JOURNAL OF VIROLOGY, June 1993, p. 3611–3614
0022-538X/93/063611-04$02.00/0

Vol. 67, No. 6

Subgenomic RNA Sequence of Human Astrovirus Supports Classification of *Astroviridae* as a New Family of RNA Viruses

STEPHAN S. MONROE,* BAOMING JIANG, SARAH E. STINE, MARION KOOPMANS, AND ROGER I. GLASS

Viral Gastroenteritis Section, Respiratory and Enteric Viruses Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333

Received 11 January 1993/Accepted 17 March 1993

We report the sequence of the subgenomic RNA of human astrovirus serotype 2. This 2,484-nucleotide RNA contains a single open reading frame, which encodes a protein with a predicted molecular mass of 88 kDa. We propose that this protein is the 90-kDa capsid precursor observed in infected cells. The deduced protein sequence does not contain conserved amino acid patterns reported for the capsid proteins of picornaviruses or caliciviruses, consistent with the classification of astroviruses as a new family of RNA viruses, designated *Astroviridae*.

Astroviruses are 28-nm nonenveloped, single-stranded RNA viruses that were initially described as particles with a "star-shaped" surface structure detected by electron microscopy in stool specimens from young children with diarrhea (1, 10). The use of sensitive enzyme immunoassays (6) and RNA probe assays (12, 21) has shown that astroviruses can be associated with up to 8.6% of episodes of diarrhea in children, more commonly than was originally suggested by electron microscopic surveys (4, 7, 9, 13). Immune electron microscopy and immunofluorescence techniques have identified five serotypes of human astroviruses, currently designated H-Ast1 to H-Ast5 (8).

Although astroviruses are known to have an RNA genome, their classification on the basis of virion composition has been hindered by conflicting reports of two, three, or four distinct capsid proteins (14, 20). To examine the mechanism of replication of astroviruses, we analyzed the synthesis of proteins and RNA during a single-cycle infection of cultured cells (14). We detected a previously unreported 90-kDa protein that, by virtue of its reactivity with hyperimmune rabbit serum, is presumed to be a capsid protein precursor. This 90-kDa precursor could be cleaved by trypsin in vitro, with the appearance of three smaller proteins (31, 29, and 20 kDa). A second observation of our in vitro studies was a previously unreported 2.8-kb RNA that is polyadenylated and that we presumed to be a subgenomic mRNA encoding the 90-kDa precursor polypeptide. Although these results prompted us to conclude that astroviruses should not be classified as members of either the *Picornaviridae* or the *Caliciviridae*, a definitive taxonomic classification for astroviruses has been slowed by the absence of genome sequence information. Recently, several partial nucleotide sequences of human astrovirus serotype 1 (H-Ast1) have been reported, including 1,034 nucleotides from the 3′ end of genomic RNA (19), a 289-nucleotide immunoreactive epitope which overlaps the 3′ end sequence (11), and two overlapping regions which hybridize only to genomic RNA (11, 18). We sequenced the entire subgenomic RNA of human astrovirus serotype 2 (H-Ast2) and compared the predicted amino acid sequence of the capsid precursor with sequences available for picornaviruses and caliciviruses to determine whether astroviruses represent a unique family of RNA viruses.

cDNA cloning and RNA blot hybridization. Cell-culture-adapted human astrovirus serotype 2 (H-Ast2) was obtained from John Kurtz (Oxford, England), was plaque purified three times before use, and was propagated in LLCMK2 cells (ATCC CCL7.1) as previously described (14). Double-stranded cDNA was synthesized from the polyadenylated fraction of RNA isolated from astrovirus-infected cells (cDNA cloning kit; Boehringer Mannheim Biochemicals) and was cloned into the pBluescript II plasmid vector (Stratagene). Recombinant clones were screened for astrovirus-specific inserts by hybridization of ^{32}P-labelled RNA transcribed in vitro from individual cDNA clones to total cytoplasmic RNA isolated from uninfected and astrovirus-infected cells.

The RNA transcripts from the insert in one cDNA clone (no. 16) hybridized to both the 7.2- and 2.8-kb viral RNAs (Fig. 1). This reactivity was first detectable at 12 h postinfection, coincident with detection of these RNAs by metabolic labeling (14). The hybridization of a cRNA probe to both virus-specific RNAs confirmed that the 2.8-kb RNA contains sequences present in the larger species, with the relative intensities indicating that the smaller RNA is present in at least a 10-fold molar excess. These observations support our earlier conclusion that the 2.8-kb RNA is a subgenomic mRNA (14).

Nucleotide sequence analysis. Sequence information for the subgenomic RNA was obtained by three approaches: (i) sequencing of supercoiled DNA from two plasmids with cDNA inserts, (ii) sequencing of RNA purified from virions, and (iii) amplification of genomic RNA by reverse transcriptase-polymerase chain reaction (PCR), followed by sequencing of the double-stranded DNA products. Plasmid DNA and PCR products were sequenced with modified T7 DNA polymerase (Sequenase2; U.S. Biochemicals). Sequence information from the 5′ end of the original clone was used to generate oligonucleotide primers for a second round

* Corresponding author.

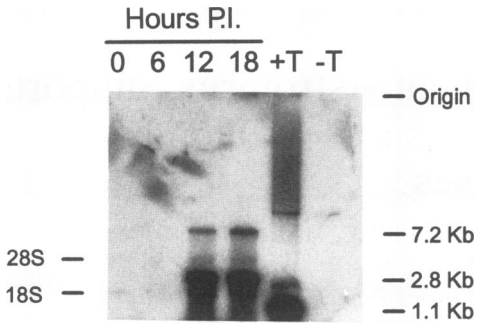

FIG. 1. Autoradiogram of RNA blot hybridization analysis. Total cytoplasmic RNA isolated from astrovirus-infected cells at the indicated times postinfection and unlabeled RNA transcribed in vitro from cDNA clone 16 were resolved in a 1.2% agarose gel, transferred to a nylon membrane, and probed with ^{32}P-labelled RNA transcribed from cDNA clone 16. The positions of the 28S and 18S ribosomal RNAs are indicated on the left. The positions of genomic (7.2 kb), subgenomic (2.8 kb), and in vitro transcript (1.1 kb) RNAs are indicated on the right. Lanes: T+, unlabeled positive-sense RNA transcribed from cDNA clone 16; T−, unlabeled minus-sense RNA transcribed from cDNA clone 16. The upper band in lane +T represents hybridization of the probe RNA to the template DNA in the transcription reaction.

of cDNA cloning from cytoplasmic RNA. PCR products were purified by gel filtration (Miniprep spun column; Pharmacia) before sequencing. RNA was sequenced using reverse transcriptase and dideoxynucleotide terminators (RNA sequencing kit; Boehringer Mannheim Biochemicals), with primers derived from the sequence of the cDNA clones. Although direct RNA sequencing often resulted in regions of the gels that were difficult to interpret, ambiguities were resolved through the use of reverse transcriptase-PCR sequencing over the same regions. The information from the three independent sequencing strategies was combined to arrive at a consensus sequence for the entire subgenomic region (5). The sequence derived from cDNA clone 16 contains a 29-nucleotide poly(A) tract immediately adjacent to the cloning linker, indicating that this cDNA insert is probably derived from the extreme 3′ end of viral RNA. The location of the 5′ end of the subgenomic RNA was estimated by primer runoff, using total cytoplasmic RNA as template (data not shown).

The consensus sequence for the unique region of the subgenomic RNA is 2,484 nucleotides long and includes the following features: (i) an 11-nucleotide 5′-untranslated region (5′-UTR); (ii) a 2,388-nucleotide open reading frame (ORF); and (iii) an 85-nucleotide 3′-UTR (Fig. 2).

Analysis of the predicted capsid precursor polypeptide. The single ORF in the subgenomic RNA encodes a 796-amino-acid polypeptide with a predicted molecular mass of 88 kDa, consistent with the estimated 90-kDa mass of the capsid protein precursor we observed in infected cells (14). At the amino terminus, the predicted polypeptide has a region of basic amino acids that may play a role as a nucleic acid binding motif. At the carboxy terminus is a region of acidic amino acids whose function is unknown.

Comparison of the H-Ast2 subgenomic RNA and deduced protein sequences to the H-Ast1 partial sequence. A comparison of the H-Ast2 RNA and deduced protein sequences to the partial sequences previously reported for H-Ast1 (11, 19) indicated regions of both similarities and differences. The nucleotide sequence immediately adjacent to the poly(A) tract, including the 3′-UTR and the last eight codons of the predicted ORF is 94% conserved, with only five differences and two single base insertions in the first 109 unique nucleotides. Four of the five differences, including two in the coding region, result in compensating changes that maintain base pairing in predicted stem-loop structures at the 3′ ends of the RNAs (22) (Fig. 3). Note that stem I includes base pairs involving the poly(A) tract. The two insertions in the H-Ast2 sequence occur in a predicted loop between conserved stems I and II. The terminator UAG codons are located in the loop at the top of stem II, between the conservative changes. The stems marked III, although similar in predicted secondary structure, are composed of dissimilar sequences. The conserved primary and secondary structures at the 3′ end of the genome may function as recognition sites during RNA replication. As a further indication that the primary sequence information in this 3′ region is conserved among astroviruses, we have recently used oligonucleotide primers derived from this region to amplify RNA from all five reference serotypes of human astrovirus (unpublished data).

In contrast to the high degree of primary sequence conservation at the 3′ end of the genome, there is only 59% nucleotide sequence identity in the consensus coding region sequence from H-Ast1 (11, 19) and the corresponding region of H-Ast2. Alignment of the 392-amino-acid partial H-Ast1 sequence with the corresponding region of the H-Ast2 amino acid sequence indicates an overall similarity of 67%, with 52% identical residues (5, 15). The proteins are more conserved at their carboxy termini, which both include the highly acidic region, with 80% similarity and 62% identity over the terminal 114 residues.

Comparison of the predicted capsid protein precursor to those from other nonenveloped RNA viruses. Recent publication of sequences from the capsid regions of several caliciviruses has allowed identification of conserved amino acid patterns within and among the caliciviruses and picornaviruses (17). One highly conserved sequence found both in calicivirus capsid proteins and in picornavirus VP3 proteins is proline-proline-glycine, or PPG. This sequence is present in the partial H-Ast1 sequence but is not present at the corresponding location in the H-Ast2 sequence. That this match probably represents random, rather than functional, homology is supported by the lack of similarity between the surrounding astrovirus protein sequences and conserved amino acid patterns on either side of the PPG in calicivirus and picornavirus capsids (16). Furthermore, we were unable to detect similarity between any of the most highly conserved calicivirus capsid protein motifs and the deduced H-Ast2 capsid protein precursor sequence.

Classification of astroviruses. The two well-characterized families of nonenveloped, monopartite RNA animal viruses are *Picornaviridae* and *Caliciviridae*. The sequence of the H-Ast2 subgenomic RNA presented in this report supports our earlier conclusion that astroviruses cannot be readily classified into either of these two families. A common feature of the replication of those positive-strand RNA viruses that encode their structural proteins at the 3′ end of the genome is the production of one or more subgenomic mRNAs during the virus infection cycle. Picornaviruses and flaviviruses, which encode their structural proteins at the 5′ end of the genome, do not synthesize subgenomic RNAs. We reported the synthesis of a smaller-than-genome-sized RNA in H-Ast2-infected cells (14) and proposed that it serves as a subgenomic mRNA. A subgenomic RNA of similar size has recently been detected in H-Ast1-infected

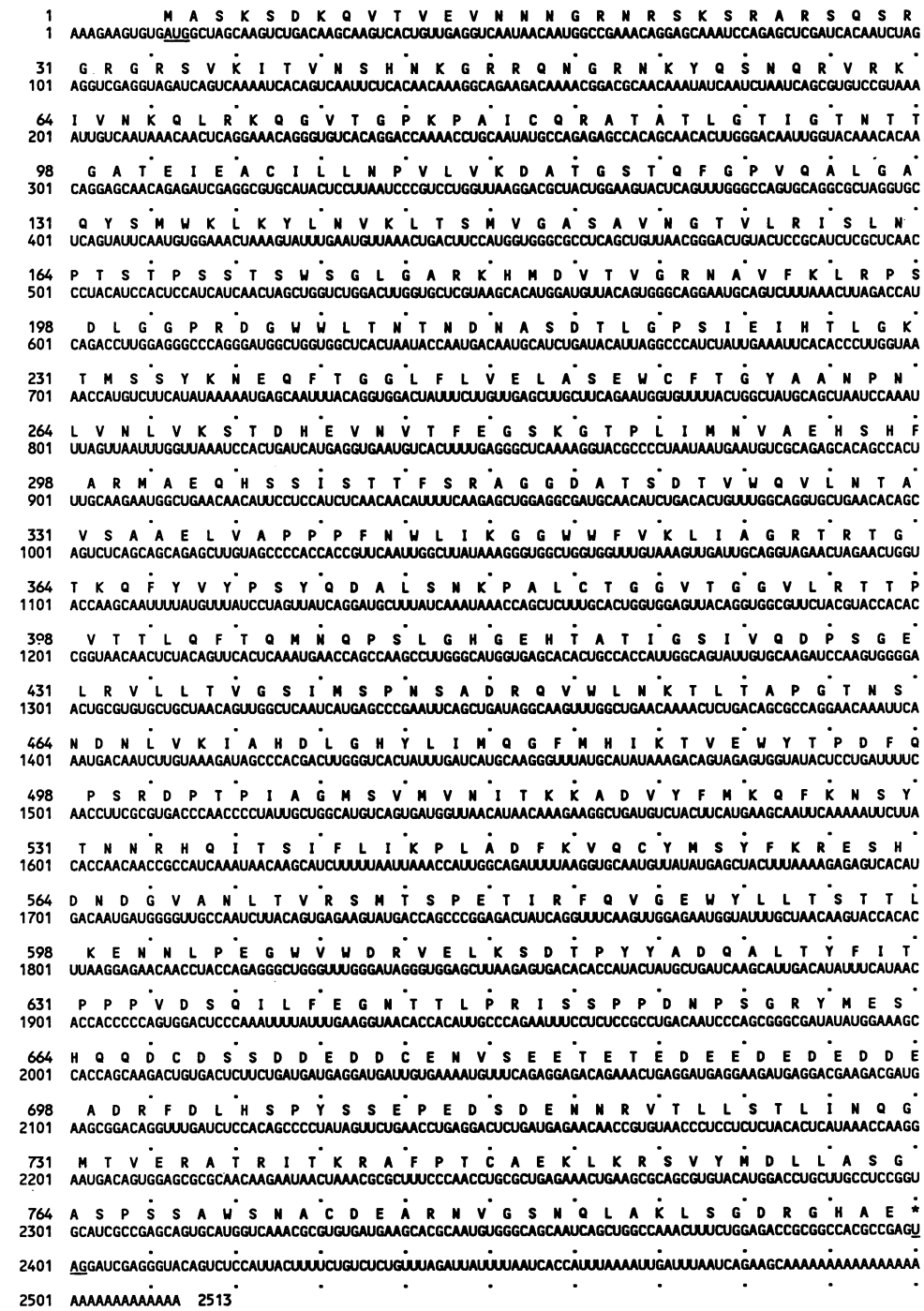

FIG. 2. Nucleotide sequence and deduced amino acid sequence of H-Ast2 subgenomic RNA. The predicted initiation and termination codons are underlined.

cells, using RNA blot hybridization (11, 18). The synthesis of a subgenomic mRNA in infected cells is inconsistent with classification of astroviruses in the family *Picornaviridae*.

Astroviruses from humans and animals contain several capsid proteins of 29 to 33 kDa, which distinguishes them from caliciviruses which have a single capsid protein of 58 to 65 kDa (3, 20). The ORF in the subgenomic mRNA of H-Ast2 encodes a protein with a predicted mass of 88 kDa, consistent with the 90-kDa precursor protein detected in astrovirus-infected cells (14). Treatment of the intracellular 90-kDa capsid precursor with trypsin in vitro resulted in the appearance of three immunoreactive proteins (31, 29, and 21 kDa) (14). Although we do not know any details of the processing of this precursor to the mature capsid proteins, we are currently attempting to express the protein encoded by the H-Ast2 subgenomic RNA in order to identify the

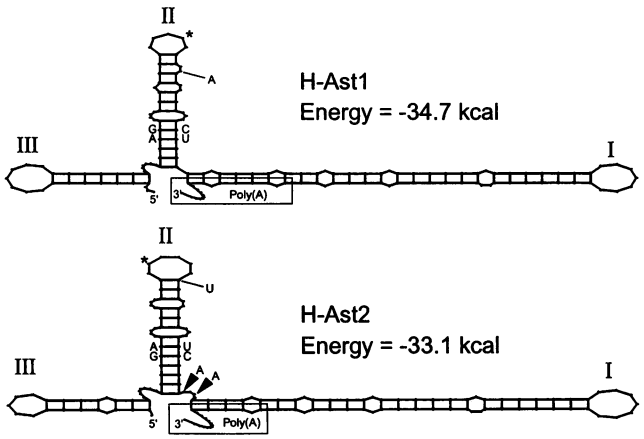

FIG. 3. Predicted secondary structure at the 3′ end of astrovirus RNA sequences. The structures were calculated by the method of Zuker and Stiegler (5, 22). The H-Ast1 structure contains a total of 154 nucleotides, including 134 bases from the reported 3′ end sequence (19) plus 20 additional adenine residues. The H-Ast2 structure contains 156 nucleotides corresponding to bases 2349 to 2504 in Fig. 2. The region of the poly(A) tract involved in stem I is boxed. The two insertions in the loop between stems I and II are shown with arrowheads. The residues within the conserved stem II that vary between the two serotypes are indicated. The terminator codons in the loop of stem II are marked with asterisks.

capsid proteins and define their antigenic sites. These differences in capsid protein composition and the absence of conserved amino acid patterns are inconsistent with classification of astroviruses in the family *Caliciviridae*. On the basis of this report and our earlier studies, astroviruses will be classified in a separate family of nonenveloped RNA viruses, *Astroviridae*, in the forthcoming Sixth Report of the International Committee on the Taxonomy of Viruses (2).

Nucleotide sequence accession number. The GenBank accession number of the H-Ast2 subgenomic RNA sequence shown in Fig. 2 is L06802.

We thank Mark Pallansch, Jon Gentsch, Olen Kew, and John O'Connor for critical review of the manuscript.

REFERENCES

1. **Appleton, H., and P. G. Higgins.** 1975. Viruses and gastroenteritis in infants. Lancet **i**:1297.
2. **Bishop, D. H. L.** 1992. Personal communication.
3. **Carter, M. J., I. D. Milton, and C. R. Madeley.** 1991. Caliciviruses. Rev. Med. Virol. **1**:177–186.
4. **Cruz, J. R., A. V. Bartlett, J. E. Herrmann, P. Cáceres, N. R. Blacklow, and F. Cano.** 1992. Astrovirus-associated diarrhea among Guatemalan ambulatory rural children. J. Clin. Microbiol. **30**:1140–1144.
5. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
6. **Herrmann, J. E., N. A. Nowak, D. M. Perron-Henry, R. W. Hudson, W. D. Cubitt, and N. R. Blacklow.** 1990. Diagnosis of astrovirus gastroenteritis by antigen detection with monoclonal antibodies. J. Infect. Dis. **161**:226–229.
7. **Herrmann, J. E., D. N. Taylor, P. Echeverria, and N. R. Blacklow.** 1991. Astroviruses as a cause of gastroenteritis in children. N. Engl. J. Med. **324**:1757–1760.
8. **Kurtz, J. B., and T. W. Lee.** 1984. Human astrovirus serotypes. Lancet **ii**:1405. (Letter.)
9. **Lew, J. F., C. L. Moe, S. S. Monroe, J. R. Allen, B. M. Harrison, B. D. Forrestor, S. E. Stine, P. A. Woods, J. C. Hierholzer, J. E. Herrmann, N. R. Blacklow, A. V. Bartlett, and R. I. Glass.** 1991. Astrovirus and adenovirus associated with diarrhea in children in day care settings. J. Infect. Dis. **164**:673–678.
10. **Madeley, C. R., and B. P. Cosgrove.** 1975. 28 nm particles in faeces in infantile gastroenteritis. Lancet **ii**:451–452. (Letter.)
11. **Matsui, S. M., J. P. Kim, H. B. Greenberg, L. V. M. Young, L. S. Smith, T. L. Lewis, J. E. Herrmann, N. R. Blacklow, K. Dupuis, and G. R. Reyes.** 1993. Cloning and characterization of human astrovirus immunoreactive epitopes. J. Virol. **67**:1712–1715.
12. **Moe, C. L., J. R. Allen, S. S. Monroe, H. E. Gary, Jr., C. D. Humphrey, J. E. Herrmann, N. R. Blacklow, C. Carcamo, M. Koch, K.-H. Kim, and R. I. Glass.** 1991. Detection of astrovirus in pediatric stool samples by immunoassay and RNA probe. J. Clin. Microbiol. **29**:2390–2395.
13. **Monroe, S. S., R. I. Glass, N. Noah, T. H. Flewett, E. O. Caul, C. I. Ashton, A. Curry, A. M. Field, R. Madeley, and P. J. Pead.** 1991. Electronmicroscopic reporting of gastrointestinal viruses in the United Kingdom, 1985–87. J. Med. Virol. **33**:193–198.
14. **Monroe, S. S., S. E. Stine, L. Gorelkin, J. E. Herrmann, N. R. Blacklow, and R. I. Glass.** 1991. Temporal synthesis of proteins and RNAs during human astrovirus infection of cultured cells. J. Virol. **65**:641–648.
15. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**:443–453.
16. **Neill, J. D.** 1992. Nucleotide sequence of the capsid protein gene of two serotypes of San Miguel sea lion virus: identification of conserved and non-conserved amino acid sequences among calicivirus capsid proteins. Virus Res. **24**:211–222.
17. **Tohya, Y., Y. Taniguchi, E. Takahashi, E. Utagawa, N. Takeda, K. Miyamura, S. Yamazaki, and T. Mikami.** 1991. Sequence analysis of the 3′-end of feline calicivirus genome. Virology **183**:810–814.
18. **Willcocks, M. M., and M. J. Carter.** 1992. Program Abstr. Third International Symposium on Positive Strand RNA Viruses, Clearwater, Fla., abstr. P2-47.
19. **Willcocks, M. M., and M. J. Carter.** 1992. The 3′ terminal sequence of a human astrovirus. Arch. Virol. **124**:279–289.
20. **Willcocks, M. M., M. J. Carter, and C. R. Madeley.** 1992. Astroviruses. Rev. Med. Virol. **2**:97–106.
21. **Willcocks, M. M., M. J. Carter, J. G. Silcock, and C. R. Madeley.** 1991. A dot-blot hybridisation procedure for the detection of astrorvirus in stool samples. Epidemiol. Infect. **107**:405–410.
22. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.