Mapping of the Amino Terminus of the H-1 Parvovirus Major Capsid Protein

PETER R. PARADISO,¹* KENNETH R. WILLIAMS,² AND ROBERT L. COSTANTINO¹

Institute for Medical Research of Bennington, Bennington, Vermont 05201,¹ and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510²

Received 1 June 1984/Accepted 4 July 1984

The amino terminus of VP2', the major capsid protein of the parvovirus H-1, was identified and mapped to the H-1 genome. The protein initiates at the start codon at nucleotide 2797 and is translated uninterrupted to the stop codon at nucleotide 4582. The primary sequence predicts a protein of 593 amino acids (65,500 daltons) with an amino acid composition which very closely matches the experimentally determined composition of the pure protein. The data suggest that the VP2' mRNA has a 5' leader sequence of ca. 650 bases and that protein translation initiates downstream from the sole splice junction.

The autonomous parvoviruses contain a single-stranded DNA genome with two apparently overlapping genes (1, 10, 11). These viruses make very efficient use of a limited coding capacity and direct the synthesis of at least three (11) and probably four viral proteins (2). Studies of the H-1 parvovirus have demonstrated that a noncapsid protein, NCVP1, is encoded by a large open reading frame in the 3' half of the viral genome (11). The capsid proteins VP1 and VP2' of H-1 are encoded in the 5' half of the viral genome (17). Cotmore et al. (2) have shown that the genome of a related virus, minute virus of mice (MVM), has a similar organization (NCVP1 and NS-1 being analogous proteins from H-1 and MVM, respectively). This group has further identified a second noncapsid protein, NS-2 (24,000 daltons) which is as yet unmapped (2).

The complete nucleotide sequences of both H-1 (11) and MVM (1) are known, and the overall homology in the coded sequences is >70% (1, 11). Two promoter sequences at map positions 4 and 39 have been identified for both genomes. It is probable that the NCVP1 (NS-1) mRNA is transcribed from the promoter at 4 map units (m.u.) and that the protein initiates at the AUG at codon 264 of H-1 and terminates at either nucleotide 2280 or 2303 (11). The coding boundaries for the capsid proteins VP1 and VP2' are less clear. It is known that virtually all of the sequences of VP2' (68,000 daltons) are contained within the sequences of VP1 (88,000 daltons) and that these proteins probably have common carboxy termini (1, 2, 10, 11, 14). Analysis of MVM mRNA has identified three major mRNAs of 4.8, 3.3, and 3.0 kilobases (kb), each of which is missing a short intron between 46 and 48 m.u. The 4.8-kb mRNA initiates at the promoter at 4 m.u. and probably encodes NS-1 (NCVP1). The 3.3-kb mRNA is similar to the 4.8-kb mRNA, except with a large intron between 10 and 40 m.u. removed. The 3.0-kb mRNA, which is the most abundant species, initiates at the promoter at 39 m.u. (10). Although the H-1 mRNAs have not been mapped in such detail, the 4.8-, 3.0-, and 2.8-kb mRNA species identified by Green et al. (3) for H-1 probably correspond to the 4.7-, 3.3-, and 3.0-kb mRNAs of MVM (10). It is not presently known which of these RNA species code for VP1 and VP2' or where either of these proteins initiates.

In the present study we isolated peptides from the amino terminus of VP2' (from H-1), taking advantage of both the ability to proteolytically cleave VP2' to VP2 in vitro and the known nucleotide sequence. VP2' initiates at the AUG at nucleotide 2797, downstream from the intron between 46 and 48 m.u.

MATERIALS AND METHODS

Virus growth and purification. H-1 virus was grown in simian virus 40-transformed human newborn kidney cells and purified by precipitation with polyethylene glycol as described previously (8). The viruses derived from these cells contain >95% VP2' as their major capsid protein, with only trace amounts of VP2 (7). Virus containing >95% VP2 as the major capsid protein was prepared in vitro by incubating the virus in 20 mM NH_4HCO_3 (pH 8.0) with trypsin (2% by weight of the capsid protein) at 37°C for 2 h (7). Such treatment of H-1 virus results only in the removal of 2,000 to 3,000 daltons of protein from VP2', thus forming VP2. VP2 and VP1 in capsids are resistant to cleavage by trypsin (4, 7).

Purification of capsid proteins. Capsid proteins were purified by preparative gel electrophoresis (5). VP2' was derived from empty capsids, and VP2 was derived from trypsin-cleaved virus. Proteins were localized in undenatured gels by staining a guide strip and collected either by electroeluting the proteins into running buffer (5) or by homogenizing the slices in running buffer and incubating at 37°C for 6 to 8 h. The acrylamide was removed by centrifugation.

Alkylation and cyanogen bromide cleavage of H-1 proteins. Whole capsid or purified protein (up to 2 mg) was dissolved in 0.5 ml of 6 M guanidine hydrochloride–2 mM EDTA–0.5 M Tris-hydrochloride (pH 8.1) and flushed with N_2 . After 30 min at 50°C, 0.1 ml of dithiothreitol (0.12 M) was added at 50°C for 4 h. Then 0.2 ml of iodoacetate (0.24 M) was added, and the pH was maintained at 8.1 for 20 min in the dark. The sample was dialyzed extensively against cold water, lyophilized, and taken up in 6 mg of cyanogen bromide per ml in 70% formic acid. The reaction was terminated after 24 h by adding 10 volumes of water and lyophilizing.

Peptide analysis. Peptides were sequenced on a solid-phase sequenator (Sequemat), and the phenylthiohydantoin derivatives were analyzed by high-performance liquid chromatography (HPLC) described by B. M. Merrill, K. R. Williams, T. W. Chase, and W. Konigsberg (J. Biol. Chem., in press). Amino acid analyses were done on a Beckman 121M Analyzer.

* Corresponding author.

† Present address: Praxis Biologics, Rochester, NY 14623.

RESULTS

Analysis of the amino-terminal cyanogen bromide fragment of VP2'. Attempts to sequence the amino terminus of VP2' and VP1 by two manual methods (13, 16) or with a spinning-cup sequencer were unsuccessful, apparently because of blocked amino acids at their respective termini. However, VP2' (68,000 daltons) can be cleaved to VP2 (65,000 daltons) in vitro by incubating virus particles with a variety of proteases (4, 7), and several lines of evidence indicated that this cleavage is at the amino terminus of VP2'. Tattersall et al. (14) initially proposed that for MVM particles the cleavage site was ca. 30 amino acids from the amino terminus. Our own experiments confirmed this prediction for H-1 by comparing VP2' and VP2 which had been cleaved at their cysteine residues after cyanylation (15). When ^{14}C -labeled sodium cyanide is used in conjunction with dithionitrobenzoate for the cyanylation, subsequent cleavage at pH 9.0 results in peptides labeled at their modified amino termini, except for the true amino-terminal peptide, which is unlabeled. When this cleavage was done on VP2' and VP2, VP2 had a unique amino-terminal peptide (identified as a Coomassie blue-positive but nonradioactive band on gel electrophoresis) of 9,000 to 10,000 daltons, which was not present in similarly cleaved VP2' (data not shown). Thus, the cleavage of VP2' to VP2 with trypsin in vitro occurs at the amino terminus of VP2', and by comparing maps of these two proteins, amino-terminal-derived peptides can be identified.

Using this information, we compared peptides from VP2' and VP2 produced by cleavage at methionine residues. Cleavage of VP2' with cyanogen bromide after carboxymethylation consistently resulted in the appearance of five

FIG. 1. Cyanogen bromide cleavage of capsid proteins. H-1 capsid proteins were cleaved with cyanogen bromide as described in the text. Lane 1, Proteins from virus which were not treated with trypsin; lane 2, trypsin-treated virus in which the VP2' was converted to VP2; lane 3, cleaved protein from empty capsids. Lane 3 was identical to lane 1, as expected.

TABLE 1. Amino acid composition of peptide D

Amino acid ^a	Experimental values ^b	Predicted no. of residues ^c
Asx (N, D)	3.19 (8.6)	8
Thr (T)	2.74 (7.4)	8
Ser (S)	2.70 (7.3)	7
Glx (Q, E)	1.99 (5.4)	5
Pro (P)	ND ^d	1
Gly (G)	6.62 (17.9)	18
Ala (A)	1.56 (4.2)	5
Val (V)	0.68 (1.8)	3
Met (M)		1
Ile (I)	1.31 (3.5)	3
Leu (L)	1.42 (3.9)	4
Tyr (Y)	0.64 (1.7)	2
Phe (F)	0.41 (1.1)	1
His (H)	0.78 (2.1)	2
Lys (K)	0.34 (0.9)	1
Arg (R)	1.14 (3.1)	3

^a Shown in parentheses is the one-letter amino acid code used in Fig. 3.

^b Experimental values are in nanomoles per milliliter with the calculated number of residues per peptide in parentheses.

^c Shown is the predicted composition of the peptide coded between nucleotides 2800 and 3018 of the H-1 genome (see Fig. 3).

^d ND, Not determined.

bands from 8,000 to 16,000 daltons (Fig. 1, lane 1, bands A through E). When the cleavage of VP2' was compared with that of VP2, peptide D was absent (lane 2), and peptide F appeared. Thus, peptide D is the amino-terminal cyanogen bromide fragment of VP2', and peptide F is derived from VP2 as a result of trypsin cleavage within peptide D.

Peptide D was subsequently purified by HPLC (data not shown) and subjected to amino acid analysis. Table 1 shows the actual composition of this peptide and a comparison to that predicted from translating the open reading frame coded between nucleotides 2800 and 3018 of H-1 DNA (see Fig. 3). The correspondence of the two places the start of VP2' at 54 m.u., considerably downstream from both the promoter at 39 m.u. and the intron at 46 to 48 m.u. The predicted peptide contains 73 amino acids and has a molecular weight of 7,216, which is slightly smaller than that predicted for peptide D by gel analysis. Although a peptide of this size cannot be matched exactly by amino acid analysis, the rather distinct high glycine content is very diagnostic. Attempts to sequence peptide D either manually (13) or with a solid-phase sequencer (Merrill et al., in press) were unsuccessful, indicating a blocked terminus as expected. Identification and placement of peptide D further confirmed that the cleavage of VP2' was amino terminal.

Analysis of peptides released when VP2' is cleaved to VP2 in vitro. To get closer to the amino terminus of VP2', we analyzed the peptides generated when full virus was cleaved with trypsin. The peptides were separated by HPLC (Fig. 2), and the indicated fractions were subjected to amino acid analysis. Only three peptides were present in significant amounts, and they matched the amino acid composition of the 20-amino-acid sequence predicted between nucleotides 2800 and 2860 (Fig. 3). The matches can be seen in Table 2. Peptide 1 contained only three amino acids, Val, Glu, and Arg, and matches amino acids 18 through 20 in Fig. 3. The composition of peptide 2 matches the sequence of residues 1 through 17 in Fig. 3. Peptide 3 was a partial cleavage product with a composition matching residues 1 through 20. When peptide 2 was coupled and put on the sequencer, no

FIG. 2. Separation of tryptic peptides from H-1 virus by HPLC. H-1 virus was treated with trypsin to convert VP2' to VP2, and the released peptides were purified by HPLC on an Ultrasphere-ODS C18 column (Altex). The sample was loaded onto the column in 0.1% trifluoroacetic acid (buffer A) and eluted with a gradient of acetonitrile in 0.1% trifluoroacetic acid (buffer B). The gradient used was 0 to 25% buffer B from 10 to 60 min, 25 to 50% buffer B from 60 to 160 min, and 50 to 70% buffer B from 160 to 180 min. All of the peaks were analyzed for amino acid content, with only peaks 1, 2, and 3 containing significant yields of amino acids. The column flow rate was 1.0 ml/min with 0.1 absorbance units full scale at 220 nm.

phenylthiohydantoin amino acids were released, thus identifying it as the blocked amino-terminal tryptic peptide. However, when peptide 2 was first cleaved with *Staphylococcus aureus* V8 protease, which cleaves at the carboxy side of glutamic acid residues, the sequence derived was X-Asn-Gln-Pro-Asp-Thr-Gly-Ile-Ala-Asn-Ala-Arg. This confirms the position of peptide 2, and its composition matches that of the predicted peptide from nucleotides 2800 to 2850. Thus VP2' initiates at the AUG codon at nucleotide 2797, and VP2 derived with trypsin in vitro begins at the TCA codon (Ser) at nucleotide 2860.

DISCUSSION

The H-1 parvovirus genome contains two promoters at 4 and 39 m.u. (11) and is highly homologous to MVM, which has the same two promoters (1, 10). In MVM, the most abundant mRNA species (R3) uses the promoter at 39 m.u. and probably encodes the major capsid protein VP2'. In H-1, the analogous transcript is the 2.8-kb mRNA seen by Green et al. (3). Based on RNA mapping (3, 10) and hybrid-arrested translation mapping (11), we will assume that the VP2' mRNA in H-1 infected cells is signaled by the promoter at nucleotide 1979 with the mRNA initiated downstream of this TATAAA signal sequence. This mRNA contains an intron between 46 and 48 m.u. (ca. 100 bases) and terminates at the polyadenylation signal at 95 m.u. (1, 10, 11). The precise boundaries of the intron are not known (3, 10). The data described here on the amino terminus of VP2', along with the previous RNA mapping studies, indicate that VP2' is

coded, uninterrupted, from nucleotide 2800 to the terminator codon at nucleotide 4573.

Figure 3 shows the DNA sequence and the predicted protein sequence of H-1 from the TATAAA signal at nucleotide 1979 (39 m.u.) through the first cyanogen bromide cleavage point of VP2' (nucleotide 3018). If we assume that the first ATG codon at nucleotide 1999 is too close to the promoter signal (12 nucleotides) to be part of the message, there are two other ATG codons before the actual start of VP2' at nucleotide 2797. The first one is at nucleotide 2289 (44.2 m.u.), and the first in-frame terminator is 22 codons away. The second is at nucleotide 2362 (45.6 m.u.), and the first in-frame terminator is 29 codons away. Neither of these ATG codons is used to initiate VP2', and it is possible that both are in the intron of the VP2' mRNA since they are within 2 m.u. of the estimated intron boundaries of 46 and 48 m.u. (10). In this regard it is interesting to note the well-conserved splice donor consensus sequence AAGGTACA at nucleotide 2281 (6). MVM DNA contains analogous ATG codons at nucleotides 2286 (45 m.u.) and 2354 (46.3 m.u.), along with a unique ATG codon at nucleotide 2332 (45.9 m.u.). However, MVM does not have an ATG codon which corresponds to the one at nucleotide 1999 in H-1.

The calculated molecular weight of VP2', based on the predicted sequence of the 593 amino acids, is 65,500. This is very close to the previous estimates of 66,000 and 68,000 daltons based on migration in sodium dodecyl sulfate-containing polyacrylamide gels (4, 9). The VP2 molecule formed in vitro with trypsin is 63,530 daltons (573 amino acids). The predicted amino acid composition of VP2' matches very closely that previously reported for VP2' based on the amino acid analysis of the pure protein (Table 3) (11).

The amino-terminal boundaries of VP1 are not yet known. The genomic organization of H-1 and MVM suggests that VP1 contains an additional 20,000 daltons (ca. 180 amino acids) in addition to the sequences present in VP2'. If VP1 mRNA were to initiate translation at nucleotide 2289, switch reading frames before the first terminator codon via a 100-base intron (10), and then continue in the same open reading frame as VP2' (to the terminator at nucleotide 4573), the translated VP1 would be 80,000 daltons. This is somewhat smaller than the previously predicted 86,000 to 88,000 daltons (4, 9), but not unreasonably so. In fact, our preliminary studies on the cyanogen bromide cleavage of VP1 would tend to support this coding scheme. Cleavage of VP1 yields a peptide that is the same size as peptide D, which is as expected if VP1 reads through the initiator methionine of VP2'. Furthermore, these studies have identified a 20,000-dalton cyanogen bromide peptide which is unique to VP1. An alternate possibility which has been suggested is that VP1 mRNA initiates at the promoter at 4 m.u. and contains a large intron between 10 and 40 m.u. along with the second intron between 46 and 48 m.u. (3, 15). By this scheme it is possible for VP1 to initiate translation near the promoter at 4 m.u. and coterminate with VP2'. However, our hybrid-arrested translation experiments (11) showed that sequences upstream of 9.5 m.u. were not translated for VP1 or VP2'. Another possibility is that H-1 VP1 initiates at nucleotide 1999, but since this start signal is not conserved in either MVM (1) or canine parvovirus (S. L. Rhode III, personal communication), this seems unlikely. Further support for the initiation of translation of VP1 at nucleotide 2289 is the strong splice donor consensus sequence AGGTAAG at nucleotide 2318 (6), before the terminator in this reading frame. We are continuing our investigation of the primary structure of VP1 to distinguish these alternatives.

ACKNOWLEDGMENTS

We thank Mary LoPresti for excellent technical assistance and Solon Rhode III for helpful discussions throughout the course of this work.

This work was supported in part by Public Health Service grant CA-25866 from the National Institutes of Health and grant PCM-8315586 from the National Science Foundation.

LITERATURE CITED

1. 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.
2. 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.
3. Green, M. R., R. M. Lebovitz, and R. G. Roeder. 1979. Expression of the autonomous parvovirus H1 genome: evidence for a single transcriptional unit and multiple spliced polyadenylated transcripts. *Cell* **17**:967-977.
4. Kongsvik, J. R., M. S. Hopkins, and K. A. O. Ellem. 1978. Two populations of infectious virus produced during H-1 infection of synchronized transformed cells, p. 505-520. *In* D. C. Ward and P. Tattersall (ed.), *Replication of mammalian parvoviruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
5. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680-685.
6. MacCumber, M., and R. Ornstein. 1984. Molecular model of mRNA splicing. *Science* **224**:402-405.
7. Paradiso, P. R. 1981. Infectious process of the parvovirus H-1: correlation of protein content, particle density, and viral infectivity. *J. Virol.* **39**:800-807.
8. Paradiso, P. R. 1983. Analysis of the protein-protein interactions in the parvovirus H-1 capsid. *J. Virol.* **46**:94-102.
9. Peterson, J., R. Dale, R. Karess, D. Leonard, and D. Ward. 1978. Comparison of parvovirus structural proteins—evidence for post-translational modification, p. 431-446. *In* D. Ward and P. Tattersall (ed.), *Replication of mammalian parvoviruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. 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.
11. Rhode, S. R., 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.
12. Tal, J., D. Ron, P. Tattersall, S. Bratosin, and Y. Aloni. 1979. About 30% of minute virus of mice RNA is spliced out following polyadenylation. *Nature (London)* **279**:649-651.
13. Tarr, G. 1977. Improved manual sequencing methods. *Methods Enzymol.* **47**:335-357.
14. Tattersall, P., A. J. Shatkin, and D. C. Ward. 1977. Sequence homology between the structural polypeptides coded for by minute virus of mice. *J. Mol. Biol.* **111**:375-394.
15. Vanaman, T. C., and G. R. Stark. 1970. A study of the sulfhydryl groups of the catalytic subunit of *Escherichia coli* aspartate transcarbamylase. *J. Biol. Chem.* **245**:3565-3573.
16. Weiner, A., T. Platt, and K. Weber. 1972. Amino terminal sequence analysis of proteins purified on a nanomole scale by gel electrophoresis. *J. Biol. Chem.* **247**:3242-3251.