Molecular cloning and characterization of hepatitis A virus cDNA

(hepatitis A virus RNA/picornavirus)

JOHN R. TICEHURST*, VINCENT R. RACANIETTOL, BAHIGE M. BAROUDY*, DAVID BALTIMORE†, ROBERT H. PURCELL‡, and STEPHEN M. FEINSTONE*

*Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20205; and †Whitehead Institute for Biomedical Research, Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Communicated by Robert M. Chanock, June 30, 1983

ABSTRACT Double-stranded cDNA was synthesized from hepatitis A virus (HAV) RNA and inserted into the Pst I site of pBR322. Restriction endonuclease digestion and cross-hybridization of fragments yielded a map of overlapping cloned cDNAs that included at least 99% of the viral genome. Molecular clones containing HAV cDNA were identified by hybridizing cloned cDNA to electrophoretically resolved RNA from uninfected and HAV-infected tissue culture cells. Cloned cDNA probes specifically hybridized to RNA from infected cells, and the predominant species identified had the characteristic genomic length of picornaviral RNA (≈7,500 nucleotides). A partial sequence from the 3' end of the genome revealed 414 bases in an open reading frame followed by two closely-spaced stop codons, a 68-base noncoding region, and a tract of poly(A).

HAV labeled in tissue culture. We report here the cloning of cDNA representing at least 99% of the genome of HAV and partial analysis of its sequence.

MATERIALS AND METHODS

Source and Propagation of HAV. The HM-175 strain of HAV was isolated from a family outbreak in Australia (11). HAV that had been passaged twice in marmosets (Saguinus mystax and Saguinus labiatus) (11) was inoculated into eight marmosets. A 20% (wt/vol) suspension of HAV-infected liver in phosphate-buffered saline was injected intravenously, and all animals were monitored by serum enzymes and immunofluorescence of liver biopsy (12). When maximum immunofluorescence was reached 10–14 days after inoculation, the marmosets were killed and their livers were immediately removed, minced, frozen in liquid N₂, and stored at −70°C until virus purification.

HAV was also passaged 6 times in marmosets and 19 times in secondary African green monkey kidney (AGMK) monolayers (11), which were used as a source of RNA for hybridization (see below). Infected cultures were harvested at 21 days when 100% of the cells exhibited maximal immunofluorescence.

Virus Purification. Minced marmoset livers (total, 170 g) were homogenized, extensively extracted with trichlorotrifluoroethane and chloroform, and treated with micrococcal nuclease (13). Virus was monitored by radioimmunoassay (14) and by electron microscopy during successive sucrose, CsCl, and sucrose gradients.

RNA Extraction and Characterization. Suspensions of purified HAV were incubated at 57°C for 15 min with 500 µg of proteinase K per ml, after which NaDODSO₄ was added to a concentration of 0.5%, and incubation was continued for an additional 30 min. After extraction with phenol and then with chloroform/isomyl alcohol (24:1, vol/vol), RNA was precipitated in ethanol, redissolved, and, after removal of a portion for analysis, reprecipitated. RNA was characterized by UV spectrophotometry and agarose gel electrophoresis after denaturation with 1 M glyoxal/50% dimethyl sulfoxide (vol/vol) (15). Other portions were used to analyze template quality and optimal conditions for cDNA synthesis as described in Results and Fig. 1.

RNA for hybridization studies was prepared from uninfected and HAV-infected (see above) AGMK cells by isolation of total cytoplasmic RNA (16). Polyadenylated RNA was isolated by oligo(dT)-cellulose chromatography (17). RNA was transferred to nitrocellulose paper after electrophoretic separation through agarose gels containing glyoxal/dimethyl sulfoxide-denatured RNA (18).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HAV, hepatitis A virus; ds cDNA, double-stranded cDNA; kb, kilobase(s).
Preparation of cDNA Clones. HAV RNA (0.8 μg) derived from marmoset liver served as a template for cDNA synthesis using reverse transcriptase (120 units/ml) for 30 min at 37°C in 160 μl that contained 50 mM Tris-HCl (pH 8.3)/10 mM MgCl₂/100 mM KCl/500 μM dATP/[α-32P]dTTP (0.025 Ci/mmol); 1 Ci = 3.7 × 10^10 Bq/dGTP/TPP/1 mM dithiothreitol/4 mM sodium pyrophosphate/[dTTP]₉₋₁₉ at 30 μg/ml/ RNasin (2,000 units/ml). After addition of EDTA to 20 mM, RNA-cDNA hybrids were isolated by phenol extraction, column chromatography, and ethanol precipitation. The RNA template was hydrolyzed in 0.3 M NaOH/0.7 M NaCl/5 mM EDTA for 2 hr at 37°C (8).

By using a modification of conditions optimized for poliovirus cDNA template (8), the second strand of cDNA was synthesized for 30 min at 37°C using the large (Klenow) fragment of Escherichia coli DNA polymerase I (28 units/ml) in 10 mM Tris-HCl (pH 7.5)/5 mM MgCl₂/5 mM dithiothreitol/30 μM dATP/[α-32P]dTTP (0.45 Ci/mmol)/dGTP/TPP; and cDNA at 1 μg/ml.

After phenol extraction, column chromatography, and ethanol precipitation, double-stranded cDNA (ds cDNA) was digested for 1 hr at 37°C using nuclease S1 at 10 units/ml (0.1 unit per ng of ds cDNA) in 30 mM NaOAc (pH 4.5)/0.3 M NaCl/3 mM ZnCl₂/5% glycerol (vol/vol), followed by addition of EDTA to 35 mM, phenol and ether extraction, and dialysis against 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA (8).

Homopolymer tails of dCMP were added to ds cDNA using terminal deoxynucleotidyl transferase (250 units/ml) for 20 min at room temperature in 100 μl that contained 0.14 M potassium cacodylate (pH 7.2)/1 mM CoCl₂/0.2 mM dithiothreitol/nuclease-free bovine serum albumin (500 μg/ml)/200 μM dCTP (8). After phenol extraction, 50% of the ds cDNA was ether-extracted and precipitated with ethanol.

The remaining tailed ds cDNA was applied to a 3-ml column of Sepharose 4B in 20 mM Tris-HCl (pH 8.0)/0.6 M NaCl/2 mM EDTA (16). The first five 65-μl fractions containing ds cDNA were pooled and precipitated in ethanol after the addition of 2 μg of yeast tRNA.

Plasmid vector pBR322, cleaved at the Pst I site and tagged with dGMP, was hybridized to equimolar amounts of both tailed ds cDNA preparations and used to transform E. coli HB101 by standard procedures (16).

Analysis of cDNA Clones. Clones containing putative HAV sequences were screened by cleaving recombinant plasmid preparations with Pst I and sizing by gel electrophoresis. Cleared cDNA inserts isolated from low melting point agarose (19) were labeled by nick-translation and used as probes in hybridization (i) to electrophoretically separated RNA bound to nitrocellulose paper (described above) for establishing the identity of cloned cDNA species, (ii) to DNA bound to nitrocellulose paper after lysis of bacterial colonies in situ for further screening, and (iii) to restriction fragments of DNA resolved by electrophoresis and bound to nitrocellulose paper for confirmation of tentative restriction maps constructed on the basis of single and double enzyme digests. Extent of RNA sequence not represented in cDNA clones was determined by primer extension of HAV-infected AGMK cytoplasmic RNA (8). Sequence determination of cloned cDNA was by the method of Maxam and Gilbert (20).

Materials. The following sources were used: AGMK cells, Flow Laboratories (Rockville, MD); micrococcal nuclease and large fragment of E. coli DNA polymerase I, Boehringer Mannheim; proteinase K, Merck (Darmstadt, Federal Republic of Germany); HPLC grade deoxynucleoside triphosphates, ICN; [α-32P]dCTP, New England Nuclear or Amersham; RNasin, JEM Research (Kensington, MD); reverse transcriptase, J. W. Beard, Life Sciences (St. Petersburg, FL), through the National Cancer Institute; nuclease S1, P-L Biochemicals; terminal deoxynucleotidyltransferase, nuclease-free bovine serum albumin, nick-translation kits, and certain restriction endonucleases, Bethesda Research Laboratories; other restriction enzymes, New England Biolabs, Sepharose 4B, Pharmacia, and oligo(dG)-tailed pBR322, New England Nuclear. 125I-Labeled anti-HAV from the HAVAB kit (Abbott) was used for detection of HAV by radioimmunoassay (14).

RESULTS

Virus Purification and RNA Characterization. Direct electron microscopic examination of purified HAV particles revealed a homogeneous population of 27-nm virions. The yield of virion RNA from eight marmoset livers was approximately 1.0 μg and the A₂₆₀/A₂₈₀ ratio was 2.0. Samples used for UV spectroscopy were also denatured and analyzed by electrophoresis through an agarose gel (15). HAV RNA contained a discrete band comigrating with poliovirus type 1 RNA (approximately 7,440 nucleotides; refs. 7–9).

Analytic cDNA Synthesis. RNAs from several sources of HAV and from poliovirus type 2 were compared for template quality in cDNA synthesis under conditions optimal for poliovirus RNA (Fig. 1). HAV RNA derived from liver or AGMK culture yielded a series of transcripts ranging in size from slightly smaller than the longest poliovirus type 2 cDNA to less than 500 nucleotides (Fig. 1, lanes a, b, and b’; lanes e and e’ show poliovirus type 2 cDNA). Presumably, degradation of HAV RNA prevented more extensive synthesis of reverse transcript amplifying the expected full length of 7,500 nucleotides. Other HAV cDNAs in Fig. 1 (lanes c and d) revealed evidence of more extensive RNA synthesis.

![Fig. 1. Alkaline agarose gel electrophoresis of cDNA from HAV RNA and poliovirus type 2 RNA. By using conditions optimal for poliovirus cDNA synthesis (8), RNAs (5 μg/ml or less) were incubated for 60 min at 42.5°C in 10 μl containing 50 mM Tris-HCl (pH 8.3)/10 mM MgCl₂/50 mM KCl/500 μM dATP/[α-32P]dCTP (2 Ci/mmol)/dGTP/TPP/0.4 mM dithiothreitol/4 mM sodium pyrophosphate/[dTTP]₁₋₁₉ at 30 μg/ml/reverse transcriptase (80 units/ml). Two reactions (cDNAs in lanes b’ and e’) also contained RNasin (2,000 units/ml). Reverse transcripts from RNA templates: lane a, HAV RNA derived from marmoset liver; lanes b and b’, HAV RNA from AGMK cells; lane c, HAV RNA from human stool; lane d, HAV RNA from AGMK cells; lanes e and e’, poliovirus type 2 RNA. HAV RNAs were isolated as described in Materials and Methods except that derived from human stool (lane c), which was extracted from purified virus (a gift from S. Locarnini), and an earlier preparation from AGMK cells (lanes d) using modifications of standard procedures (5). Migration of λ HindIII fragments (in kb) is indicated on the left.](image-url)
degradation, but all gave a similar banding pattern that was different from that of poliovirus cDNAs.

The effect of varying several chemical constituents and physical parameters on yield and length of HAV cDNA was analyzed by alkaline agarose gel electrophoresis and incorporation of \(^{32}\text{P}\)dCTP into trichloroacetic acid-precipitable product (data not shown). Incubation for 30 min, or with reverse transcriptase at 120 units/ml, or with 100 mM KCl increased size and quantity of cDNA when compared to that shown in Fig. 1 (lane b'). Denaturation of HAV RNA was not attempted because quantity was limited and, in earlier experiments, heat or methylmercury treatment of poliovirus RNA decreased yield and size of cDNA.

Preparation of cDNA Clones. A significant portion of HAV cDNA transcribed under preparative conditions was 3,000–7,500 nucleotides long (Fig. 2, lane a). However, a wide size range of ds cDNA molecules was produced (Fig. 2, lane b), and most of the preparation was less than 2,000 nucleotides long after nuclease S1 digestion (lane c). Approximately 200 ng of ds cDNA was synthesized. Ten nanograms of ds cDNA selected for large size by gel filtration (Materials and Methods) yielded 232 tetracycline-resistant E. coli transformants. Cleavage with Pst I demonstrated inserts of 1,000 base pairs or more in 43 of the 232 recombinant plasmids, designated pHAV<sub>LB</sub>. An additional 2,710 clones were obtained from 9 ng of unfraccionated ds cDNA. From this group, only pHAV<sub>L</sub>, 1307 (described below) was extensively characterized.

Identity of Cloned cDNA. The identity of inserted DNAs in recombinant plasmids was established by hybridization to RNA bound to nitrocellulose paper after gel electrophoresis (Fig. 3). A nick-translated probe prepared from the insert of pHAV<sub>L</sub> 39 specifically hybridized to RNA from HAV-infected AGMK cells (Fig. 3, lanes c' and e). Similar results (not shown) were obtained when the inserts of pHAV<sub>L</sub> 93 or pHAV<sub>L</sub> 225 were used as probes. The predominant band identified had the size expected for genomic HAV RNA and comigrated with poliovirus type 2 RNA. Diffuse hybridization to lanes containing RNA from infected cells was probably due to RNA degradation. Nick-translated pBR322 did not hybridize to any RNA species from either HAV-infected or uninfected AGMK cells (data not shown), thereby eliminating the possibility that a small amount of pBR322 contaminating the insert probes was responsible for specific hybridization. None of the pHAV<sub>L</sub> probes tested to date hybridized to poliovirus RNA.

Restriction Map of Cloned HAV cDNAs. The restriction map shown in Fig. 4 is based on data obtained from digests and on hybridization of labeled inserts to fractionated DNA (data not shown). Hybridization of an insert fragment from pHAV<sub>L</sub> 228 [corresponding to the region from 2.4 to 3.0 kilobases (kb) in Fig. 4] to DNA from bacterial colonies was used to select pHAV<sub>L</sub> 1307 (and pHAV<sub>L</sub> 12 and pHAV<sub>L</sub> 153, described in the legend to Fig. 4) for further analysis. Plasmids pHAV<sub>L</sub> 113, pHAV<sub>L</sub> 1307, pHAV<sub>L</sub> 228, pHAV<sub>L</sub> 148, and pHAV<sub>L</sub> 207 overlap to generate a map of about 7.4 kb, which represents at least 99% of the genome of HAV. Inserts from other clones (as indicated in Fig. 4) provided confirmation of overlap regions except for that between pHAV<sub>L</sub> 228 and pHAV<sub>L</sub> 148. Nick-translated pHAV<sub>L</sub> 228 insert hybridized to the 1.5-kb Pst I fragment of pHAV<sub>L</sub> 148. The overlap as drawn is based on the assumption that there is only one HincII site, present in both clones, in the area about 4 kb from the 5' terminus of the genome. No other restriction sites in this region were found by digestion with >30 other enzymes that have five- and six-base recognition sites. As noted in the legend to Fig. 4, heterogeneity in one restriction site (Nde I) was detected in the region about 2.5 kb from the 5' end of the viral RNA.

Primer Extension of Cloned cDNA. A labeled Pst I/BamH I fragment (0.2–0.7 kb in Fig. 4) from pHAV<sub>L</sub> 113 was used as a primer for cDNA synthesis. When the template was cytoplasmic RNA from HAV-infected AGMK cells, the primer was extended to give a discrete band (data not shown). Assuming that the 5' termini from HAV RNAs of different passage levels are conserved, we calculated that cloned cDNA extended to within 50 bases of the 5' end of the genome. There was no primer extension when RNA from uninfected AGMK cells was used as a template.

Partial Sequence of the HAV Genome. The sequence of cloned cDNA corresponding to the 3' end of HAV RNA is presented in Fig. 5. There is an open reading frame that extends from the beginning of the sequence to a pair of nonsense codons that are separated by six bases (bases 415–426). The second nonsense codon is followed by 51 bases that do not contain....
an open reading frame initiated by an ATG, followed by poly(A), presumed to represent the 3' end of the viral genome. The longest possible polypeptide encoded in the second and third reading frames is 29 amino acids.

Preliminary sequence data (unpublished results) indicate that an open reading frame begins with an ATG approximately 750 bases from the 5' end of HAV RNA and continues for at least 378 bases. Numerous stop codons in the other reading frames preclude a significant translation product.

**DISCUSSION**

Molecular cloning of cDNA is a useful approach for the study of RNA viruses such as HAV. Analysis of virion structure and HAV replication has been limited by the persistence of cellular functions in HAV-infected tissue culture, the long time required for virus replication, and the small quantities of HAV obtainable. The availability of cloned HAV cDNA, made possible by using efficient conditions for ds cDNA synthesis that were well characterized for another picornavirus RNA (8), should be useful for deducing RNA and amino acid sequences and elucidating the events of replication.

In an effort to produce cDNA clones that contained only HAV sequences, HAV RNA was extracted from highly purified virions. HAV RNA probably represents a very small proportion of nucleic acid from potential sources. The virus appears, from electron microscopic (21) and biophysical data (unpublished data), to be associated with intracellular membranous vesicles in vivo and in vitro. It is possible that nonviral nucleic acid species could contaminate HAV preparations by copurifying with the virus or associated membranes. In earlier experiments the template for cDNA was RNA extracted from HAV that had been incompletely purified from marmoset liver. The cDNA clones derived from this RNA were found to contain nonviral sequences that were presumably of host origin (unpublished data). In the experiments described here, liver homogenates were hydrocarbon-extracted until an aqueous–organic interphase layer was no longer detected. Micrococcal nuclease digestion was used to eliminate unencapsidated nucleic acids. In addition, any viral probe (labeled RNA or cDNA) used to screen clones might be as contaminated with nonviral sequences as the template used for cloning. Instead of using such a probe, pHAV inserts were characterized by restriction analysis, cross-hybridization, and hybridization to HAV RNA contained in RNA from a heterologous source.

It would be difficult at present to directly determine the sequence of HAV RNA as a reference, but analysis of cloned cDNA sequences described here is consistent with a genome that has the structural organization of other picornaviruses but does not show significant homology to members of this group (7–9, 22). Although there is only one known serotype of HAV, a restriction site variation between HAV cDNA inserts was detected. It is impossible to determine whether the alteration is due to RNA sequence heterogeneity or an error in reverse transcription. However, it would not be surprising to find sequence heterogeneity, because the HAV used in this study was not cloned. The Nde I site variation (2.5 kb from the 5' end of the genome) occurs in an area that corresponds to the sequence coding for the NH₂ terminus of VP1 in poliovirus RNA (7–9). This polypeptide is thought to be important in picornaviral surface structure and in immunity (23–25). In other picornaviruses, strain and serotype differences are often manifested by changes in VP1 RNA sequence (9, 26, 27).

It may be feasible to use HAV cDNA as a probe to detect viral or virus-related sequences from a variety of specimens. Experiments (data not shown) using a dot blot assay (18) indicate that cloned HAV cDNA is a sensitive and specific probe for detecting viral RNA molecules in cytoplasmic RNA from tissue culture.

Because many of the same limitations for obtaining the RNA
Biochemistry: Tiedehurst et al.

The nucleotide sequence of cloned cDNA corresponding to the HAV genome 3' terminal sequence was determined for the insert from pHAV39 (20), starting near the Auc I/Xho I site (Fig. 4). Possible translation products are listed in all three reading frames below the nucleotide sequence, with the frame probably used for viral protein uppermost. Following poly(A) bases (478-492), poly(G) links cloned HAV cDNA to pBR322 (data not shown).

of HAV applies to its proteins, there is no information about amino acid sequences and little is known about the immunologic determinants that elicit the polyclonal antibody response to infection. Antisera directed against specific HAV polypeptides have not yet been identified. It may be possible to prepare such antisera from synthetic peptides or from cDNAs expressed in prokaryotes (28) or eukaryotes (29) as has been done for other picornaviruses (24, 25). Immunogenic peptides or polypeptides would have obvious potential for vaccine development.

The support and contributions of Charles Burkholder, Robert Chanock, Roger Deely, Dave Hoggan, Bill Hoyer, Ching-Juh Lai, Tom Miele, Kathy Mifialik, Terry Popkin, Arnold Rabson, Johanna Sears, Margaret Stewart, and especially Taylor Chestnut are greatly appreciated. We also thank Linda Jordan for editorial assistance. This work was supported by Grant AI-08388 from the National Institute of Allergy and Infectious Diseases and Grant CA-14051 from the National Cancer Insti-

stitute (core grant to S. Luria). V.R.W. was supported by a postdoctoral fellowship from the National Institute of Allergy and Infectious Diseases. D.B. is an American Cancer Society Research Professor.


