Analysis of Astrovirus Serotype 1 RNA, Identification of the Viral RNA-Dependent RNA Polymerase Motif, and Expression of a Viral Structural Protein

TERRY L. LEWIS,^{1,2} HARRY B. GREENBERG,^{1,2,3,4} JOHN E. HERRMANN,⁵
LUCINDA S. SMITH,⁴ AND SUZANNE M. MATSUI^{1,2,4*}

Program in Cancer Biology,¹ Department of Medicine (Gastroenterology),² and Department of Microbiology and Immunology,³ Stanford University School of Medicine, Stanford, California 94305; Veterans Affairs Medical Center, Palo Alto, California 94304⁴; and Department of Medicine (Infectious Diseases), University of Massachusetts Medical Center, Worcester, Massachusetts 01655⁵

Received 17 May 1993/Accepted 11 August 1993

We report the results from sequence analysis and expression studies of the gastroenteritis agent astrovirus serotype 1. We have cloned and sequenced 5,944 nucleotides (nt) of the estimated 7.2-kb RNA genome and have identified three open reading frames (ORFs). ORF-3, at the 3' end, is 2,361 nt in length and is fully encoded in both the genomic and subgenomic viral RNAs. Expression of ORF-3 in vitro yields an 87-kDa protein that is immunoprecipitated with a monoclonal antibody specific for viral capsids. This protein comigrates with an authentic 87-kDa astrovirus protein immunoprecipitated from infected cells, indicating that this region encodes a viral structural protein. The adjacent upstream ORF (ORF-2) is 1,557 nt in length and contains a viral RNA-dependent RNA polymerase motif. The viral RNA-dependent RNA polymerase motifs from four astrovirus serotypes are compared. Partial sequence (2,018 nt) of the most 5' ORF (ORF-1) reveals a 3C-like serine protease motif. The ORF-1 sequence is incomplete. These results indicate that the astrovirus genome is organized with nonstructural proteins encoded at the 5' end and structural proteins at the 3' end. ORF-2 has no start methionine and is in the -1 frame compared with ORF-1. We present sequence evidence for a ribosomal frameshift mechanism for expression of the viral polymerase.

Astrovirus was first described as a human pathogen in 1975 during an outbreak of gastroenteritis among newborn infants (27). Since then, five serotypes of human astrovirus have been identified (23) and adapted to growth in tissue culture (25, 43). Immunologic reagents derived from astroviruses grown in cell culture have been used in epidemiologic studies to define the role of astroviruses in gastroenteritis (13, 14). In a recent study from Thailand, the incidence of astrovirus diarrhea in children attending outpatient clinics was 8.6%, far greater than the incidence of enteric adenovirus (2.6%) gastroenteritis and nearly half that of rotavirus (19%) (15). Similar values were obtained in a study from Guatemala (5, 6). In a longitudinal study in Atlanta, astrovirus was the most common viral pathogen detected in fecal specimens of human immunodeficiency virus-infected patients with diarrhea (12%) (11).

The genome of human astrovirus is estimated to be a 7.2-kb RNA molecule of positive sense. In addition, a ~2.8-kb subgenomic RNA that is coterminal with the 3' end of the genome is synthesized in astrovirus-infected cells (29, 32). The complete subgenomic RNA from astrovirus serotype 2 (31) and the 3'-terminal 1,034 nucleotides (nt) from serotype 1 (42) have been cloned and sequenced and do not have significant homology to other RNA viruses. Matsui et al. (29) identified two immunoreactive regions of the virus. One of the clones that defines an immunoreactive epitope overlaps with the 3'-terminal sequence of Willcocks and Carter (42) and hybridizes to both viral genomic and subgenomic poly(A)⁺ RNAs from infected cells on Northern (RNA) blots (29). The three clones defining the other immunoreactive region hybridize

only to viral genomic RNA. The astrovirus sequence that is currently available in the literature is limited and not sufficient to provide detailed insight into its molecular biology or relatedness to other viruses.

Astrovirus protein composition is not well characterized. Reports vary on the size and number of astrovirus structural proteins (12, 24, 40, 43). Monroe et al. (32) described an ~90-kDa, astrovirus-specific protein that is synthesized during infection and speculated that this may be encoded by the subgenomic RNA. They also suggested that this protein may undergo cleavage by trypsin into three smaller proteins that are similar in size to previously described astrovirus surface proteins. Information regarding the astrovirus nonstructural proteins is not currently available.

Recently we cloned and sequenced 5,944 nt of the ~7.2-kb astrovirus genome. We now report the results of a detailed analysis of this astrovirus sequence. The genome organization was elucidated from the location of the viral RNA-dependent RNA polymerase (RDRP) motif-encoding region and determination that the subgenomic RNA does encode a viral structural protein. The available data also suggest a -1 frameshift translation strategy for expression of the gene encoding the astrovirus RNA-dependent RNA polymerase.

MATERIALS AND METHODS

Cells, viruses, and RNA isolation. Human astrovirus serotypes 1, 2, 4, and 5 were propagated in LLCMK₂ cells (ATCC CCL 7.1). RNA for Northern blot hybridization and reverse transcriptase (RT)-PCR was obtained from infected cell lysates as described previously (29, 32). Total cellular and viral RNA was extracted from these lysates by using a one-step guanidinium-phenol extraction protocol (4).

* Corresponding author. Mailing address: Division of Gastroenterology, P304, Stanford University School of Medicine, Stanford, CA 94305-5487. Phone: (415) 493-5000, Ext. 3121. Fax: (415) 852-3259.

Primers. All oligonucleotide primers were synthesized on an ABI model 394 DNA synthesizer, using phosphoramidite chemistry. The primer sequences were derived from published astrovirus serotype 1 sequences (29, 42). The first primers used were 5'-end primer A (5'-CCTTGCCGTAAGTTTGTGAGT-3') (29) and 3'-end primer B (5'-TTTGCTTCTGATTAAATCAA-3') (42).

cDNA synthesis, PCR amplification, and cloning. First-strand cDNA synthesis was performed with oligo(dT), random primers, or primer B. Approximately 3 µg of total RNA from infected cells was annealed with 0.5 µg of primer at 94°C, 56°C, 42°C, and room temperature for 10 min each. First-strand cDNA synthesis and subsequent PCR were performed as previously described (37). PCR products were analyzed on 1% agarose gels and purified by using Gene CleanII (Bio 101). The double-stranded cDNA products were verified to be astrovirus specific by hybridization to Northern blots as previously described (29, 31). The purified PCR products were either sequenced directly (described below) or cloned into M13 mp18, M13 mp19, or pBluescript KS(-). DNA was made blunt ended with T4 DNA polymerase (Promega) and ligated to *Sma*I-digested vector as described previously (39).

First-strand cDNA ligations followed by PCR were used to obtain 5'-end, double-stranded cDNA products (16, 26). All PCR products were cloned and analyzed as described above.

Astrovirus serotype 1-specific primers were used to amplify specific regions of other available serotypes of astrovirus as described below.

Sequencing. The PCR products were sequenced either directly or after cloning. The cloned PCR products were sequenced with the Sequenase 2.0 kit (U.S. Biochemical) according to the manufacturer's recommendations. Areas that were difficult to interpret because of compression of bands were resolved by sequencing with dITP, a nucleotide analog of dGTP, or with direct PCR sequencing. Briefly, for the latter method, the PCR products used for direct sequencing were separated on a 1% low-melting-point agarose gel, excised, and purified by using Magic PCR preps (Promega). Both strands of the eluted PCR preps were directly sequenced with the *fmol* DNA sequencing kit (Promega), using the direct incorporation protocol and 10 µCi of [³⁵S]dATP (Amersham). Following the thermocycling process, 1 U of terminal transferase (Gibco-BRL) was added in the presence of 1 mM deoxynucleoside triphosphates, 0.1 M potassium cacodylate (pH 7.2), 2 mM CoCl₂, and 200 mM dithiothreitol, and the mixture was incubated for 1 h at 37°C before termination of the reaction. The reaction products were separated on a 6% acrylamide-7 M urea buffer gradient gel, and autoradiograms were obtained (28). Sequence alignments and comparisons were made with the assistance of the University of Wisconsin Genetics Computer Group software (7).

In vitro transcription and translation. The 3'-terminal 2,361 nt defining the open reading frame (ORF) of the viral subgenomic RNA was cloned into pBluescript KS(-) as described above. The orientation of the insert was determined by *Bam*HI restriction enzyme digestion and confirmed by PCR with insert- and vector-specific primers. The coupled transcription-translation was done by using the TnT kit (Promega) with T3 RNA polymerase under conditions recommended by the manufacturer. Astrovirus serotype 1-specific proteins were produced and labeled in vitro with [³⁵S]methionine.

Metabolic labeling of astrovirus serotype 1 proteins in cell culture. LLCMK₂ cell monolayers in six-well plates were infected with astrovirus serotype 1 in medium containing trypsin (5 µg/ml). After 1 h, complete medium with 2% fetal bovine serum was added. At 12 h postinfection, methionine-

free minimal essential medium and 20 µCi of [³⁵S]methionine were added to each well. Control wells were mock infected with trypsin-containing medium and otherwise treated identically to infected wells. Infection with label incorporation was allowed to proceed for 6 h (18 h postinfection).

Radioimmunoprecipitation. Cells were lysed in high-salt radioimmunoprecipitation assay (RIPA) buffer (10 mM Tris-HCl [pH 7.5], 150 mM NaCl, 500 mM KCl, 5 mM EDTA, 2% Triton X-100) on ice for 30 min. The cell debris was removed by centrifugation for 20 min in a microcentrifuge. Then 900 µl of RIPA buffer was added to 100 µl of lysate, followed by either 1 µl of polyclonal antiserum produced in rabbits to astrovirus serotype 2 or 1 µl of astrovirus group-reactive monoclonal antibody 8E7 (13). Hyperimmune rotavirus antiserum and a rotavirus monoclonal antibody were used as negative controls (38). Immunoprecipitation of astrovirus-specific protein(s) synthesized in vitro was accomplished by adding 1 ml of RIPA buffer and 1 µl of antibody to 10 µl (20%) of the TnT reaction as described above. After extensive washing, the precipitates were resuspended in dithiothreitol (100 mM)-containing buffer and separated on a sodium dodecyl sulfate (SDS)-12% polyacrylamide minigel. The gel was fixed, enhanced with salicylic acid, dried, and autoradiographed.

Nucleotide sequence accession numbers. GenBank nucleotide sequence accession numbers are as follows: L23508 and L23511, suspected astrovirus frameshift region from serotypes 2 and 5, respectively; L23509, L23510, and L23512, astrovirus RDRP motifs from serotypes 2, 4, and 5, respectively; and L23513, 5944 nucleotides from astrovirus serotype 1.

RESULTS

Molecular cloning. Clones spanning 5,944 nt of the astrovirus type 1 genome were obtained by using astrovirus-specific primers from previously published partial sequences (29, 42). Previous reports had also demonstrated that cells infected with either astrovirus serotype 1 or astrovirus serotype 2 produce two populations of poly(A)⁺ RNA with overlapping 3'-end sequence (29, 32). Thus, selection of primer sequence and primer orientation was based on the postulate that probes representing published clones that hybridize to viral genome-length RNA alone are located toward the 5' end and those hybridizing to both viral genomic and subgenomic RNA are found toward the 3' end. Clones that bridged the gap between known sequences were generated by using multiple primer pairs in numerous RT-PCRs. In addition, a fortuitous mispriming provided a 1.2-kb PCR product that extended toward the 5' end from primer A. First-strand cDNA ligation elucidated an additional 316 nt 5' to the 1.2-kb PCR product. All clones were verified to be astrovirus specific by hybridization to Northern blots of RNA from infected, but not uninfected, cells. As expected, the two patterns of hybridization described above were observed (data not shown) (29). Clones from an average of three separate RT-PCRs were sequenced in both directions. PCR sequencing of nearly all of the 5,944 bases was used to verify the clone sequences and to interpret compressed regions. This extensive sequencing, using a variety of strategies, helped to decrease the likelihood of errors due to *Taq* DNA polymerase misincorporation and to obtain a consensus sequence of the viral RNA population.

Sequence analysis. Three ORFs were identified: ORF-1 (nt 1 to 2018), terminating with TAG at 2019 to 2021, ORF-2 (nt 1949 to 3505), terminating with TAG at 3506 to 3508, and ORF-3 (nt 3501 to 5861), initiating with ATG at 3501 to 3503 and terminating with TAG at 5862 to 5864 (Fig. 1). The 3'

FIG. 1. Schematic representation of ORFs identified in the 5,944 nt of astrovirus type 1 sequence. ORF-1 is incomplete. ORF-2 does not contain an obvious start methionine. The 3' nontranslated region consists of 80 nt preceding the poly(A) tail.

noncoding region consists of 80 nt preceding the poly(A) tail, as previously described (31, 42). The ORF-1 sequence is incomplete, and the start ATG has not yet been determined. There is no identifiable start ATG in ORF-2 until 453 bases into its coding sequence. This observation is discussed in greater detail below. All sequences presented in this report have been submitted to the GenBank data base (accession numbers are listed in Materials and Methods) and will not be duplicated here. All references to nucleotide number are based on GenBank accession number L23513.

(i) **Subgenomic RNA encodes a single long ORF.** ORF-3-specific probes hybridize to both viral genomic and subgenomic RNAs from infected cells. ORF-3 and ORF-2 overlap by 5 nt (nt 3501 to 3505). ORF-3 consists of 2,361 nt encoding 787 amino acids and predicts a protein of 87 kDa. The 3'-terminal 1,034 bases are 97% homologous to the published astrovirus serotype 1 sequence (42). The sequence obtained in the current study also has a 3-nt insertion (GAC) at nt 5562 to 5564. The astrovirus serotype 1 used in this study had been passed several times in cell culture, in contrast to the very low passage virus that was cloned by Willcocks and Carter (42), which may have been responsible for these differences. The recently published astrovirus serotype 2 subgenomic sequence is 71% homologous at the nucleotide level and 80% similar with 70% identity at the amino acid level to the astrovirus serotype 1 sequence determined in this study (31). The type 2 ORF is 27 nt longer, with greater similarity to the type 1 sequence at the 5' end and diverging at the 3' end. The 3' nontranslated region is highly conserved between the two serotypes. No significant homology to other available sequences in GenBank was found. The 65 amino acids at the amino terminus of the deduced protein are highly basic, and the approximately 100 amino acids near the C terminus are highly acidic.

(ii) **In vitro expression of the 3' ORF.** ORF-3 was cloned and expressed in pBluescript KS(–) in a coupled transcription-translation system. An intense band corresponding to a protein of 87 kDa in mass and several bands of lower intensity were seen on denaturing gels (Fig. 2). Astrovirus-specific antibodies (serum obtained from a rabbit immunized with gradient-purified astrovirus particles [lane 2] and monoclonal antibody 8E7 [lane 4]) immunoprecipitated the predominant 87-kDa protein. Prior studies had demonstrated that monoclonal antibody 8E7 was directed to an antigenic structure found on the astrovirus virion (13). None of the smaller bands were immunoprecipitated by astrovirus-specific antibodies, suggesting that a conformational epitope may be lost as the protein degrades or that internal initiation at other out-of-frame methionines does not produce epitopes recognized by the antibodies used. Control antibodies (rotavirus-specific polyclonal and monoclonal antibodies; lanes 3 and 5, respectively) did not immunoprecipitate any of the translation products. In addition, radiolabeled cell lysates of infected and uninfected LLCMK₂ cells were tested for immunoprecipitation with as-

FIG. 2. In vitro transcription and translation of astrovirus serotype 1 ORF-3. The coupled transcription-translation reaction was performed with T3 RNA polymerase and labeled with [³⁵S]methionine in cell-free rabbit reticulocyte lysates. The translated proteins were analyzed by electrophoresis on an SDS-12% polyacrylamide minigel. Lane 1 shows untreated labeled translation products. An aliquot of the translation products shown in lane 1 was immunoprecipitated with hyperimmune polyclonal astrovirus serotype 2 antiserum (lane 2), hyperimmune polyclonal rotavirus antiserum (lane 3), astrovirus-specific monoclonal antibody 8E7 (lane 4), and rotavirus VP7-specific monoclonal antibody M60 (lane 5). Sizes are indicated in kilodaltons.

trovirus- and rotavirus-specific polyclonal antibodies (data not shown). An 87-kDa protein was immunoprecipitated from infected cells with the astrovirus-specific antiserum. This protein comigrated with the immunoprecipitated astrovirus protein expressed in vitro. These results provide proof for the hypothesis (1, 35) that the subgenomic RNA encodes a single, large viral structural protein.

(iii) **Identification of a viral RNA-dependent RNA polymerase.** Amino acid sequence analysis of ORF-2 demonstrates the presence of conserved sequences consistent with the RDRP motif. Nucleic acid probes representing the regions of the viral genome encoding this viral polymerase motif hybridize only to viral genomic RNA on Northern blots. When this motif was aligned with other known viral RNA polymerases, the astrovirus serotype 1 deduced amino acid sequence showed greatest similarity with the polymerase motifs of viruses in the poliovirus-like group of the plus-stranded RNA viruses (data not shown) (36).

A 600-nt fragment (nt 2701 to 3300) encoding the RDRP motifs of astrovirus serotypes 2, 4, and 5 was cloned and sequenced by using astrovirus serotype 1-specific primers for RT-PCR. Astrovirus serotype 3 grows poorly in culture and was not used in any of these studies. All four astrovirus serotypes have a high degree of similarity at the amino acid level, with identity in the conserved areas of the motif (Fig. 3). There is greater variation at the nucleotide level, ranging from 85.8% for serotype 5 to 93.7% for serotype 2 when compared with serotype 1. Most nucleotide changes were third-base changes, and amino acid substitutions were generally conservative. There is no identifiable start methionine for ORF-2.

(iv) **5' nonstructural protein region (ORF-1).** This 5' ORF is incomplete, and the start methionine has not yet been identified. This region contains sequences consistent with a 3C-like serine protease motif (amino acids 192 to 327) that has been identified in other viruses (8–10). No viral helicase has been identified in the available sequence. The ORF-1 stop codon (TAG) was found in multiple RT-PCR products from numerous RNA preparations of astrovirus serotype 1. These products were sequenced on both strands from cloned products and by

FIG. 3. Nucleic acid and amino acid comparisons among astrovirus serotypes 1, 2, 4, and 5 over a 600-nt region encompassing the RDRP motif (nt 2701 to 3300). Astrovirus serotype 1-specific primers were used for RT-PCR of viral RNA (forward, 5'-TCACCAATGGAAGGCGGCTT; reverse, 5'-GGAGTGATTCAAGATCAGGT). The conserved amino acids defining the motif are shown in bold. Dashes indicate nucleic acid or amino acid identity.

direct PCR sequencing. The sequence in this region was clearly readable and contained no compressions. Since this result suggested a less than obvious strategy for viral polymerase expression, presence of the stop codon was further verified by direct PCR sequencing of this region from astrovirus serotypes 2 and 5 that were available in our laboratory. Sequence comparisons over 283 nt that spanned the overlap between ORF-1 and ORF-2 revealed 93% identity between astrovirus serotypes 1 and 2, 92% identity between serotypes 1 and 5, and 95% identity between serotypes 2 and 5 (Fig. 4). This analysis confirmed the presence of the termination codon in all three

serotypes at the same position. Attempts to amplify this region in astrovirus serotype 4 were unsuccessful.

DISCUSSION

The findings in this study provide further characterization of astrovirus genomic organization and suggest possible replication strategies. Three potential ORFs were identified. Partial sequence of ORF-1, the 5'-most ORF found in this study, indicates that this ORF encodes a viral 3C-like serine protease motif. ORF-2, which does not contain an obvious start methi-

FIG. 4. Nucleic acid and amino acid comparisons of a 283-nt region overlapping ORF-1 and ORF-2 from astrovirus serotypes 1, 2, and 5. Astrovirus serotype 1-specific primers were used for RT-PCR of viral RNA (forward, 5'-ATACCGGTTGACTAAGGCAGA; reverse, 5'-AGACCAAGGAGATCGTCCCT). The suspected heptameric shift sequence and stem-loop bases are underlined. The shift sequence, A AAA AAC, conforms to the 7-nt shift sequence, X XXY YYZ. The GGGCCCC may form a stem-loop with the downstream sequence, GGGGCC. The TAG is the stop codon in ORF-1.

online, encodes the viral RDRP motif. ORF-1 and ORF-2 can be localized to a region of the viral genome that is upstream (5') of the subgenomic RNA. ORF-2 is in the -1 translation frame with respect to ORF-1. ORF-3 encodes a viral structural protein, as shown by immunoprecipitation analysis.

Two lines of evidence indicate that ORF-3 encodes the viral structural protein. In the first set of experiments, RNA from ORF-3 was transcribed and translated in vitro. The initiation codon (ATG) at nt 3501 to 3503 was selected as the most likely start site for translation of the subgenomic message on the basis of the following observations: (i) hybridizations of Northern blots with probes encompassing this region helped to approximate the location of the 5' end of the subgenomic RNA and (ii) this ATG was in a favorable context for initiation, according to Kozak's rules (22). There are two upstream, in-frame methionines that do not meet either of these criteria. The predominant product of the in vitro transcription-translation experiments was an 87-kDa protein that was immunoprecipitated with both astrovirus particle-specific hyperimmune sera and a monoclonal antibody documented to be reactive with virion.

Additional experiments confirmed that an 87-kDa protein is synthesized by serotype 1 astrovirus during infection of susceptible cells at 18 h postinfection. This protein is also recognized by the hyperimmune sera. Prior observations that astroviruses are composed of multiple structural proteins ranging in size from ~20 to 36.5 kDa suggest possible posttranslational cleavage of a single precursor protein (32). Further studies will be necessary to determine the size and number of structural proteins synthesized in an infected cell over time. ORF-3 has recently been cloned into a baculovirus transfer vector and cotransfected with baculovirus DNA. Plaque-purified recombinant viruses were identified by PCR of the viral DNA. These recombinants were shown to express an 87-kDa protein in a Western blot (immunoblot) assay with an astrovirus monoclonal antibody (data not shown). Work is in progress to characterize this product further.

ORF-2 contains the viral polymerase motif and does not hybridize to viral subgenomic RNAs from infected cells. When aligned with polymerases of other RNA viruses, the astrovirus polymerase motif is most similar to those from the poliovirus-like group (36). The conserved amino acids are identical or similar to those of other enteric and hepatic viruses, including

feline calicivirus (FCV), Norwalk virus, and hepatitis A virus (19, 33, 34). Comparison of RNA polymerase motifs is believed to be a direct way to evaluate evolutionary relatedness among different viruses. Such analysis has allowed the placement of picornaviruses and caliciviruses in viral polymerase supergroup I as described by Koonin (21). The sequence data obtained in our current study suggest that astroviruses may be yet another member of this supergroup.

A region of 600 nt encoding the RDRP motif from three of the other available human astrovirus serotypes demonstrated a high degree of homology at the nucleotide and amino acid levels (Fig. 3). Serotype 1 astrovirus is most closely related to serotype 2 and most divergent from serotype 5. Most nucleotide differences occur in the third base position of codons and do not alter the amino acid sequence. These astrovirus polymerase motif sequences may prove useful in the development of nucleic acid diagnostic reagents, including probes and primers for RT-PCR.

Thus, the astrovirus genome is organized with nonstructural proteins encoded at the 5' end and structural protein(s) encoded at the 3' end, as judged from sequence analysis of available clones, Northern blot hybridization data, and results of expression studies of ORF-3. This genomic organization is similar to that of the nonenveloped members of the family *Caliciviridae* in which the 3C protease and polymerase genes precede the ORF encoding structural proteins (19, 20, 26, 30, 34, 41). The entire sequence of astrovirus must be elucidated before the presence or absence of other protein motifs can be assessed with confidence.

The expression strategy for the production of the viral polymerase was not immediately apparent since this ORF does not contain an early methionine and is not in the same frame as ORF-1. In fact, the first in-frame ATG is located 453 nt into the ORF-2 sequence, well into the region encoding the recognizable RDRP motif, making an internal ribosome entry site unlikely. ORF-1, as presented here, also does not contain an initiation codon, but it will most likely be identified as more 5'-end sequence becomes available. ORF-2 is in the -1 translational reading frame, if ORF-1 is assumed to be in the 0 frame. One possible mechanism of translation of the polymerase gene in this situation is ribosomal frameshifting. So far, the only evidence that this strategy is used by astrovirus comes from sequence analysis. Ribosomal frameshifting is found in

FIG. 5. Predicted RNA secondary structures at the suspected astrovirus ribosomal frameshift site. The shift heptamer, A AAA AAC, is identified by a bar over the sequence, and the 0 frame stop codon is underlined.

many retroviruses and coronaviruses as a translational control mechanism for modulation of viral polymerase and/or protease expression (2, 3, 17, 18, 44). One requirement for ribosomal frameshifting is the presence of a shift sequence defined as X XXY YYZ, where X and Y may be the same nucleotide and the codons represent the 0 reading frame or, in this case, ORF-1. A second requirement is that there is potential for nearby downstream RNA sequences to form secondary structures such as stem-loops or pseudoknots. Analysis of the astrovirus sequence in the region where ORF-1 and ORF-2 overlap reveals a potential heptameric shift sequence at nt 1967 to 1973 (A AAA AAC; Fig. 5) followed by a stem-loop structure beginning 6 nt downstream. The G-C-rich regions predicted to form the stem of the stem-loop structure are indicated in Fig. 5. These sequences are highly conserved in three available astrovirus serotypes tested (serotypes 1, 2, and 5). There is 100% homology over the 43 nt potentially containing sequences critical for ribosomal frameshifting.

Two different translation products can be predicted, depending on whether ribosomal frameshifting occurs. If secondary structure does not form while the ribosome traverses the shift sequence, ORF-1 is translated completely to its termination codon at nt 2019 to 2021 and only the ORF-1 product is produced. If secondary structures form and transiently impede ribosomal translation along the shift sequence, the ribosomes may shift to the -1 position, as a result of the nature of the shift sequence, and then resume translation in the -1 frame that encodes the viral polymerase motif. In this case, the ORF-1 stop codon would not be encountered and ribosomal frameshifting would result in the production of a polyprotein that is composed of the ORF-1 product and the ORF-2 product, the viral RNA-dependent RNA polymerase. This could potentially encode a polymerase protein of ~58 kDa, depending on where proteolytic cleavage occurs to release the polymerase protein from the polyprotein. Work is currently in progress to prove this speculation as a potential strategy for modulation of astrovirus polymerase protein synthesis. A similar mechanism has been proposed for the translation of the FCV ORF-3 (35) but has not been proven to date. FCV ORF-3 encodes a protein of unknown function, in a part of the genome that is different from the polymerase-encoding region described for coronaviruses (2) and retroviruses (3) and hypothesized for astroviruses (see above). In addition, the proposed FCV shift sequence, G GAU UUA, does not conform as well to the X XXY YYZ consensus sequence (18).

The astroviruses have some features in common with the caliciviruses. Both are nonenveloped viruses with similar genome organization and a replication strategy that includes the production of a subgenomic viral RNA species. However, as astroviruses are analyzed further, several differences have emerged. These include the size and number of structural

proteins and possibly a novel mechanism for translating the polymerase that is similar to the strategy used by retroviruses and coronaviruses. The results obtained in the current study provide evidence that astroviruses have unique features that distinguish them from viruses in other viral families. These results provide support for the recent classification of astroviruses into a new family, *Astroviridae* (31).

ACKNOWLEDGMENTS

We thank Michel Bremont, Patrick Brown, Philip Dormitzer, Stephen Dunn, Kirk Fry, Jungsuh Kim, and Gregory Reyes for many helpful discussions.

This work was supported by the office of Research and Development, Department of Veterans Affairs (merit review grant to S.M.M.) and by PHS grants 2 T32 CA09302-16, awarded by the National Cancer Institute, and DK 38707, awarded to the Stanford Digestive Disease Center.

REFERENCES

1. Boga, J. A., M. S. Marin, R. Casais, M. Prieto, and F. Parra. 1992. In vitro translation of a subgenomic mRNA from purified virions of the Spanish field isolate AST/89 of rabbit hemorrhagic disease virus (RHDV). *Virus Res.* **26**:33-40.
2. Brierley, I., P. Digard, and S. C. Inglis. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**:537-547.
3. Chamorro, M., N. Parkin, and H. E. Varmus. 1992. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc. Natl. Acad. Sci. USA* **89**:713-717.
4. Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**:156-159.
5. Cruz, J. R., A. V. Bartlett, J. E. Herrmann, P. Caceres, N. R. Blacklow, and F. Cano. 1992. Astrovirus-associated diarrhea among Guatemalan ambulatory rural children. *J. Clin. Microbiol.* **30**:1140-1144.
6. Cruz, J. R., P. Caceres, F. Cano, J. Flores, A. Bartlett, and B. Torun. 1990. Adenovirus types 40 and 41 and rotaviruses associated with diarrhea in children from Guatemala. *J. Clin. Microbiol.* **28**:1780-1784.
7. Devereaux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **17**:387-395.
8. Godney, E. K., L. Chen, S. N. Kumar, S. L. Methven, E. V. Koonin, and M. A. Brinton. 1993. Complete genomic sequence and phylogenetic analysis of the lactate dehydrogenase-elevating virus (LDV). *Virology* **194**:585-596.
9. Gorbalenya, A. E., A. P. Donchenko, V. M. Blinov, and E. V. Koonin. 1989. Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases: a distinct protein superfamily with a common structural fold. *FEBS Lett.* **243**:103-114.
10. Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov. 1988. Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. *FEBS Lett.* **236**:287-290.
11. Grohmann, G. S., R. I. Glass, H. G. Pereira, S. S. Monroe, A. W. Hightower, R. Weber, and R. T. Bryan. 1993. Enteric viruses and diarrhea in HIV-infected patients. *N. Engl. J. Med.* **329**:14-20.
12. Herrington, A. J., E. W. Gray, and D. R. Snodgrass. 1981. Purification and characterization of ovine astrovirus. *J. Gen. Virol.* **53**:47-55.
13. Herrmann, J. E., R. W. Hudson, D. M. Perron-Henry, J. B. Kurtz, and N. R. Blacklow. 1988. Antigenic characterization of cell-cultivated astrovirus serotypes and development of astrovirus-specific monoclonal antibodies. *J. Infect. Dis.* **158**:182-185.
14. 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.

15. **Herrmann, J. E., D. N. Taylor, P. Echverria, and N. R. Blacklow.** 1991. Astroviruses as a cause of gastroenteritis in children. *N. Engl. J. Med.* **324**:1757–1760.
16. **Hofmann, M. A., and D. A. Brian.** 1991. A PCR-enhanced method for determining the 5' end sequence of mRNAs. *PCR Methods Appl.* **1**:43–45.
17. **Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus.** 1988. Signals for ribosomal frameshifting in the rous sarcoma virus. *Cell* **55**:447–458.
18. **Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus.** 1988. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature (London)* **331**:280–283.
19. **Jiang, X., D. Y. Graham, K. Wang, and M. K. Estes.** 1990. Norwalk virus genome cloning and characterization. *Science* **250**:1580–1583.
20. **Jiang, X., M. Wang, D. Y. Graham, and M. K. Estes.** 1992. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *J. Virol.* **66**:6527–6532.
21. **Koonin, E. V.** 1991. The phylogeny of RNA-dependent RNA polymerase of positive-strand RNA viruses. *J. Gen. Virol.* **72**:2197–2206.
22. **Kozak, M.** 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**:8125–8148.
23. **Kurtz, J. B., and T. W. Lee.** 1984. Human astrovirus serotypes. *Lancet* **ii**:1405.
24. **Kurtz, J. B., and T. W. Lee.** 1987. Astroviruses: human and animal. *CIBA Found. Symp.* **128**:92–107.
25. **Lee, T. W., and J. B. Kurtz.** 1981. Serial propagation of astrovirus in tissue culture with the aid of trypsin. *J. Gen. Virol.* **57**:421–424.
26. **Lambden, P. R., E. O. Caul, C. R. Ashley, and I. N. Clarke.** 1993. Sequence and genome organization of a human small round-structured (Norwalk-like) virus. *Science* **259**:516–519.
27. **Madeley, C. R., and B. P. Cosgrove.** 1975. Viruses in infantile gastroenteritis. *Lancet* **ii**:124.
28. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Matsui, S. M., J. P. Kim, H. B. Greenberg, L. 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.
30. **Meyers, G., C. Wirblich, and H. J. Thiel.** 1991. Rabbit hemorrhagic disease virus-molecular cloning and nucleotide sequencing of a calicivirus genome. *Virology* **184**:664–676.
31. **Monroe, S. S., B. Jiang, S. E. Stine, M. Koopmans, and R. I. Glass.** 1993. Subgenomic RNA sequence of human astrovirus supports classification of *Astroviridae* as a new family of RNA viruses. *J. Virol.* **67**:3611–3614.
32. **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.
33. **Najarian, R., D. Caput, W. Gee, S. J. Potter, A. Renard, J. Merryweather, G. Van Nest, and D. Dina.** 1985. Primary structure and gene organization of human hepatitis A virus. *Proc. Natl. Acad. Sci. USA* **82**:2627–2631.
34. **Neill, J. D.** 1990. Nucleotide sequence of a region of the feline calicivirus genome which encodes picornavirus-like RNA-dependent RNA polymerase, cysteine protease and 2C polypeptides. *Virus Res.* **17**:145–160.
35. **Neill, J. D., I. M. Reardon, and R. L. Heinrikson.** 1991. Nucleotide sequence and expression of the capsid protein gene of feline calicivirus. *J. Virol.* **65**:5440–5447.
36. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**:3867–3874.
37. **Qi, F. X., J. R. Ridpath, T. L. Lewis, S. R. Bolin, and E. S. Berry.** 1992. Analysis of the bovine viral diarrhea virus genome for possible cellular insertions. *Virology* **189**:285–292.
38. **Shaw, R. D., P. T. Vo, P. A. Offit, B. S. Coulson, and H. B. Greenberg.** 1986. Antigenic mapping of the surface proteins rhesus rotavirus. *Virology* **155**:434–451.
39. **Shepard, A. R., and N. L. Eberhardt.** 1992. A simple step to reduce background in *E. coli* transformations of blunt-ended plasmid ligation products. *BioTechniques* **13**:40–42.
40. **Shimizu, M., J. Shirai, M. Norita, and T. Yamane.** 1990. Cytopathic astrovirus isolated from porcine acute gastroenteritis in an established cell line derived from porcine embryonic kidney. *J. Clin. Microbiol.* **28**:201–206.
41. **Tam, A. W., M. M. Smith, M. Guerra, C. C. Huang, D. W. Bradley, K. E. Fry, and G. R. Reyes.** 1991. Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology* **185**:120–131.
42. **Willcocks, M. M., and M. J. Carter.** 1992. The 3' terminal sequence of a human astrovirus. *Arch. Virol.* **124**:279–289.
43. **Willcocks, M. M., M. J. Carter, F. R. Laidler, and C. R. Madeley.** 1990. Growth and characterization of human faecal astroviruses in a continuous cell line. *Arch. Virol.* **113**:73–82.
44. **Wilson, W., M. Braddock, S. E. Adams, P. D. Rathjen, S. M. Kingsman, and A. J. Kingsman.** 1988. HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. *Cell* **55**:1159–1169.