Synthesis of the Infectious Pancreatic Necrosis Virus Polyprotein, Detection of a Virus-Encoded Protease, and Fine Structure Mapping of Genome Segment A Coding Regions

ROY DUNCAN, EVA NAGY, PETER J. KRELL, AND PETER DOBOS*

Department of Microbiology, College of Biological Science, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Received 6 March 1987/Accepted 5 August 1987

Infectious pancreatic necrosis virus (IPNV) is an economically important fish pathogen and is the prototype of the newly established Birnaviridae virus family (4). Other members of the family include infectious bursal disease virus of domestic fowl and Drosophila X virus (DXV) of Drosophila melanogaster (10). Birnaviruses possess a bisegmented, double-stranded RNA genome contained within a medium-sized, unenveloped, isosahedral capsid (11). Birnavirus gene expression involves the production of four unrelated major gene products which undergo various posttranslational cleavages to produce three to five viral structural proteins (1, 17, 25, 26). The largest protein, VP1, is encoded by the smaller segment B genomic RNA, whereas the remaining three major gene products are encoded by the larger segment A RNA (1, 21, 23, 25).

Genome segment A is 3,097 base pairs long and contains a single large open reading frame (ORF) of 2,916 bases capable of encoding a 106,000 molecular-weight polyprotein (106K polyprotein) (13). The ORF is bracketed by 5' and 3' terminal flanking regions (119 and 62 bases, respectively), which presumably contain sites for ribosome and RNA polymerase binding. A second small ORF in reading frame 2 (444 bases), starting with an initiation codon at 68 bases, which overlaps with the 5' end of the polyprotein ORF (Fig. 1) and could encode a 17K, arginine-rich polypeptide was also identified. Neither of these predicted gene products has ever been detected in vitro or in vivo.

Recently, Hudson et al. (17) reported that the sequence of genome segment A of infectious bursal disease virus, like that of IPNV, contains a single large ORF. Thus the polyprotein mode of gene expression may be common among birnaviruses.

The order of the coding regions of the three major gene products of IPNV segment A (VP2, VP3, and NS) on the large ORF is 5'-pVP2-NS-VP3-3' (13, 27). The coding regions on the IPNV Sp strain, infectious bursal disease virus, and Drosophila X virus large RNA segments are in the same order (16, 17, 26). The boundaries of the coding regions of pVP2, NS, and VP3 are not known.

The discovery of the large ORF on segment A suggested that the three major gene products of this genome segment must arise by cotranslational proteolytic cleavage of a precursor polypeptide. These cleavages occur both in infected cells and during cell-free translations primed with denatured, viral genomic RNA. The location of the cleavage sites and the identity of the protease(s) involved are not known. Since the cleavages occur in vitro, it has been proposed that at least one of the virus-encoded proteins may be a protease, possibly the only nonstructural protein, the NS polypeptide (17).

In the case of IPNV, previous in vitro translations strongly suggested that the three major gene products of IPNV genome segment A (the precursor to VP2 [pVP2], VP3, and the nonstructural polypeptide, NS) are produced by independent initiation of translation on a polycistronic mRNA (23, 27). It would seem that, at least in vitro, IPNV is capable of generating viral polypeptides both by polyprotein processing and by independent translation initiation at internal sites. A similar situation has been reported to occur for poliovirus gene expression (12, 28).

The purpose of this study was threefold: (i) to identify and isolate the IPNV polyprotein, (ii) to determine whether any of the three major gene products of segment A possesses proteolytic activity and is involved in polyprotein processing, and (iii) to localize the pVP2 and NS carboxy-terminal cleavage sites and the internal translation initiation sites used for NS and VP3 production in vitro. The detection and purification of the polyprotein should facilitate further studies aimed at determining the series of events involved in structural protein formation and would confirm the sequence prediction of a polyprotein mode of gene expression.
We report here the identification and gel purification of the IPNV polyprotein. Partial peptide mapping indicates that it contains the coding regions of the major gene products and confirms that IPNV utilizes a polyprotein mode of gene expression. We present evidence indicating that NS is a virus-encoded protease responsible for the cleavage which generates pVP2 from the polyprotein. We have also localized the cleavage site which generates the carboxy terminus of NS and located what we believe represent the internal translation initiation sites used to generate the amino termini of NS and VP3 in vitro.

MATERIALS AND METHODS

Cells and virus. IPNV Jasper strain (serotype 1) was grown in CHSE-214 cell monolayers and purified as previously described (10).

Viral RNA. IPNV double-stranded genomic RNA was isolated from purified virions by disruption with 1% sodium dodecyl sulfate (SDS) and incubation with 1 mg of proteinase K per ml at 37°C for 4 h followed by extensive phenol and phenol-chloroform extractions. Purified RNA was ethanol precipitated in the presence of 2.5 M ammonium acetate at −20°C and suspended in sterile TE buffer at pH 7.5 (10 mM Tris hydrochloride, 1 mM EDTA). The concentration was determined by measuring the A_{260} and the quality of the RNA was checked by denaturation with 10 mM methyl-mercury hydroxide followed by electrophoresis in 1% agarose gels containing 0.5 μg of ethidium bromide per ml.

Uninfected and infected cell lysates. [35S]methionine-labeled, uninfected cells and IPNV-infected cells were prepared by a 4-h pulse with 10 μCi of [35S]methionine (New England Nuclear Corp., Boston, Mass.) per ml starting at 8 h postinfection.

IPNV cDNA production. Purified IPNV genomic RNA was denatured in 90% dimethyl sulfoxide (50°C for 45 min), tailed with poly(A) polymerase (Bethesda Research Laboratories, Gaithersburg, Md.), and used as templates for cDNA production with avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc., St. Petersburg, Fla.) as previously described (6). Plus- and minus-strand cDNA was size fractionated on 5 to 20% sucrose gradients (105,000 × g, 4.5 h, 20°C). Fractions containing cDNA greater than 1,500 bases long were pooled. The plus- and minus-strand cDNAs were hybridized, and truncated molecules were extended by using reverse transcriptase and the Klenow fragment of DNA polymerase I by standard procedures (22). Double-stranded cDNA was oligo(dC) tailed and ligated into oligo(dG)-tailed pBR322 at the PstI site (22). Recombinant plasmid DNAs containing cDNA inserts were characterized by Southern blotting (32) and probing with single-stranded IPNV cDNA. Segments A-specific clones were identified by Northern blotting (RNA blotting) of IPNV genomic RNA (33) and probing with nick-translated cDNA clones (29). The two largest segment A cDNAs were subcloned into M13mp19 and sequenced by a modified dideoxy chain-termination procedure (3, 9). These clones were subsequently subcloned into plasmid transcription vectors.

Transcription plasmid subcloning. IPNV cDNA fragments from M13mp19 cDNA clones were subcloned into pGEM1 and pGEM2 plasmid vectors (Promega, Madison, Wis.) in TB1 cells (Bethesda Research Laboratories). The clones were identified according to the vector and the approximate size of the insert (in hundreds of base pairs). Clones pG1.31 and pG2.31 contained full-length IPNV cDNA inserts (1 to 3,097 base pairs) cloned into the HindIII-XbaI sites in pGEM1 and pGEM2, respectively. Clone pG1.27 contained a 2,746-base-pair cDNA insert (bases 351 through 3097) in the EcoRI-XbaI sites of pGEM1. Clone pG1.8 contained the 3′-terminal 746-base PstI fragment cloned into the PstI site of pGEM1, whereas pG1.22 contained the 5′-terminal 2,276-base PstI fragment. Clone pG2.21 was a 2,193-base 5′-terminal clone up to the Sall site cloned into the HindIII-Sall sites of pGEM2. All of the

![Figure 1](http://jvi.asm.org/)

**FIG. 1.** Distribution of termination codons (vertical lines) in all three reading frames of both the plus- and minus-strand RNAs of genome segment A of IPNV. The top diagram represents a partial physical map of genome segment A cDNA, and the approximate locations of the pVP2-, NS-, and VP3-coding regions are indicated above the map. The figure represents the plus strand in the 5′-to-3′ direction. Abbreviations: P, PstI; B, BamHI; E, EcoRI; PV, PvuII; X, XhoI; S, SalI.
clones were physically mapped to ensure the correct orientation and size of the insert, and the identity of the clones was confirmed by Southern blotting and probing with nick-translated IPNV cDNA cloned into M13mp19.

**mRNA production with T7 polymerase.** Uncapped mRNA was prepared in vitro by using T7 RNA polymerase (Bethesda Research Laboratories and Promega Biotec) to generate runoff transcripts from linearized, cDNA-containing vectors. Vectors were linearized by restriction endonuclease digestion (10 U/μg of DNA), followed by digestion with 1 mg of proteinase K per ml and 1% SDS for 30 min at 37°C and subsequent phenol extraction and ethanol precipitation. pG1.31, pG1.22, and pG1.8 were linearized with HindIII, pG2.31 and pG1.27 were linearized with Xbal, and pG2.21 was linearized with Sall. Linearized plasmid DNA was suspended at 0.1 mg/ml in diethylpyrocarbonate (Sigma Chemical Co., St. Louis, Mo.)-treated, distilled H2O and used as templates for T7 polymerase. The 100-μl reaction mixtures contained 40 mM Tris hydrochloride (pH 7.5), 6 mM MgCl2, 2 mM spermidine, 10 mM NaCl, 10 mM dithiothreitol, 0.5 mM each ribonucleotide, 1.5 U of plasmid DNA, and 15 U of T7 RNA polymerase. Reactions were incubated for 90 min at 37°C, and the DNA template was digested for 15 min at 37°C with 0.16 μg of RNase-free DNase I (Bethesda Research Laboratories). The RNA was phenol and phenol-chloroform extracted, followed by three cycles of ethanol precipitation in the presence of 2.5 M ammonium acetate. The mRNA was suspended at 1 mg/ml in diethylpyrocarbonate-treated, distilled H2O and stored in working samples at −70°C. The concentration was checked by A260, and the quality of the RNA was analyzed by methyl mercury denaturation (10 mM for 15 min at room temperature) and agarose gel electrophoresis. The yield was between 2 and 10 μg of RNA per μg of DNA.

**In vitro translations.** Methylmercury hydroxide-denatured genomic RNA and T7 transcripts were translated in vitro in cell-free rabbit reticulocyte lysates (New England Nuclear or Promega Biotec). Other in vitro translation systems were also tried (e.g., rabbit reticulocyte lysates from Bethesda Research Laboratories or Amersham Corp. [Arlington Heights, Ill.] or a cell-free wheat germ system), but we were unable to translate denatured genomic viral RNA with these systems. The 25-μl reactions were prepared according to the manufacturers’ specifications and contained 40 μCi of [35S]methionine (New England Nuclear), 1 U of plasmid RNase inhibitor per μl, and 100 to 200 ng of T7 transcripts or 500 ng of denatured, genomic RNA. Exogenous K+ and Mg2+ (as the acetate salts) were added to the New England Nuclear lysates to final concentrations of 50 and 2 mM, respectively. Reactions were incubated at 37°C for 90 min and terminated by the addition of 1/5 volume of 5× electrophoresis sample buffer (10% β-mercaptoethanol, 5% SDS, 50% glycerol, 0.5 M Tris hydrochloride [pH 6.8], 0.1% bromophenol blue) and boiled for 3 min. The products were fractionated on discontinuous 11% acrylamide gels containing 0.24% bisacrylamide and 1% SDS (5) at a constant voltage of 15 V/cm. Molecular weight markers (Sigma) were included in the gel and visualized by staining with Coomassie blue and destaining (19). Markers and their approximate molecular weights were phosphorylase b (94,000), bovine serum albumin (67,000), ovalbumin (43,000), carbonic anhydrase (30,000), soybean trypsin inhibitor (20,100), and lysozyme (14,400).

**Purification of viral polypeptides.** [35S]Methionine-labeled viral polypeptides from infected cell lysates or from in vitro translations were fractionated by SDS-polyacrylamide gel electrophoresis (PAGE) and localized by autoradiography. Specific polypeptides were excised from unstained, dried gels and recovered from reswollen gel slices by electrophoresis in Laemmli running buffer (19) with ISC0 sample concentrator cups (Instrumentation Specialties Co., Lincoln, Neb.) for 4 h at 150 V at room temperature with the addition of 20 μg of carrier bovine serum albumin. The electroeluted polypeptides were dialyzed against TE (pH 7.5), and the amount of trichloroacetic acid-precipitable radioactivity present in each sample was determined. The purity of the polypeptides was checked by SDS-PAGE (11% polyacrylamide gels).

**Partial proteolysis with NCS.** The purified polypeptides were digested with N-chlorosuccinimide (NCS: Aldridge Chemical Co., Inc., Milwaukee, Wis.), which cleaves proteins at tryptophan residues under partial proteolysis conditions (20). Each reaction contained 150 μl of 35S-labeled polypeptides (approximately 20 to 40 cpn, 1 to 10 μg of protein) in 10 mM Tris hydrochloride (pH 7.5), 0.15 g of urea, 150 μl of glacial acetic acid, and 0.03 M NCS. Reactions were incubated at room temperature for 30 min. The reactions were stopped, and the peptides were precipitated by the addition of an equal volume of 20% trichloroacetic acid and overnight incubation at −20°C. The precipitate was recovered by centrifugation for 30 min at 4°C in a microfuge; the pellets were washed twice with ethanol-ether (1:1), suspended in 50 μl of electrophoresis sample buffer, and boiled. Peptides were fractionated by SDS-PAGE (15% polyacrylamide gels); the gels were stained with Comassie blue, destained, treated with En3Hance (New England Nuclear) for fluorography, dried, and exposed on Kodak XAR X-ray film at −70°C for 3 to 10 days. The same molecular weight markers were used as for 11% gels with the addition of RNase A (13,700 molecular weight) and RNase T1 (11,100 molecular weight). The gel molecular weight estimates of the NCS-treated peptides were reduced by 5% to account for peptide oxidation (20). The 5% estimate was determined by comparing the mobilities of untreated, purified polypeptides with the mobilities of the corresponding NCS-treated, undigested polypeptides.

**Computer analysis of the cDNA sequence.** The cDNA sequence was analyzed on an IBM-PC computer by using the software programs of Schwindinger and Warner (31) and Mount and Conrad (24).

**RESULTS**

**Isolation of the IPNV polyprotein.** To identify the IPNV polyprotein we subcloned full-length clones and various truncated segment A-specific cDNA clones from M13mp19 into the transcription vectors pGEM1 and pGEM2. Linearized vectors were used as templates to produce the runoff transcripts, which were translated in vitro; the products were analyzed by SDS-PAGE (Fig. 2). Translation of the minus-sense mRNA (from pG1.31, bases 1 through 3097) produced no polypeptides other than those produced by endogenous mRNA (data not shown). Translation of full-length, plus-sense mRNA (from pG2.31, bases 1 through 3097) produced several polypeptides not present in endogenous translations (lanes 31 in Fig. 2). Polypeptides were detected which comigrated with pVP2, VCY, and NS present in-infected cells (lane 1C) and with polypeptides produced by the in vitro translation of denatured, genomic RNA (lane DS). In addition, seven to
nine other polypeptides of various sizes were produced which did not correspond to any previously described IPNV gene products. They are identified by their gel molecular weight estimates in Fig. 2.

The largest product detected was a polypeptide with an estimated molecular weight of 101,000 (PP in Fig. 2), in good agreement with the size estimate of the predicted IPNV polyprotein (106,000 molecular weight). This polypeptide migrated above VP1, the gene product of genome segment B, which is the largest previously identified IPNV gene product. Some translations primed with denatured, genomic RNA produced a very faint product which comigrated with the 101K polypeptide, although it was not usually evident.

Translation of the 5′-terminal truncated mRNA (from pG1.27, bases 351 through 3097) yielded the same polypeptides as those obtained with full-length mRNA, except for pVP2 and the 101K polypeptide (Fig. 2). Transcripts derived from this clone contained a 5′-terminal deletion, to the internal EcoRI site at 351 base pairs, and lacked methionine codons 1 and 2 (Met-1 and Met-2) of the large ORF (see Fig. 4A for the locations of the in-phase methionine codons). These results confirmed the mapping of pVP2 to the 5′ end of the RNA and indicated that the 101K product was probably the predicted IPNV polyprotein which must initiate at Met-1 or Met-2 of the large ORF. The 67K polypeptide could not always be easily detected from pG1.27, and its intensity varied considerably in different preparations, even with pG2.31. Low-molecular-weight (14,000 to 20,000) polypeptides present also in the 5′-truncated mRNA translation (pG1.27) could not have arisen from translation of the second small ORF in reading frame 2. We have not been able to detect the predicted 17K product of the small ORF in vivo or in vitro.

The relative amounts of the various translation products varied with the source of the RNA, the translation kit, and the K+ concentration (Fig. 2). Translations with the New England Nuclear kits were routinely done at a K+ concentration of 50 mM to enhance the translation of uncapped mRNA (2).

Denaturation of the T7 mRNA with 10 mM CH3HgOH before translation or the addition of placental RNase inhibitor to 3 U/ul in the translation mixture did not affect the polypeptide profile (data not shown). Incubation of the translation mixture at 30°C, which reportedly can increase the production of larger polypeptides (7), did not affect the quality or quantity of the translation products.
polypeptide presumably cleavage to produce the amino terminus of VP3 must occur close to the amino-terminal methionine residue of VP3. Since we were unable to obtain amino-terminal amino acid sequences from virion-derived VP3 due to a block in amino-terminal amino acid sequence determination, we attempted to map the amino terminus of VP3 by identifying the internal translation initiation site with a truncated cDNA clone.

The 5'-proximal methionine codon in clone pG1.8 (bases 2351 through 3097) was met-18 (2,364 bases in the large ORF), which is contained in a perfect Kozak consensus sequence for a eucaryotic translation initiation site (CCAC CATGG) (18) and seemed a likely candidate for the VP3 translation initiation site.

Translation of pG1.8-derived transcripts produced one dominant translation product in large amounts which as a 29K polypeptide is smaller by approximately 2,000 daltons than authentic VP3 (Fig. 4A), but with a similar NCS digestion profile of VP3 (Fig. 4B). This polypeptide was purified from gels and digested with NCS. The peptide profile was compared with the VP3 peptide profile (Fig. 4B). This translation product clearly represented a truncated VP3. The X’s in Fig. 4B indicate the two peptides which migrated faster than the corresponding peptides of VP3. The

Peptide mapping of in vitro-produced IPNV gene products. Major products of in vitro translation of T7 transcripts or denatured genomic RNA and pVP2, VP3, and NS from IPNV-infected cells pulse-labeled with [35S]methionine were fractionated by SDS-PAGE and recovered by electroelution as described in Materials and Methods. Samples of the purified polypeptides were partially digested with 0.03 M NCS (cleaves at tryptophan residues) and then analyzed by SDS-PAGE on 15% polyacrylamide gels. Figure 3 illustrates the peptide profiles of pVP2, VP3, and NS obtained from T7 transcripts, denatured genomic RNA, and IPNV-infected cells compared with the peptide profile of the 101K polypeptide from T7 transcripts. Each profile of the three major IPNV gene products was unique except for the presence of a 14,000- to 15-molecular-weight peptide in both the NS and VP3 patterns (Fig. 3, lanes NS and VP3).

The same nine major peptides of pVP2 and the three major cleavage peptides of VP3 were all present in the putative polyprotein cleavage pattern. Extra bands for the polypeptide presumably represent peptides which span the pVP2-NS-VP3 junctions.

The NCS-treated peptide patterns of each of the major gene products were identical regardless of the source of the polypeptide (e.g., NS from infected cells and from in vitro translation of T7 mRNA and genomic RNA yielded the same two peptides with identical mobilities after NCS cleavage). This demonstrated that the cDNA clones were capable of directing the production of authentic viral gene products and that the IPNV polyprotein was accurately cleaved in vitro to generate the carboxy termini of both pVP2 and NS.

Mapping the amino terminus of VP3. The similarity of the NCS-treated peptide patterns of VP3 produced in vivo or in vitro indicated either that the in vitro internal translation initiation site of VP3 also functions in vivo or that

FIG. 3. NCS-generated partial peptide profiles of selected IPNV gene products separated by SDS-PAGE (15% polyacrylamide gels). [35S]methionine-labeled pVP2, NS, and VP3 were obtained from either IPNV-infected cells (lanes I), in vitro translations primed with denatured genomic RNA (lanes D), or pG2.31 T7 transcripts (lanes T). The polyprotein (PP) was purified from pG2.31 T7 transcript translations. NCS indicates the migration of purified pVP2, NS, and VP3 before NCS treatment. Numbers down the right side represent molecular weight markers (in thousands). The locations of uncleaved pVP2, NS, and VP3 are indicated on the left side of the corresponding lanes. The nine major partial peptides of pVP2 and the three major peptides of VP3 that were evident in the PP profile are indicated by horizontal lines. The three dots in lane NS-T indicate the locations of contaminating VP3 peptides.

FIG. 4. (A) In vitro translation products of T7-derived transcripts of clones pG2.31 and pG1.8. END represents endogenous products, and INF represents IPNV-infected cell polypeptides. Molecular weight markers (in thousands) are indicated on the left. The major IPNV translation products are labeled, and the approximate molecular weight (in thousands) of the major pG1.8 product is indicated. The line drawing shows the region of the cDNA contained in clone pG1.8. Numbers under the physical map represent the 24 in-phase methionine codons present in the large ORF. (B) NCS-treated peptide profiles of purified VP3 produced by pG2.31 (lane 31) and pG1.8 (lane 8) separated by SDS-PAGE (15% polyacrylamide gel). The approximate molecular weights of the peptides are indicated (in thousands) and have been corrected for oxidation (see Materials and Methods). The peptides of pG1.8 VP3 that migrated differently than their counterparts from pG2.31 are indicated with X’s.
location of the four smaller peptides in the VP3 coding region could be

produce nine tryptophan-labeled partial cleavage products with NCS. With Met-17 as a start site, these peptides would have molecular weights of 36,500 (uncut VP3), 22,000, 21,000, 15,400, 11,300, 7,000, 5,300, 3,600, and 1,700. On our

15% gel system, the four smaller peptides would not be

FIG. 5. IfN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions are indicated on the right. The

underlined sequences are those used in the coding regions and should be

of PTVN cDNA. Met-18, 3660 DUNCAN ET

the cDNA strap strand. A cDNA plus-strand sequence from 1,416 bases to the 5' and 3' ends of the minus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

3000 2600 2200 1800 1400 1000 600 200 0

underlined sequences are those used in the coding regions and should be

of PTVN cDNA. Met-18, 3660 DUNCAN ET

the cDNA strap strand. A cDNA plus-strand sequence from 1,416 bases to the 5' and 3' ends of the minus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be

FIG. 5. IFN-polymerase (PNV) polymerase-encoding segment (3,097 bases). The 5' and 3' ends of the minus-strand segment are from 1,416 bases to the 5' and 3' ends of the plus-strand segment. The numbers at the ends of the two vectors are those used in the coding regions and should be

underlined sequences are those used in the coding regions and should be

15% gel system, the four smaller peptides would not be

underlined sequences are those used in the coding regions and should be
The sequences are aligned with the common AACAA sequence in the 5'-to-3' direction of the plus cDNA strand. Vertical lines indicate exact matches, and Y and R indicate conserved pyrimidine and purine residues, respectively. The conserved AACAA and ATC sequences are highlighted, and initiation codons are boxed. Horizontal lines indicate single base shifts. The numbers indicate the nucleotide position from the 5' end of the plus cDNA strand.

Peptides with gel estimated molecular weights of 24,000, 22,000, and 15,000 are present in both the pG1.31 and pG1.8 VP3 products (Fig. 4B). These peptides of VP3 presumably correspond to the 23.1K, 21.5K, and 15.4K internal peptides. The 30K (uncut VP3) polypeptide and the 12K peptide showed faster electrophoretic mobilities (lane 8) than the corresponding peptides of full-length VP3 (lane 31), with estimated molecular weight reductions of 1,000 to 1,500. These peptides presumably represent the amino-terminal peptides (26.7K and 11.3K), indicating that VP3 produced by pG1.8 was amino truncated. This indicated that Met-18 is not the internal translation initiation site used for VP3 production in vitro.

These data suggested that Met-16 or Met-17 (2307 or 2328 bases, respectively) must represent the internal translation initiation site for VP3. The NCS peptide profiles were inconsistent with translation initiation at any other methionine codon. Translation initiation at Met-16 or Met-17 would produce VP3 polypeptides with molecular weights of 27,400 or 26,700, respectively. Since gel molecular weight estimates are not sufficiently precise to distinguish between these two possibilities, we cannot conclusively state whether Met-16 or Met-17 is used for VP3 production in vitro. We compared the sequence surrounding Met-1, the start site of pVP2 and probably the polyprotein, with the sequence surrounding Met-17 (Fig. 6). The sequences flanking Met-1 and Met-17 have approximately a 55% sequence homology, whereas there is less than 17% sequence homology between Met-1 and Met-16-flanking sequences with no conserved groups of nucleotides (data not shown). In view of the sequence conservation around Met-17 we propose that Met-17 is the most likely internal translation initiation site for VP3 production in vitro.

Detection of a viral protease and fine-structure mapping of its coding region. To map the boundaries of the NS-coding region, we prepared two cDNA fragments containing 3'-terminal plus-strand truncations, pG2.21 (1 to 2,198 bases) and pG1.22 (1 to 2,281 bases) (Fig. 7), mRNA produced from these clones should yield pVP2 and either a full-length or carboxyl-terminal truncated NS depending on where the NS-coding region ends.

Figure 7 shows the translation products produced by these clones. Clone pG2.21 gave an unexpected polypeptide pattern. The dominant product had a gel molecular weight
estimate of 80,000, and none of the other polypeptides produced corresponded to any of the IPNV gene products produced by full-length RNA (clone pG2.31). We interpreted these results to mean that the carboxy-truncated polypeptide produced by pG2.21 was incorrectly processed or not cleaved at all. The dominant 80K product presumably represented the unprocessed, truncated polypeptide, which, from the cDNA sequence, should have a molecular weight of 75,300.

Clone pG1.22 contained the identical cDNA insert as pG2.21, except that it was extended by 84 bases at the 3’ end of the plus strand. This clone should produce the pVP2-NS part of the polypeptide, with the NS containing 31 more amino acids than the NS of pG2.21. Clone pG1.22 directed the synthesis of normal amounts of pVP2, which was the dominant translation product, and minute amounts of a polypeptide that comigrated with NS on 10% polyacrylamide gels (Fig. 7). In addition, the 49K and 37K polypeptides, which are also produced by genomic RNA in vitro (Fig. 2), were produced in normal amounts.

The pVP2, 49K, 37K, and NS polypeptides from pG1.22 and the 80K product from pG2.21 were gel purified and digested with NCS. The peptide profiles were compared to the profiles of the corresponding polypeptides produced by full-length RNA (pG2.31) (Fig. 8). The peptide profile of the 80K product of pG2.21 contained the nine major peptides of pVP2 in addition to some extra higher- and lower-molecular-weight peptides, indicating that it represented the unprocessed, truncated polypeptide.

The pVP2, 49K, and 37K polypeptides produced by pG1.22 were indistinguishable from the corresponding polypeptides produced by pG2.31 (Fig. 8). Since the NS of pG1.22 was slightly smaller than authentic NS by approximately 500 daltons, the carboxy terminus of the NS-coding region must be encoded by the region immediately after the PstI site at 2276 bases, the 3’ end of pG1.22. The data also indicated that NS is involved in the cleavage that generates the carboxy terminus of pVP2, which is presumably the same terminus as the termini of the 49K and 37K polypeptides, suggesting that NS is a virus-encoded protease. The production of the pVP2, 49K, and 37K polypeptides was dependent on the presence of a functional NS gene product. An analysis of the 49K and 37K NCS-treated peptide patterns (Fig. 8) revealed that these two polypeptides are most likely derived by cleavage of the 95K and 89K polypeptides at the pVP2-NS cleavage site (data not shown). The 95K and 89K polypeptides are derived by translation initiation at downstream methionine codons (see pG1.27 in Fig. 2), most probably Met-3 and Met-5 (543 and 696 bases, respectively), and thus represent amino-terminally truncated polyproteins. The 49K and 37K polypeptides would therefore represent amino-terminally truncated pVP2 polypeptides.

Knowing the location of the carboxy terminus of NS, we used gel molecular weight estimates of NS to determine which methionine codon serves as the internal translation initiation site for NS production in vitro.

Gel molecular weight estimates of NS range from 27,000 to 29,000. Using the internal NCS-digested peptides of VP3 (Fig. 4B) as internal molecular weight standards on a 15% polyacrylamide gel and extrapolating, we estimated the molecular weight estimate for NS of 26,000, slightly smaller than the estimate of 28,000 molecular weight obtained with external standards. There are only two in-phase methionine codons (Met-10 and Met-11) (1,515 and 1,629 bases, respectively) that could serve as initiation codons to generate a polypeptide in the expected size range (27,700 and 23,800 molecular weight, respectively). Met-9, the next upstream methionine codon, would produce a 34.5K NS polypeptide. Examination of the sequence from Met-10 (1,515 bases) to the 3’ end of the NS-coding region (approximately 2,294 bases) indicates the presence of two Trp codons at 1,638 and 1,887 bases (Fig. 5). Initiation at Met-10 would produce four [35S]methionine-labeled, NCS-treated peptides detectable in our gel system with molecular weights of 27,700, 23,300, 14,300, and 13,700. Conversely, initiation at Met-11 would produce three peptides with molecular weights of 23,800, 23,300, and 14,300. Assuming that the 23.8K and 23.3K peptides were not resolved and represented the uncleaved 28K NS polypeptide in Fig. 8, the NCS cleavage pattern was only consistent with internal translation initiation at Met-11.

**DISCUSSION**

We report here the successful isolation of the IPNV polyprotein, showing that IPNV utilizes polypeptide production and processing as a means of gene expression. The use of cDNA clones to produce mRNA in vitro with T7 polymerase and the translation of this RNA allowed the isolation and characterization of the IPNV polyprotein.

We also present the first direct evidence that the NS region of the polyprotein has proteolytic activity and is responsible for the cleavage between pVP2 and NS. The active site of the protease is most likely located at the carboxy terminal, since the deletion of the carboxy-terminal 32 to 35 amino acids of NS (clone pG2.21) resulted in incomplete cleavage of the pVP2-NS truncated polyprotein.

We have no direct evidence showing that the cleavage between NS and VP3 is also mediated by the NS protease. If, however, NS is autocatalytic, it might recognize the same dipeptide at the pVP2-NS and NS-VP3 cleavage site. The only dipeptide occurring once at either side of the NS sequence is a Tyr-Leu dipeptide (Fig. 5). Further work should enable us to ascertain whether it constitutes the cleavage site.

The polypeptide and the other high-molecular-weight in vitro products (Fig. 2) were not detected in translations of genomic RNA. These products should be produced by genomic RNA in vitro, since genomic RNA directs the production of pVP2, 49K, and 37K, which are cleavage products of these larger precursors. This may reflect the presence of a slightly altered NS with reduced protease activity produced by the T7 polymerase transcripts. The cDNA sequence, obtained from two independently isolated clones, revealed two base discrepancies between the two clones at positions 1816 and 2190. A- to G- transitions were discovered which result in amino acid changes in the NS-coding region (Glx-Gly and Gly-Arg). The genomic RNA has not been sequenced over these regions, and we do not know which amino acid reflects the correct sequence of the polypeptide in vivo. The clones used in this study were obtained from the clone containing a glutamic acid codon at 1,816 bases and a glycine codon at 2,190 bases. If either of these codons did not represent the correct sequence of the genomic RNA, then the T7-derived transcripts could conceivably produce an NS with reduced proteolytic activity.

We have obtained in localization the boundaries of the NS- and VP3-coding regions. The VP3-coding region presumably ends at the termination codon of the polypeptide ORF at 3,035 bases. We have determined that ribosomes preferentially use Met-16 or Met-17 to initiate VP3 translation in vitro. The sequences flanking Met-17 show sequence conservation with the sequences flanking Met-1 (Fig. 6), the
start site of the polyprotein. In particular, 5 to 10 bases upstream there is a conserved 5'-ATC-3' triplet, the same sequence found preceding the vesicular stomatitis virus ribosome-binding sites (15). Approximately 10 to 20 bases upstream there is a conserved 5'-AACAA-3' sequence. At about the same distance upstream of the vesicular stomatitis virus ribosome-binding sites for the N and P genes, there is a conserved 5'-AACAG-3' sequence (15). Sargan et al. (30) determined a consensus eucaryotic translation initiation site sequence based on an analysis of 162 mRNAs (ATTCAC C(X)8ATG). The Kozak consensus sequence (CCAC CATG) is actually a subset of this sequence. We were interested to note the presence of the ATC triplet in this consensus sequence, which also precedes the vesicular stomatitis virus N- and P-protein translation initiation sites (15) and is found in the sequence preceding Met-1 and Met-17 of IPNV (Fig. 6). Due to the sequence surrounding Met-17, we believe that Met-17 is the most likely candidate for the VP3 internal translation initiation site.

Similarly, the data indicate that Met-11 is the only methionine codon which would produce an NS, terminating within three or four amino acids of the region encoded by the Ps I site, at 2,276 bases, with a predicted NCS cleavage pattern consistent with that observed for both in vivo- and in vitro-derived NS.

We found that the production of NS from either T7 transcripts or genomic RNA was highly variable, whereas VP3 was always the dominant product produced. This suggests that NS production may be more sensitive to environmental conditions. A similar situation has been reported to occur during in vitro translation of the Sendai virus P/C mRNA (8), where three alternate translation initiation sites on this mRNA were shown to function in vivo and in vitro at different levels. These workers reported that the ratios of some of the products synthesized in vitro varied with the source of the reticulocyte lysate, and they proposed that the salt concentration and possibly RNA secondary structure might be involved in ribosome initiation. A similar situation might explain the variable production of NS in vitro.

The production of several high-molecular-weight polyproteins from segment A indicated that ribosomes were initiating at multiple sites on the IPNV RNA (Fig. 2). None of these high-molecular-weight products has yet been detected in vivo, and their biological significance remains unclear. Similarly, it is presently unclear whether the internal translation initiation sites we have mapped in this study, which are used for NS and VP3 production in vitro, also function in vivo. This is an important question, since it has profound implications for both IPNV gene expression and eucaryotic translation in general.

There are several reports in the literature of other viruses that appear to initiate translation at methionine codons located a considerable distance from the 5' end of the RNA. The vesicular stomatitis virus phosphoprotein (P protein) mRNA also produces a 7K protein from the 3'-proximal region of the RNA by internal translation initiation at a downstream, in-phase methionine codon (15). This site also appears to function in vivo. The adenovirus type 2 polymerase gene encodes the 140K DNA polymerase (AdPol) and a 62K polypeptide produced by internal translation initiation in vitro at the 11th or 12th in-phase methionine codon (1,638 or 1,655 bases). Evidence suggests this site may function in vivo as well (14).

A situation very similar to our results reported here involves the production in vitro of two polyproteins (105K and 95K) from cowpea mosaic virus genome segment M RNA by translation initiation at two sites (34). Recent evidence obtained with antisemum raised against synthetic peptides present near the amino terminus of both polyproteins indicates that the downstream initiation site also functions in vivo (34). We are currently using monoclonal antibodies in an attempt to determine whether the alternate initiation sites we have detected in vitro also function in vivo.

Dorner et al. (12) and Phillips and Emmert (28) demonstrated that the poliovirus P2-3b and P3-1b regions could be translated independently of the polyprotein in rabbit reticulocyte lysates through the use of internal ribosome-binding sites. It was proposed that certain cell factors necessary for the efficient utilization of the polyprotein initiation codon may be present at suboptimal levels in reticulocyte lysates such that alternate initiation sites may be utilized.

Our results indicate that IPNV genome segment A may be similar to poliovirus RNA in that certain factors are present at suboptimal levels in reticulocyte lysates and that ribosome initiation may occur at multiple internal sites. This may explain the appearance of the 95K, 89K, 75K, and 67K polypeptides in vitro translations.

ACKNOWLEDGMENTS

This work supported by the Natural Sciences and Engineering Research Council of Canada.

We thank P. Hudson for the amino acid sequence comparisons of infectious bursal disease virus and IPNV and for making data available before publication.

LITERATURE CITED


