JOURNAL OF VIROLOGY, May 1985, p. 630–633
0022-538X/85/050630-04$02.00/0

Vol. 54, No. 2

Nucleotide Sequence of the Coat Protein Gene of Canine Parvovirus

SOLON L. RHODE III

The Eppley Institute, University of Nebraska Medical Center, Omaha, Nebraska 68105

Received 5 November 1984/Accepted 15 January 1985

The nucleotide sequence of the canine parvovirus (CPV2) from map units 33 to 95 has been determined. This includes the entire coat protein gene and noncoding sequences at the 3′ end of the gene, exclusive of the terminal inverted repeat. The predicted capsid protein structures are discussed and compared with those of the rodent parvoviruses H-1 and MVM.

Canine parvovirus (CPV2) is a member of the autonomously replicating parvoviruses, which is a group of animal viruses with icosahedral protein capsids and a small, nonsegmented, linear, single-stranded DNA genome (1, 3). CPV2 is a pathogen of dogs that is antigenically and biologically related to feline panleukopenia virus (FPV) (9, 15, 19). CPV2 and FPV produce similar diseases in their respective hosts. They are most pathogenic in young animals, in which they produce an enteritis and leukopenia. We have made a preliminary characterization of CPV2, in which we found many similarities between CPV2 and the rodent parvoviruses represented by H-1 (13). H-1 DNA was found to hybridize to CPV2 DNA. In this study, the nucleotide sequence of CPV2 from map units (m.u.) 33 to 95 is presented. This region of the viral genome completely encloses the coat protein gene for capsid proteins VP1 and VP2, and the predicted structures of these genes will be compared with those of two rodent parvoviruses, H-1 and MVM.

The DNA sequence of CPV2 is presented in Fig. 1 and extends from the *Sau*3A site at m.u. 33 to the *Hae*III site at m.u. 95 (12). Figure 2 summarizes the sequencing strategy. Like the rodent parvoviruses H-1 and MVM, there is an open reading frame extending from m.u. 33 to 44, ending at nucleotide 604 of Fig. 1 (2, 12, 18). This region is ca. 85% homologous to H-1 or MVM in its predicted amino acid sequence. It is assumed that these sequences code for the same noncapsid protein, NCVP1, as in the two rodent viruses (6, 18). Embedded within the NCVP1 gene and at the same position as in H-1 and MVM is a TATAAA sequence at nucleotide 318 which is suggestive of a promoter. There is substantial evidence that the 5′ end of the most abundant rodent parvovirus mRNA species that codes for the major capsid protein VP2 maps to this site (7, 16). Unlike the rodent parvoviruses, there is no CCAAT at −87 nucleotides to the TATA in CPV2; instead, the sequence TGAAT is found at this position (nucleotide 231). CPV2 and MVM, but not H-1, contain methionine AUG codons (at 525 in CPV2) in the area from 20 nucleotides 3′ to the TATA box to the end of the NCVP1 reading frame. These are not consensus initiation sequences, and all of them are followed closely by termination codons (10). Thus, they should not be expected to interfere with the initiation on the downstream AUG codons (11). The first ATG in H-1 distal to the putative cap site is at the position of the ATG (nucleotide 613) in CPV2 that is within a highly conserved section of sequence following the NCVP1 gene terminator at 604. This ATG has the most preferred flanking sequences for initiation of translation, AXXAUGG (10). The 10 codons immediately following are highly conserved, with four of five nucleotide changes (compared with H-1) being in redundant positions and the fifth resulting in a conservative switch from lysine to arginine. The CPV2 sequence then diverges considerably from that of H-1 and MVM at nucleotide 651. This region contains terminators for all three reading frames in all of these viruses. Therefore, if we contend that the AUG at 613 codes for the amino terminus of the coat protein VP1, the region immediately after 651 must be an intron. In support of this, the sequence at 642 to 648, AGGTAAG, is a consensus splice donor site, except it has AG dinucleotides preceding it within 20 nucleotides (4). This is also in agreement with the mapping of a splice site to this region in MVM by S1 nuclease analysis (16).

Direct analysis of the H-1 capsid protein VP2 by protein sequencing and amino acid analysis of mapped peptide fragments has placed the amino terminus for VP2 at the ATG at m.u. 54 or nucleotide 1114 in Fig. 1 (14). This ATG is conserved in CPV2, H-1, and MVM, and it is within the large open reading frame which extends from nucleotide 689 to 2867 of Fig. 1. This region was shown to code for both VP1 and VP2 in H-1 (8). Thus, VP2 is embedded completely within the larger VP1, and its molecular weight is predicted to be 65,000 (65K), close to the estimate of 67K by polyacrylamide gel electrophoresis (13). I suggest that VP1 is translated by initiating translation with the AUG codon at 613, splicing the mRNA at 643, and rejoining the large open reading frame at an undetermined point after nucleotide 689. Two possible acceptor sites are found at 713 (AG/GACTT) or 728 (AG/GTTAT). The former gives a protein of 80.7K, close to the gel estimate of 82K.

The mRNA for VP2 has a very long leader sequence, since it begins at ca. nucleotides 340 to 350 (7) and has a small splice at m.u. 45. Either the translation apparatus passes over the AUG at 613 (and those preceding it in CPV2 and MVM) or, more likely, this AUG is removed by splicing with a different splice donor site. The sequence AGGTACG beginning at 606 is close to the consensus sequence for splice donor sites and would very neatly remove the offending AUG. It is also sufficiently close to the splice donor site at 643 that previous studies with the S1 nuclease mapping technique may have missed resolving the different boundaries (7, 16). This arrangement is highly conservative for coding sequences, as the VP1 gene begins only 6 nucleotides 3′ to the end of the NCVP1 gene. In addition, within those 6 nucleotides is a putative splice donor site that may excise the initiation codon for VP1 and so allow VP2 translation to begin at the next available AUG at nucleotide 1114. The predicted introns are then only 70 and 106 nucleotides in length, assuming the acceptor site is at 713. We are conducting experiments to determine whether the divergent se-

FIG. 1. The nucleotide sequence of CPV2 from the *Sau*3A site at m.u. 33 to the *Hae*III site at m.u. 95. The virus used was canine parvovirus strain 780929 obtained from D. L. Carmichael (Cornell University, Ithaca, N.Y.). Virus was propagated, titers were determined, and replicative form DNA was prepared as previously described (13). Fragments of CPV2 replicative form DNA were cloned into M13 vectors and sequenced by the dideoxynucleotide method as previously described (18). The open reading frames for the VP1 and VP2 capsid proteins as discussed in the text are presented in codon format with the assigned amino acid above. The closed arrowheads indicate proposed splice donor sites, and the open arrowhead is a proposed splice acceptor site. The arrow indicates the start of VP2 translation.

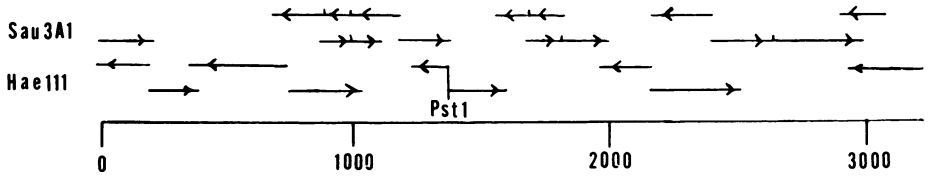

FIG. 2. The sequencing strategy used is diagramed. The bulk of the sequence was determined with *Sau*3A fragments cloned into M13mp8 and *Hae*III fragments cloned with *Eco*RI linkers into M13mp7. The fragments bounded by the indicated *Pst*I site were cloned into M13mp7-01.

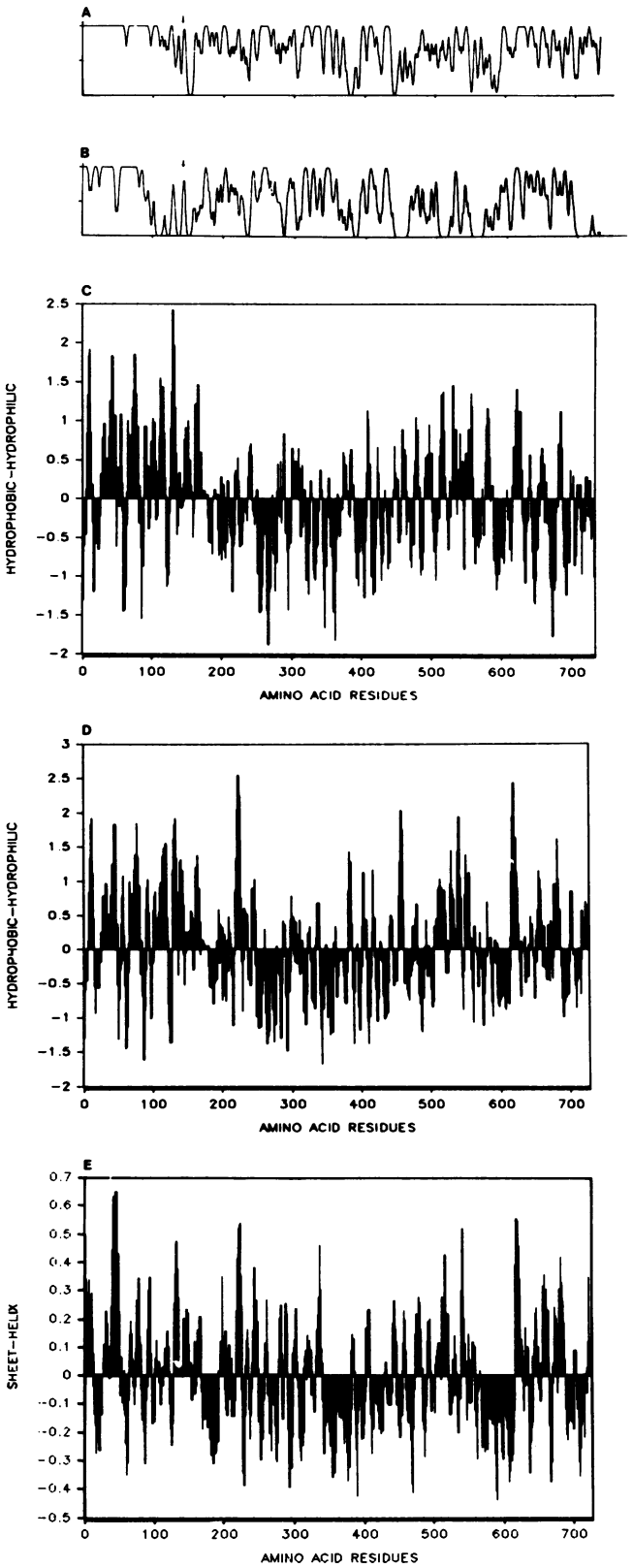

FIG. 3. (A) A comparison of the coat protein primary structure of H-1 with MVM via a function that we devised that generates a value of 1 for complete homology in a window of seven residues and 0 for no matches in the same window. The function assigns a value of 0.3 for the center position in the seven-residue window, and values of 0.2, 0.1, and 0.05 for the flanking residues on both sides. The structure for VP1 is the one assumed to be most likely as discussed in the text. The arrow indicates the start point of VP2. (B) The homology of H-1 VP1 to CPV2 VP1 as in (A). (C) The hydrophilicity of the H-1 VP1 calculated by the method of Hopp and Woods (8). Hydrophilic values are positive and above the zero line. (D) The hydrophilicity of the CPV2 VP1. (E) The predicted conformation of CPV2 VP1 calculated by the method of Chou and Fasman (5), with the values averaged for a moving window of six residues.

quences within this intron region play a regulatory role in the transcription of the promoter at m.u. 38.

The primary amino acid structure of VP2 was predicted by the nucleotide sequence, and the complete structure of VP1 will also be known once the exact splicing arrangements have been determined (Fig. 1). We compared the structures of VP1 (VP2) of H-1 and CPV and H-1 and MVM for the predicted structures of VP1 by using the splice sites discussed above (Fig. 3). The upper two panels plot a function that measures homology of amino acid sequence in a moving window of seven residues, with a value of 1 for a complete match and a value of 0 for no matches. Figure 3A compares H-1 with MVM and Fig. 3B compares H-1 with CPV2. It is readily apparent that the amino-terminal portions of the reading frame are highly homologous. The VP2 protein begins at residue 138 for H-1 in Fig. 3 (at the arrow), a region that is homologous in both MVM and CPV2. H-1 is more homologous to MVM than CPV2, as expected, and the comparison of these two patterns shows that the areas of high divergence tend to occur in the same positions. Figure 3C and D represents functions which describe the hydrophilicity of local areas of the H-1 or CPV2 capsid proteins, and Fig. 3E is a graphic of the predicted sheet or alpha helix structures. In general, the more highly conserved domains are the hydrophobic ones, and the hydrophilic domains tend to be poorly conserved. All three of these viruses have a highly conserved glycine-rich region at residues 165 to 182 in which 13 of 18 amino acids are glycine. Glycine is a strong breaker of the alpha helix (5), so this region is expected to be a random coil. Following this section, the middle portions of the polypeptide are rich in hydrophobic amino acids, and the Hopp and Woods plot reflects this, with the function being largely negative from ca. residue 180 to 400. It will be of interest to determine which domains are on the external surface of the virion and which define the major antigenic epitopes of the parvovirus capsids. The trypsin-sensitive site of H-1 full capsids has been mapped to the two arginine residues corresponding to arginine residues 154 and 157 in CPV2 (14). Since this region is conserved, it is likely that this is a surface domain in CPV2 as well.

The sequences determined here do not include the inverted repeat at the viral 5′ terminus. They do include the bulk of the 5′ noncoding region. A possible polyadenylate signal, AATAAA, occurs at nucleotide 3193, and this is the homologous position of a similar signal in H-1 and MVM. As in the parvovirus H-1, there is a tandem repeat in the

```
CPV: TAACATACTTACTATGTTTTTATGTTTA-TT-AGATG--TTTTAAGATTAATTAAA
     ****  **  ********* * *****  ** * **     ****   *** ** *
H-1: TAACCAACCAACTATGTTTCTCTGTTTGCTTCACATAATACTTAAAC-TAACTAGA

CPV: TTACAGCATAGATTTAGT-ACTTGTATTTGATATAGGATTTAGAAGGTTTGTTATA
      **** **** *  **   ****      * *** *  *** * **       * ***
H-1: CTACAACATAAAAATATACACTT---AATAATAGATTATTAAAAATAAC-ATAATA

CPV: TGGTATACAATAACTGTAAGAAATAGAAGAACATTTAGATCATAGTTAGT-AGTTT
     *****     ********** *****  ***** *** *** *    **** ****
H-1: TGGTAGGT--TAACTGTAAAAAATAATAGAACTTTTGGATAAATA-TAGTTAGTTG

CPV: TGTTTTATAAAATGTA-TTGTAAAACTATTAATGTATGTTGTTATGGTGTGGGTGG
        *********  *   *****          *           **** ********
H-1: G--TTTATAAAAAGATTTTGTA---------TTTGGG-------TGGT-TGGGTGG

CPV: TTGGTTGGTTTGCCCTTAGAAT--ATGTTAAGGACCAAAAAAATCAATAAAAGACA
     *********   ******** *  ****** ************* ******* *
H-1: TTGGTTGGTACTCCCTTAGACTGAATGTTAGGGACCAAAAAAAT-AATAAAATAAT

CPV: TTTAAAACTAAATGGCC
     *  **    * * ** *
H-1: TAAAAT-GAACAAGGAC
```

FIG. 4. A comparison of the noncoding sequences following the coat protein genes of CPV2 and H-1. The sequences are listed from the TAA terminator codon to the GGCC *Hae*III site of CPV2, which is just inboard of the 5'-terminal inverted repeat. Each virus has had one copy of its internal repeat removed for simplicity. *, Matching nucleotides.

noncoding 5' end of CPV2 that is 60 base pairs long. Unlike H-1, this repeat begins in the coat protein exon at nucleotide 2844. Since the parvoviruses H-3 (17) and MVM (2) do not have such a tandem repeat, it is clearly not essential for replication. They may serve a passive function of adjusting the DNA length to an optimal size for packaging in their respective capsids. If we compare the noncoding sequences of H-1 and CPV2 with only one copy of the repeated sequences of each, we find regions that are highly homologous, suggesting important regulatory functions for this area (Fig. 4). The overall homology for this region is 62%. The origin of replication maps to the right-end noncoding region, but the mechanism of initiation is not known (17).

I thank Russel Doolittle and Patrick Iverson for computer analyses of protein primary structures and Jeanne Helft for technical assistance.

This work was supported in part by Public Health Service grant CA-26801 from the National Institutes of Health and grant DMB-8444778 from the National Science Foundation.

LITERATURE CITED

1. **Appel, J. J. G., F. W. Scott, and L. E. Carmichael.** 1979. Isolation and immunization studies of a canine parvo-like virus from dogs with haemorrhagic enteritis. Vet. Rec. **105**:156–159.
2. **Astell, C. R., M. Thomson, M. Merchlinsky, and D. C. Ward.** 1983. The complete DNA sequence of minute virus of mice, an autonomous parvovirus. Nucleic Acids Res. **11**:999–1018.
3. **Carmichael, L. E., J. C. Joubert, and R. V. Pollock.** 1980. Heamagglutination by canine parvovirus: serologic studies and diagnostic applications. Am. J. Vet. Res. **40**:784–791.
4. **Cech, T. R.** 1983. RNA splicing, three themes with variations. Cell **34**:713–716.
5. **Chou, P. Y., and G. D. Fasman.** 1977. Prediction of the secondary structure of proteins from their amino acid sequence. J. Mol. Biol. **115**:135–175.
6. **Cotmore, S. F., L. J. Sturzenbecker, and P. Tattersall.** 1983. The autonomous parvovirus MVM encodes two nonstructural proteins in addition to its capsid polypeptides. Virology **129**:333–343.
7. **Green, M. R., R. M. Lebovitz, and R. G. Roeder.** Expression of the autonomous parvovirus H1 genome: evidence for a single transcriptional unit and multiple spliced polyadenylated transcripts. Cell **17**:967–977.
8. **Hopp, T. P., and V. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:3824–3828.
9. **Johnson, R. J., and P. B. Spradbrow.** 1979. Isolation from dogs with severe enteritis of a parvovirus related to feline panleukopenia virus. Aust. Vet. J. **55**:151.
10. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
11. **Liu, C.-C., C. C. Simonsen, and A. D. Levinson.** 1984. Initiation of translation at internal AUG codons in mammalian cells. Nature (London) **309**:82–85.
12. **McMaster, G. K., J.-D. Tratschin, and G. Siegl.** 1981. Comparison of canine parvovirus with mink enteritis virus by restriction site mapping. J. Virol. **38**:368–371.
13. **Paradiso, P. R., S. L. Rhode III, and I. I. Singer.** 1982. Canine parvovirus: a biochemical and ultrastructural characterization. J. Gen. Virol. **62**:113–125.
14. **Paradiso, P. R., K. R. Williams, and R. L. Costantino.** 1984. Mapping of the amino terminus of the H-1 parvovirus major capsid protein. J. Virol. **52**:77–81.
15. **Parrish, C. R., L. E. Carmichael, and D. F. Antczak.** 1982. Antigenic relationships between canine parvovirus type 2, feline panleukopenia virus and mink enteritis virus using conventional antisera and monoclonal antibodies. Arch. Virol. **72**:267–278.
16. **Pintel, D., D. Dadachanji, C. R. Astell, and D. C. Ward.** 1983. The genome of minute virus of mice, an autonomous parvovirus, encodes two overlapping transcription units. Nucleic Acids Res. **11**:1019–1038.
17. **Rhode, S. L., III, and B. Klaassen.** 1982. DNA sequence of the 5' terminus containing the replication origin of parvovirus replicative form DNA. J. Virol. **41**:990–999.
18. **Rhode, S. L., III, and P. R. Paradiso.** 1983. Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybrid-arrested translation. J. Virol. **45**:173–184.
19. **Tratschin, J.-D. G. K. McMaster, B. Kronauer, and B. Siegl.** 1982. Canine parvovirus: relationship to wild-type and vaccine strains of feline panleukopenia virus and mink enteritis virus. J. Gen. Virol. **61**:33–41.