Porcine Parovirus: DNA Sequence and Genome Organization

By ANA I. RANZ, JUAN J. MANCLÚS, ESMERALDA DÍAZ-AROCA AND JOSÉ I. CASAL*

Inmunología y Genetica Aplicada, S.A. (Ingenasa), Hermanos García Noblejas, 41-2º 28037 Madrid, Spain

(Accepted 15 June 1989)

SUMMARY

We have determined the nucleotide sequence of an almost full-length clone of porcine parovirus (PPV). The sequence is 4973 nucleotides (nt) long. The 3' end of virion DNA shows a Y-shaped configuration homologous to rodent paroviruses. The 5' end of virion DNA shows a repetition of 127 nt at the carboxy terminus of the capsid proteins. The overall organization of the PPV genome is similar to those of other autonomous paroviruses. There are two large open reading frames (ORFs) that almost entirely cover the genome, both located in the same frame of the complementary strand. The left ORF encodes the non-structural protein NS1 and the right ORF encodes the capsid proteins (VP1, VP2 and VP3). Promoter analysis, location of splicing sites and putative amino acid sequences for the viral proteins show a high homology of PPV with feline panleukopenia virus and canine paroviruses (FPV and CPV) and rodent parovirus. Therefore we conclude that PPV is related to the Kilham rat virus (KRV) group of autonomous paroviruses formed by KRV, minute virus of mice, Lu III, H-1, FPV and CPV.

INTRODUCTION

Porcine parovirus (PPV) is a major cause of reproductive failure in swine, resulting in foetal death and mummification, still births, and delayed return to oestrus (Joo & Johnson, 1976; Mengeling, 1978). PPV is an autonomous replicating parovirus, containing a ssDNA molecule of approximately 5000 nucleotides (nt) (Molitor et al., 1984); only the minus (genomic) strand is packaged into virions. Four virus-specific proteins have been described: three capsid proteins (A or VP1, B or VP2 and C or VP3, of Mr values 83000, 64000 and 60000, respectively) and one non-structural protein (NS1; Mr 84000) (Molitor et al., 1983, 1985).

DNA sequences of AAV2 [serotype 2 of adeno-associated virus (AAV)] (Srivastava et al., 1983), minute virus of mice (MVM) (Astell et al., 1986), H-1 (Rhode & Paradiso, 1983), canine parovirus (CPV) (Reed et al., 1988), feline panleukopenia virus (FPV) (Carlson et al., 1985), bovine parovirus (BPV) (Chen et al., 1986) and B19 (human parovirus B19) (Shade et al., 1986) have been reported. These studies indicate that autonomous paroviruses show several common features. There are two large open reading frames (ORFs), the mRNAs from both ORFs are polyadenylated and 3'-coterminal at about map unit (m.u.) 95, the left ORF encodes non-capsid proteins which are necessary for viral DNA replication and the right ORF encodes the major capsid proteins of the virus as a nested set.

PPV is a difficult virus to propagate in vitro due to the high tendency of this virus to produce defective interfering particles (Choi et al., 1987), particularly when used at a high m.o.i. or when it is highly passaged. These defective particles contain random genome deletions or duplications. To learn more about the genetic strategy and organization of PPV and its relationship to other paroviruses we decided to clone and sequence it.

In this paper we report the nucleotide sequence of PPV. We have used the NADL-2 strain (Mengeling & Cutlip, 1976), a tissue culture-adapted strain, which shows the presence of two different replicative forms (RFs) of DNA (Molitor et al., 1984). One of the RFs, NADL-2, is...
infective and apparently identical to the RF from the NADL-8 strain (a highly pathogenic isolate). The other RF, NADL-2*, is not infective and shows a deletion of 300 nt near the 5' terminus. We have cloned and sequenced the NADL-2 RF. The sequence shows major homologies with FPV and rodent parvovirus which would be useful in the construction of molecular probes for clinical detection, design of vaccines and in establishing the evolutionary relationship between parvoviruses. This is the first report concerned with the entire PPV genome.

METHODS

Materials. Restriction and DNA modification enzymes were purchased from Boehringer Mannheim or New England Biolabs, and deoxynucleotides and dideoxynucleotides were purchased from Pharmacia. [α-35S]dATP was obtained from New England Nuclear. Chemical reagents for oligonucleotide synthesis were purchased from Applied Biosystems and other chemical reagents were obtained from Sigma. For cloning and sequencing the plasmid DNAs used were pUC18 and the phages M13mp18 and mp19 (Messing, 1983).

Virus and cells. The swine testis (ST) cell line was a gift from Dr L. Enjuanes (Centro de Biologia Molecular, Madrid, Spain). The cells were grown in Dulbecco’s modified Eagle’s medium supplemented with 10% foetal calf serum and antibiotics. All virus passages were done on ST (Pirtle, 1984). Viral strains used were NADL-2 (ATCC VR-742) a tissue culture-adapted virus (Mengeling & Cutlip, 1976), and NADL-8 (a gift from Dr W. Mengeling, National Animal Disease Center, Ames, Ia., USA), a highly pathogenic strain of PPV. To minimize the presence of defective particles we used a low m.o.i. and a low passaged virus as the starting material for all the cloning experiments.

DNA isolation. RF DNA was extracted from infected cells at 48 h post-infection by trypsinizing the cells and then centrifuging. The method was basically similar to that described by Molitor et al. (1984). Cells were lysed with 4 volumes of lysis buffer (0-75% SDS, 20 mM-Tris-HCl pH 7.5, 10 mM-EDTA and 250 μg/ml proteinase K) for 2 h at 37 °C, followed by overnight incubation on ice. Chromosomal DNA and particulate cell debris were eliminated by centrifugation (Kontron, TST 55.5, 30000 r.p.m., 90 min, 4 °C). Supernatants were precipitated with isopropanol, resuspended in Tris-EDTA (TE) buffer (10:1), treated with RNase A and phenol-extracted. Further purification was achieved by electrophoresis of the RF DNA in 1% low melting point (LMP) agarose (FMC) gels in Tris-borate-EDTA (TBE) buffer. DNA was extracted from LMP agarose gel slices as previously described (Langridge et al., 1980).

DNA cloning and transformation. The virus genome orientation is given according to the convention of Armentrout et al. (1978), with the 3' end of the minus-strand DNA to the left. The cloning strategy for the PPV genome will be described elsewhere (J. J. Manclis et al., unpublished). Briefly, purified RF DNA was digested with appropriate restriction enzymes (PstI and EcoRI) and the individual bands were separated on 1% LMP agarose gels in TBE buffer and eluted as described (Langridge et al., 1980). The fragments containing the 3' and 5' ends were treated respectively with mung bean nuclease and the Klenow fragment to leave blunt ends (Maniatis et al., 1982). These DNA fragments were ligated with appropriately restricted pUC18 DNA and transformed into Escherichia coli JM109 cultures as described by Hanahan (1983). All transformation mixtures were plated onto LB plates containing 100 μg of ampicillin per ml and X-gal as the indicator. Cloned fragments were then assembled in a clone called pPPV-10 (Fig. 1), which contained a complete copy of the viral genome since it was able to infect ST cells by transfection (J. J. Manclis et al., unpublished results).

Sequential PPV DNA. The strategy for sequencing PPV DNA was to clone five large fragments, by digestion of pPPV-10, into M13mp18 or M13mp19 double-stranded RF DNA by using combinations of appropriate restriction enzymes (Fig. 1). Ligated DNA was transformed into JM109 cells, and white plaques were selected for single-stranded template preparation. The template DNAs were then sequenced by using the dideoxynucleotide chain termination method (Sanger et al., 1977) and PPV sequence-specific oligonucleotides as primers. Oligonucleotides were synthesized on a 381A apparatus (Applied Biosystems). The oligonucleotides were made consecutively at 250 bp intervals, once the sequence had been read through (Fig. 1). This process considerably speeds the sequencing. The sequence of each fragment was determined on a minimum of two separate gels and two different templates.

Computer program and analysis. The DNA sequence analysis was performed using the Beckman Microgenie Software (Queen & Korn, 1984) and the Sequence Analysis Package (Stephens, 1985) on an IBM PC AT computer.

RESULTS

Nucleotide sequence of the 5' end

The cloning of the 5' end was done by filling in PPV RF DNA with the Klenow fragment and then digesting with EcoRI. The fragments generated, containing the blunt 5' end of RF DNA and an internal cohesive restriction enzyme site, were directionally cloned into pUC18, and then
Porcine parvovirus nucleotide sequence

Fig. 1. Sequencing strategy and restriction map of PPV. The PPV genome was cloned in pUC18 to provide a clone called pPPV-10. Five fragments of pPPV-10 covering the whole genome of PPV were cloned in M13mp18 or -mp19 for sequencing.

(a)

(b)

Fig. 2. Terminal structures of PPV DNA. (a) 3' End nucleotide sequence (minus strand); (b) 5' end nucleotide sequence (minus strand).

into M13mp18 and mp19 (Fig. 1). DNA sequence data and comparison with available data from rodent parvoviruses (Astell et al., 1986) indicate the possible absence of some nucleotides (70 to 80) in the palindrome region with a U-shaped configuration which forms the 5' end. Fig. 2(b) shows a significant homology of PPV with the MVM 5'-terminal sequence (Astell et al., 1986) (23 of 40 nt) in the sequenced side of the loop. Since the sequenced 5' end is derived from a clone, pPPV-10, it shows only the flip orientation. A 127 bp direct repeat begins at nt 4519 (Fig. 3), five bases before the stop codon of the VP1 and VP2 coding sequences, and finishes at nt 4772. This perfect repeat in the 5' end is a feature present in CPV (Reed et al., 1988) and also in the rodent
paroviruses (Astell et al., 1986). Inside the 127 bp repetition there are also small 24 bp repetitions (nt 4538 to 4562 and nt 4612 to 4635), although the matching is not as perfect (20 of 24 nt).

**Nucleotide sequence of the 3' end of the minus strand**

The 3' end of the genome was sequenced from clone pPPV-3 in M13mp18 (Fig. 1), which contains a 270 bp insert between the 3' end and the PstI site. This region of the DNA contains a highly secondary structure making resolution difficult. This sequence is slightly shorter (102 nt compared to 120) but very much resembles that of CPV and rodent parvoviruses (Hauswirth, 1984; Reed et al., 1988) forming an Y-shaped configuration (Fig. 2a). The total length of the 3' end was 102 nt with 82 nt making up the stem structure and 20 nt formed the Y structure (Fig. 2a). Several regions of the Y structure are well conserved in rodent parvoviruses (Astell et al., 1986) and PPV, but not so well in CPV (Reed et al., 1988): nt 2 to 19 of the rodent parvoviruses and nt 2 to 16 of PPV (13 of 16 match) and nt 28 to 44 of the rodent parvoviruses and nt 27 to 40 of PPV (13 of 16 match). Like all the paroviruses sequenced to date, except B19 (Shade et al., 1986), the 3' end is not related in sequence to the 5' end.

**Genome organization and assignment of PPV genes**

Clones PPV-1, PPV-2 and PPV-4 were used to determine the sequence of the major portion of the genome (Fig. 1). PPV DNA-specific oligonucleotides were synthesized to cover the entire genome. The average distance between them was 250 nt. In a few cases it was necessary to reduce the distance between oligonucleotides to get a good sequence. The sequence and single letter amino acid translation of PPV are shown in Fig. 3. The nucleotide sequence of PPV was 4973 bases long, which is the shortest sequence reported to date for autonomous paroviruses. However if we add 70 to 80 nt accounting for the probable loss in the process of cloning of the 5' end we would finish with about 5053 nt, a length similar to that reported for rodent paroviruses (5151 nt) (Astell et al., 1986).

The final DNA sequence of the PPV genome was analysed by the computer programs already mentioned (Queen & Korn, 1984; Stephens, 1985). The potential coding domains for both the complementary strand (C strand, plus polarity) and the viral strand (V strand, minus polarity) are shown in Fig. 4. There are two predominant ORFs (A and B), both occurring in frame 3 of the C strand (Fig. 4). No ORFs of significant size were found in the minus strand of the virus.

A computer search of the C strand for possible promoter regions was done by using the information on eukaryotic promoters described by Bensinhom et al. (1983). These features include an enhancer region (E) about 100 bp upstream of the cap site; a G + C-rich activator region (A) approximately 50 to 75 bp upstream of the cap site and an A + T-rich domain (TATA box), which usually lies about 30 ± 5 bp upstream of the cap site and positions the RNA polymerase for initiation of transcription. Several possible promoters were characterized searching for the consensus sequence TATA\_A. Three possible promoters were localized, one at m.u. 3-7 (P3-7) (TATAAA, nt 183), a second at m.u. 38 (P38) (TATAT, nt 1923) and a third at m.u. 46 (P46) (TATAAA, nt 2329) which contained all the promoter components. Promoters P3-7 and P38 are both analogous in map position to other parovirus promoters (Astell et al., 1986; Reed et al., 1988) and probably initiate, respectively, transcription of ORFs A and B. Unfortunately, no RNA mapping data are available for PPV and we therefore cannot correlate our results with transcription data.

A characteristic signal at the 3' end of eukaryotic mRNAs is the sequence AAUAAA about 20 nt before the poly(A) tract (Wickens & Stephenson, 1984), an essential but not sufficient signal for polyadenylation. Several additional signals have been suggested, such as downstream G/T clusters (Birnstiel et al., 1985) or a CAYUG sequence upstream or downstream from the poly(A) site (Berget, 1984). A search of the C strand of PPV shows 16 possible polyadenylation signals. Most of them (nine of 16) are located in the 5' end, downstream from the ORF B stop codon at nt 4524. Only these sites fulfil all the theoretical requirements and they are probably functional. The others were located within the coding region of ORF A (nt 791, 918, 1041, 1152, 1581, 1991 and 3191) and possibly are not functional as occurs with AAV2 (Srivastava et al., 1983) and FPV (Carlson et al., 1985).
Porcine parovirus nucleotide sequence

Fig. 3. The nucleotide sequence of the cloned PPV genome. The nucleotide sequence corresponds to the complementary (plus) strand (5' to 3'). Below the nucleotide sequence are the one-letter amino acid sequences of putative polypeptides corresponding to the major ORFs. Proposed mRNA 5' donor splice junctions are overlined and the 3' acceptor splice junctions are underlined. Polyadenylation signal regions are indicated as A+. The sequence is complete except for approx. 80 nt at the right end.

Fig. 4. Genomic organization of the complementary (C) (plus-polarity) and viral (V) (minus-polarity) strands. Each line designates the stop codon position of each frame in both strands.
Table 1. Location of 5' donor and 3' acceptor splicing sites*

<table>
<thead>
<tr>
<th>5' Donors</th>
<th>Location (nt)</th>
<th>3' Acceptors</th>
<th>Location (nt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGGTATAT</td>
<td>580</td>
<td>GTTCAGGT</td>
<td>1651</td>
</tr>
<tr>
<td>CAAGTGACT</td>
<td>607</td>
<td>TGCCAGGT</td>
<td>1831</td>
</tr>
<tr>
<td>ATGGTGAGT</td>
<td>817</td>
<td>TACCAGGT</td>
<td>2297</td>
</tr>
<tr>
<td>AAGGTAAAG</td>
<td>1681</td>
<td>CTACAGGT</td>
<td>2911</td>
</tr>
<tr>
<td>CAGGTGATT</td>
<td>1834</td>
<td>CCACAGGA</td>
<td>3574</td>
</tr>
<tr>
<td>AAGGTAAGG</td>
<td>2260</td>
<td>GCTCAGGT</td>
<td>3783</td>
</tr>
<tr>
<td>GAGTTAAGG</td>
<td>2293</td>
<td>CACACAGGC</td>
<td>4020</td>
</tr>
<tr>
<td>GAGGTAAGA</td>
<td>2658</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAGGTAGTT</td>
<td>2914</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATGGGAGGT</td>
<td>3134</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAGGTAAGG</td>
<td>3786</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAAGGTAAT</td>
<td>3982</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The consensus sequences used for the 5' donor was AAGGTAAGT and for the 3' acceptor (Py)₆ XCAGG₆ (Mount, 1982). The computer searched for a maximum of two mismatches in nine nucleotides. The regions underlined are homologous to the consensus sequences.

By analogy with other parvoviruses (MVM, H-1, CPV, FPV), amino acid homologies observed for the putative proteins, and also by expression studies of chimeric gene fusions containing PPV DNA fragments in a prokaryotic host (Halling & Smith, 1985) we deduce that the left ORF of PPV codes for the non-structural protein NS1 and the right ORF codes for the capsid proteins. NS1 is encoded by the left half of the genome in AAV, H-1, MVM, BPV and CPV and this multifunctional polypeptide seems necessary for viral DNA replication (Cotmore & Tattersall, 1987). NS1 has been detected as an immunoprecipitation product of PPV-infected cell lysates, with an M₉ of 84000 (Molitor et al., 1985). The capsid proteins are always encoded by the right ORF as a nested set. Usually there are two encoded capsid proteins. In the case of PPV we have three capsid proteins VP1, VP2 and VP3 of M₉ 83000, 64000 and 60000, respectively (Molitor et al., 1983). VP3 comes from proteolytic cleavage of VP2 and is the most abundant protein in viral capsids.

There is an ATG codon located at nt 279, which could serve as an initiator for the NS1 protein. This codon conforms to the Kozak consensus sequence (AnnATGG) for initiation (Kozak, 1983) and is the first in-frame ATG in the right ORF, downstream from the TATA sequence. Therefore we assign this codon as the initiation site for NS1. There is also strong homology with the initial sequence of amino acids for the non-structural proteins of rodent paraviruses (Fig. 5). However the polypeptide encoded by the left ORF would be of M₉ 75307.

To explain an M₉ of 84000 for NS1, calculated by gel electrophoresis, we should accept that the observed difference is due to the extent of phosphorylation of the NS1 (Jongeneel et al., 1986). In fact, Molitor et al. (1985) have described the presence of forms with slower electrophoretic mobilities for the PPV NS1 which are probably due to a different level of phosphorylation.

The initiation codon for VP1 also conforms to Kozak's rule and is found at nt 2268. This codon is the first ATG in-frame after the P38 and is soon followed by the in-frame termination codon at 2331. The exact splicing pattern of the mRNA for VP1 is not certain; a computer search for the possible splice donor (consensus sequence AAGGT₆AGT) and acceptor [consensus sequence (Py)₆XCAGG₆] sites (Mount, 1982) for VP1 and VP2 revealed several sites, which are shown in Table 1. One donor site lies immediately upstream (nt 2260) of the proposed ATG start codon for VP1 (nt 2268); another donor lies a few bases downstream (nt 2293). A total of seven possible acceptor sites had greater than 65% homology with the consensus sequence and retained the core CAGG region (Table 1). By homology with other paraviruses the most probable acceptor site used by capsid proteins' mRNAs should be located at nt 2374 (CTCTACCAGGT). The first donor mentioned could serve as a splicing site for the transcription of the VP2-specific mRNA removing the ATG of the VP1 encoding mRNA and
Porcine parvovirus nucleotide sequence

PPV: AGNYSEELVATNLWDDQKEAFSWFTQKQLNLGKMRWNNYNKDDTAINL
CPV: XSNCVTEVMEGVKLXKEAESFVFCNGQVNLGKDRWNNYNKTPILLSL
MVM: AGNYSEELVATNLWDDQKEAFSWFTQKQLNLGKMRWNNYNKDDTAINL

PPV: ORQAETDIDCATNWESESLSLLTQVVFDFSLVLKKCLFEPLQKNSPQDFIQ
CPV: ORQAETDIDCATNWESESLSSLQVVFDFSLVLKKCLFEPLQKNSPQDFIQ
MVM: ORQAETDIDCATNWESESLSSLQVVFDFSLVLKKCLFEPLQKNSPQDFIQ

PPV: HEDKTLKCHVLLGCKQQGKXPLENLLWSRWLLGMDPCLPTVERIKREIA
CPV: HEWGDQWCHVLLGCKQQGKXPLENLLWSRWLLGMDPCLPTVERIKREIA
MVM: HEWGDQWCHVLLGCKQQGKXPLENLLWSRWLLGMDPCLPTVERIKREIA

PPV: ELEWVTLLYKHTQKTKYTMHGFGNMIAYFYKKKIKGERCYCFFDSSGWM
CPV: ELEWVTLLYKHTQKTKYTMHGFGNMIAYFYKKKIKGERCYCFFDSSGWM
MVM: ELEWVTLLYKHTQKTKYTMHGFGNMIAYFYKKKIKGERCYCFFDSSGWM

PPV: TNFLKGERHLVSLYQDPTETTVTQAPAKQOTKKEVSKCGLDLWVK
CPV: TNFLKGERHLVSLYQDPTETTVTQAPAKQOTKKEVSKCGLDLWVK
MVM: TNFLKGERHLVSLYQDPTETTVTQAPAKQOTKKEVSKCGLDLWVK

PPV: TVSDWMMTPDPSYIEMMGSCGGNLKTLNLELTLARTKTAQKILILEAKSAT
CPV: TVSDWMMTPDPSYIEMMGSCGGNLKTLNLELTLARTKTAQKILILEAKSAT
MVM: TVSDWMMTPDPSYIEMMGSCGGNLKTLNLELTLARTKTAQKILILEAKSAT

PPV: NISVTIRKCMFHLNYYKRCQFNMCHVLLGCKQQGKRMMFRHGPASTGSIIAQANL
CPV: NISVTIRKCMFHLNYYKRCQFNMCHVLLGCKQQGKRMMFRHGPASTGSIIAQANL
MVM: NISVTIRKCMFHLNYYKRCQFNMCHVLLGCKQQGKRMMFRHGPASTGSIIAQANL

PPV: GNVGCYNAANVFPPNDCTKNKLIWIEEAGNFGQNFQKAICSGQTIRIDQKGGKSIQI
CPV: GNVGCYNAANVFPPNDCTKNKLIWIEEAGNFGQNFQKAICSGQTIRIDQKGGKSIQI
MVM: GNVGCYNAANVFPPNDCTKNKLIWIEEAGNFGQNFQKAICSGQTIRIDQKGGKSIQI

PPV: EPTPVMTTNEIWIYRIGCEPERHTQQIRDLMLNVLKLPGDFGLLEETWPLICA
CPV: EPTPVMTTNEIWIYRIGCEPERHTQQIRDLMLNVLKLPGDFGLLEETWPLICA
MVM: EPTPVMTTNEIWIYRIGCEPERHTQQIRDLMLNVLKLPGDFGLLEETWPLICA

PPV: WLVKQYCTMSVWHSWPDWENPAVTPVTPTINETDISSWSVADNY
CPV: WLVKQYCTMSVWHSWPDWENPAVTPVTPTINETDISSWSVADNY
MVM: WLVKQYCTMSVWHSWPDWENPAVTPVTPTINETDISSWSVADNY

PPV: RACF QNCAPTINIS
CPV: RACF EQOLEDFRDDIS
MVM: RACF GA EPLKDFEPLNIS

Fig. 5. Homology of the translated left ORFs between parvoviruses PPV, CPV and MVM. Homologous regions are enclosed by boxes.
allowing the next ATG at nt 2787 to be the initial ATG for VP2. The second donor could be used as a splicing site for the VP1-specific mRNA yielding a protein of \( M_r \) 83000, very close to that estimated by gel electrophoresis (Molitor et al., 1983). The predicted \( M_r \) for VP2 would be 64416, similar to that estimated previously.
Porcine parvovirus nucleotide sequence

Sequence homology between PPV and other parvoviruses

There is a striking overall DNA homology between MVM, CPV and PPV (approx. 63%). This homology is maintained, or even improved, at the amino acid sequence level. The amino acid homologies among MVM, CPV and PPV NS1 and VP2 proteins are depicted in Fig. 5 and 6. The left ORF is highly conserved between these parvoviruses with a homology of approx. 70%. Homology with human parvovirus B19 (Shade et al., 1986) and BPV (Chen et al., 1986) is much lower at approx. 20%. Even so, there are certain regions of the PPV DNA sequence within the NS1 coding region that are highly conserved in all the parvoviruses (amino acids 389 to 408 of PPV have 90% or greater homology with all the other parvoviruses) (Fig. 5). The right ORF is not as conserved as the left ORF, showing a homology of approx. 50 to 60% with FPV and rodent parvoviruses and around 20% with other autonomous parvoviruses. VP1-specific sequences are well conserved with respect to MVM and CPV. The amino terminus of VP1, formed by 13 amino acids, share the same sequence in all these viruses. There is also a high homology (73%) between MVM and CPV in the first 120 amino acids of VP1. No significant homologies with BPV or B19 are found in this VP1-specific sequence. Four small blocks of strong amino acid homologies are found in the VP2-specific sequence (a glycine-rich region at the beginning of the protein, TPWS at nt 3105, YNNDLTA at nt 3279 and PIWXK at nt 4161). These four blocks are well conserved in all the autonomous parvoviruses except the glycine-rich region in B19 (Shade et al., 1986).

DISCUSSION

An almost full-length copy of the PPV genome was sequenced from clones made from RF DNA. The copy contained all the transcription signals and coding sequences of PPV. Therefore we can compare our data with those described to date for other parvoviruses.

The 3' end sequence was cloned intact (except for possibly 4 nt) as indicated by the restriction map and DNA sequence analysis and it resembles the Y-stem structure of CPV and rodent parvoviruses (Astell et al., 1986; Reed et al., 1988). The 5' end was not a complete copy; we believe that 70 to 80 nt are lost from one of the sides of the 5' loop, probably due to a cloning artefact. Even so, the absence of these nucleotides seems not to affect the replication ability of this copy of the PPV genome in the plasmid pPPV-i0 as it was able to infect ST cells by transfection and generate mature virions (J. J. Manclls et al., unpublished results). We have observed a perfect 127 bp duplication in the 5' end, which partially covers the carboxy terminus of the VP1–VP2 coding region. This direct repeat is the largest one reported to date in a parvovirus genome doubling in size the direct repeat described in MVM (65 nt) (Astell et al., 1986). It has been suggested that these sequence duplications may be caused by high passage of virus in tissue culture to accommodate the viral genome into the capsid. In the NADL-2 strain of PPV the presence of two forms of RF DNA (differing by approximately 300 bp) has been described (Molitor et al., 1984). However this deletion, which involves the loss of the SacI site and one of the BglII sites, does not coincide in position with the repetition, being situated approximately between nt 3900 and 4200. We did not observe any significant feature in this region to predict a possible deletion signal.

Similar to all the parvoviruses described to date, there are two large ORFs located in the C strand. PPV presents multiple promoter-like sites, but by homology with other parvoviruses (Astell et al., 1986; Reed et al., 1988) we assign two main promoters located at m.u. 3-7 and 38 that initiate the transcription of left and right ORFs respectively. No TATA-like sequences have been observed in the 5' end, therefore differing from FPV and CPV. The viral genome contains 16 potential polyadenylation signals, eight of them inside the 127 bp duplication, a much higher number than any other reported to date for autonomous parvoviruses. It is noteworthy that a similar phenomenon has been described for MVM (Astell et al., 1986). In that case two polyadenylation sites are repeated twice. In PPV, with a double length in the duplication (127 nt rather than 65 nt), there are four polyadenylation sites repeated twice. The functional significance of this cluster of poly(A) signals is not well understood.
A. I. RANZ AND OTHERS

The predicted splice positions are analogous to those described for MVM, H-1 and CPV. Unlike B19, BPV or MVM, these transcripts splice within the same frame and not in different frames. We propose two different donor sites: one at nt 2260, which removes the VP1 ATG used by the mRNA encoding VP2 and the other at nt 2293, which would be used by VP1 mRNA. These two donors would use the same acceptor site at nt 2374.

We have shown that the putative amino acid sequence for the PPV NS1 protein is highly homologous to CPV and the rodent parvoviruses (Fig. 5). Also, PPV retains the GKRN region, common to all the parvoviruses, which may be used as a diagnostic probe for parvovirus identification. This sequence conforms also to the consensus sequence G(X)_{4}GKT/ S(X)_{4}I/L/V, which has been recognized as a feature of purine triphosphate binding sites present in proteins of different organisms, suggesting the presence of ATPase or GTPase activities in the NS1 protein of parvoviruses.

With regard to MVM, H-1, CPV, and FPV, VP1-specific sequences also show a high homology (approx. 73%). The amino terminus of the VP1 protein contains the basic, proline-rich sequence MAPPAKRAKR, which has been implicated in the translocation of the virus to the cell nucleus (Cotmore & Tattersall, 1987). VP2 is not as highly conserved, but also shows a good homology (approx. 50 to 60%) and contains several regions common to all the parvoviruses. The sequence PIW, conserved among all the paroviruses, only shows slight homology as shown below:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AAV</td>
<td>VYLQGP</td>
<td>IWAKI</td>
</tr>
<tr>
<td>BPV</td>
<td>JSYRNP</td>
<td>IWVKE</td>
</tr>
<tr>
<td>MVM</td>
<td>VYPQGQ</td>
<td>IWDKEL</td>
</tr>
<tr>
<td>PPV</td>
<td>VFPNGQ</td>
<td>IWDKEL</td>
</tr>
</tbody>
</table>

Since this region is not present in the defective genome NADL-2*, which has been reported to be non-infective, its role could be important in producing intact viral structural proteins and therefore mature virions.

The glycine-rich region, located 70 nt downstream from the initiation codon of the VP2 protein, has been implicated in the VP2 to VP3 cleavage (Paradiso, 1984). It has been suggested (Rhode, 1985) that a run of glycines distorts the α-helical structure generating a possible site for proteolytic cleavage, generating the smallest capsid protein, VP3. TPW and YNN regions (Chen et al., 1986) are also found in PPV making these regions good candidates to search for universal probes for detecting autonomous parvoviruses. The functional implications of these regions are not known at the present time.

Finally these results are in good agreement with the antigenic relationship among autonomous parvoviruses previously reported (Mengeling et al., 1986) and confirm the inclusion of PPV in the KRV-type group (which also includes parvoviruses MVM, H-1, Lu III, FPV and CPV), making the relationship between PPV and other autonomous parvoviruses such as B19 or BPV very distant.

We thank Ana I. Sebastián for assistance in preparation of the manuscript. This work was supported in part by grants from the Comisión Interministerial para la Ciencia y Tecnología and from the Unión Española de Explosivos Rio Tinto.

REFERENCES


Astell, C. R., Gardiner, E. M. & Tattersall, P. (1986). DNA sequence of the lymphotropic variant of minute virus of mice, MVM (i), and its comparison with the DNA sequence of the fibrotopic prototype strain. Journal of Virology 57, 656-659.


Porcine parvovirus nucleotide sequence


(Received 8 March 1989)