JOURNAL OF VIROLOGY, Sept. 1994, p. 5588–5595
0022-538X/94/$04.00+0

Vol. 68, No. 9

The Human Astrovirus RNA-Dependent RNA Polymerase Coding Region Is Expressed by Ribosomal Frameshifting

BEATE MARCZINKE,[1] ALISON J. BLOYS,[1] T. DAVID K. BROWN,[1] MARGARET M. WILLCOCKS,[2]† MIKE J. CARTER,[2]† AND IAN BRIERLEY[1]*

Division of Virology, Department of Pathology, University of Cambridge, Cambridge CB2 1QP,[1] and Department of Virology, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH,[2] United Kingdom

Received 15 April 1994/Accepted 27 May 1994

The genomic RNA of human astrovirus serotype 1 (HAst-1) contains three open reading frames (ORFs), 1a, 1b, and 2. ORF 1b is located downstream of, and overlaps, 1a, and it has been suggested on the basis of sequence analysis that expression of ORF 1b is mediated through −1 ribosomal frameshifting. To examine this possibility, a cDNA fragment containing the 1a-1b overlap region was cloned within a reporter gene and placed under the control of the bacteriophage SP6 promoter in a recombinant plasmid. Synthetic transcripts derived from this plasmid, when translated in the rabbit reticulocyte lysate cell-free system, specified the synthesis of polypeptides whose size and antibody reactivity were consistent with an efficient −1 ribosomal frameshift event at the overlap region. The HAst-1 frameshift signal has two essential components, a heptanucleotide slippery sequence, A_6C, and a stem-loop structure in the RNA. The presence of this structure was confirmed by complementary and compensatory mutation analysis and by direct structure probing with single- and double-stranded RNA-specific reagents. The HAst-1 frameshift signal, like that present at the overlap of the *gag* and *pro* genes of the retrovirus human T-cell lymphotrophic virus type II, does not involve the formation of an RNA pseudoknot.

Astroviruses are small (28 nm), nonenveloped viruses often associated with diarrhea in young animals or human infants (39). They possess a positive-sense, single-stranded RNA genome of about 7 kb in length. Recently, the complete genomic sequence of an isolate of human astrovirus serotype 1 (HAst-1) has been determined (37). The genome comprises 6,813 bases and contains three sequential open reading frames (ORFs) designated ORFs 1a, 1b, and 2 (Fig. 1). The 3′-most coding region, ORF 2, encodes the viral structural proteins (38) and is thought to be expressed from a subgenomic mRNA (27). ORFs 1a and 1b contain amino acid motifs indicative of nonstructural proteins, and a characteristic YGDD motif found in the RNA-dependent RNA polymerases of a variety of RNA viruses (21) is located in ORF 1b. The mechanism of expression of this ORF is uncertain, since it overlaps 1a by 71 nucleotides (nt) and is in the −1 reading frame with respect to 1a. An examination of the sequence information present within the overlap region of the two ORFs has raised the possibility that 1b is expressed as a translational fusion with the upstream 1a ORF following a −1 ribosomal frameshift event within the 1a-1b overlap region (20, 24, 37).

Several eukaryotic viruses are known to use an efficient −1 ribosomal frameshifting mechanism to control expression of their replicases (reviewed in reference 1). Ribosomal frameshifting was first described in 1985 as the mechanism by which the Gag-Pol polyprotein of the retrovirus Rous sarcoma virus is expressed from the overlapping *gag* and *pol* ORFs (19). Related frameshift signals have since been documented in an increasing number of systems including several other retroviruses, a number of eukaryotic positive-strand RNA-containing viruses, a double-stranded RNA virus of *Saccharomyces cerevisiae*, and some plant RNA viruses (reviewed in reference 14). The phenomenon is not restricted to viruses; frameshift signals of the retrovirus type have been found recently in a number of *Escherichia coli* insertion elements and in a cellular gene, the *dnaX* gene of *E. coli* (35). Work from several groups has shown that the signals for the frameshift event reside in the nucleotide sequence of the RNA around the site at which frameshifting occurs and are composed of two elements. The first is a heptanucleotide stretch containing two homopolymeric triplets and conforming to the motif XXXYYYN (where X can be A, U, or G; Y can be A or U; and N can be A, U, or C). This has been termed the slippery sequence and is the actual site where the ribosome changes into the −1 reading frame (16). The slippery sequence in itself is insufficient to promote efficient frameshifting, and additional information downstream is required. In most of the systems studied to date, this information is in the form of an RNA pseudoknot, an unusual kind of RNA structure which is formed when the nucleotides in the loop of a hairpin-loop structure base-pair with a region outside the loop forming an additional stem (30). A computer-aided study by ten Dam and colleagues (34) revealed that of 37 established or potential ribosomal frameshift sites, 26 of these had the potential to form an RNA pseudoknot structure downstream of the slippery sequence.

Analysis of the sequence of the ORF 1a-1b overlap region of human astrovirus serotype 1 (24, 37) (Fig. 1) and serotype 2 (20) has revealed a number of features suggestive of a ribosomal frameshifting signal, including a potential slippery sequence, A_6C, positioned 6 nt upstream of a 7-nt inverted repeat which could potentially fold into a GC-rich hairpin-loop structure. However, there are also a number of unusual features, including a second potential slippery sequence, GA_5C, located 8 nt upstream of the A_6C sequence. Furthermore, although several nucleotide stretches are present down-

* Corresponding author. Mailing address: Division of Virology, Department of Pathology, University of Cambridge, Tennis Court Rd., Cambridge CB2 1QP, United Kingdom. Phone: 223 336914. Fax: 223 336926. Electronic mail address: ib103@mole.bio.cam.ac.uk.

† Present address: School of Biological Sciences, University of Surrey, Guildford, Surrey GU2 5XH, United Kingdom.

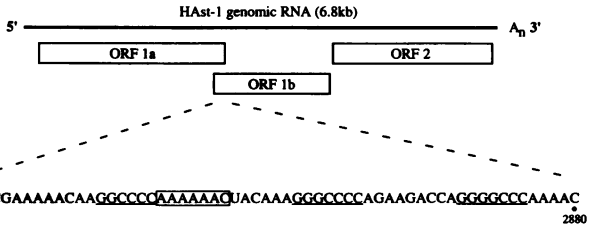

FIG. 1. Genomic organization of HAst-1. The 1a and 1b ORFs overlap by 71 nt with 1b in the −1 reading frame with respect to 1a. Features present within the overlap region which may be involved in ribosomal frameshifting are highlighted. These include the two potential slippery sequences GA_5C (shaded) and A_6C (shaded and boxed) and three GC-rich regions (underlined) which may participate in the formation of RNA secondary structures. The termination codon for the 1a ORF is located at position 2891 to 2893 (37).

stream of the hairpin loop, stretches which could potentially base-pair to nucleotides within the loop to generate an RNA pseudoknot, the first of these that would generate a pseudoknot typical of previously described frameshift sites is located over 600 nt downstream of the putative hairpin loop. Such a long-range RNA-RNA interaction would generate a pseudoknot with a loop 2 far greater in size than any previously reported in naturally occurring frameshift signals.

In this report, we demonstrate that the ORF 1a-1b overlap region of the Newcastle isolate of human astrovirus serotype 1 (37) is able to direct −1 ribosomal frameshifting, albeit at a comparatively low efficiency, with approximately 1 ribosome in 20 changing frame within the overlap region. Our results indicate that the frameshift takes place at the slippery sequence A_6C and that the GA_5C sequence does not contribute detectably. The minimal sequence required to direct frameshifting in a heterologous genetic context consists of the A_6C sequence and the downstream hairpin loop. Site-directed mutagenesis experiments and structure mapping techniques were employed to investigate the folding of the RNA at the frameshift site. Our results indicate that the astrovirus frameshift signal involves a hairpin loop rather than an RNA pseudoknot structure.

MATERIALS AND METHODS

Construction of plasmids. The basic plasmids used in this study are detailed in Fig. 2. Plasmid pTZ19R clone 23B4 was digested with *Bam*HI and *Nco*I, and a 1,509-bp DNA fragment containing the putative astrovirus 1a-1b frameshift region was isolated (HAst-1 sequence information from position 2,681 bp to 4,190 bp [37]). Following end filling with the *E. coli* DNA polymerase Klenow enzyme, the fragment was inserted into *Bgl*II-digested, end-filled plasmid pFScass 5 (5) to create plasmid pAV1. In this plasmid, the HAst-1 sequence information is inserted into the influenza A/PR8/34 PB2 gene (42) at position 483, such that the upstream PB2 sequence is in frame with the HAst-1 1a gene and is under the control of the bacteriophage SP6 promoter. Plasmid pAV1 and mutant derivatives contain the bacteriophage f1 intergenic region (9) in a nonessential region of the vector and may therefore be converted to a single-stranded form by superinfection with bacteriophage R408 (31). Plasmid pAV14 was constructed by inserting a pair of complementary synthetic oligonucleotides (5′-TCGAGCCAAAAAACTACAAAGGGCCCCAGAAGA CCAAGGGGCCCAGATAGCC-3′ [52-mer] and 5′-GGCT ATCTGGGCCCCTTGGTCTTCTGGGGCCCTTTGTAGT

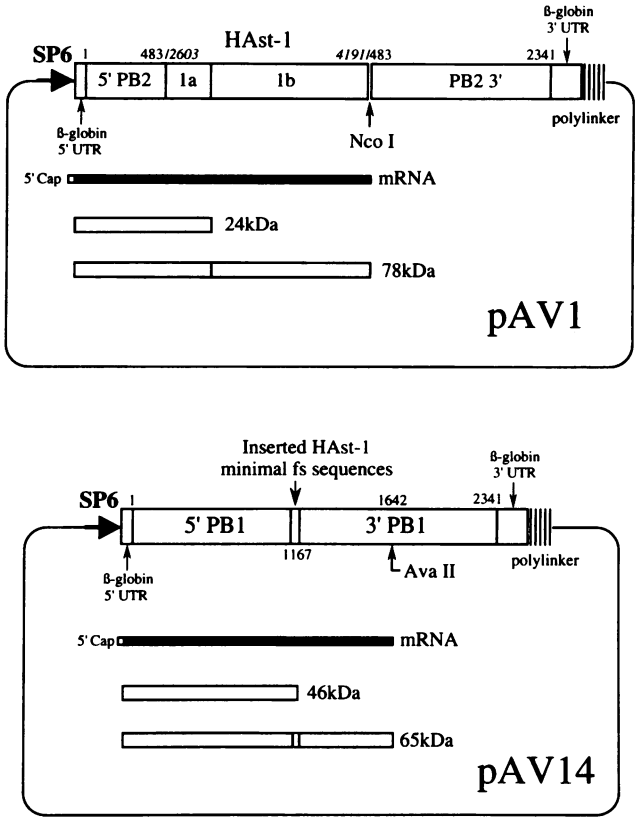

FIG. 2. The basic plasmids employed in this study. Plasmid pAV1 contains a reporter gene, the influenza virus PB2 gene, flanked by β-globin noncoding regions under the control of the bacteriophage SP6 promoter. The HAst-1 sequence information was introduced at the *Bgl*II site of the PB2 gene (position 483 [4]) such that the 5′ portion of the PB2 gene was in frame with the 1a ORF. The diagram shows the predicted sizes of protein products which would be expected following ribosomal frameshifting within the 1a-1b overlap during translation of *Nco*I-digested pAV1 template. Plasmid pAV14 contains a reporter gene, the influenza virus PB1 gene, flanked by β-globin noncoding regions under the control of the bacteriophage SP6 promoter. A pair of complementary synthetic oligonucleotides encoding the putative HAst-1 minimal frameshift region (see Materials and Methods and other text) were inserted into the *Xho*I and *Pvu*II sites of this plasmid (position 1167 [33]). The diagram shows the predicted sizes of protein products which would be expected following ribosomal frameshifting within the inserted oligonucleotides during translation of *Ava*II-digested pAV14 template.

TTTTTGGC-3′ [48-mer]) encoding the putative HAst-1 minimal frameshift region into the *Xho*I and *Pvu*II sites of the A/PR8/34 PB1 reporter gene of plasmid pPSO (33). In pAV14, ribosomes are able to translate the entire PB1 gene if a −1 frameshift occurs within the inserted oligonucleotide sequences. All plasmid junctions were confirmed by dideoxy sequencing (32) of single-stranded templates rescued from *E. coli* JM101 (41).

Site-directed mutagenesis. Site-specific mutagenesis was carried out by a procedure based on that of Kunkel (22) as described previously (4).

In vitro transcription. Plasmids were prepared for in vitro transcription as described previously (4). In vitro transcription reactions were carried out essentially as described by Melton et al. (26) and included the synthetic cap structure m^7G

(5′)ppp(5′)G (New England Biolabs) to generate capped mRNA. Product RNA was recovered by a single extraction with phenol-chloroform (1:1) followed by ethanol precipitation in the presence of 2 M ammonium acetate. The RNA pellet was dissolved in water, and remaining unincorporated nucleoside triphosphates were removed by Sephadex G-50 chromatography. RNA was recovered by ethanol precipitation, dissolved in water, and checked for integrity by electrophoresis on 1.5% agarose gels containing 0.1% sodium dodecyl sulfate (SDS). If the RNA transcripts were to be used in structure mapping experiments, the template DNA was degraded by treatment with *E. coli* RNase-free DNase I (Boehringer Mannheim; 7.5 U/μg of template DNA; 37°C, 30 min) immediately following the transcription reaction.

In vitro translation. In ribosomal frameshift assays, serial dilutions of purified mRNAs were translated in rabbit reticulocyte lysates as described previously (4). Translation products were analyzed on SDS–12.5% or SDS–15% (wt/vol) polyacrylamide gels according to standard procedures (12). The relative abundance of nonframeshifted or frameshifted products on the gels was estimated by scanning densitometry of direct autoradiographs and adjusted to take into account the differing methionine contents of the products. Scans were performed on exposures that were in the range in which film response to excitation was linear. Frameshift efficiencies were calculated from those dilutions of RNA at which translation was highly processive (RNA concentrations of 10 to 25 μg of RNA per ml of reticulocyte lysate). Radioimmunoprecipitation of translation products was as described previously (3).

RNA structure mapping. RNA structure mapping was carried out largely as described by Harrison and Lever (13), with the RNA modification conditions of Christiansen et al. (7). RNA transcripts were prepared from *Nco*I-digested pAV1. Prior to biochemical analysis, the RNA was heated at 55°C for 10 min and cooled slowly to room temperature over a 30- to 45-min period in the relevant structure mapping buffer. The following RNA modifications were made, with approximately 1 μg of RNA per reaction mixture: RNase V_1 (0 to 0.5 U) at 0°C for 30 min; RNase T_1 (0 to 40 U) at 0°C for 30 min; dimethyl sulfate (DMS) (0 to 0.3% [vol/vol]) at 30°C for 20 min; 2-keto-3-ethoxybutyraldehyde (kethoxal) (0 to 2 mg/ml) at 30°C for 20 min; and diethylpyrocarbonate (DEPC) (0 to 2.5% [vol/vol]) at 30°C for 20 min. For the RNases, the cleavage reactions were stopped by extraction with phenol-chloroform, and the modified RNA was precipitated with 3 volumes of ethanol in the presence of 0.3 M sodium acetate (pH 4.8) and 10 μg of yeast tRNA carrier. For chemical modification, the modified RNA was precipitated directly by addition of 3 volumes of ethanol and 10 μg of carrier tRNA, redissolved in water, and re-ethanol precipitated. In the case of DMS, the reaction was terminated by addition of one-fifth of a volume of DMS stop buffer (1 M Tris-acetate [pH 7.5], 1 M 2-mercaptoethanol, 1.5 M sodium acetate, 0.1 mM EDTA) prior to ethanol precipitation. For kethoxal, the first ethanol precipitation was carried out in the presence of 6 mM boric acid, to help stabilize the modified bases. Sites of RNA modification were mapped by primer extension with avian myeloblastosis virus reverse transcriptase (AMV RT) and an oligonucleotide primer complementary to HAst-1 nt 2958 to 2974. In this procedure, 0.5 pM modified RNA and 50 pM primer were annealed in a buffer containing 15 mM Tris (pH 7.5) and 25 mM KCl by heating at 70°C for 1 min and cooling slowly to 42°C. The primer was extended at 42°C in a buffer containing an additional 7.5 mM Tris (pH 8.3) (at 42°C)–30 mM KCl and also 7.5 mM $MgCl_2$–6 mM dithiothreitol–3 μM dATP–0.3 mM each dCTP, dGTP, and dTTP–2 μCi of ^{35}S-dATP (400 Ci/mmol)–5 U of RNase inhibitor (RNA Guard, Pharmacia)–5 U of AMV RT (Boehringer Mannheim). After 15 min, the dATP concentration was adjusted to 0.3 mM, and the incubation was continued for a further 15 min. Formamide was added to 30% (vol/vol), and the reaction mixture was heated to 80°C for 10 min and loaded onto an 8% acrylamide–7 M urea gel. Primer extension products were visualized by autoradiography of dried gels. RNA sequencing ladders were prepared by the dideoxy chain termination method of Sanger et al. (32) with pAV1-*Nco*I RNA template and AMV RT.

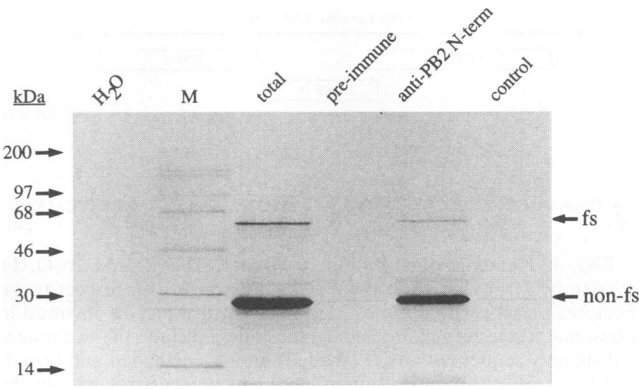

FIG. 3. Ribosomal frameshifting within the 1a-1b overlap of HAst-1. Reticulocyte lysate translation products synthesized in response to RNA transcribed from *Nco*I-digested pAV1 were analyzed directly (total) or immunoprecipitated with either a preimmune serum, a hyperimmune serum raised against the N-terminal 161 amino acids of PB2 (3), or a control antiserum (a polyclonal antiserum to the influenza PA protein). Polypeptides were labelled with [^{35}S]methionine, separated on an SDS–15% polyacrylamide gel, and detected by autoradiography. H_2O indicates a no-RNA control translation. M, molecular mass standards. Non-fs and fs indicate the nonframeshifted and frameshifted species, respectively.

RESULTS

A ribosomal frameshift signal is present at the HAst-1 1a-1b junction. We tested the possibility that the HAst-1 1b ORF was expressed by −1 ribosomal frameshifting with plasmid pAV1 (Fig. 2; see Materials and Methods). This plasmid contains the influenza virus A/PR8/34 PB2 gene (under the control of a bacteriophage SP6 RNA polymerase promoter) into which has been cloned a region of cDNA from the HAst-1 genome which includes the 1a-1b overlap region, with the 1a ORF fused in frame with the 5′ portion of the PB2 gene. The plasmid was linearized with *Nco*I and transcribed with SP6 RNA polymerase. Translation of the resulting RNA in the rabbit reticulocyte cell-free translation system was predicted to generate a 27-kDa polypeptide, the product of ribosomes which translate the upstream PB2 sequences and terminate at the HAst-1 1a termination codon, and in addition, following a −1 frameshift within the 1a-1b overlap region, a 78-kDa frameshift product. As can be seen in Fig. 3, translation of the mRNA derived from pAV1-*Nco*I gave rise to two distinct products of approximately 28 and 64 kDa. Both products could be immunoprecipitated with a polyclonal rabbit antiserum raised against the N-terminal portion of the PB1 gene (but not by preimmune serum or a hyperimmune serum raised against a different protein), confirming that the larger species is a PB2-1a-1b fusion protein produced by −1 ribosomal frameshifting. The apparent molecular weight of the nonframeshifted product was

very close to that predicted from the coding sequence, but that of the frameshift product was lower than expected (discussed below). Taking into account the relative number of methionine codons in the nonframeshifted and frameshift species (as predicted for a 78-kDa product), we estimate that 5% of ribosomes change frame within the overlap region.

Signals for frameshifting in HAst-1. Most of the viral frameshift sites studied to date have two major components, a heptanucleotide slippery sequence of the general organization XXXYYYZ and a region of mRNA secondary structure, usually in the form of an RNA pseudoknot. The HAst-1 1a-1b overlap region, however, contains two potential slippery sequences, GA_5C and A_6C, each of which is followed by a GC-rich stretch which could potentially be involved in the formation of RNA secondary structure. Although the GA_5C sequence deviates from the XXXYYYZ consensus, studies on the relative frameshift efficiency of a variety of sequences placed upstream of an RNA pseudoknot based on that of the coronavirus infectious bronchitis virus (5) have indicated that slippery sequences of the order GXXYYYZ (where X and Y are A or U) can direct efficient frameshifting, at least in the context of the infectious bronchitis virus-based pseudoknot. To determine which of the sequences is employed by HAst-1, two in-frame deletion mutants of pAV1 were prepared, in which sequence information was deleted from the beginning of the inserted HAst-1 information to either just upstream (pAV5) or just downstream (pAV6) of the A_6C motif. Synthetic mRNA transcribed from these plasmids was tested for frameshifting by in vitro translation (Fig. 4). In each case, a nonframeshifted product of reduced size was observed, consistent with the introduction of the deletion into the constructs. Frameshifting was, however, seen only in construct pAV5, where it occurred at the wild-type level. In pAV6, frameshifting was abolished. These observations support the idea that frameshifting in HAst-1 occurs at the A_6C motif, since the GA_5C motif can be removed without apparent effect (in pAV5), yet removal of the A_6C motif (in pAV6) abolishes frameshifting. Furthermore, we found that changing the sequence of the putative slip-site from A_6C to AAAAGAC in another mutant construct (pAV3) was also inhibitory to frameshifting (Fig. 4). It is well documented that mutations in that portion of the slippery sequence which would be decoded in the ribosomal aminoacyl site prior to tRNA slippage (i.e., XXXYYYZ) are greatly inhibitory to the frameshift process (4, 5, 8, 16). The demonstration that pAV5 can direct efficient frameshifting indicates that the particular GC-rich stretch immediately following the GA_5C sequence is unlikely to be a component of an RNA secondary structure involved in the frameshift process. Thus, the likely site of the ribosomal frameshift and the 5′ boundary of the frameshift signal in HAst-1 is the A_6C sequence.

Several authors have noted that downstream of the A_6C sequence is a potential hairpin-loop structure (20, 24, 37). It is likely that this structure is involved in astrovirus frameshifting since an A_6C stretch alone is insufficient to promote high-level frameshifting in the reticulocyte lysate system (18) and, furthermore, the first base (G) of the potential hairpin loop is located 6 nt downstream of the A_6 sequence, a distance typical for a *cis*-acting RNA structure involved in frameshifting (34). We tested the importance of this region by creating a series of complementary and compensatory base changes within the predicted arms of the stem by mutagenesis and tested for frameshifting. In pAV7 and 8, in which the three bases at the bottom of arms 1 and 2, respectively, were changed to their complementary nucleotides, frameshifting was reduced some four- to fivefold, as would be expected if the stem contributed to the frameshift process. In the double mutant construct,

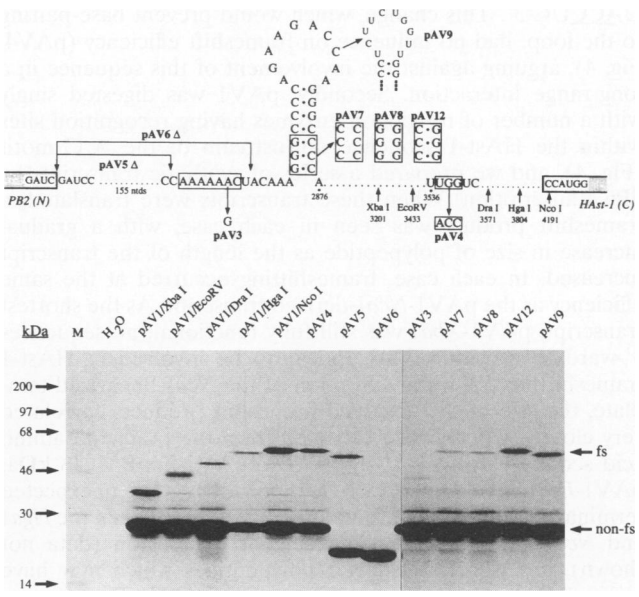

FIG. 4. Signals for ribosomal frameshifting in HAst-1. The upper panel shows the position of specific deletions and point mutations created within the 1a-1b junction region of pAV1 by site-directed mutagenesis. The diagram shows part of the sequence of the RNA around the frameshift site, including the suspected slippery sequence A_6C (boxed), the predicted stem-loop structure, and a potential pseudoknot-forming sequence, UUGGUC, located downstream (see text). In pAV5 and pAV6, the region between the arrowheads was deleted (Δ), but the 1a reading frame was maintained. In pAV9, the sequence of the loop was changed to the complementary sequence, but the remaining sequences were unchanged. Relevant restriction enzyme cut sites in the parental plasmid pAV1 are marked with their genomic position (37). The lower panel shows reticulocyte lysate translation products synthesized in response to RNA transcribed from *Xba*I-, *Eco*RV-, *Dra*I-, *Hga*I-, or *Nco*I-digested pAV1 template and, in addition, from *Nco*I-digested mutant derivatives of pAV1. Polypeptides were labelled with [^{35}S]methionine, separated on an SDS–15% polyacrylamide gel, and detected by autoradiography. H_2O indicates a no-RNA control translation. M, molecular mass standards. Non-fs and fs indicate the nonframeshifted and frameshifted species, respectively.

pAV12, in which both changes were made and the hairpin-loop structure was restored, frameshifting was at the wild-type level. This supports the idea that a hairpin forms downstream of the HAst-1 slippery sequence and is essential for efficient frameshifting. As the majority of RNA structures associated with frameshift sites are RNA pseudoknots, we wished to test whether the HAst-1 hairpin is itself part of a higher-order structure. We examined the RNA sequence downstream of the hairpin for contiguous stretches of 5 or more nt which could potentially base-pair with nucleotides present in the loop of the stem-loop and form the second stem of an RNA pseudoknot. Within the region of HAst-1 cloned into pAV1, only two candidate sequences were found, namely, 5′ UUGGUC 3′, located 670 nt downstream of the hairpin loop (genomic position 3536 to 3541), and 5′ UUGGUU 3′, some 1,000 nt distant (genomic position 3864 to 3869). Although these downstream sequences are considerably more distant from the hairpin loop than any yet seen in RNA pseudoknots involved in frameshifting, we tested their possible involvement in two ways. Firstly, site-directed mutagenesis was employed to change the sequence of the first (and most stable) potential pseudoknot forming the sequence from 5′ UUGGUC 3′ to 5′

UACCUC 3′. This change, which would prevent base-pairing to the loop, had no influence on frameshift efficiency (pAV4, Fig. 4), arguing against the involvement of this sequence in a long-range interaction. Secondly, pAV1 was digested singly with a number of restriction enzymes having recognition sites within the HAst-1 sequence downstream of the A_6C motif (Fig. 4), and we prepared a series of synthetic transcripts by SP6 transcription. When these transcripts were translated, a frameshift product was seen in each case, with a gradual increase in size of polypeptide as the length of the transcript increased. In each case, frameshifting occurred at the same efficiency as the pAV1-*Nco*I-derived transcript. As the shortest transcript, pAV1-*Xba*I was still fully functional; no sequences 3′-wards of position 3203 appear to be involved in HAst-1 frameshifting. With the exception of the *Nco*I-linearized template, the size of the observed frameshift products correlated very closely with the size expected from the predicted amino acid sequence (pAV1-*Xba*I, 39 kDa; pAV1-*Eco*RV, 48 kDa; pAV1-*Dra*I, 54 kDa; pAV1-*Hga*I, 63 kDa). No unexpected termination codons were found to be present between the *Hga*I and *Nco*I sites upon resequencing of this region (data not shown), nor were any shorter polypeptides which may have arisen from proteolytic processing of the pAV1-*Nco*I frameshift product apparent on the gels. Thus, it is not clear why the pAV1-*Nco*I frameshift product has a lower apparent molecular weight than that predicted. It may well be that the protein simply migrates anomolously in SDS-polyacrylamide gels.

Since no convenient restriction sites closer to the stem-loop were present, we tried an alternative strategy to test whether the nucleotides in the loop of the hairpin loop were base-paired by changing the sequence of the loop to its complementary sequence (5′ TCTTCTGGTT 3′, pAV9). If the loop is base-paired, then by changing its sequence composition, any base-pairing ought to be abolished. However, this change was found to have no effect on the frameshift efficiency of pAV9 transcripts, supporting the idea that the loop nucleotides are unpaired and also indicating that the primary sequence of the loop nucleotides is unimportant in HAst-1 frameshifting. This suggested that HAst-1 frameshifting requires only the A_6C slippery sequence and the downstream hairpin loop. To test this, a pair of complementary synthetic oligonucleotides containing these sequences were inserted into the frameshift reporter plasmid pPSO (to create pAV14 [Fig. 2]) and frameshifting was examined by in vitro transcription and translation following digestion of the plasmid with *Ava*II (Fig. 5). In pAV14, the sizes of the predicted nonframeshifted and frameshift products are, respectively, 46 and 65 kDa. As can be seen in Fig. 5, both products were observed, and after correcting for the relative methionine content of the two species, the frameshift efficiency was found to be 5%. A variant of pAV14 was also prepared (pAV13), in which the tyrosine codon (UAC) immediately downstream of the A_6C sequence was changed to a termination codon (UAG). In pAV13, the nonframeshift product was somewhat smaller than that of pAV14, since 1a termination occurs earlier on the mRNA, but the frameshift efficiency was unchanged. In this construct, frameshifting must take place upstream of the UAG codon and almost certainly occurs at the A_6C sequence. As these experiments were carried out in a heterologous genetic context (the PB1 gene), the signals for HAst-1 frameshifting are contained within the inserted oligonucleotides, supporting the idea that the only requirements are the A_6C sequence and the hairpin loop downstream.

Structure of the RNA at the HAst-1 frameshift site. The mutagenesis experiments described above provide evidence that HAst-1 frameshifting occurs at an A_6C stretch and that a downstream stem-loop structure plays a key role in the process. These studies, although supportive, do not rule out formally the possibility that a pseudoknot forms and is essential for frameshifting in HAst-1, since in the complementary loop construct pAV9 and the minimal HAst-1 frameshift constructs pAV13 and -14, we could not exclude the possibility that chance base-pairing had occurred between the loop and regions elsewhere. We decided, therefore, to probe the structure of the RNA in this region by biochemical methods (10). For this analysis, the 2-kb pAV1-*Nco*I SP6-derived transcript was employed. In the first step, RNA was either untreated or subjected to limited RNase hydrolysis or chemical modification with structure-specific probes. After the reaction was stopped, the RNA was phenol-chloroform extracted and hybridized with an oligonucleotide complementary to HAst-1 nt 2958 to 2974, which was used as a primer for AMV RT. In treated samples, enzymatic cleavage or chemical modifications cause RT to terminate or pause as it moves along the template, generating a series of cDNAs whose lengths define the sites of cleavage-modification. Untreated controls are included to identify those cDNA fragments which arise from natural RT termination or pausing sites in the RNA (strong stops), which can arise from RNA nicks, secondary structures, or sequence-dependent termination sites. In order to determine the sites of cleavage-modification produced by the various treatments, the labelled primer extension products were analyzed on 8% denaturing polyacrylamide gels alongside size markers prepared by dideoxy sequencing of pAV1-*Nco*I RNA with AMV RT. A representative selection of the individual biochemical analyses is shown in Fig. 6, panels A, B, and C, and a diagrammatic summary is shown in panel D. All modifications were mapped in at least two independent experiments. The single-strand-specific probes employed were RNase T_1, which cleaves internucleotide bonds 3′ of unpaired G residues; kethoxal, which reacts with unpaired G residues at N-1 and N-2; DMS, which methylates unpaired A at N-1 and unpaired C at N-3; and DEPC, which reacts with the N-7 group of unpaired A residues (and may also react with unpaired G, C, and U under certain conditions of pH and ionic strength). To

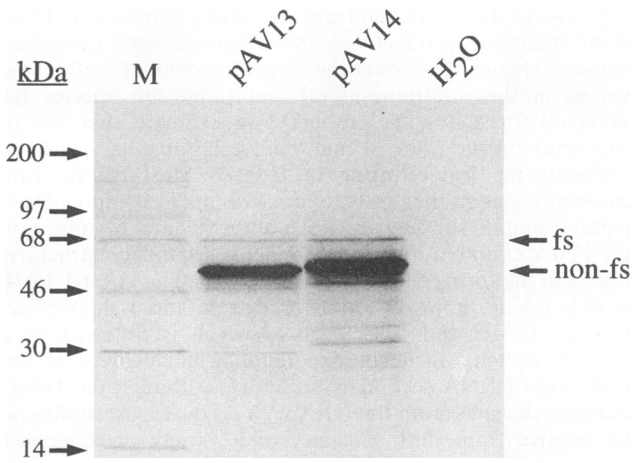

FIG. 5. The figure shows reticulocyte lysate translation products synthesized in response to RNA transcribed from *Ava*II-digested pAV13 or pAV14. Polypeptides were labelled with [^{35}S]methionine, separated on an SDS–12.5% polyacrylamide gel, and detected by autoradiography. H_2O indicates a no-RNA control translation. M, molecular mass standards. Non-fs and fs indicate the nonframeshifted and frameshifted species, respectively.

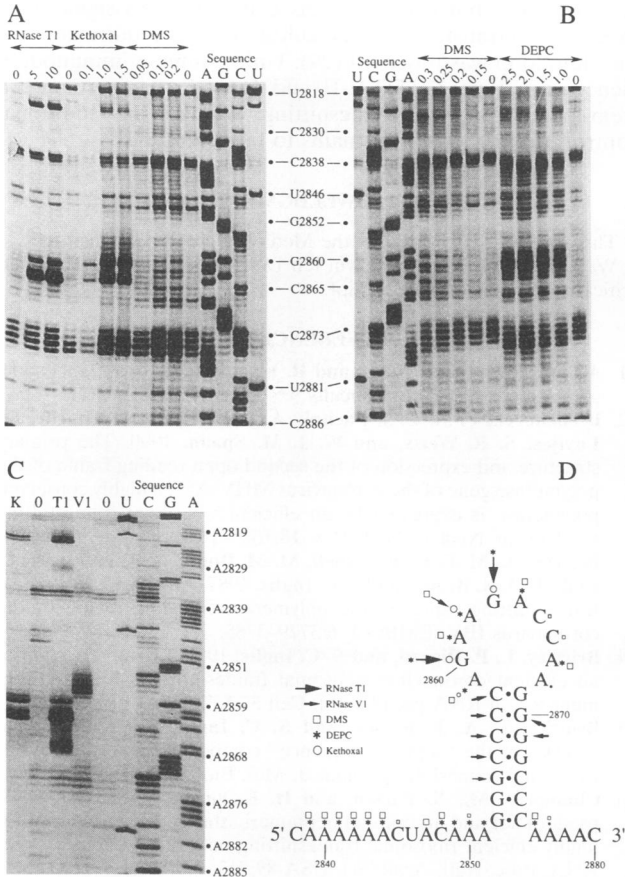

FIG. 6. Structure of the RNA at the HAst-1 frameshift site. A pAV1-*Nco*I SP6-derived mRNA transcript was subjected to limited enzymatic or chemical probing, and the sites of modification were mapped by primer extension with AMV RT as described in Materials and Methods and in other text. Panels A, B, and C show representative examples of primer extension products analyzed on 8% acrylamide–7 M urea gels. Uniquely modified nucleotides were identified by their absence in untreated control lanes (0). Modification results in the cessation of AMV RT elongation at the nucleotide immediately preceding the modified position. Treatments were as follows: panel A, RNase T_1 (units per microgram of RNA), kethoxal (milligrams per milliliter), and DMS (% [vol/vol]); panel B, DMS (% [vol/vol]) and DEPC (% [vol/vol]); and panel C, K (kethoxal, 1 mg/ml), T_1 (RNase T_1, 10 U/μg of RNA), and V1 (RNase V_1, 0.25 U/μg of RNA). The results of the analysis are summarized in panel D. Smaller symbols indicate a weaker reactivity with a particular nucleotide.

probe double-stranded regions, RNase V_1 was employed, which cleaves internucleotide bonds in helical regions. RNase V_1 is not base specific; it requires 4 to 6 nt that are in helical conformation, whether base paired or single stranded and stacked. Bands in the untreated control lanes derived from natural stops and included a number which mapped to residues at the base of the predicted stem (C-2873 to 2875). Such stops in untreated samples were not seen with the corresponding bases of the other arm of the predicted stem (G-2852 to 2854), and this is probably related to the fact that the C's are encountered first by the AMV RT. Perhaps a proportion of RT molecules pause or terminate as they encounter the stem (at the C's), but those RT molecules which do not pause have no difficulty in reverse-transcribing the other arm of the stem, since the stem is already unwound. The presence of these bands has hampered our interpretation of how well this part of the putative stem is recognized by the modifying reagents, but despite this, the study has provided clear biochemical evidence that a stem-loop structure is present downstream of the HAst-1 frameshift site and that the loop is unpaired. With respect to the single-strand-specific reagents, RNase T_1 did not cleave the G residues within the predicted stem region (G-2852 to 2854, G-2869 to 2872) but reacted strongly with the two loop G's (G-2860 and G-2863) (Fig. 6A), and similarly kethoxal reacted with the two loop G's but not with the stem G's. DMS modified all the loop A's (except A-2868) and C's (weakly) but did not modify the stem C's. DEPC reacted with the loop A (A-2868 weakly) and G residues (G-2860 and G-2863) but did not show detectable reactivity with the stem G's. The double-strand-specific reagent RNase V_1 was less useful in the structure mapping because of the natural stops present in arm 2 of the stem (see particularly Fig. 6A and B) but was noticeably unreactive against loop nucleotides with the exception of a weak signal from A-2859, the first loop base. Additionally, the enzyme cleaved with specificity in the stem region, at (almost exclusively) C residues in arm 1 and at G-2870 and G-2871 at the top of arm 2. There were a number of unusual reactivities. In occasional RNase T_1 digests (Fig. 6C), a rather blurred band appeared which corresponded approximately to A-2868 and G-2869, suggesting occasional reactivity at the top of the stem, but in most reactions, this blurred band was not seen (Fig. 6A) and may be an artifact. Two additional and unusual modifications were reproducibly seen, namely, at A-2862 with RNase T_1 and at A-2859 and A-2862 with kethoxal. We do not know why G-specific reagents should show reactivity with certain loop A's, but in each case, the reactive bases are adjacent to G residues, perhaps allowing some nonspecific reactions to occur. In general, however, the results of the structure mapping experiments support strongly the conclusions of the mutagenesis studies that the HAst-1 frameshift signal includes a stem-loop structure but not an RNA pseudoknot.

DISCUSSION

Sequencing studies of human astrovirus serotypes have revealed that the virus genome contains three ORFs, 1a, 1b, and 2, and it has been suggested that expression of the 1b ORF occurs by a −1 ribosomal frameshifting mechanism (20, 24, 37). We have tested this hypothesis by cloning the 1a-1b overlap region from HAst-1 into a heterologous reporter gene construct and tested for frameshifting by translating synthetic mRNAs derived from this plasmid in the rabbit reticulocyte lysate in vitro translation system. Our results show that frameshifting in HAst-1 occurs with an efficiency of about 5%; astroviruses are therefore a member of the growing list of positive-strand RNA viruses which utilize high-efficiency ribosomal frameshifting as a mechanism of gene expression. If the frameshift efficiency of 5% determined in vitro is reflected in virus-infected cells, then on the basis of the nucleotide sequence of the 1a and 1b ORFs (37), we would predict the synthesis of a 108-kDa 1a product and a 169-kDa 1a-1b fusion product in a ratio of approximately 20:1.

The HAst-1 frameshift signal has two components, a heptanucleotide slippery sequence, A_6C-, and a stem-loop structure in the RNA. The presence of the hairpin loop was confirmed by complementary and compensatory mutation analysis and by direct structure probing with single- and double-strand-specific chemical and enzymatic reagents. A potential slippery sequence, GA_5C, located upstream of the A_6C sequence is not required for frameshifting, and the 1a termination codon downstream of the hairpin loop is also

uninvolved. Indeed, there do not appear to be any sequence requirements other than the slip sequence and hairpin loop; a 37-nt stretch containing just these components is entirely sufficient for frameshifting in a heterologous genetic context. From the mutational analysis described, and from previous studies of the process, it is highly likely that the A_6C sequence is the actual site of the frameshift, with the ribosome-bound peptidyl-$tRNA^{Lys}$ and aminoacyl-$tRNA^{Asn}$ slipping simultaneously from A-AAA-AAC in the zero reading frame back to AAA-AAA in the −1 frame during the frameshift as predicted from the model of Jacks and colleagues (16).

In a computer analysis of documented or predicted ribosomal frameshift signals, ten Dam and colleagues (34) identified some 37 gene overlaps, of which 6 were suspected to employ the slippery sequence A_6C. Of these, two systems have been investigated in detail, namely, those at the *gag-pro* overlap regions of the retroviruses mouse mammary tumor virus (6, 18, 28) and human T-cell leukemia virus type II (HTLV-II [11, 25]). Site-directed mutagenesis studies have demonstrated a requirement for an RNA pseudoknot structure in mouse mammary tumor virus *gag-pro* frameshifting, but in HTLV-II, a simpler GC-rich hairpin loop appears to be both necessary and sufficient for frameshifting in vitro. In both systems, the stability of the respective structures is important in frameshifting. In mouse mammary tumor virus, mutations predicted to destabilize greatly either of the two stem components of the pseudoknot reduce the frameshift efficiency in vitro by about 10-fold (from 20% to 2%). Similarly, in HTLV-II, mutations which reduced the stability of the predicted hairpin loop from −16.4 to −10.5 kcal (−68.6 to −43.9 kJ)/mol) (11) (calculated according to the method of Turner et al. [36]) abolished frameshifting at the *gag-pro* overlap. As evidenced from the mutagenesis experiments and direct RNA structure probing described here, the HAst-1 stem-loop does not fold into an RNA pseudoknot structure and thus more closely resembles the frameshift signal of HTLV-II. The frameshift efficiency of the *gag-pro* signal of HTLV-II, estimated to be around 13% in reticulocyte lysates (25), is somewhat greater than that measured here for the HAst-1 signal (5%), but this is consistent with the lower predicted stability of the HAst-1 stem-loop (−13.6 kcal [−56.9 kJ]/mol).

What sets the level of frameshifting at a particular site is still far from clear. Experiments in which the slippery sequence of the coronavirus infectious bronchitis virus (UUUAAAC) was replaced by a wide variety of naturally occurring slippery sequences have provided some evidence that an RNA pseudoknot can enhance the level of frameshifting seen for a particular slip sequence (5). However, among those sites known, or suspected, to involve pseudoknots, there is considerable variation in frameshift efficiency, from the relatively low-level shifts of the yeast double-stranded L-A virus (1.8% [8]) and Rous sarcoma virus (5% [16]) to the higher-level shifts (30%) of the coronaviruses (2, 3, 15, 23). It seems likely, therefore, that a combination of factors will be involved in setting the frameshift efficiency, including the type of slippery sequence, the distance between the slippery sequence and the downstream RNA structure, and the nature of this structure in terms of size, stability, and complexity. Recent work with the frameshift signal at the *gag-pol* overlap of the human immunodeficiency virus type 1 (29) has highlighted an additional complication, namely, the possibility that differences may be seen between frameshift signals studied in vitro and in vivo. The *gag-pol* overlap of human immunodeficiency virus type 1 contains a U_6A slippery sequence and a stable hairpin loop downstream (17). Translation studies have indicated only a modest involvement of this structure during frameshifting in vitro (29, 40), but when mRNAs containing this signal were expressed in transfected tissue culture cells, a requirement for the stem-loop was observed (29). Unlike human immunodeficiency virus type 1, HTLV-II and HAst-1 clearly require a stem-loop for in vitro frameshifting. Whether any additional controls operate in vivo remains to be determined.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council U.K.

We would like to thank Edwin ten Dam for useful comments and critical reading of the manuscript.

REFERENCES

1. **Atkins, J. F., R. B. Weiss, and R. F. Gesteland.** 1990. Ribosome gymnastics—degree of difficulty 9.5, style 10.0. Cell **62**:413–423.
2. **Bredenbeek, P. J., C. J. Pachuk, A. F. H. Noten, J. Charité, W. Luytjes, S. R. Weiss, and W. J. M. Spaan.** 1990. The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. Nucleic Acids Res. **18**:1825–1832.
3. **Brierley, I., M. E. G. Boursnell, M. M. Binns, B. Bilimoria, V. C. Blok, T. D. K. Brown, and S. C. Inglis.** 1987. An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J. **6**:3779–3785.
4. **Brierley, I., P. Digard, and S. C. Inglis.** 1989. Characterisation of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell **57**:537–547.
5. **Brierley, I., A. J. Jenner, and S. C. Inglis.** 1992. Mutational analysis of the "slippery sequence" component of a coronavirus ribosomal frameshifting signal. J. Mol. Biol. **227**:463–479.
6. **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.
7. **Christiansen, J., J. Egebjerg, N. Larsen, and R. A. Garrett.** 1990. Analysis of rRNA structure: experimental and theoretical considerations, p. 229–252. *In* G. Spedding (ed.), Ribosomes and protein synthesis, a practical approach. Oxford University Press, Oxford.
8. **Dinman, J. D., T. Icho, and R. B. Wickner.** 1991. A −1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. Proc. Natl. Acad. Sci. USA **88**:174–178.
9. **Dotto, G. P., V. Enea, and N. D. Zinder.** 1981. Functional analysis of bacteriophage f1 intergenic region. Virology **114**:463–473.
10. **Ehresmann, C., F. Baudin, M. Mougel, P. Romby, J.-P. Ebel, and B. Ehresmann.** 1987. Probing the structure of RNAs in solution. Nucleic Acids Res. **15**:9109–9128.
11. **Falk, H., N. Mador, R. Udi, A. Panet, and A. Honigman.** 1993. Two *cis*-acting signals control ribosomal frameshift between human T-cell leukemia virus type II *gag* and *pro* genes. J. Virol. **67**:6273–6277.
12. **Hames, B. D.** 1991. An introduction to polyacrylamide gel electrophoresis, p. 1–91. *In* B. D. Hames and D. Rickwood (ed.), Gel electrophoresis of proteins—a practical approach. IRL Press, Oxford.
13. **Harrison, G. P., and A. M. L. Lever.** 1992. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. J. Virol. **66**:4144–4153.
14. **Hatfield, D. L., J. G. Levin, A. Rein, and S. Oroszlan.** 1992. Translational suppression in retroviral gene expression. Adv. Virus Res. **41**:193–239.
15. **Herold, J., T. Raabe, B. Schelle-Prinz, and S. G. Siddell.** 1993. Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. Virology **195**:680–691.
16. **Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus.** 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell **55**:447–458.
17. **Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus.** 1988. Characterisation of ribosomal frameshifting in HIV-1 *gag-pol* expression. Nature (London) **331**:280–283.

18. **Jacks, T., K. Townsley, H. E. Varmus, and J. Majors.** 1987. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. Proc. Natl. Acad. Sci. USA **84:**4298–4302.
19. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science **230:** 1237–1242.
20. **Jiang, B. M., S. S. Monroe, E. V. Koonin, S. E. Stine, and R. I. Glass.** 1993. RNA sequence of astrovirus: distinctive genome organisation and a putative retrovirus-like ribosomal frameshifting signal that directs the viral replicase synthesis. Proc. Natl. Acad. Sci. USA **90:**10539–10543.
21. **Kamer, G., and P. Argos.** 1984. Primary structural comparison of RNA-dependent RNA polymerases from plant, animal and bacterial viruses. Nucleic Acids Res. **12:**7269–7282.
22. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA **82:**488–492.
23. **Lee, H.-J., C.-K. Shieh, A. E. Gorbalenya, E. V. Koonin, N. La Monica, J. Tuler, A. Bagdzhadzhyan, and M. M. C. Lai.** 1991. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology **180:**567–582.
24. **Lewis, T. L., H. B. Greenberg, J. E. Herrmann, L. S. Smith, and S. M. Matsui.** 1994. Analysis of astrovirus serotype 1 RNA, identification of the viral RNA-dependent RNA polymerase motif and expression of a viral structural protein. J. Virol. **68:**77–83.
25. **Mador, N., A. Panet, and A. Honigman.** 1989. Translation of *gag*, *pro*, and *pol* gene products of human T-cell leukemia virus type 2. J. Virol. **63:**2400–2404.
26. **Melton, D. A., P. A. Krieg, M. R. Robagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridisation probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:** 7035–7056.
27. **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.
28. **Moore, R., M. Dixon, R. Smith, G. Peters, and C. Dickson.** 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events required for translation of *gag* and *pol*. J. Virol. **61:**480–490.
29. **Parkin, N. T., M. Chamorro, and H. E. Varmus.** 1992. Human immunodeficiency virus type 1 *gag-pol* frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. J. Virol. **66:**5147–5151.
30. **Pleij, C. W. A., and L. Bosch.** 1989. RNA pseudoknots: structure, detection and prediction. Methods Enzymol. **180:**289–303.
31. **Russel, M., S. Kidd, and M. R. Kelley.** 1986. An improved filamentous helper phage for generating single-stranded plasmid DNA. Gene **45:**333–338.
32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
33. **Somogyi, P., A. J. Jenner, I. Brierley, and S. C. Inglis.** 1993. Ribosomal pausing during translation of an RNA pseudoknot. Mol. Cell. Biol. **13:**6931–6940.
34. **ten Dam, E. B., C. W. A. Pleij, and L. Bosch.** 1990. RNA pseudoknots. Translational frameshifting and readthrough on viral RNAs. Virus Genes **4:**121–136.
35. **Tsuchihashi, Z.** 1991. Translational frameshifting in the *Escherichia coli* dna X gene *in vitro*. Nucleic Acids Res. **19:**2457–2462.
36. **Turner, D. H., N. Sugimoto, and S. M. Freier.** 1988. RNA structure prediction. Annu. Rev. Biophys. Chem. **17:**167–192.
37. **Willcocks, M. M., T. D. K. Brown, C. R. Madeley, and M. J. Carter.** The complete sequence of a human astrovirus. J. Gen. Virol., in press.
38. **Willcocks, M. M., and M. J. Carter.** 1993. Identification and sequence determination of the capsid protein gene of human astrovirus serotype 1. FEMS Microbiol. Lett. **114:**1–8.
39. **Willcocks, M. M., M. J. Carter, and C. R. Madeley.** 1992. Astroviruses. Rev. Med. Virol. **2:**97–106.
40. **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.
41. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
42. **Young, J. F., U. Desselberger, P. Graves, P. Palese, A. Shatzman, and M. Rosenberg.** 1983. Cloning and expression of influenza virus genes, p. 129–138. *In* W. G. Laver (ed.), The origin of pandemic influenza viruses. Elsevier Science, Amsterdam.