



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Map Location of Lactate Dehydrogenase-Elevating Virus (LDV) Capsid Protein (Vp1) Gene

E. K. GODENY,* D. W. SPEICHER,† AND M. A. BRINTON*¹

*Georgia State University, Biology Department, University Plaza, Atlanta, Georgia 30303;
and †Wistar Institute, 36th and Spruce Streets, Philadelphia, Pennsylvania 19104

Received November 28, 1989; accepted April 18, 1990

Lactate dehydrogenase-elevating virus (LDV) is currently classified within the Togaviridae family. In an effort to obtain further information on the characteristics of this virus, we have begun to sequence the viral RNA genome and to map the virion structural protein genes. A sequence of 1064 nucleotides, which represents the 3' terminal end of the genome, was obtained from LDV cDNA clones. A 3' noncoding region of 80 nucleotides followed by two complete open reading frames (ORFs) were found within this sequence. The two ORFs were in different reading frames and overlapped each other by 11 nucleotides. One ORF encoded a protein of 170 amino acids and the other ORF, located adjacent to the 3' noncoding region of the viral genome, encoded a 114 amino acid protein. Thirty-three N-terminal residues were sequenced directly from purified LDV capsid protein, Vp1, and this amino acid sequence mapped to the ORF adjacent to the 3' noncoding region. The presence of overlapping ORFs and the 3' terminal map position of Vp1 indicate that LDV differs significantly from the prototype alpha togaviruses. © 1990 Academic Press, Inc.

The Togaviridae family consists of small, spherical, enveloped viruses with icosahedral nucleocapsids and single-stranded RNA genomes of positive polarity. Recent sequence analyses indicate that many of the viruses initially classified in this family differ from the prototype alpha togaviruses in their genome organizations and replication strategies (1-4).

Lactate dehydrogenase-elevating virus (LDV) is currently classified as a togavirus (1). The LDV particle has a diameter of 50-55 nm and the diameter of the nucleocapsid has been estimated to be 30-35 nm (5, 6). The genome of LDV is a single-stranded RNA molecule of positive polarity (5) which contains a poly(A) tract at its 3' terminus (7, 8). The estimated molecular weight of the genome is 5×10^6 Da (5, 9). LDV particles are composed of at least three structural proteins: the capsid protein, Vp1, with a molecular weight of 15,000 Da; a nonglycosylated envelope protein, Vp2, with a molecular weight of 18,000 Da; and an envelope glycoprotein, Vp3, which exhibits a heterogeneous migration pattern on SDS-PAGE with an estimated molecular weight range between 24,000 and 44,000 Da (5, 10). It is not known whether the Vp3 region on the gel contains more than one protein or various differentially glycosylated forms of the same protein.

Although *in vivo* replication of LDV is highly efficient, it is difficult to produce virus in tissue culture; no cell line has yet been found which can efficiently support LDV replication. LDV replicates in primary murine cell cultures containing macrophages (11), but only a small

subpopulation (6-20%) of cells in these macrophage cultures are permissive for LDV replication (12). Since such a small proportion of cells are infected in the macrophage cultures, it has not been possible to detect intracellular viral components in cell culture extracts. Because of the technical difficulties inherent in studying LDV replication in tissue culture, neither the gene order nor the replication strategy of this virus have yet been delineated. In order to obtain the information needed to definitively classify LDV, we have begun to map the structural proteins of the neurotropic isolate of LDV (LDV-C; 13) on the viral genome.

LDV-C (approximately 2×10^{11} ID₅₀) was purified from blood plasma taken from 50 CD-1 mice (Charles River Breeding Laboratories, Boston, MA) as previously described (8). The viral structural proteins were separated by SDS-PAGE and transferred electrophoretically to a polyvinylidene difluoride (PVDF) membrane (Millipore Corp., Bedford, MA) using a modified Towbin Tris-glycine buffer (12.5 mM Tris, 96 mM glycine, pH 8.3) containing no methanol (14). After staining with Coomassie blue, the Vp1 protein band, which migrated with an apparent molecular weight of 14,000 Da (data not shown), was excised from the PVDF membrane and analyzed in an Applied Biosystems Model 475A protein sequencer. Automated protein sequence analysis was performed in the gas phase mode with on-line PTH analysis using a Model 120A analyzer as previously described (15). The following N-terminal 33 amino acid sequence was obtained for Vp1:

S Q N K K K G G Q N K G A N Q Q L N Q-
L I S A L L R N A G Q N K G.

¹ To whom reprint requests should be addressed.

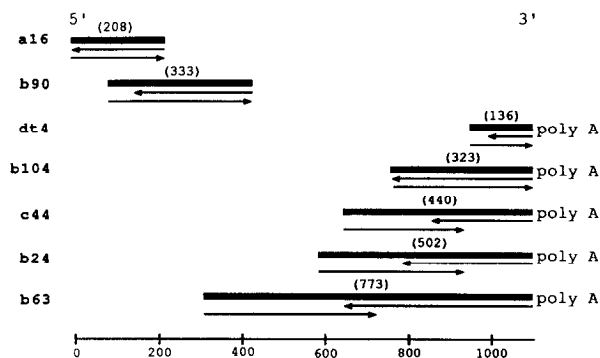


FIG. 1. Alignment of seven LDV-C cDNA clones representing the 3' terminus of the viral genome. The numbered scale depicts the length of the sequence (in nucleotides) beginning at the 5' end. The length of the heteropolymeric sequence of each clone (in nucleotides) is shown in parentheses above the respective clone. Arrows indicate the direction and extent of sequence analysis for each clone.

No homologous sequence was found in a search of the National Biomedical Research Foundation Protein Sequence data base.

We previously reported an LDV-C cDNA clone, dt4, which was synthesized by oligo-deoxythymidine priming of the viral RNA and represents the 3' terminus of the LDV-C genomic RNA (8). The genomic RNA of LDV-C was recloned as previously described (8) using calf thymus (ct) pentameric DNA for priming. Both strands of the double-stranded cDNA clones were sequenced by the dideoxy chain termination method as previously described (8) until complementary overlapping sequences were obtained within each clone. As shown in Fig. 1, four ct-primed clones (b24, b63, b104, and c44) were found to contain long poly(A) tracts at one end. These four clones completely overlap the unique sequence of the dt4 clone. The longest poly(A) tract found among these clones was 52 nucleotides in length which is very close to the length of the 3' terminal poly(A) tract (approximately 50 nucleotides) previously estimated directly from the LDV genomic RNA (7). Two additional clones, b90 and a16, further extended the 5' end of this sequence (Fig. 1).

The sequence obtained from the seven DNA clones extends 1064 nucleotides beyond the 3' terminal poly(A) tract of the LDV-C genome. This sequence, which has been converted to the viral RNA sequence, is shown in Fig. 2. Because of the high mutation rate characteristic of RNA virus replication, multiple clones were sequenced in order to obtain the majority nucleotide for any given position. With the exception of the reported change in the dt4 clone (8) at position 976 of this sequence, identical nucleotide sequences were found in all clones at all positions represented by three or more clones. Although the regions between nucleotides 89 and 207, 292 and 421, and 565 and 657 were

each sequenced from only two overlapping clones, the sequences obtained were identical except at position 311. At position 311, clone b63 contained uridine, whereas clone b90 contained cytosine (Fig. 2). The regions between nucleotides 1 through 88, 208 through 291, and 422 through 564 have not yet been confirmed in overlapping clones.

When this 1064 nucleotide sequence was translated using the Sequence Analysis Package of the Wisconsin Genetics Computer Group (16), two complete ORFs were found in different reading frames (Fig. 2). No ORF of significant size was found in the third reading frame. One ORF begins with a start codon (AUG) at nucleotide 135 and ends with an ochre termination codon (UAA) at nucleotide 648. This ORF encodes a protein of 170 amino acids (denoted as VpX, Fig. 2). The nucleotide ambiguity at position 311 does not change the encoded amino acid. Although the identity of the encoded protein is not yet known, the amino acid sequence of this protein does not contain potential N-linked glycosylation sites and is not sufficient in length to be the gene encoding the envelope glycoprotein Vp3. We have not yet obtained sufficient data to determine whether this ORF encodes the Vp2 protein.

The 5' end of the second ORF overlaps the 3' end of the ORF described above by 11 nucleotides and is in a different reading frame (Fig. 2). The sequence of the overlap region between these two ORFs was confirmed in four clones. The second ORF begins with a start codon at position 637 and ends with a single termination codon (UAG) at nucleotide 982. The 3' non-coding region of the LDV genomic RNA is 80 nucleotides in length (Fig. 2).

The N-terminal amino acid sequence obtained directly from the LDV-C Vp1 protein was found to map to the 5' terminus of the second ORF (Fig. 2). This ORF encodes a 114 amino acid protein which would have a molecular weight of approximately 12,200 Da. The estimated molecular weight of Vp1 is 15,000 Da (5, 10). The amino acid sequence of Vp1 indicates that, like other RNA virus capsid proteins, the LDV capsid protein is a basic protein: 16% of the residues in this sequence are basic amino acids (lysine or arginine) at pH 7, while only 3% of the residues are acidic (aspartic acid or glutamic acid). Consistent with its amino acid composition, Vp1 migrated to the upper pH range (pH \geq 8; data not shown) when electrophoresed on a two-dimensional gel (17).

Another partial ORF is located at the 5' end of the nucleotide sequence. This ORF is at least 144 nucleotides in length and is in a different frame from the adjacent ORF, but in the same frame as the capsid protein ORF. The 3' terminus of this ORF overlaps the 5' end of the adjacent ORF by 10 nucleotides (Fig. 2). Since the

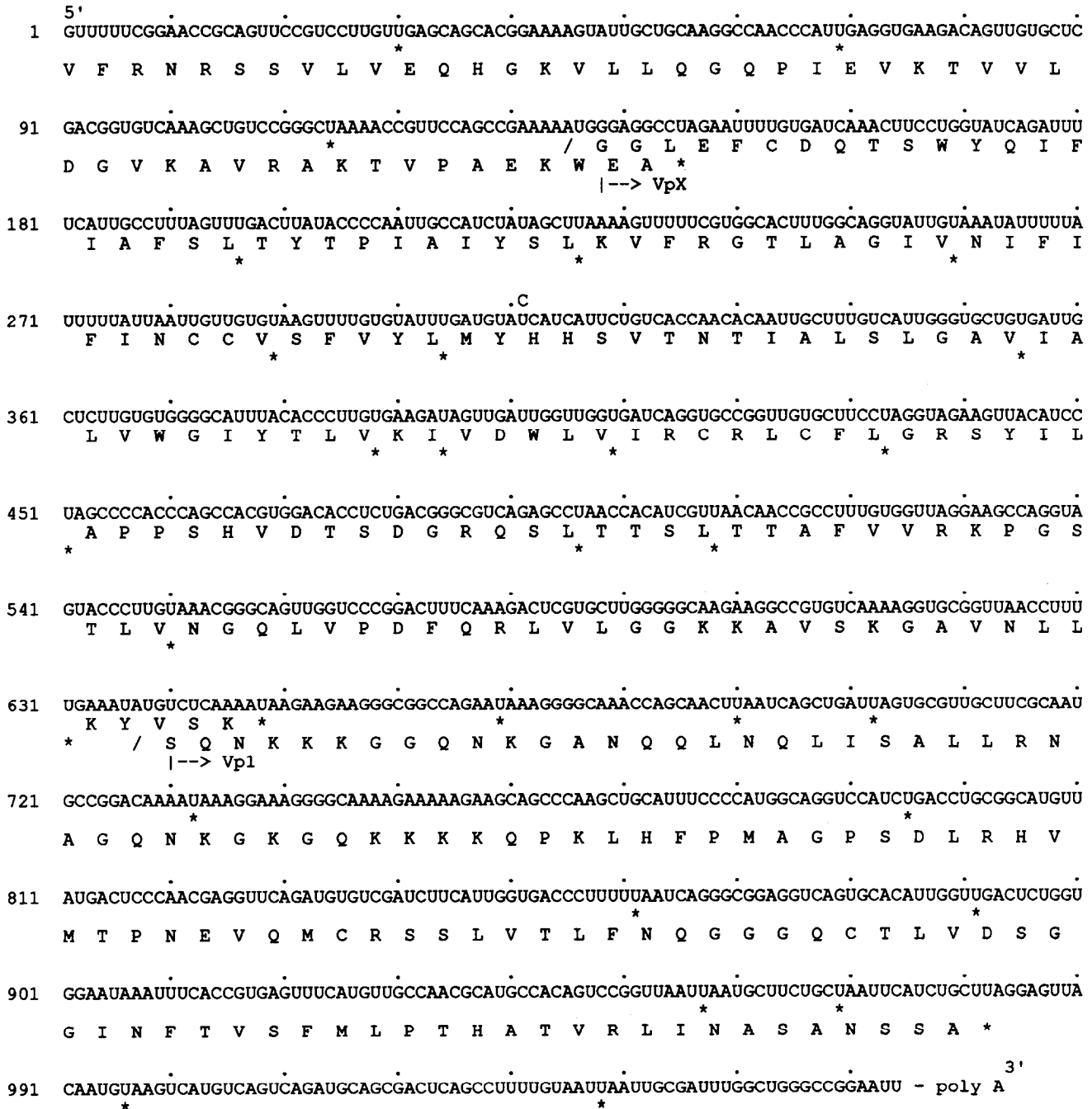


FIG. 2. LDV-C RNA sequence obtained from the seven cDNA clones shown in Fig. 1. The amino acids encoded by the ORFs are shown below the nucleotide sequence: / indicates start codons and * indicates termination codons. The termination codons found in the third reading frame (which did not exhibit an ORF) are not shown.

sequence of the 5' end of this ORF is incomplete, we do not yet know the length nor identity of the protein encoded by this ORF. There is one potential N-linked glycosylation site near the 5' end of this partial ORF.

Although morphologically similar to the togaviruses and flaviviruses, the presence of multiple ORFs in the LDV genome and the 3' terminal location of the capsid protein gene suggest that LDV is neither a togavirus

nor a flavivirus. The genomes of the viruses within these two families contain one or two long ORFs which encode polyproteins and the capsid protein genes of these viruses are located at the 5' ends of their respective ORFs.

The genome structure of LDV resembles that of equine arteritis virus (EAV), which is the only member of the genus *arteriviruses* within the *Togaviridae* family

(7). Although this virus is still classified as a togavirus, it differs significantly from the alpha and rubi togaviruses: (a) the EAV proteins are encoded by multiple ORFs; (b) the capsid protein gene maps to the 3' terminus of the coding region on the genome (4); and (c) the EAV proteins are translated from six subgenomic mRNAs (19). The presence of overlapping reading frames each beginning with a start codon suggests that LDV proteins are also translated from subgenomic mRNAs. However, due to the difficulty of detecting intracellular viral components in LDV-infected tissue culture extracts, no LDV subgenomic mRNAs have yet been observed. EAV and LDV are also morphologically similar in both virion size and nucleocapsid structure. Although the LDV structural proteins are similar in size to those of EAV (20), no serologic cross-reactivity has been found between these two viruses (21).

The LDV genome structure is also similar to that of the coronaviruses (18). However, LDV has an icosahedral nucleocapsid, whereas the coronaviruses have helical nucleocapsids. Nonetheless, the properties of the LDV genome suggest that LDV belongs to a recently proposed virus superfamily (4) consisting of the coronaviruses, the toroviruses, and the arteriviruses. Further characterization of LDV is necessary to facilitate the classification of this virus and to determine the degree of similarity it shares with EAV and the coronaviruses.

ACKNOWLEDGMENTS

This work was supported by Public Health Service Grant NS 19013 from NINCDS. We thank Michelle Gonda and Janice Dispoto for technical assistance. We also thank Kaye Speicher, Kevin Beam, and Clement Purcell in the Protein Microchemistry Facility at the Wistar Institute for protein sequence analysis and technical assistance.

REFERENCES

- WESTAWAY, E. G., BRINTON, M. A., GAIDAMOVICH, S. YA., HORZINEK, M. C., IGARASHI, A., KAARIAINEN, L., LVOV, D. K., PORTERFIELD, J. S., RUSSELL, P. K., and TRENT, D. W., *Togaviridae. Intervirology* **24**, 125-139 (1985).
- WESTAWAY, E. G., BRINTON, M. A., GAIDAMOVICH, S. YA., HORZINEK, M. C., IGARASHI, A., KÄÄRIÄINEN, L., LVOV, D. K., PORTERFIELD, J. S., RUSSELL, P. K., and TRENT, D. W., *Flaviviridae. Intervirology* **24**, 183-192 (1985b).
- COLLETT, M. S., MOENNIG, V., and HORZINEK, M. C., *J. Gen. Virol.* **70**, 253-266 (1989).
- SPAAN, W. J. M., DEN BOON, J. A., BREDEBEEK, P. J., CHIRNSIDE, E. D., NOTEN, A. F. H., SNIJDER, E. J., DE VRIES, A. A. F., and HORZINEK, M. C., *In "New Aspects of Positive Strand RNA Viruses"* (M. A. Brinton and F. X. Heinz, Eds.), pp. 12-19. ASM Press, Washington, DC, 1990.
- BRINTON-DARNELL M., and PLAGEMANN, P. G. W., *J. Virol.* **16**, 420-433 (1975).
- HORZINEK, M. D., VAN WIELINK, P. S., and ELLENS, D. J., *J. Gen. Virol.* **26**, 217-226 (1975).
- BRINTON, M. A., GAVIN, E. I., and FERNANDEZ, A. V., *J. Gen. Virol.* **67**, 2673-2684 (1986).
- GODENY, E. K., WERNER, M. R., and BRINTON, M. A., *Virology* **172**, 647-650 (1989).
- DARNELL, M. B., and PLAGEMANN, P. G. W., *J. Virol.* **10**, 1082-1085 (1972).
- MICHAELIDIS, M. C., and SCHLESINGER, S., *Virology* **55**, 211-217 (1973).
- ROWSON, K. E. K., and MAHY, B. W. J., *J. Gen. Virol.* **66**, 2297-2312 (1985).
- TONG, S. L., STUEKEMANN, J., and PLAGEMANN, P. G. W., *J. Virol.* **22**, 219-227 (1977).
- MARTINEZ, D., BRINTON, M. A., TACHOVSKY, T. G., and PHELPS, A. H., *Infect. Immun.* **27**, 979-987 (1980).
- TOWBIN, H., STAHELIN, T., and GORDON, J., *Proc. Natl. Acad. Sci. USA* **76**, 4350-4354 (1979).
- SPEICHER, D., *In "Techniques in Protein Chemistry"* (T. Hugli, Ed.), pp. 24-35. Academic Press, San Diego, 1989.
- DEVEREUX, J., HAEBERLI, P., and SMITHIES, O., *Nucleic Acids Res.* **12**, 387-395 (1984).
- O'FARRELL, P. H., *J. Biol. Chem.* **250**, 4007-4021 (1975).
- LAI, M. M. C., *In "Coronaviruses; Adv. Exp. Med. Biol. 218"* (M. M. C. Lai and S. A. Stohman, Eds.), pp. 7-13. Plenum, New York, 1987.
- VAN BERLO, M. F., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M., *Virology* **118**, 345-352 (1982).
- BRINTON, M. A., *In "The Togaviruses: Biology, Structure, Replication"* (R. W. Schlesinger, Ed.), pp. 623-666. Academic Press, San Diego, 1980.
- VAN BERLO, M. F., ZEEGERS, J. J. W., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M., *Zentralbl. Veterinaermed Reihe B* **38**, 297-304 (1983).