Diagnosis of Hepatitis A Virus Infection: a Molecular Approach

Omana V. Nainan,† Guoliang Xia,* Gilberto Vaughan,†,‡ and Harold S. Margolis†,‡

Division of Viral Hepatitis, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333; Institute of Diagnosis and Epidemiologic References, Mexico City 11340, Mexico; and Pediatric Dengue Vaccine Initiative, International Vaccine Institute, Seoul, Korea

INTRODUCTION
The disease described as “jaundice” in ancient Greek, Roman, and Chinese literature probably was viral hepatitis. A viral etiology was postulated as the cause of certain forms of jaundice as early as 1912, and the term “infectious hepatitis” was used because the disease often occurred in epidemics (43). Hepatitis A, a term first introduced by Krugman et al. in 1967 (138), is now known to be caused by infection with hepatitis A virus (HAV), one of five viruses, each belonging to a different family, whose primary site of replication is the liver.

Early epidemiological studies further characterized hepatitis into infectious and serum forms, based on patterns of disease transmission (17, 82, 102, 103, 104). Epidemiologic and transmission studies with humans showed that infectious hepatitis, or hepatitis A, was transmitted primarily by the fecal-oral route (136, 137, 138). In 1973, HAV was identified in the stools of infected persons (80), which eventually led to development of diagnostic tests, propagation in cell culture, molecular characterization, and development of a vaccine (80, 193).

HEPATITIS A VIRUS INFECTION
Clinical Features
Infection by HAV is generally self-limited and can produce effects that range from a lack of symptoms to death from...
Fulminant hepatitis. The likelihood of clinically apparent disease associated with HAV infection increases with age. In children <6 years of age, most infections (70%) are asymptomatic (98), and if illness does occur, it is usually anicteric. Among older children and adults, infection is usually symptomatic, with jaundice occurring in >70% of patients (141). After an average incubation period of 28 days (range, 15 to 50 days), most HAV-infected persons developed nonspecific constitutional signs and symptoms followed by gastrointestinal symptoms. Typically, these include fever, malaise, anorexia, nausea, abdominal discomfort, dark urine, and jaundice, all of which usually last <2 months. There is no evidence of chronic liver disease or persistent infection following infection. However, 15 to 20% of the patients may have prolonged or relapsing disease lasting up to 6 months (93, 218), and HAV has been detected in serum for as long as 6 to 12 months after infection (19).

Fulminant hepatitis is a rare complication of hepatitis A. The risk of acute liver failure ranges from 0.015 to 0.5%, and the highest rates occur among young children and older adults with underlying chronic liver disease (2, 31). In a large epidemic in Shanghai, China, involving over 300,000 adolescents and young adults, 47 deaths (0.015%) were reported (100, 263). Other than older age and underlying chronic liver disease, no other host factors have been identified as predisposing patients to fulminant hepatitis A (250). Among persons referred to a tertiary care center with acute hepatic failure in the United States, 20 of 295 (6.8%) had hepatitis A (210). No single viral factor has been associated with fulminant disease.

Reported findings among persons with fulminant disease include lower viral load (198) and nucleotide and/or amino acids substitutions in the 5′ untranslated region (5′UTR), P2 region, and P3 region of the HAV genome (86; CDC, unpublished data).

Pathogenesis and Natural History of HAV Infection

The pathogenetic events that occur during HAV infection have been determined in experimental infection of nonhuman primates and natural infection of humans (Fig. 1). HAV is primarily hepatotropic; it replicates in the liver, produces a viremia, and is excreted in bile and shed in the stools of infected persons. Feces can contain up to 10^9 infectious virions per gram and is the primary source of HAV infection (221, 241). Peak fecal excretion, and hence infectivity, occurs before the onset of jaundice, symptoms, or elevation of liver enzymes (221, 241) and declines after jaundice appears. Compared to adults, children and infants can shed HAV for longer periods, i.e., up to several months after the onset of clinical illness (35, 241). Fecal shedding of HAV has been shown to occur as late as 6 months after diagnosis of infection in premature infants (205).

Viremia occurs within 1 to 2 weeks after HAV exposure and persists through the period of liver enzyme elevation, based on studies in humans and experimentally infected chimpanzees (8, 48, 91, 137, 145, 157). Virus concentrations in serum are 2 to 3 log_{10} units lower than those in stool (19, 48, 145). An analysis of serum specimens collected prospectively during human and chimpanzee HAV infection showed that HAV RNA was
present for ~3 to 4 weeks before the onset of jaundice and that virus concentrations were highest during the period that precedes onset of liver enzyme elevations (19). Viremia may be present for a much longer period during the convalescent phase of hepatitis A than was previously appreciated (19, 84, 264), although virus concentration is lowest during this period.

The virus is also shed in saliva in most hepatitis A patients. In experimentally infected marmosets (156, 187), the viral load appears to be 1 to 3 log10 units lower than that found in serum (156). However, no epidemiological data suggest that saliva is a significant source of HAV transmission.

A humoral immune response to HAV structural proteins occurs prior to onset of symptoms. Immunoglobulin M (IgM) antibodies to HAV (IgM anti-HAV) are detectable at or prior to onset of clinical illness, decline in about 3 to 6 months, and become undetectable by commercially available diagnostic tests (127). IgG antibodies to HAV (IgG anti-HAV) appear soon after IgM, persist for years after infection, and confer lifelong immunity (220). IgA is also produced during infection for a limited time (4, 152). The role of IgA antibodies in the response against HAV is still unknown. Unlike other Picornaviridae family members, HAV does not seem to elicit an effective intestinal immune response (229).

IgG and IgA anti-HAV are detected in serum, saliva, urine, and feces (39, 60, 115, 123, 151, 152, 177, 178, 183, 214). Saliva tests have been reported as an alternative to conventional serum testing for anti-HAV due to their simplicity of sample collection (115, 177, 178). Several studies have demonstrated the benefits of implementing saliva testing as screening tool in outbreak investigations and epidemiological studies (110, 115, 140). However, the sensitivity of detecting anti-HAV in saliva is 1 to 3 log10 units lower than that with serum (156, 177, 178; CDC, unpublished data).

Antibodies against nonstructural proteins are also produced, although their role in maintenance of immunity is probably less important than that of antibodies to capsid antigens due to their low concentration and lack of neutralization capacity. Antibodies to nonstructural proteins have been detected in humans and experimentally infected chimpanzees but are absent in vaccinated individuals (126, 202). However, because of what appears to be a variable host antibody response during HAV replication, a diagnostic test for these antibodies, which could be used to complement current anti-HAV testing and differentiate previously infected from vaccinated persons, has not been developed (201, 233).

**HEPATITIS A VIRUS**

HAV is a nonenveloped RNA virus 27 to 32 nm diameter in size, with an icosahedral symmetry, which belongs to the genus Hepatovirus of the Picornaviridae family. Unlike other members of the family, HAV requires a long adaptation period to grow in cell culture, replicates slowly, and rarely produces a cytopathic effect (57, 107, 146). HAV is stable in the environment for at least 1 month (162) and is more resistant to heating and chlorine inactivation than is poliovirus. Inactivation of HAV
Table 1. HAV genome organization and open reading framea

<table>
<thead>
<tr>
<th>HAV genome segment</th>
<th>Nucleotides</th>
<th>Polypeptide amino acids</th>
<th>Known function(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5′UTR 1–735 (741)b</td>
<td>Unclear, translation initiation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1 735–3029 1–765</td>
<td>Capsid Proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A 735–803 1–23</td>
<td>VP4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1B 804–1469 24–245</td>
<td>VP2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1C 1470–2207 246–491</td>
<td>VP3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1D 2208–3029 492–765</td>
<td>VP1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2 3030–5000 766–1422</td>
<td>Nonstructural Proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2A 3030–3242 766–836</td>
<td>Morphogenesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2B 3243–3959 837–1087</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2C 3996–5000 1088–1422</td>
<td>RNA synthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3 5001–7415 1423–2227</td>
<td>Nonstructural proteins and VPG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3A 5001–5222 1423–1496</td>
<td>Pre-VPG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B 5223–5291 1497–1519</td>
<td>VPG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3C 5292–5948 1520–1738</td>
<td>Protease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D 5949–7415 1739–2227</td>
<td>RNA polymerase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3′UTR 7416–7478</td>
<td>Unknown, translation terminator, poly(A)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Reference HAV strain, HM175 (subgenotype 1B, accession no. M14707) (46, 107).
b Downstream AUG codon at nt 741 to 743 preferred for translation initiation.

requires heating foods to >85°C for 1 min, and disinfection of surfaces requires 1 min of contact with a 1:100 dilution of sodium hypochlorite (i.e., household bleach), whereas poliovirus is inactivated at 72°C for 15 s and by treatment with a 1:125 dilution of bleach for 30 s (159, 234, 257).

Genomic Organization

HAV has a positive-polarity, single-stranded 7.5-kb genome (45, 46, 171) that is organized similarly to those of the other picornaviruses (Fig. 2). The 5′UTR is 734 to 740 nucleotides (nt) long and has a covalently linked virus-specific protein (VPg) rather than a cap structure (259). Translation occurs in a cap-independent fashion under control of an internal ribosome entry site located within the 5′UTR (21). The translation terminator sequence, the 3′UTR of 40 to 80 nucleotides, has a poly(A) tract (46) (Table 1). The remainder of the genome is composed of a single open reading frame with three distinct regions (P1, P2, and P3) and is translated as a single polyprotein of 2,225 to 2,227 amino acids (10, 46). This polyprotein subsequently undergoes cleavage mediated by a viral protease (3Cpro) (211), which results in production of four capsid and several nonstructural proteins (Fig. 2; Table 1).

HAV Genetic Diversity

HAV displays a high degree of antigenic (amino acid) and genetic (nucleotide) conservation throughout the genome (45, 46, 147, 200, 207). However, enough genetic diversity exists to define several HAV genotypes and subgenotypes (200). The entire nucleic acid sequences of several HAV strains have been determined by molecular cloning (10, 45, 46, 96, 171, 182, 243, 244), and a large number of HAV isolates have been characterized by sequencing of short genome segments. The genomic regions most commonly used to define HAV genotypes include (i) the C terminus of the VP3 region (116), (ii) the N terminus of the VP1 region (5, 112, 199, 203), (iii) the 168-bp junction of the VP1/P2A regions (116, 200, 203), (iv) the 390-bp region of the VP1-P2B regions (112, 170), and (v) the entire VP1 region (52, 167) (Fig. 2; Table 2).

When sequence variation within the VP1/P2A junction is used to define genotypes and subgenotypes, genotypes have >15% nucleotide variation between isolates and subgenotypes have 7 to 7.5% nucleotide variation (200). Seven HAV genotypes have been identified; four genotypes (I, II, III, and VII) are of human origin, and three (IV, V, VI) are of simian origin. Genotypes II and VII were initially defined based on a single isolate for each (34, 200). However, further investigations have reclassified genotype VII as a subgenotype of genotype II (52, 154). Genotypes I and III are the most prevalent genotypes isolated from humans.

The three simian genotypes were each defined by unique nucleotide sequences from the P1 regions of HAV strains recovered from species of Old World monkeys. In addition, all simian HAVs have a distinct signature sequence at the VP1-P2B regions (112, 170), and (v) the entire VP1 region (52, 167) (Fig. 2; Table 2).

Antigenicity and Serotype

Only a single serotype of HAV exists, despite genetic heterogeneity at the nucleotide level. Individuals infected by HAV in one part of the world are protected from reinfection by HAV from other parts of the world. Immune globulin preparations containing anti-HAV, irrespective of their geographic origin, appear to provide protection from disease, and vaccines
prepared from virus isolates originating in Australia or Costa Rica protect from infection worldwide (113, 172). The antigenic structure of the virus is relatively simple, with a restricted number of overlapping epitopes combining to form a single dominant antigenic site that interacts with virus-neutralizing antibodies. These epitopes are highly conformational and are formed by amino acid residues located on more than one capsid protein (168, 185, 186, 228). Convalescent-phase sera obtained from hepatitis A patients are reactive primarily to VP1 and to a lesser extent to VP0 and VP3 (254). Recombinant VP1 fusion protein expressed in *Escherichia coli* reacted with rabbit anti-HAV serum (180). However, chimpanzees immunized with this recombinant protein produced antibodies that reacted with VP1 of only denatured and not intact virus, and the animals were not protected when they were challenged with wild-type HAV (CDC, unpublished data). Empty particles appear to be antigenically indistinguishable from infectious, RNA-containing virions, suggesting that antigenicity may depend on assembly of the major capsid proteins or smaller capsid precursors. An accurately processed and assembled recombinant HAV polyprotein has been produced, which was able to elicit neutralizing antibodies detected by commercial assays (139, 262).

Naturally occurring antigenic variants of HAV have been observed only among strains isolated from Old World monkeys (167, 248). These viruses are genetically distinct from human HAV isolates and are not recognized by certain monoclonal antibodies produced against human HAV (128, 167). However, simian HAV binds human polyclonal anti-HAV, and chimpanzees immunized with these viruses had an antibody response that was protective against infection with human HAV challenge (78; CDC, unpublished data). Recently, human HAV isolates with capsid amino acid substitutions and deletions in the immunodominant antigenic site were reported (52, 206). However, it is not clear if capsid antigenicity or virus neutralization was affected by these changes.

**DIAGNOSTIC APPROACHES TO HAV DETECTION**

**Detection of HAV-Specific Antibodies**

The humoral immune response plays the pivotal role in the diagnosis of HAV infection and the differentiation of hepatitis A from other types of viral hepatitis. There are a number of commercially available assays for the detection of IgM and total anti-HAV (71, 122, 188). IgM, IgA, and IgG anti-HAV are usually present at the onset of symptoms (Fig. 1). Since hepatitis due to HAV infection is clinically indistinguishable from disease caused by other hepatitis viruses (i.e., HBV, HCV, HDV, and HEV), serologic testing is required to make the diagnosis (230).

**Antibodies to structural proteins.** Diagnostically, IgM anti-HAV has been used as the primary marker of acute infection (58); it is comprised mainly of antibodies against capsid proteins. A number of methods have been used to detect this virus-specific antibody class, including radioimmunoassay (194), immunochemical staining (109), enzyme-linked immunosorbent assay (61), immunoblotting (254), and dot blot immunogold filtration (214). IgM anti-HAV enzyme immunoassays are available commercially (188) The commercially available diag-

<table>
<thead>
<tr>
<th>Region (reference[s])</th>
<th>Primer type&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Primer&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Nucleotide sequence (5’→3’)</th>
<th>Position&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>C terminus of VP3 (116)</td>
<td>2020P</td>
<td>ACAGGTATAAACAAAGTCAG</td>
<td>2020–2037</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2226N</td>
<td>CTCCAGAATCATCTCC</td>
<td>2226–2211</td>
<td></td>
</tr>
<tr>
<td>N terminus of VP1 (22, 112, 199, 202)</td>
<td>External</td>
<td>2133P</td>
<td>GTGAATGTTTATCTTTCAGCAAT</td>
<td>2133–2155</td>
</tr>
<tr>
<td></td>
<td>2451N</td>
<td>GATCGATGTAGTCAGGATCTTC</td>
<td>2451–2429</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2172P</td>
<td>GCTCCTTATATCTGCTAGGAT</td>
<td>2172–2195</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2415N</td>
<td>CAGGAAATGTCGATGTCATTCTC</td>
<td>2415–2392</td>
<td></td>
</tr>
<tr>
<td>VP1/P2A junction (22, 116, 200)</td>
<td>External</td>
<td>2950P</td>
<td>TTGTCTGTGCAGAAACATTGAG</td>
<td>2950–2972</td>
</tr>
<tr>
<td></td>
<td>3308N</td>
<td>AGTCAACTTCTTCTAGCAAAACTT</td>
<td>3308–3285</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2984P</td>
<td>TCCAGAGCTCCATGAGA</td>
<td>2984–3001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3217N</td>
<td>AGGGGCTGAAAGTACTTATTGGA</td>
<td>3217–3193</td>
<td></td>
</tr>
<tr>
<td>VP1-P2B (112, 170)</td>
<td>External</td>
<td>2870P</td>
<td>GACAGATTCTACATTTTGGATG</td>
<td>2870–2893</td>
</tr>
<tr>
<td></td>
<td>3381N</td>
<td>CCATTTCAAGAGTCCACACT</td>
<td>3381–3360</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2896P</td>
<td>CTATTCAGAATTTCTCAGAAT</td>
<td>2896–2918</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3289N</td>
<td>AACCTTATTATTTTCAGTCTT</td>
<td>3289–3268</td>
<td></td>
</tr>
<tr>
<td>Entire VP1 (52, 167)</td>
<td>External</td>
<td>2167P</td>
<td>GTTTTTGCTTCTTATCTGCTAGT</td>
<td>2167–2192</td>
</tr>
<tr>
<td></td>
<td>3308N</td>
<td>GCATACACTCTTCTAGCAAAACTT</td>
<td>3308–3285</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2172P</td>
<td>GCTCCTTATATCTGCTAGGAT</td>
<td>2172–2195</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3125N</td>
<td>CCTGATTCTATGATCTTCT</td>
<td>3125–3106</td>
<td></td>
</tr>
<tr>
<td>VP3-P2B (CDC, unpublished data)</td>
<td>External</td>
<td>2020P</td>
<td>ACAGGTATAAACAAAGTCAG</td>
<td>2020–2037</td>
</tr>
<tr>
<td></td>
<td>3381N</td>
<td>CCTTCAGAAGAGTCCACACT</td>
<td>3381–3360</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2133P</td>
<td>GTGAATGTTTATCTTTCAGCAAT</td>
<td>2133–2155</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3289N</td>
<td>AACCTTATTATTTTCAGTCTT</td>
<td>3289–3268</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> External primers were used for first-round amplification, and internal primers were used for second-round nested amplification.

<sup>b</sup> P, positive-sense primer; N, negative-sense primer.

<sup>c</sup> Positions relative to the genome of HAV strain HM175 (accession no. M14707).
Antigen Detection

Cell culture propagation. HAV has been grown in several cell types of human and nonhuman origins, including primary and secondary African green monkey kidney cells (59, 253) and fetal rhesus monkey kidney cells (83). In contrast to most picornaviruses, HAV of human origin requires an extensive adaptation period before it grows in cell culture, and once adapted, HAV produces a persistent infection and becomes attenuated, as shown by not producing disease in experimentally inoculated nonhuman primates (81). In addition, relatively low concentrations of virus and viral antigen are produced compared to other picornaviruses. Mutations in viral nucleic acid may play a major role in the adaptation of HAV in cell culture (47, 76, 77, 88) and attenuation (79, 81). HAV replicates in cell culture without cytopathic signs of infection and without apparent host cell damage. Because of the lack of a cytopathic effect in cell culture, immunological assays are required to detect HAV antigen (217). Methods commonly used to quantitate infectivity include radioimmunofocus assay (144), fluorescent focus assay, in situ radioimmunooassay (217), and in situ hybridization (121).

Cytopathic variants of HAV have been observed in selected cell culture systems. These cytopathic viruses produce an acute rather than persistent infection (55, 173). The replication cycle of these variants is shorter (2 to 3 days) than that observed for noncytopathic HAV, and they produce a much higher viral yield (55).

Detection in clinical and environmental samples. HAV was first visualized in fecal extracts by electron microscopy using homologous antisera (80), and similar virus-like particles were observed in the sera and livers of marmosets experimentally infected with human HAV (192). HAV antigen has been detected in stool, cell culture, and environmental samples by using radioimmunoassays and enzyme immunoassays (107, 165). Viremia during HAV infection has been documented both by transmission studies (138) and as a result of outbreaks of posttransfusion hepatitis A (176). However, detection of antigen in blood has been difficult (106) because fibronectin can bind to HAV and mask antigenic determinants required for immunological detection (213). HAV capsid polypeptides and viral RNA have been detected in IgM circulating immune complexes isolated from experimentally infected chimpanzees (158).

Molecular Detection Methods

Nucleic acid detection techniques are more sensitive than immunoassays for viral antigen to detect HAV in samples of different origins (e.g., clinical specimens, environmental samples, or food). HAV has been detected with techniques such as restriction fragment length polymorphism (94), single-strand conformational polymorphism (85, 94), Southern blotting (25, 27, 169, 208), nucleic acid sequencing-based amplification (118, 119, 120), nucleic acid hybridization (266), and reverse transcription-PCR (RT-PCR) and antigen capture RT-PCR (56, 57, 116, 189, 199, 200). Amplification of viral RNA by RT-PCR is currently the most sensitive and widely used method for detection of HAV RNA.

Purification of viral RNA from clinical and environmental samples is the first step in RT-PCR, and different extraction platforms are often required, depending on the source of the specimen. Early protocols for RNA extraction from serum or stool included protease K digestion followed by phenol-chloroform extraction and ethanol precipitation (167, 199). Subsequently, products which used guanidinium thiocyanate-phenol-chloroform (41) became commercially available for extraction of RNA and total nucleic acids (40) and increased the sensitivity and specificity of HAV detection (65, 112, 170). Antigen capture RT-PCR (116) and magnetic beads coated with anti-HAV (124) have been used to separate virus from potential inhibitors of reverse transcription and PCRs that are often found in environmental and stool samples. Automated RNA extraction protocols have been applied to HAV detection. These have increased specimen throughput, are reproducible and reliable, and have detection sensitivity similar to that of manual extraction methods (142; CDC, unpublished data).

The efficiency of reverse transcription of RNA to cDNA can be reduced by the presence of inhibitors in the source material. Engineered reverse transcriptase with no RNase activity (RNase H) is believed to increase transcription efficiency, although naturally occurring enzymes continue to be used for identification of a variety of targets, including HAV, with high specificity and sensitivity (169, 226, 227). Although thermostable enzymes may improve efficiency of the RT reaction by reducing secondary structure and improving priming, their effect on HAV detection has not been evaluated. Specific or random primers can be used for the reverse transcription reaction (226, 227). Random primers along with specific primers have been routinely used in our laboratory for the detection of HAV because of the increased sensitivity and specificity (CDC, unpublished data).

 Primer pairs (Table 2; Fig. 2) spanning the desired genomic region are used to amplify cDNA. Nested PCR, where products obtained from first-round PCR are used as a template for a second round of PCR, has been used to amplify HAV from clinical and environmental samples where the viral load is...
expected to be low (19, 72, 92, 149, 169). Analysis of the PCR product by probe hybridization also has been shown to increase the sensitivity of detection (62, 95, 116).

The development of single-step RT-PCR methods, in which reverse transcription and PCR are performed together, has considerably reduced the time and handling during cDNA synthesis (72, 149). However, this method appears to reduce detection sensitivity by up to 1 log unit compared to the two-step RT-PCR method (CDC, unpublished data).

Multiplex RT-PCR, where genome sequences of more than one organism are amplified simultaneously, provides the most efficient way to detect multiple agents in clinical and environmental samples compared to conventional RT-PCR amplification for each agent. This method has been developed for simultaneous detection of HAV and HEV (125) and of HAV, rotavirus, and poliovirus (92, 247).

Real-time PCR, which has revolutionized nucleic acid detection by its high speed, sensitivity, and reproducibility and minimization of contamination, has been applied to the detection and quantification of HAV (53). Real-time PCR has been introduced in our laboratory for rapid analysis of specimens in outbreak situations, with <36 h required from amplification to nucleic acid sequence results. The increased speed of real-time PCR is mainly due to reductions in amplification cycles, elimination of post-PCR detection procedures, and availability of devices for sensitive detection of amplified products. Small amplicons are recommended, and this may also play a role in the speed; however, it has been shown that decreasing the size of the PCR fragment does not necessarily improve PCR performance (174). Molecular beacons (1, 155, 240) and TaqMan-based assays (53) are a few of the chemistries that have been used for real-time PCR detection of HAV. Real-time PCR chemistries use primers and probes with separate fluorogenic labels; the generation of a fluorescent signal depends on the interaction between PCR products and/or probes (24, 53, 155, 240). A DNA-binding fluorophore, SYBR green, has been widely used because of its simplicity (129, 131, 166). The advantage of SYBR green over labeled probe-based detection platforms is the detection of highly variable genome regions for which probe design is often difficult (134). SYBR green interference with nucleic acid sequencing can be eliminated by performing a nested PCR without the fluorophore.

### Nucleic Acid Sequencing

Nucleic acid sequencing is performed on PCR products to confirm their specificity and provide the ultimate means to identify and characterize the organism. Nucleic acid sequencing of selected genomic regions of HAV has been used to determine the genetic relatedness of isolates (112, 116, 170, 199, 200). The original nucleic acid sequencing methodology described by Sanger et al. (209), which required independent labeling reactions for each nucleotide and conventional gel electrophoresis, has been replaced by high-throughput methods, including fluorescent dyes for label terminators and capillary arrays for electrophoresis, which have improved sequencing speed and accuracy (160, 196).

### Molecular Detection from Water and Food

HAV is stable in the environment, especially when associated with organic matter, and is resistant to low pH and heating (107). These characteristics facilitate the likelihood of transmission by contaminated food and water and also improve the likelihood of detection in environmental samples, including water and sewage (15).

Because HAV grows so slowly in cell culture and because of the generally low levels of contamination in environmental samples, virus detection became feasible only with the availability of sensitive nucleic acid detection methods. However, HAV detection in food has not been included as a part of routine analysis of these outbreaks in most parts of the world (9). The primary reason is that because of the long disease incubation period, implicated foods usually have been consumed or discarded by the time the outbreak is recognized (135). The same holds true for detection of HAV in waterborne outbreaks, unless there is ongoing contamination.

**Detection in food.** HAV has been successfully isolated from bivalve mollusks (72, 73, 206), oysters (44, 67), mussels (36), and clams (18, 132, 236, 237). Methods for extraction of HAV RNA from food samples have included the use of guanidine isothiocyanate followed by several precipitation steps with polyethylene glycol (PEG) or ultracentrifugation (72, 73, 237). The addition of viral concentration steps (e.g., high pH or PEG precipitation) followed by RNA extraction and purification using poly(dT) magnetic beads has been shown to increase the sensitivity of HAV detection from bivalve mollusk samples (133).

**Detection in water.** Concentration by filtration is the classic approach for viral detection from water and wastewater, including the use of beef extracts to elute virus from filters and PEG for concentration (223). However, these components can interfere with RT-PCR (212). Ultrafiltration has been used as an alternate method for virus concentration, and vortex flow filtration followed by microcentrifugation has been used to overcome the limitations of conventional concentration methods (246).

Nucleic acid hybridization assays using labeled probes were initially used to detect HAV in contaminated water (121). However, this method had low sensitivity and required several logs of virus for detection. Today, PCR is widely used for HAV detection in environmental samples (62, 95, 124, 212, 246).

### Molecular Epidemiology of Hepatitis A

Molecular biomarkers have been used to determine the genetic relatedness of organisms, both to identify and track modes and chains of transmission and to characterize the evolution of organisms in host populations. New approaches from disciplines such as bioinformatics and molecular evolution have added to the understanding of both the epidemiology of agents within host populations and the dynamics of genetic changes within the agent itself. Molecular epidemiology has played an important role in improving the understanding of HAV infection by identifying sources of infection and the dynamics of virus evolution.
Overview of Hepatitis A Epidemiology

Worldwide, the endemcity of HAV infection varies according to regional hygienic standards; the highest prevalence of infection occurs in regions with the lowest socioeconomic levels (12, 99). The United States generally has a low endemcity of HAV infection, although high rates of infection occur in certain populations (11, 28, 32). Approximately every decade, epidemics of hepatitis A have occurred in the United States; the last occurred in 1995. Although the incidence of hepatitis A has declined over the past decade, primarily because of hepatitis A immunization (13, 14, 23, 163, 252, 255), it remains a frequently reported disease. In 2003, 7,653 cases of hepatitis A were reported to CDC (30), which, when corrected for underreporting and asymptomatic infections (7, 255), represents an estimated 36,700 cases and 79,600 infections.

Modes of Transmission and Sources of HAV Infection

The most common reported source of infection is household or other close contact with an infected person (11, 28, 32). Other potential sources of infection include men having sex with men (MSM), travel to countries where HAV is endemic, and illicit drug use. Contaminated food and water are an infrequent source of infection, although they have been associated with outbreaks. On rare occasions, HAV infection has been transmitted by transfusion of blood or blood products (145, 224).

Personal contact. Most transmission occurs among close contacts in households and extended-family settings (225); this mode accounts for at least 25% of infections. No identifiable source for infection is reported for 40 to 50% of cases, and personal contact with an unidentified source shedding HAV is likely to explain most of these cases. A study in Salt Lake City, Utah, showed that 98 of 390 (25%) household contacts of 167 infected cases without an identified infection source had serologic evidence of recent HAV infection (225). Children have the highest incidence of infection, and infected young children with their less scrupulous hygiene probably serve as a major source of transmission (98, 222). Children appear to excrete virus longer than adults (114, 205, 218, 265), which may facilitate transmission, although good epidemiologic studies are lacking in this area.

MSM. Although early studies did not show a higher prevalence of HAV infection among MSM (238), prospective studies showed a high incidence of infection (49, 54). In addition, hepatitis A outbreaks among MSM have been reported in the United States and Europe (22, 105, 170, 235). HAV isolates with nucleotide sequence identity have been observed within several outbreaks among MSM and between outbreaks which occurred in different geographic locations (22, 105, 170, 235).

Illicit drug use. Outbreaks of HAV infection among injection drug users (IDUs) have been reported in North America and Scandinavia (97, 101, 143). Several routes of transmission are likely to occur, and these include a combination of person-to-person and percutaneous spread. Injection of drugs is often a group activity, and poor personal hygiene among illicit drug users may be a source of HAV contamination of drugs or drug paraphernalia, as well as direct person-to-person transmission. Fecal contamination of drugs by rectal transportation has been reported but is probably a rare source of infection (143). Percutaneous transmission of HAV by needle sharing can also occur (97).

The complexity of hepatitis A transmission among injection and noninjection drug users is illustrated by an outbreak among methamphetamine users and their contacts in Polk County, Florida, during 2001 and 2002 (252). The significance of IDUs as the source of the outbreak was generally underestimated, since case patients were not willing to admit illicit drug use. However, HAV sequence analysis showed identity or near identity among cases, including cases with no direct links to each other and whose disease onsets varied over time. These findings indicated a limited number of HAV introductions with subsequent person-to-person transmission among IDU networks and their contacts (97, 170). Similar studies in Scandinavia have shown that HAV strains among IDUs involved in outbreaks were significantly different from those of the infected non-drug-using population (97, 143, 232).

International travel. Persons from regions of low endemcity traveling to regions of high HAV endemcity are at substantial risk for acquiring infection (28, 231). In the United States, about 10% of hepatitis A patients have reported recent travel outside the United States as their only possible source of infection (11, 32, 170). HAV sequences from case patients with history of travel are closely related to the sequence patterns of isolates from countries where the travel occurred (170, 232; CDC, unpublished data).

Food and water. HAV transmission has been associated with contaminated food and water (63, 112, 175, 263). Outbreaks associated with consumption of mussels (70), clams (18, 100), contaminated lettuce (204), ice slush beverages, raw oysters (67), frozen strawberries (112, 175), blueberries (27), raspberries (195, 197), green onions (3, 63, 260), and other salad items (108, 153, 179) have been reported. The potential for extensive disease transmission is illustrated by the largest recorded hepatitis A outbreak, which occurred from consumption of sewage-contaminated clams and caused illness in 300,000 persons in Shanghai, China (263), and by outbreaks that extended to multiple states in the United States through widespread distribution of HAV-contaminated food (3, 112, 260). Although food- and waterborne outbreaks are often newsworthy, they account for only 2% of the total reported cases in the United States (28, 32).

Waterborne transmission of hepatitis A appears to be less common than transmission by food, and during the past 10 to 15 years contaminated water has not been identified as a source for any cases of hepatitis A in the United States (28, 58). Inadequate sewage management has played an important role in most waterborne outbreaks, which have generally involved contaminated groundwater obtained from wells or rivers (16, 68, 164, 181) and drinking water (125, 190).

Unfortunately, because of the long incubation period of HAV infection, virus detection in food is difficult, unless some of the food was kept or contamination is ongoing. Even more difficult to detect are sporadic cases associated with contaminated food. Only recently have such transmission chains been identified, and then only through the use of molecular epidemiologic investigations (3, 18, 112, 206, 260).
The determination and analysis of molecular biomarkers (e.g., nucleic acid sequencing or phylogenetic analysis) have been used to determine the relatedness of strains from two or more infected persons. These analyses have been greatly facilitated by (i) the ability to easily recover HAV from the serum, rather than stool, of infected persons and (ii) molecular methods (e.g., PCR, DNA sequencing, and sequence analysis) which have been adapted to achieve the rapid throughput of a large number of specimens. These recent advances have resulted in molecular epidemiology becoming an essential tool in the investigation of certain types of hepatitis A outbreaks and in epidemiologic studies of HAV transmission.

The key elements of a molecular epidemiologic analysis include determination of the genotypes and genetic relatedness of HAV isolates obtained from infected persons and a rigorous epidemiologic investigation. To facilitate the molecular analyses of HAV, RNA amplification and nucleic sequencing should be performed on one of the genome regions for which large amounts of comparative data exist (e.g., the VP1-P2B region) (Table 2; Fig. 2).

**Distribution of HAV genotypes.** Nucleotide sequence data indicate that HAV genotypes have unique geographic distributions (Table 3; Fig. 3). Genotype I is most prevalent worldwide, and subgenotype IA is more common than IB. Because genotype I is so common, genotyping alone rarely can be used to identify the source of an HAV outbreak or chain of transmission.

Countries and their commonly found genotypes are shown in Table 3. Subgenotypes IA and IB are most often found in North and South America, Europe, China, and Japan (200). For subgenotype IIIA, the prototype strain, PA21 (20), was originally isolated from captured Panamanian owl monkeys and thought to be a simian virus. However, once nucleotide sequence analysis was performed many years later, the unique nucleotide and amino acid sequence patterns which differentiate human from simian HAV (i.e., “simian signature”) (167, 200) were not found in this strain. Over the ensuing years, subgenotype IIIA isolates have been recovered only from humans in many parts of the world (Table 3).

Cocirculation of multiple genotypes or subgenotypes has been reported in some regions of the world. Cocirculation of subgenotypes IA and IB is reported in South Africa (242), Brazil (251), and Israel (CDC, unpublished data), and subgenotypes IA and IIIA are reported in India (111) and the Central Asian Republics of the former Soviet Union (CDC, unpublished data). Both subgenotypes IA and IB have been isolated from the United States; however, most IB isolates were found among infected travelers returning from other countries (170; CDC, unpublished data).

**Genetic relatedness of HAV.** Hepatitis A virus displays a high degree of antigenic and genetic conservation (146, 147, 200, 207) and does not appear to accumulate the high frequency of genetic changes seen in many RNA viruses. Although precise data on the genome mutation rate are not available, it appears to be lower than those of other RNA viruses (10^{-4} to 10^{-5} substitution per base per round of copy) (75). Even in extended common-source outbreaks, viruses isolated from the first cases are genetically the same as those isolated from the last cases (63, 112, 170, 175). However, even genetic heterogeneity exists in several HAV genome regions to differentiate the relatedness of isolates circulating within and between communities over time, including within subgenotypes (50, 51, 52, 161, 170).

Initially, a 168-nucleotide fragment of the VP1/P2A junction was used for genotype analysis (116, 200). However, this sized fragment was shown not to sufficiently differentiate genotype or relatedness among HAV isolates (52, 206). Currently, the majority of molecular epidemiologic studies conducted at CDC use a 390-nucleotide-long fragment from the VP1-P2B region which includes 2A, one of the most variable regions of the genome (112, 170; CDC, unpublished data) (Table 2; Fig. 2). In a comparative analysis of 240 HAV isolates from a number of communities, using the 350-nucleotide fragment of the VP1-
FIG. 3. Phylogenetic analysis of 3,582 HAV isolates in the VP1/P2B region (315-bp fragment). The dendrogram (unweighted-pair group method using average linkages) was created by using Kimura’s two-parameter model. Isolates were obtained from the United States (n = 2,190) and other parts of the world. The U.S. specimens included specimens from hepatitis A cases from CDC’s Sentinel Counties Study of Acute Viral
P2B region and a 900-nucleotide fragment from the entire VP1 region, genotype assignment was concordant in all instances and the longer fragment increased the likelihood of identifying a difference between two isolates by only 3% (CDC, unpublished data).

A phylogenetic analysis, restricted to a 315-nucleotide fragment of VP1-P2B, of over 3,000 HAV isolates from the United States and 12 other countries is shown in Fig. 3. The majority of isolates (85%) were obtained from investigations conducted by CDC’s Division of Viral Hepatitis; the remaining sequences were obtained from GenBank (accession numbers AB020564 to AB020569, AF050223 to AF050238, AF386846 to AF386888, AF396391 to AF396408, AJ296172, AJ299460 to AJ299467, AJ505561 to AJ505625, AY294047 to AY294049, AY322842 to AY323047, AY875649 to AY875672, and AY753408 to AY753530). Most isolates (80%) belong to subgenotype IA, 17% belong to subgenotype IB, and 3% belong to genotype IIIA. However, these genotype proportions probably do not represent the true worldwide distribution of HAV genotypes, since this is essentially a convenience sample which contains a disproportionate number of isolates from North America. Within a genotype or subgenotype, there was a clustering of HAV isolates by location or geographical regions. For example, isolates from North America, South America, Europe, and Middle East tended to cluster together. Phylogenetic analysis also showed distinct clusters of HAV isolates within subgenotype IA and defined three major “clusters,” i.e., US-IA1, US-IA2, and US-IA3 (Fig. 3). Nucleic acid sequences obtained from HAV isolates from the interior of Mexico and the U.S.-Mexican border were closely related to isolates in US-IA1, while US-IA2 isolates were predominantly from IDUs and US-IA3 isolates were from MSM.

Applications of Molecular Epidemiologic Investigations

The long incubation period of hepatitis A, the epidemic but often sporadic nature of disease transmission, and the often unapparent links between persons involved in outbreaks have made molecular markers important tools for epidemiologic investigations of this infection. The use of nucleotide sequence patterns allows the investigator to determine whether viruses from the same or different locations are related to each other. Case patients with identical sequence patterns and similar epidemiologic characteristics usually suggest a common-source exposure (112, 170). However, there have been instances where sequence analysis suggested that temporarily related common-source outbreaks might have been related, when in fact they represented different transmission incidents. This occurred because of the somewhat limited variability of the HAV nucleic acid sequence and the inability of the relatively short sequenced fragments to precisely differentiate the genetic relatedness of HAV isolates. In these instances, outbreak strains had identical or nearly identical sequences which represented similar geographical sources of the infecting virus. However, the epidemiologic investigation showed that the outbreaks were not related (3, 63; CDC, unpublished data).

Molecular epidemiologic investigations have to be planned and conducted with equal attention being paid to the methodological requirements of both disciplines: epidemiology and molecular biology. The use of molecular biomarkers alone will not fix a flawed epidemiologic investigation, and patterns of transmission suggested by patterns of HAV genetic relatedness may not be identified if the epidemiologic information was not collected at the time of the initial investigation.

Population-based molecular epidemiologic studies have shown that HAV is often transmitted within networks of persons with similar risk factors for infection (97, 143, 170, 232, 252). Certain sequence patterns appear to cluster among persons with particular risk factors and may be used as an indicator of the source of infection. For instance, in the United States, outbreaks among MSM have often shared one sequence pattern in subgenotype IA, and most cases associated with injection drug use have shared certain “signature” patterns (170, 249).

Molecular epidemiology has become a particularly powerful tool for the investigation of food-borne outbreaks of disease. The close or identical genetic relatedness of isolates from cases in different locales has provided the link to what previously would have been considered sporadic, yet independent, outbreaks. In addition, the availability of a large and ever-growing database of HAV sequences has allowed the identification of the geographic source of virus contamination (3, 112, 260). Recognition that geographically separated outbreaks may be linked by a food source and identification of the potential geographic origin of the contamination have accelerated the public health response to these outbreaks.

PREVENTION OF HEPATITIS A

The availability of vaccines to provide long-term immunity against HAV infection has the potential to significantly reduce disease incidence and possibly eliminate infection transmission (13, 28, 107, 159). A dramatic effect of widespread childhood hepatitis A vaccination has been observed in the United States. Significant disease reductions have occurred in vaccinated populations that historically had the highest disease incidence, namely, American Indians, Alaskan Natives, and persons living in the western United States (14, 23, 30, 32, 163, 215, 255).

The significant reduction in hepatitis A incidence in the United States due to immunization has also changed the epidemiology of this infection. Prior to childhood hepatitis A vaccination, the highest disease incidence occurred among children (28, 29, 159, 215). Now the highest incidence of dis-
case is among adults, a substantial proportion of whom have risk factors for infection, such as IDUs and MSM (11, 28, 32, 170, 252). In addition, childhood immunization appears to have blunted or possibly even eliminated the cyclical nature of hepatitis A in the United States (13, 29, 163). However, at this time, hepatitis A immunization appears not to have changed the epidemiology of sporadic cases of the disease.

What might become the role of molecular diagnostics in the era of hepatitis A immunization? One role would be the continued use of molecular epidemiology to identify chains of transmission, including the extent of outbreaks among IDUs and MSM. Another would be the expanded use of molecular epidemiology to investigate small outbreaks or sporadic cases in different locales and determine if they are related, such as through the distribution of contaminated food. In addition, most food-borne outbreaks should be investigated to determine the genetic relatedness of HAV isolates to improve the identification of the source of contamination and the extent of the distribution of contaminated food. The use of molecular epidemiology to investigate hepatitis A outbreaks has provided a powerful tool which has improved the public health response to these situations. However, it has also meant that forensic practices (e.g., a documented chain of specimen custody) must be used during the investigation, because these data have eventually been entered into evidence in court cases.

Molecular diagnostics should become more widely used to assess the effectiveness of hepatitis A vaccination as disease incidence declines to very low levels. Because of the high proportion of asymptomatic HAV infections, nucleic acid amplification techniques will be required to determine the extent to which unidentified infection occurs. In addition, it will be very important to characterize the genetic makeup of HAV from immunized persons who may subsequently become infected. Such investigations would identify the possible establishment of antibody-resistant mutants, which may have a selective advantage. Such investigations would identify the possible establishment of antibody-resistant mutants, which may have a

ACKNOWLEDGMENTS

The following people or groups provided specimens or sequences that have not been previously published: Daniel Shouval, Nili Daudi (Cairo University, Cairo, Egypt), Michael O. Favorov (CDC-CAR, Liver Unit, Hadassah University Hospital, Jerusalem, Israel), Ala Blumer, G. D. Young.

REFERENCES


Downloaded from http://cmr.asm.org/ by guest on January 11, 2021.


Schultheiss, T., Y. Y. Kusov, and V. Gauss-Muller. 1994. Proteinase 3C of hepatitis A virus (HAV) cleaves the HAV polyprotein P2-P3 at all sites within the HAV genome. J. Virol. 68:908–912.


