Journal of General Virology (1993), **74**, 2539–2547. *Printed in Great Britain*

RNA editing in Newcastle disease virus

Michael Steward, I. Barry Vipond,† Neil S. Millar‡ and Peter T. Emmerson*

Department of Biochemistry and Genetics, The Medical School, University of Newcastle upon Tyne NE2 4HH, U.K.

The co-transcriptional editing of the Newcastle disease virus (NDV) P gene has been studied by sequence analysis of cloned viral genomic RNA and mRNA. Evidence has been obtained for the specific insertion of non-templated G nucleotides, the consequence of which is the generation of three populations of P gene-derived mRNAs. The three populations encode proteins (P, V and W) which have a common N-terminal region, but which utilize three different reading frames at their C termini. Paradoxically, NDV edits its P gene mRNA by the insertion of non-templated G residues in a manner similar to Sendai and measles viruses (P → V editing) despite its apparent closer evolutionary relationship to the simian virus type 5, mumps and related group of viruses which edit a V genomic sequence to generate an mRNA to encode a functional P protein (V → P editing).

Introduction

Newcastle disease virus (NDV), the aetiological agent of fowl pest of poultry, is the type species of the paramyxovirus genus. The negative-stranded RNA genome contains six genes encoding six major structural proteins (3′ NP-P-M-F-HN-L 5′) (Millar & Emmerson, 1988). The paramyxovirus genome is used with great efficiency, with over 95% of the RNA encoding the viral proteins. An apparently universal feature of paramyxoviruses is the use of alternative reading frames and RNA editing within their P genes resulting in increased coding capacity of the genome (for review, see Kolakofsky *et al.*, 1990). The translation products of mRNAs transcribed from alternative open reading frames (ORFs) within the P gene appear to be non-structural viral proteins present in the infected cell but absent from the virus particle. In the case of NDV, two major non-structural proteins have been identified, being defined variously as 28K to 33K and 36K to 38K (Chambers & Samson, 1982) on the basis of their apparent mobility in SDS–PAGE.

The majority of the non-structural proteins of paramyxoviruses appear to be derived from the viral P gene (Lamb & Paterson, 1990), which encodes the structural phosphoprotein, an essential component of the viral transcriptive and replicative complex (Hamaguchi *et al.*, 1983). The first non-structural proteins to be described (C and C′) were identified in Sendai virus and were shown to be the products of ORFs that initiated close to the start of the P ORF, but in the +1 reading frame (Giorgi *et al.*, 1983) by a process of alternative transcriptional initiation (Kozak, 1981). C proteins have since been described in a variety of other paramyxoviruses and morbilliviruses such as human parainfluenza virus type 3 (hPIV3), measles virus and canine distemper virus (CDV).

An alternative method for the generation of non-structural proteins involves the co-transcriptional editing of P gene mRNA transcripts (Thomas *et al.*, 1988; Cattaneo *et al.*, 1989). Non-templated G residues are inserted at a specific site within the P gene, the consequence of which is the utilization of alternative reading frames during translation (Kolakofsky *et al.*, 1990). Furthermore, the editing has been found to occur in two distinct patterns. For Sendai virus (Vidal *et al.*, 1990*a*), measles virus (Cattaneo *et al.*, 1989) and related viruses, an mRNA with a single inserted G residue is the predominant species after the unedited message. However in simian virus type 5 (SV5) (Thomas *et al.*, 1988), mumps virus (Elliott *et al.*, 1990) and related viruses the majority of the edited mRNA contains two inserted residues compared with the virion (v) RNA.

The two mRNAs in each case code for proteins which are amino co-terminal, but divergent at the C termini. The longer protein in each case is the viral structural P protein, whereas the shorter protein is referred to as

† Present address: Department of Biochemistry, Centre for Molecular Recognition, University of Bristol, University Walk, Bristol BS8 1TD, U.K.

‡ Present address: AFRC Laboratory of Molecular Signalling, University of Cambridge, Department of Zoology, Downing Street, Cambridge CB2 3EJ, U.K.

The nucleotide sequence reported in this paper has been submitted to the EMBL database, and assigned the accession number Z26249.

0001-1903 © 1993 SGM

protein V, and may have an inhibitory role in paramyxovirus replication (Curran *et al.*, 1991*a*, *b*). The V protein shows a striking degree of similarity between paramyxoviruses. A highly conserved cluster of cysteine residues is present at the C terminus which may act as a metal-binding domain analogous to a zinc finger structure (Berg, 1986). In the case of Sendai and measles viruses, it is the edited transcript which encodes the V protein whereas for SV5 and mumps virus the unaltered mRNA codes for the V protein, and the edited (+2) mRNA encodes the P protein.

The mechanism of RNA editing by the insertion of non-templated nucleotides appears to be unique to paramyxoviruses, and is thought to occur by the polymerase complex 'stuttering' on a region of template vRNA and reiteratively reading a template base. The question of how certain viruses insert predominantly one G and others two has been addressed with regard to the stability of base pairing between the nascent mRNA molecule and the vRNA on rearward slippage by one or two bases (Vidal *et al.*, 1990*b*).

By analogy with other paramyxoviruses, two sites have been proposed as putative editing sites within the NDV P gene (Cattaneo *et al.*, 1989; Daskalakis *et al.*, 1992). We have examined the editing of the P gene mRNA in NDV strain Ulster 2C, and obtained evidence that editing during transcription generates three distinct populations of mRNAs: those encoding the structural P protein (unedited), those with a +1 frameshift, encoding the V ORF, which has previously been suggested to be the 36K non-structural protein (Samson *et al.*, 1991), and those with a +2 frameshift which, by analogy to Sendai virus, we have called the W ORF, and which we suggest may encode the 33K non-structural protein.

Methods

Cells and viruses. NDV strain Ulster 2C was obtained from Dr A. C. R. Samson of this department. Chick embryo monolayer cell cultures were grown in medium 199 supplemented with 5% fetal calf serum. The virus was used to infect the chick embryo secondary fibroblast cells at an m.o.i. of approximately 10, and the infection was allowed to proceed for 14 h. Infected monolayers were rinsed with Tris–saline (pH 7·5) prior to suspension of the cells in the respective lysis buffers provided with the mRNA purification systems.

cDNA synthesis and cloning. Total mRNA was purified on oligo(dT)–cellulose with the aid of the QuickPrep system (Pharmacia) and the Micro-Fast Track kit (Invitrogen). First-strand cDNA synthesis was performed in the presence of methylmercuric hydroxide to eliminate potential RNA secondary structures with the cDNA cycle kit (Invitrogen) using an oligo(dT) primer. The cDNA generated was used directly for amplification by 25 rounds of PCR (92 °C, 90 s; 55 °C, 120 s; 72 °C, 180 s) using *Taq* XL polymerase (Northumbria Biologicals). A region of 435 base pairs was amplified between primers $PV8^+$ (positions 374 to 390, 5′ CACCCAGGCTGCGGACG 3′) and $PV7^-$ (809 to 792, 5′ GTCAACTTTACTTACCC 3′) and cloned directly into the vector pCRII (Invitrogen). Competent *Escherichia coli* INVαF cells were transformed with aliquots of the ligation mix, and selection was carried out on LB/ampicillin/X-Gal plates. Recombinant colonies identified by a white or pale blue phenotype were screened by *Eco*RI digestion of mini-prep DNA prepared by the method of Holmes & Quigley (1981). The plasmid DNA was sequenced directly after alkaline denaturation with the aid of the Sequenase II system (U.S. Biochemicals).

Genomic vRNA preparation and cloning. Genomic RNA from the Ulster 2C strain of NDV was purified as described previously (Millar *et al.*, 1988) and reverse-transcribed by using random hexanucleotide primers and avian myeloblastosis virus reverse transcriptase. cDNA:RNA hybrids were poly(dG)-tailed and cloned into pBR322 which had been poly(dC)-tailed accordingly. Transformed *E. coli* were screened for the presence of P gene-specific inserts by colony hybridization (Grunstein & Hogness, 1975), using a region of the NDV Beaudette C strain P gene as a probe. Six positive clones were isolated (163, 305, 476, 480, 555 and 569) which covered the P gene, with the entire gene represented by at least two cDNA clones. The termini of the clones were mapped by sequencing from pBR322 $PstI^{+/-}$ primers (Pharmacia).

Positive clones were sequenced as denatured plasmid DNA either with the Klenow fragment of DNA polymerase I, or with T7 DNA polymerase (Sequenase). Sequence data were assembled using Staden software running on an IBM PC, and analysed with the PC/GENE package. The CLUSTAL module for multiple sequence alignments (Higgins & Sharpe, 1988) used the following parameter settings: K-tuple value for proteins 1; gap penalty 5; window size 10; filtering level 2·5; open gap cost 10; unit gap cost 10.

Results

Sequence analysis of the NDV P gene

The nucleotide sequence of the P gene of the Ulster 2C strain of NDV was determined from a series of overlapping cDNA clones (Fig. 1). The P gene is 1451 bases long, and contains consensus NDV transcriptional initiation and termination signals at the 5′ and 3′ ends respectively (underlined). The P gene is separated from the NP gene by two bases (CA), and the M gene by a single base (A). The translation of the sequence into all three potential reading frames shows that there is a large ORF extending from the AUG codon at position 84 to the UAA at 1271. There are also several shorter reading frames, in particular in the +1 reading frame between nucleotides 472 and 803. The potential translation product of this reading frame bears a high similarity to the C terminus of the V protein found in other paramyxoviruses. The AUG codon at position 84 represents the start codon of the P gene, and is in good context to act as an initiator codon in a eukaryotic cell (Kozak, 1981). There is also, 17 nucleotides downstream from the P gene AUG, a second initiation codon in the +1 reading frame. This may be considered to be analogous to the C protein initiation codon first seen in Sendai virus, and subsequently found in certain other paramyxoviruses (Giorgi *et al.*, 1983). However, in the case of NDV, the reading frame opened by this codon

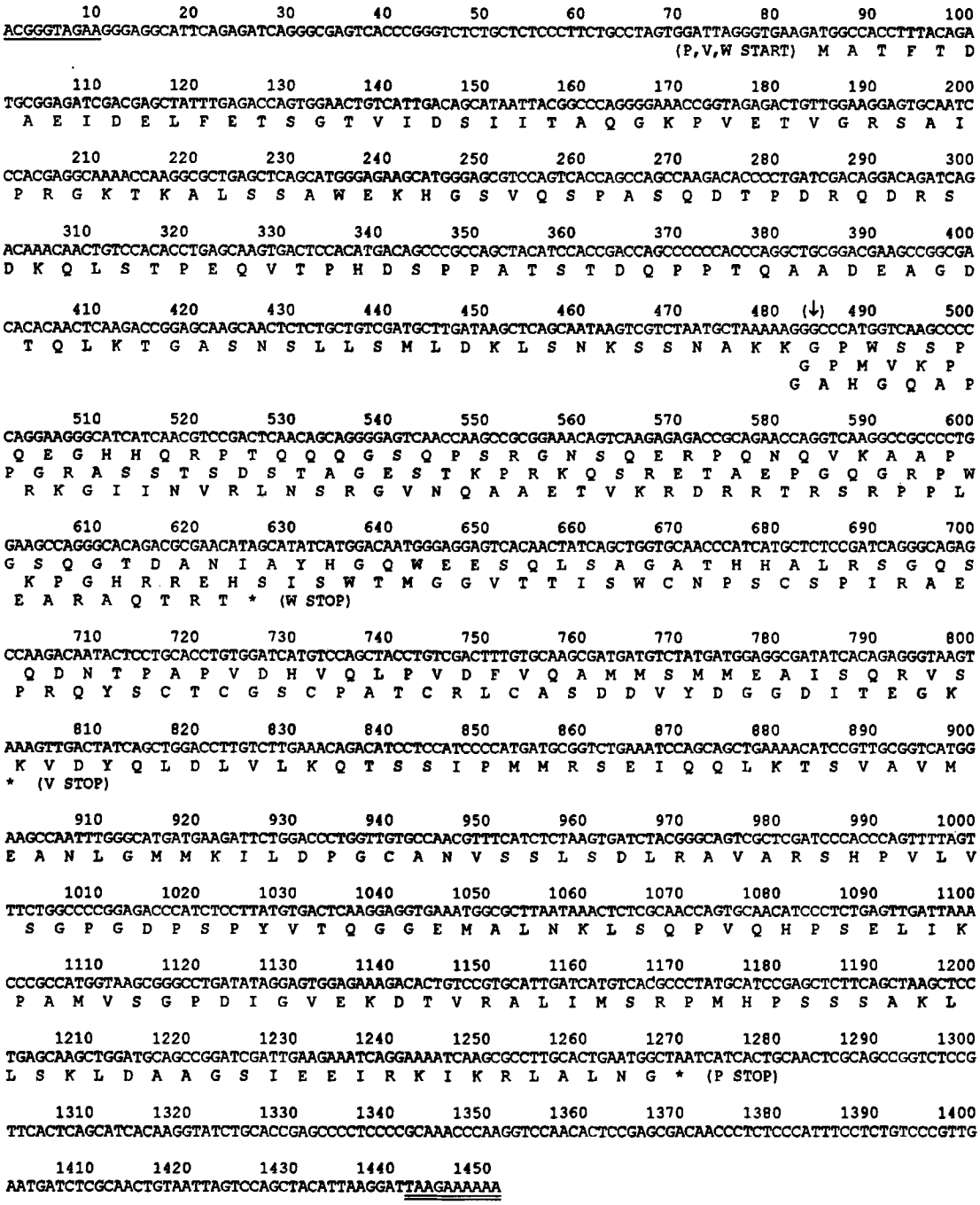

Fig. 1. Nucleotide sequence of the P/V gene of NDV strain Ulster 2C. The sequence is shown as cDNA in mRNA sense. The translation product of the major open reading frame (P) is shown below the sequence, and the predicted protein products arising from translation of the edited reading frames +1 (V) and +2 (W) are also shown. The position of the editing event is shown by the arrow (↓) above position 484. The consensus transcriptional start (single underline) and termination (double underline) signals are also shown.

extends for only 14 amino acids, and is therefore unlikely to encode a significant protein product.

Two positions had been proposed as possible sites for the transcriptional editing of NDV P mRNA. The first site, originally suggested by Cattaneo *et al.* (1989) on the basis of similarity with the measles virus sequence is at 509 in the Ulster sequence. The second site, suggested by Daskalakis *et al.* (1992) is at position 484. Analysis of nucleotide sequence around the two sites amongst the four strains of NDV for which P gene sequence is available [D26 (Ishida *et al.*, 1986); AV (McGinnes *et al.*, 1988); Beaudette C (Daskalakis *et al.*, 1992) and Ulster

```
                    A                         B
        3'          ↓                         ↓          5'
Ulster  UUACGAUUUUUCCCGGGUACCAGUUCGGGGGUCCUUCCCGUAGUAGUUGC
D26     -----------------------C-----A--U-----------------
A-V     -----------------------C--A-A---------U----------A
BC      ----------------A------C--------U--C---U-G--C-----
      471       480       490       500       510
```

Fig. 2. Alignment of the sequences of four NDV P genes around the editing site. The sequence is presented as genomic (negative) sense, and is numbered according to the Ulster sequence. The other sequences presented are D26 (Ishida *et al.*, 1986), Australia-Victoria (A-V) (McGinnes *et al.*, 1988) and Beaudette C (BC) (Daskalakis *et al.*, 1992). Identical nucleotides are represented by a dash (–). A and B refer to actual and proposed editing sites respectively.

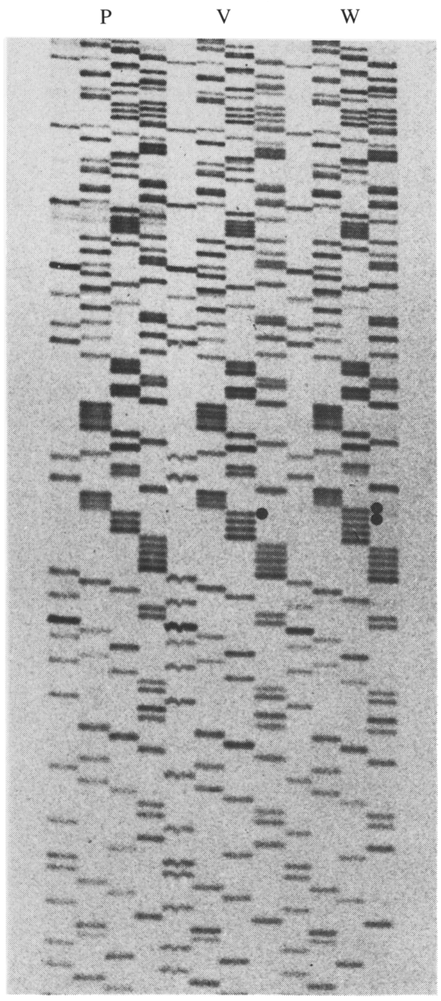

Fig. 3. Autoradiograph of a sequencing gel showing nucleotide sequence at the insertion site of three different clones. The clones shown are unaltered (P), +1G (V) and +2Gs (W), lanes are loaded TCGA. The positions of the extra bases are indicated by dots next to the relevant bands.

2C (this work)] suggests that there is a greater level of similarity around the site at position 484 (Fig. 2, 'A') than at position 509 (Fig. 2, 'B'). The Australia-Victoria (AV) strain, in particular, lacks the motif of three C bases on the genome at this point which characterizes the insertion point in the majority of other viruses where the phenomenon has been observed.

To determine whether the P gene mRNA of NDV is modified at one or both of the proposed sites, a cDNA library was constructed from mRNAs that had been isolated from chick embryo fibroblasts infected with the Ulster 2C strain of NDV, as described in Methods. Owing to the close proximity of the two proposed editing sites (positions 484 and 509) editing at either could be studied by nucleotide sequence analysis from a single oligonucleotide primer. Eighty-two clones were analysed by nucleotide sequencing reactions which were primed using oligonucleotide PV8$^+$. A typical sequence is shown in Fig. 3, where three clones, specific for the P, V and W reading frames are shown.

Full sequencing reactions were performed on several of the clones, but in later experiments only the G termination reaction was performed on each clone. Of the 82 clones, 50 (61%) were unaltered and therefore represented mRNAs that would encode the P protein. The other 32 clones studied showed insertion of non-templated G residues at position 484. None of the clones appeared to be modified at position 509. Of the altered clones, 22 (27% of total clones) contained a single inserted G residue, and therefore encoded the V protein. A further seven clones (8·5%) contained two extra G residues, accessing the third reading frame, with the potential to encode a protein analogous to the W protein of Sendai virus (Curran *et al.*, 1991*a*). A single clone was identified with three inserted G residues, and two were isolated with five added bases. No clones were found with four added bases or with greater than five. The frequency of misinsertion in NDV is therefore similar to that observed for Sendai virus (Vidal *et al.*, 1990*a*).

A conserved sequence precedes the editing site in several paramyxoviruses and may act as a signal for mRNA editing (Vidal *et al.*, 1990*b*). A pyrimidine-rich motif which is seen on the NDV genome preceding the insertion site (3′ ACGAUUUUU 5′) bears a limited degree of similarity to the polyadenylation signal (3′ AAUCUUUUUU 5′) found at the 5′ end of all NDV genes. A much closer similarity between these two sequences occurs in SV5 (Thomas *et al.*, 1988) and hPIV3 (Galinski *et al.*, 1992). The sequence that surrounds the insertion site in several paramyxoviruses is shown in Fig. 4. The insertion site is marked by a high level of conservation; a consensus sequence (A/U)UU(C/U)-UCCC occurs in all paramyxoviruses (Fig. 4). All viruses that insert predominantly one G (P → V editing) have the conserved sequence UUUUCCC whereas all viruses that insert predominantly two Gs (V → P editing) have the sequence AUUCUCCC.

```
                    ↓
NDV       AUUCAGCAGAUUACGAUUUUUCCCGGGUACCAGUU
SV5       CAAGGGGGUAGCUAAAAUUCUCCCCGUCCCUAUGG
Mumps     GCGGCCACGGUCUUAAAUUCUCCCCCCCCGGCCCU
hPIV2     GGGUUCUGGGGUUGAAAUUCUCCCCCCCUCGAUUA
hPIV4A    UACUGUGGUCCUAUAAAUUCUCCCCCCUUAUAUCU
hPIV3     CUACUGUUCCCUUAAUUUUUUCCCCCCAACCUUUC
Sendai    UCUCUGGCUGAGUUGUUUUUUCCCGUAUCCUCUCU
CDV       CCUCACACCUAGGUAAUUUUUCCCCGUGUCUUCUC
Measles   AAGGCCCUGUGGGUAAUUUUUCCCGUGUCUCGCGU
Consensus           UUUUUCCC
                    A  C
```

Fig. 4. Alignment of the sequences at the editing sites of several paramyxovirus P genes. The sequences are shown in genomic (negative) sense. The predicted point of insertion is marked by the arrow.

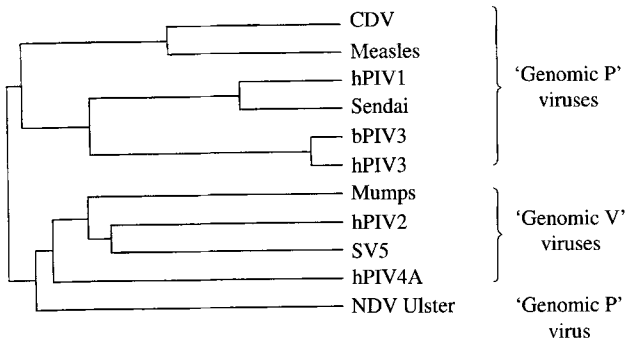

Fig. 5. Dendrogram generated by CLUSTAL alignment of the P proteins from 10 paramyxoviruses. The sequences are those of CDV (Barrett *et al.*, 1985), measles virus (Bellini *et al.*, 1985), hPIV1 (Matsuoka *et al.*, 1991), Sendai virus (Giorgi *et al.*, 1983), bPIV3 (Sakai *et al.*, 1987), hPIV3 (Galinski *et al.*, 1986), mumps virus (Takeuchi *et al.*, 1988), hPIV2 (Southern *et al.*, 1990), SV5 (Thomas *et al.*, 1988), hPIV4A (Kondo *et al.*, 1990) and NDV strain Ulster 2C (this work). Alignments were performed with parameters described in Methods.

Analysis of the predicted P, V and W protein sequences

The NDV Ulster 2C phosphoprotein is a 395-amino acid polypeptide with an M_r of 42241. As observed with other paramyxovirus phosphoproteins, this figure is at variance with the M_r determined by SDS–PAGE analysis, where the protein appears to have an M_r of between 53K and 56K. This phenomenon in the NS (P) protein of the rhabdovirus vesicular stomatitis virus (Gallione *et al.*, 1981) has been attributed to the presence of highly acidic regions. It has previously been suggested that the V ORF of NDV encodes the 36K non-structural protein (Samson *et al.*, 1991), and a similarly aberrant mobility is seen for the V protein which appears to have an M_r of 36K on gels, whereas an M_r of 25379 is predicted from the sequence. Both the P and V proteins possess relatively high proportions of serine and threonine, as may be expected for proteins known to be highly phosphorylated *in vivo*. The V protein, in common with other paramyxovirus V proteins, is cysteine-rich within its unique C-terminal region, seven cysteine residues being present within 30 amino acids. In contrast, the P protein contains only a single cysteine residue (at position 286). The presence of only a single cysteine residue would appear to preclude the possibility that the P protein forms a disulphide-linked trimer, as has previously been suggested (Smith & Hightower, 1981). The presence of P protein dimers and of heteromultimeric forms of the P and V protein cannot, however, be ruled out.

The translation product of an mRNA edited by the insertion of two non-templated G residues in Ulster strain P gene mRNA is a polypeptide of 181 amino acids and calculated M_r of 19397. By analogy with Sendai virus, we have called this ORF 'W'. Unlike the P and V proteins, the length of the W ORF varies in each strain of NDV for which the P gene sequence is available. In D26, this ORF adds only six amino acids onto the N terminus of the P protein resulting in a hybrid protein of 145 amino acids, whereas in Beaudette C and AV, the equivalent ORFs encode 221 and 228 amino acids respectively. It is possible that the W ORF encodes the 33K non-structural protein (Chambers & Samson, 1982). The 33K protein is highly basic, a feature consistent with the high number of basic amino acid codons in the unique 3′-terminal region of the W ORF.

Comparison of the NDV P protein sequence with other paramyxovirus P proteins

The P and V proteins from NDV strain Ulster 2C show over 90% similarity to their counterparts from strains D26 (94·4%), AV (92·2%) and Beaudette C (91·9%). When all four are aligned simultaneously, there is 85% similarity amongst the P proteins, and 80% amongst the V proteins (not shown). In all alignments, the similarity is strongest towards the C termini of the proteins, reflecting the pattern seen when the Ulster 2C sequence is compared with those from other paramyxoviruses detailed below.

The NDV Ulster P protein was compared with the equivalent proteins from other paramyxoviruses on the basis of amino acid sequence using the CLUSTAL module of the PC/GENE analysis software. As shown in Fig. 5, the paramyxovirus P proteins fall into two distinct subgroups, in which NDV is grouped with mumps virus, SV5, hPIV2 and hPIV4. A distinct subgroup is also formed by CDV, measles virus, bovine PIV3 (bPIV3), hPIV3, hPIV1 and Sendai virus. These viruses form similar subgroups when aligned on the basis of other viral protein sequences (data not shown). Sendai and related viruses all encode the P protein from unedited transcripts, whereas SV5 and related viruses require the insertion of two non-templated G residues into mRNA to generate the P protein. Somewhat surprisingly, NDV,

Fig. 6. Alignment of the extreme C-terminal regions of nine paramyxovirus V proteins. The number of residues from each sequence is indicated. Identical residues are marked by an asterisk and those conservatively substituted are indicated by a dot.

despite showing much greater relatedness to SV5 etc., encodes P from an unedited transcript.

When the entire P proteins of NDV, SV5, hPIV2, hPIV4, mumps virus, Sendai virus, measles virus, bPIV3 and CDV are aligned, the initiating methionine is the only amino acid found to be conserved (not shown). However, in a consensus alignment of 70 amino acids from the C terminus of the V proteins from the above viruses, there are 17 perfectly conserved residues, of which seven are cysteine residues (Fig. 6). The high degree of conservation in the C terminus of the V proteins suggests that it may interact with either a highly conserved protein found widely distributed amongst paramyxovirus host cells, as proposed by Curran *et al.* (1991*a*), or with an invariant macromolecule such as nucleic acid.

Discussion

The P gene mRNA of NDV has been shown to be modified during transcription by the addition of non-templated G residues at position 484. The analysis of 82 clones has shown that the RNA editing occurs at a single unique position within the gene, and that a proposed editing site (Cattaneo *et al.*, 1989) 25 bases downstream at position 509 remains unaltered in all cases studied. There is therefore a high degree of specificity in the selection of the site at which the insertion occurs. The precise nature of the signal is unclear, although the insertion of the non-templated G residues is thought to be due to a slippage or 'stuttering' of the transcriptase enzyme. The eight nucleotides preceding the insertion site in all NDV strains (3′ UUUUUCCC 5′, genomic sense) are identical to those in Sendai virus, and conform to a previously identified consensus sequence (3′ UUYU-CCC 5′) where Y is a pyrimidine base (Vidal *et al.*, 1990*b*). However the site at 509 (3′ UCCUUCCC 5′), proposed by Cattaneo *et al.* (1989), does not conform to this consensus and does not appear to be used for editing.

Temporary pausing (stuttering) of the transcriptive complex at the editing site is thought to allow the rearward slippage of the nascent mRNA on the vRNA template. The stability of the resultant duplex has been suggested as a means by which certain viruses insert a single G residue predominantly, and others insert two predominantly (Vidal *et al.*, 1990*b*). In the case of NDV, the slippage back by one base results in a single unstable base pairing (U:G) in the four bases prior to the insertion. A second slippage creates two such mismatches, and may therefore be considered to be a less favourable arrangement. Despite this, the frequency of insertion of multiple bases is greater than 10%. In several paramyxoviruses, the translation products of mRNAs with insertions other than to produce the V/P protein are significant non-structural proteins. For example, in bovine parainfluenza virus 3, the insertion of one and two Gs occurs with equal frequency, giving access to the V (+1G) and D (+2G) reading frames (Pelet *et al.*, 1991), whereas in certain strains of mumps virus the insertion of a single G residue gives rise to an mRNA that encodes the NS1 protein.

In NDV the majority of the transcripts with multiple insertions contain two extra G residues, and therefore give access to the third ORF beyond the insertion point. This, in turn, raises the possibility that such mRNAs may be encoding other NDV non-structural proteins. It has already been shown that the NDV 36K protein has the properties expected of a V protein (Samson *et al.*, 1991). However the origin of the 33K protein is unclear. Analysis of the potential ORFs that are accessed by the insertion of two Gs into P mRNA from the four sequenced strains of NDV shows that the length of the ORF thus created differs in each strain. In the case of Ulster, the reading frame for the W protein extends to 181 amino acids, whereas those for D26, Beaudette C and AV are of 141, 221 and 228 amino acids, respectively. The putative W protein in NDV strain Ulster has a highly basic C terminus, with 11 lysine and arginine residues amongst the unique 42 amino acids. It is notable that the W ORF actually represents the 3′ terminus of a shorter reading frame within the mRNA. Given the polycistronic nature of the P mRNA in other paramyxoviruses, the highly basic translation product of this ORF may be produced in NDV-infected cells.

The P protein of NDV bears limited but distinct similarity to the equivalent proteins from SV5, mumps, hPIV2 and hPIV4A and -4B, as well as the recently reported simian virus 41 sequence (Kawano *et al.*, 1993) (not shown). Of these viruses, all but NDV encode the V protein from an unedited transcript of the P gene, and insert predominantly two G residues to accomplish the switch to the P reading frame. The P proteins in these viruses are also of similar size, between 390 and 400 amino acids long. NDV, however, is similar to Sendai virus, measles virus etc. in that it encodes the P gene from the unaltered transcript, despite showing considerably less similarity to these viruses on the basis of

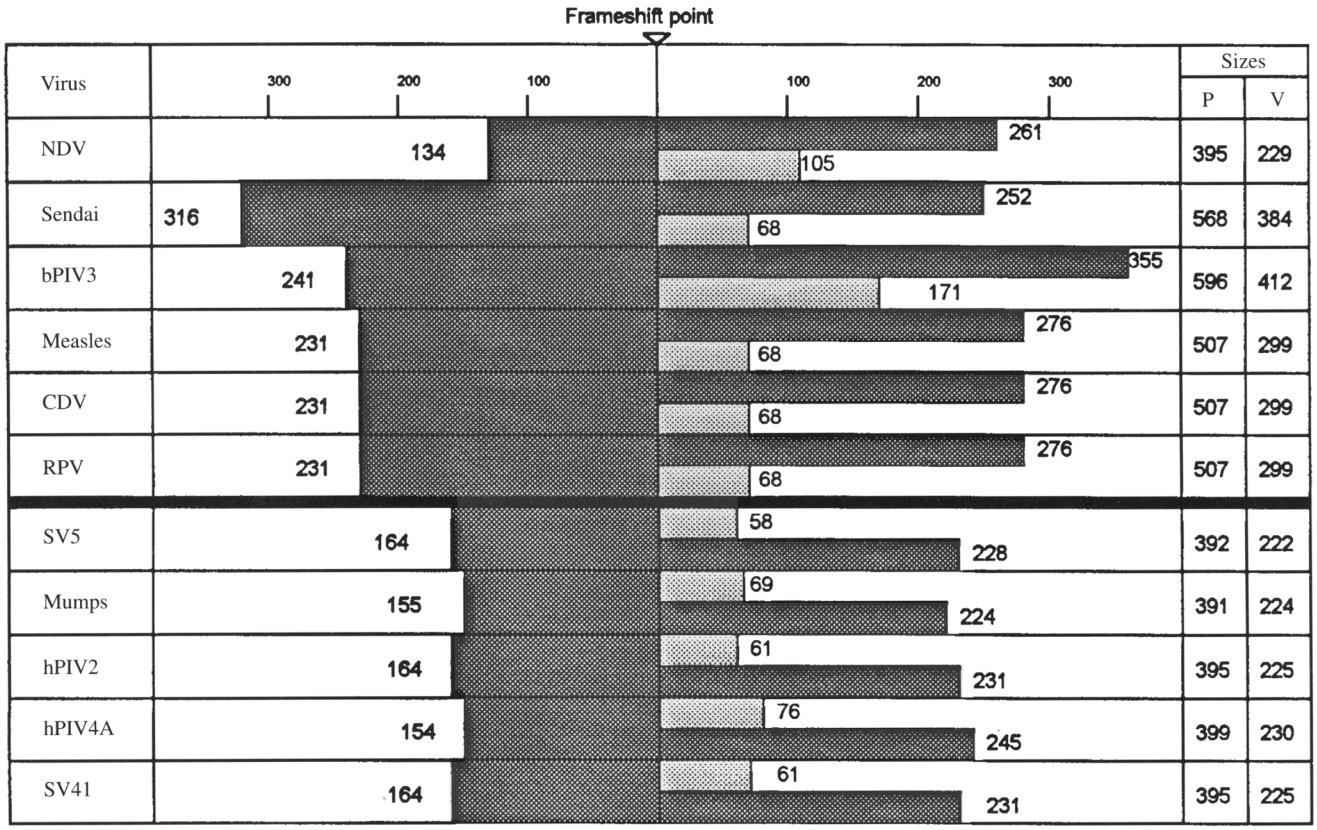

Fig. 7. Chart showing the comparative coding strategies within the P/V genes of several paramyxoviruses. The sequences are aligned at their respective frameshift points and the lengths of the amino and carboxyl termini are indicated. In each case, the P protein is indicated by the dark bar, and the unique region of the V protein by the light shading. The sequences above the horizontal bar represent viruses where the insertion event results in protein V and those below where protein V is specified from the unaltered transcript. The lengths of the respective amino acid sequences are shown in the right-hand columns. Only viruses known to encode a V protein are shown. In addition to those shown in Fig. 5, the coding schemes of rinderpest virus (RPV) (Yamanaka *et al.*, 1992) and SV41 (Kawano *et al.*, 1993) are shown.

amino acid sequence homology. The P proteins from these viruses are all larger (over 500 amino acids long) than those from the previously mentioned group. However, when the different P and V proteins are aligned around the point where the transcriptional editing occurs (Fig. 7), it is noticeable that the extra length in the majority of cases is within the N termini of the proteins. The viruses that encode the longer P proteins also encode C proteins from the P gene, and it is possible that the maintenance of the C ORF has prevented reduction in the size of the P ORF in these viruses.

Although the function of the proteins encoded by the V and W ORFs of NDV has not been established, work with Sendai virus (Curran *et al.*, 1991*a*) suggests that the V and W proteins act as inhibitors of replication. This is proposed to be due to the V and W proteins sharing the L protein-binding activity of P in their N termini, but lacking the nucleocapsid-binding domain which has been localized to the C terminus of the P protein (Ryan & Kingsbury, 1988; Ryan & Portner, 1990; Ryan *et al.*, 1991). The role of the highly conserved cysteine-rich C-terminal domain of the V protein has yet to be explained. Regularly spaced cysteine residues such as these are often found in proteins involved in protein–nucleic acid or protein–protein interactions. In the case of proteins that interact with nucleic acids, the cysteine residues are often co-ordinated around a central zinc atom to form a zinc finger (Berg, 1986). The NDV V protein has no classical zinc finger motif of either the C2H2 or C3HC4 type. However, the first four cysteines may fall into a subtype of the former group where the second pair of residues are also cysteines rather than the more common histidine.

Despite the highly conserved nature of the V protein, it is not ubiquitous amongst paramyxoviruses. Members of the pneumovirus family lack multiple cistrons within the P genes. The paramyxovirus hPIV1 lacks the contiguous V ORF, and hPIV3, despite possessing the V ORF, does not edit P mRNA to access the reading frame. However, the mRNA of hPIV3 is edited at an alternative site to create a hybrid protein, called D, which

may fulfil a role similar to that of V in other paramyxoviruses. In addition, both hPIV1 and -3, in common with other Sendai-like paramyxoviruses encode C proteins, which are absent from SV5-like viruses, including NDV.

The authors would like to thank Dr A. C. R. Samson of this department for his help with the preparation of viral mRNA, and for helpful discussion. This work was supported by grants from the Wellcome Foundation and the AFRC.

References

BARRETT, T., SHRIMPTON, S. B. & RUSSELL, S. E. H. (1985). Nucleotide sequence of the entire protein coding region of canine distemper virus polymerase associated (P) protein in mRNA. *Virus Research* **3**, 367–372.

BELLINI, W. J., ENGLUND, G., ROZENBLATT, S., ARNHEITER, H. & RICHARDSON, C. D. (1985). Measles virus P gene codes for two proteins. *Journal of Virology* **53**, 908–919.

BERG, J. (1986). Potential metal binding domains in nucleic acid binding proteins. *Science* **232**, 485–487.

CATTANEO, R., KAELIN, K., BACZKO, K. & BILLETER, M. A. (1989). Measles virus editing provides an additional cysteine-rich protein. *Cell* **56**, 759–764.

CHAMBERS, P. & SAMSON, A. C. R. (1982). Non-structural proteins in Newcastle disease virus-infected cells. *Journal of General Virology* **58**, 1–12.

CURRAN, J., BOECK, R. & KOLAKOFSKY, D. (1991*a*). The Sendai virus P gene expresses both an essential protein and an inhibitor of RNA synthesis by shuffling modules via mRNA editing. *EMBO Journal* **10**, 3079–3085.

CURRAN, J., DEMELO, M., MOYER, S. & KOLAKOFSKY, D. (1991*b*). Characterisation of the Sendai virus V protein with an anti-peptide antiserum. *Virology* **184**, 108–116.

DASKALAKIS, S., MENKE, J. B., STRIPP, B. R. & STONE, H. O. (1992). Nucleotide sequence of the phosphoprotein (P) gene of Newcastle disease virus (strain Beaudette C). *Nucleic Acids Research* **20**, 616.

ELLIOTT, G. D., YEO, R. P., AFZAL, M. A., SIMPSON, E. J. B., CURRAN, J. A. & RIMA, B. K. (1990). Strain-variable editing during transcription of the P gene of mumps virus may lead to the generation of non-structural proteins NS1 (V) and NS2. *Journal of General Virology* **71**, 1555–1560.

GALINSKI, M. S., MINK, M. A., LAMBERT, D. M., WECHSLER, S. L. & PONS, M. W. (1986). Molecular cloning and sequence analysis of the human parainfluenza 3 virus mRNA encoding the P and C protein. *Virology* **155**, 46–60.

GALINSKI, M. S., TROY, R. M. & BANERJEE, A. Y. (1992). RNA editing in the phosphoprotein gene of the human parainfluenza virus type 3. *Virology* **186**, 543–550.

GALLIONE, C. J., GREENE, J. R., IVERSON, L. E. & ROSE, J. K. (1981). Nucleotide sequences of the mRNAs encoding the vesicular stomatitis virus N and NS proteins. *Journal of Virology* **39**, 529–535.

GIORGI, C., BLUMBERG, B. M. & KOLAKOFSKY, D. (1983). Sendai virus contains overlapping genes expressed from a single mRNA. *Cell* **35**, 829–836.

GRUNSTEIN, M. & HOGNESS, D. S. (1975). Colony hybridization: a method for the isolation of cloned cDNAs which contain a specific gene. *Proceedings of the National Academy of Sciences, U.S.A.* **72**, 3961–3965.

HAMAGUCHI, M., YOSHIDA, T., NISHIKAWA, K., NARUSE, H. & NAGAI, Y. (1983). Transcriptive complex of Newcastle disease virus. I. Both L and P proteins are required to constitute an active complex. *Virology* **128**, 105–117.

HIGGINS, D. G. & SHARPE, P. M. (1988). CLUSTAL: A package for performing multiple sequence alignments on a microcomputer. *Gene* **73**, 237–244.

HOLMES, D. S. & QUIGLEY, M. (1981). The rapid boiling method for the preparation of plasmid DNA. *Analytical Biochemistry* **114**, 193–197.

ISHIDA, N., TAIRA, H., OMATA, T., MIZUMOTO, K., HATTORI, S., IWASAKI, K. & KAWAKITA, M. (1986). Sequence of 2617 nucleotides from the 3′ end of Newcastle disease virus genome RNA and the predicted amino acid sequence of the viral NP protein. *Nucleic Acids Research* **14**, 6551–6564.

KAWANO, M., TSURUDOME, M., OKI, N., NISHIO, M., KOMADA, H., MATSUMURA, H., KUSAGAWA, S., OHTA, H. & ITO, Y. (1993). Sequence determination of the P gene of simian virus 41: presence of irregular deletions near the RNA-editing sites of paramyxoviruses. *Journal of General Virology* **74**, 911–916.

KOLAKOFSKY, D., VIDAL, S. & CURRAN, J. (1990). Paramyxovirus RNA synthesis and P gene expression. In *The Paramyxoviruses*, pp. 215–233. Edited by D. W. Kingsbury. New York: Plenum Press.

KONDO, K., BANDO, H., TSURUDOME, M., KAWANO, M., NISHIO, M. & ITO, Y. (1990). Sequence analysis of the phosphoprotein (P) genes of human parainfluenza type 4A and 4B viruses and RNA editing at transcript of the P genes: the number of G residues added is imprecise. *Virology* **178**, 321–326.

KOZAK, M. (1981). Mechanism of mRNA recognition by eukaryotic ribosomes during initiation of protein synthesis. *Current Topics in Microbiology and Immunology* **93**, 81–123.

LAMB, R. A. & PATERSON, R. G. (1990). The nonstructural proteins of paramyxoviruses. In *The Paramyxoviruses*, pp. 181–214. Edited by D. W. Kingsbury. New York: Plenum Press.

MATSUOKA, Y., CURRAN, J., PELET, T., KOLAKOFSKY, D., RAY, R. & COMPANS, R. W. (1991). The P gene of human parainfluenza virus type 1 encodes P and C proteins but not a cysteine rich V protein. *Journal of Virology* **65**, 3406–3410.

MCGINNES, L. W., MCQUAIN, C. & MORRISON, T. G. (1988). The P protein and the nonstructural 38K and 29K proteins of Newcastle disease virus are derived from the same open reading frame. *Virology* **164**, 256–264.

MILLAR, N. S. & EMMERSON, P. T. (1988). Molecular cloning and nucleotide sequencing of Newcastle disease virus. In *Developments in Veterinary Virology: Newcastle Disease*, pp. 79–97. Edited by D. J. Alexander. Boston: Kluwer Academic Publishers.

MILLAR, N. S., CHAMBERS, P. & EMMERSON, P. T. (1988). Nucleotide sequence of the fusion and haemagglutinin–neuraminidase glycoprotein genes of Newcastle disease virus, strain Ulster: molecular basis for variations in pathogenicity between strains. *Journal of General Virology* **69**, 613–620.

PELET, T., CURRAN, J. & KOLAKOFSKY, D. (1991). The P gene of bovine parainfluenza virus 3 expresses all three reading frames from a single mRNA editing site. *EMBO Journal* **10**, 443–448.

RYAN, K. W. & KINGSBURY, D. W. (1988). Carboxy-terminal region of Sendai virus P protein is required for binding to viral nucleocapsids. *Virology* **167**, 106–112.

RYAN, K. W. & PORTNER, A. (1990). Separate domains of Sendai virus P protein are required for binding to viral nucleocapsids. *Virology* **174**, 515–521.

RYAN, K. W., MORGAN, E. M. & PORTNER, A. (1991). Two noncontiguous regions of Sendai virus P protein combine to form a single nucleocapsid binding domain. *Virology* **180**, 126–134.

SAKAI, Y., SUZU, S., SHIODA, T. & SHIBUTA, H. (1987). Nucleotide sequence of the bovine parainfluenza 3 virus genome: its 3′ end and the genes of NP, P, C and M proteins. *Nucleic Acids Research* **15**, 2927–2944.

SAMSON, A. C. R., LEVESLEY, I. & RUSSELL, P. H. (1991). The 36K polypeptide synthesized in Newcastle disease virus-infected cells possesses properties predicted for the hypothesized 'V' protein. *Journal of General Virology* **72**, 1709–1713.

SMITH, G. W. & HIGHTOWER, L. E. (1981). Identification of the P proteins and other disulphide linked and phosphorylated proteins of Newcastle disease virus. *Journal of Virology* **37**, 256–267.

SOUTHERN, J. A., PRECIOUS, B. & RANDALL, R. E. (1990). Two nontemplated nucleotide additions are required to generate the P mRNA of parainfluenza virus type 2 since the genome encodes protein V. *Virology* **177**, 388–390.

TAKEUCHI, K., HISHIYAMA, M., YAMADA, A. & SUGIURA, A. (1988). Molecular cloning and sequence analysis of the mumps virus gene encoding the P protein: mumps virus P gene is monocistronic. *Journal of General Virology* **69**, 2043–2049.

THOMAS, S. M., LAMB, R. A. & PATERSON, R. G. (1988). Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell* **54**, 891–902.

VIDAL, S., CURRAN, J. & KOLAKOFSKY, D. (1990*a*). Editing of the Sendai virus P/C mRNA by G insertion occurs during mRNA synthesis via a virus encoded activity. *Journal of Virology* **64**, 239–246.

VIDAL, S., CURRAN, J. & KOLAKOFSKY, D. (1990*b*). A stuttering model for paramyxovirus P mRNA editing. *EMBO Journal* **9**, 2017–2022.

YAMANAKA, M., DALE, B., CRISP, T., CORDELL, B., GRUBMAN, M. & YILMA, T. (1992). Sequence analysis and editing of the phosphoprotein (P) gene of rinderpest virus. *Virology* **190**, 553–556.

(*Received 17 June 1993; Accepted 3 August 1993*)