 of NDV and 137 of RSV) to the ends of the polypeptides, three of the RSV cysteines correspond directly to paramyxovirus cysteines (NDV positions 338, 347 and 362), five are close to paramyxovirus cysteines (NDV positions 199, 370, 394, 401 and 424) but the remaining four cysteines in the RSV F₁ polypeptide have no counterparts in NDV, SV5 or Sendai virus. The single conserved cysteine present in the NDV (position 76), SV5 and Sendai virus F₂ polypeptides must cross-link to F₁ and therefore fulfills the same function as one of the two cysteines in the RSV F₂. Shioda et al. (1986) drew similar conclusions about Sendai virus, SV5 and RSV but have aligned the polypeptides differently. In addition, Spriggs et al. (1986) have shown a relationship between the parainfluenza virus 3 and RSV F amino acid sequences by a statistical analysis of the sequences.

The site of cleavage of the signal peptide from the NDV F polypeptide can be predicted using the rules proposed by von Heijne (1983). These suggest that the hydrophobic signal region would be removed by cleavage after a small neutral residue (such as serine 31), which is preceded by a large uncharged residue (asparagine 30), another small neutral residue (alanine 29) and a glycine or proline (proline 28). This would generate an N-terminal isoleucine on NDV F₀ and leaves a problem in explaining the nature of the blocked N terminus of the protein. The site of cleavage of the signal peptide has been located after a cysteine (residue 25) for Sendai virus by gas–liquid chromatography and mass spectrometry of the blocked N terminus of the F glycoprotein (Blumberg et al., 1985), generating glutamine as the N-terminal amino acid of F₀. Glutamine can spontaneously cyclize to block the N terminus of Sendai virus F₀, but an N-terminal isoleucine on NDV F₀ would not. Acetylation is the commonest blocking modification at the N termini of proteins in eukaryotic cells, but is normally considered a modification of proteins in the cytosol and is not known to occur at N-terminal isoleucine (Tsunasawa & Sakija, 1984). Acetylation can, however, occur in the endoplasmic reticulum and secretory granules of rat pituitary cells (Glembotski, 1982). The exact N terminus of NDV F₀ can only be determined by direct chemical analysis.

The NDV F amino acid sequence contains five potential asparagine-linked glycosylation sites: the sequence Asn–Asn–Thr at residues 541 to 543 is in the presumed cytoplasmic tail of the polypeptide and would not be glycosylated, The F glycoproteins of several strains of NDV contain only high-mannose type glycans (Schwalbe & Hightower, 1982; Diabete et al., 1984; McGinnes et al., 1985). Preliminary studies suggest that the mature F glycoprotein is glycosylated with high-mannose glycans in strain Beaudette C, sequenced in the work described in this communication. The evidence for high-mannose glycans (Chambers, 1982) includes the high ratio of mannose to glucosamine in F, the absence of sulphation from F, and the small, relatively homogeneous glycopeptides generated from F. Since high-mannose glycans always contain two glucosamine residues (Kornfeld & Kornfeld, 1976; Li et al., 1978), the ratio of tritiated glucosamine in F₁ to that in F₂ should reflect the ratio of glycans attached to F₁ and F₂. For NDV strain Hickmann, the ratio is approximately 4:1 (Scheid & Choppin, 1977: Table 4). In NDV strain Beaudette C, there are four potential carbohydrate attachment sites in F₁ and one in F₂ (Fig. 1), which may, therefore, all be filled with high-mannose type glycans.

The estimate of five glycans on NDV F is consistent with estimates of the molecular weight of unglycosylated F synthesized in tunicamycin-treated cells where the signal sequence is probably cleaved but glycosylation is inhibited [55000 mol. wt. (55K) to 56K], and that of fully glycosylated F₀ (66K: Morrison & Simpson, 1980), assuming that high-mannose type glycans add about 2K to apparent molecular weight on SDS gels (Horisberger et al., 1980). If it is assumed that the signal sequence cleavage occurs at serine 31 as discussed above, that carboxypeptidase removes the basic peptide at the C terminus of F₁ (residues 112 to 116, causing a loss of 0.8K from F₁), then all five glycosylation sites are filled with high-mannose glycans adding 2K to the molecular weight and that fatty acid acylation (discussed below) adds a negligible mass to F (the mass increment would be 0.25K for each palmitate residue) the molecular weights of F₁ and F₂ would be 54.7K and 10.3K respectively, in agreement with those predicted on the basis of mobilities on SDS gels (52K to 56K and 10K: Scheid & Choppin, 1977). On these assumptions, the molecular weight of F₀ would be 55.8K or 65.8K in the
unglycosylated or glycosylated form respectively, which is in agreement with the estimates derived from SDS gels by Morrison & Simpson (1980).

Fatty acid acylation of the NDV F but not the HN glycoprotein has been reported (Schmidt, 1982; Chatis & Morrison, 1982). The G glycoprotein of VSV is fatty acid acylated, probably on a cysteine in the cytoplasmic tail of the protein near to the membrane. After initial attachment to cysteine, fatty acid residues could be transesterified to amino acids with hydroxyl functional groups (Rose et al., 1984; Magee et al., 1984). It may, therefore, be significant that the NDV F glycoprotein has a cysteine (residue 523) in the presumed transmembrane hydrophobic region near the cytoplasmic face, but the NDV HN does not (Millar et al., 1986). The RSV F glycoprotein also has a cysteine residue at the cytoplasmic face of the presumed transmembrane region that could be a site for fatty acid acylation (Collins et al., 1984).

It is possible, although at present speculative, to attempt to relate the amino acid sequences at the cleavage sites of the F0 glycoproteins of paramyxoviruses and the HA0 glycoproteins of influenza viruses to the variation in virulence of NDV isolates. NDV strain Beaudette C is of moderate virulence (Lancaster & Alexander, 1975) and F0 is readily cleaved in chick embryo cells (Samson et al., 1980). For SV5, F0 is also readily cleaved in chick embryo cells, whereas F0 of Sendai virus is not (Peluso et al., 1977; Lamb et al., 1976). In F0 of NDV strain Beaudette C and in SV5, the N terminus of F1 is preceded by a highly basic peptide (residues 112 to 116 in NDV: see Fig. 4), whereas the N terminus of F1 of Sendai virus is preceded by a single basic amino acid. For avian influenza viruses, the HA0 glycoproteins of virulent strains are readily cleaved in chick embryo cells, with the cleavage site preceded by a highly basic peptide, but the HA0 glycoproteins of avirulent strains are not cleaved and the cleavage site is preceded by a single basic amino acid (Bosch et al., 1979, 1981). Thus, as suggested for SV5 (Paterson et al., 1984) the cleavage of paramyxovirus F0 glycoproteins in chick embryo cells, which correlates with virulence for NDV isolates (Nagai et al., 1976), may be dependent on the presence of a highly basic peptide at the cleavage site. The basic peptide may subsequently be removed from both HA0 of influenza viruses and F0 of NDV by carboxypeptidase activity (Garten et al., 1981, 1982; Kohama et al., 1981). On the basis of the migrations of F0 and disulphide-linked F1,2 in isoelectric focusing, Kohama et al. (1981) have suggested that in the F0 glycoproteins of avirulent strains of NDV, the cleavage site is preceded by a less basic region than that present in virulent strains. In contrast, in the outbreak of H5N2 type avian influenza in Pennsylvania, U.S.A. in 1983, the transition from avirulence to virulence in the virus was associated with the loss of a potential glycosylation site from HA0, but no change in the amino acid sequence at the cleavage site. This was, however, thought to render HA0 more susceptible to proteolysis by removing a steric hindrance to the cleavage (Kawaoka et al., 1984). The Sendai virus F polypeptide has seven extra amino acids including a potential glycosylation site (Asn–Asp–Thr, residues 104 to 106) immediately preceding the cleavage site, relative to NDV (Fig. 4). It is possible that the extra amino acids or glycosylation at this position would hinder cleavage of Sendai virus F0. It will, therefore, be of interest to determine how the F polypeptides from avirulent strains of NDV differ from that of Beaudette C.

We thank Kate Brown and Tony Pickard for technical assistance. This work was supported by the North of England Cancer Research Campaign. Neil Millar was supported by a Science and Engineering Research Council studentship.

REFERENCES


(Received 11 June 1986)