A hepatitis E virus variant from the United States: molecular characterization and transmission in cynomolgus macaques

James C. Erker, Suresh M. Desai, George G. Schlauder, George J. Dawson and Isa K. Mushahwar

Virus Discovery Group, Experimental Biology Research, Abbott Laboratories, Dept 90D, Bldg L3, 1401 Sheridan Road, North Chicago, IL 60064, USA

The partial sequence of a hepatitis E virus (HEV-US1) isolated from a patient in the United States (US), suffering from acute viral hepatitis with no known risk factors for acquiring HEV, has been reported. These sequences were significantly different from previously characterized HEV isolates, alluding to the existence of a distinct human variant. In this paper, we report the near full-length sequences of HEV-US1 and a second US isolate (HEV-US2). HEV-US2 was identified in a US patient suffering from acute viral hepatitis. These sequences verify the presence of a new HEV strain in North America and provide information as to the degree of variability between variants. The HEV-US nucleotide sequences are 92% identical to each other and only 74% identical to the Burmese and Mexican strains. Amino acid and phylogenetic analyses also demonstrate that the US isolates are genetically distinct, suggesting the presence of three genotypes of HEV. Serum from the second US patient induced hepatitis following inoculation into a cynomolgus macaque. Within 2–4 weeks, HEV-US2 RNA was detectable in both the serum and faecal material coinciding with elevated serum alanine transaminase levels. Infection resolved as antibody titres increased 8 weeks post-inoculation.

Introduction

Acute viral hepatitis due to hepatitis E virus (HEV) infection is a serious problem in developing countries where this virus is endemic. In these areas, HEV epidemics and sporadic infections, as a result of faecal contamination of drinking water, are common (Bradley et al., 1986; Bradley, 1990; Mushahwar & Dawson, 1997). Serological and molecular analyses have suggested that outbreaks in various developing countries have been associated with isolates which are genetically related to the prototype Burmese variant. Full-length Burmese-like HEV sequences have been cloned from patients in Burma (Tam et al., 1991), India (Chauhan et al., 1994), China (Aye et al., 1992; Bi et al., 1993; Yin et al., 1994) and Pakistan (Tsarev et al., 1992). In addition, many partial sequences from these regions, Central Asia and Northern Africa have also been reported (Chatterjee et al., 1997; Tsarev et al., 1992; van Cuyck-Grande et al., 1997).

These sequences share approximately 89% regional nucleotide identity and greater than 92% identity across the entire genome. A second HEV strain has been associated with a single outbreak in Mexico in 1986 (Huang et al., 1992). The Mexican isolate shares approximately 75% nucleotide identity to the Burmese-like isolates.

Two immunodominant epitopes have been identified in HEV, one at the carboxyl end of open reading frame 2 (ORF2) and one at the carboxyl end of ORF3 (Huang et al., 1992; Khudyakov et al., 1993; Yarbough et al., 1991). Immunoassays using Burmese and Mexican derived recombinant proteins and peptides show a considerable degree of cross-reactivity (Mast et al., 1998; Yarbough et al., 1991). Commercial and experimental immunoassays are often based on the Burmese sequence alone and their performance is largely dependent on this cross-reactivity to identify all HEV related cases (Bradley, 1995). These assays identify anti-HEV in greater than 75% of acute hepatitis cases during epidemics and studies have suggested that up to 30% of healthy individuals living in areas where HEV is endemic are antibody positive (Benjelloun et al., 1997; Goldsmith et al., 1992; Mushahwar et al., 1993; Thomas et al., 1993). For the most part, antibodies to HEV have been...
identified in 0.03–2% of volunteer blood donors in developed countries (Dawson et al., 1992; Mast et al., 1997; Paul et al., 1994). However, higher prevalences of 9–21% have also been reported (Psichogiou et al., 1996; Quirina et al., 1996; Thomas et al., 1997). It is unclear whether the relatively high positivity rates among donors are due to travel to HEV endemic areas, to cross-reactivity with a related antigen, to subclinical HEV infection or to exposure to a divergent strain of HEV.

Recently, HEV related sequences were reported in a patient (USP-1) suffering from acute viral hepatitis in the United States (US). This patient had no recent travel outside the US or other known risk factors associated with HEV infection. Assays based on the Burmese strain of HEV were utilized to evaluate the presence of anti-IgG and anti-IgM directed against HEV, identifying only anti-IgG class antibodies (Kwo et al., 1997). RT–PCR studies verified the presence of a unique isolate of HEV. Approximately 65% of the viral genome has been reported, encompassing the 5’ and 3’ ends of ORF1 and the entire ORF2, ORF3 and 3’ non-translated region (NTR). While the clinical symptoms were typical of acute hepatitis due to HEV infection, the HEV-US1 sequences possessed only 77% nucleotide identity to the Burmese-like and Mexican isolates. Phylogenetic analysis indicated that this isolate was distinct from the Burmese-like and Mexican isolates and represented a unique variant or strain of HEV. Interestingly, HEV-US1 shares 92% nucleotide identity with the partial HEV-US2 sequences recently isolated from swine in the US (Meng et al., 1997; Schlauder et al., 1998).

Because currently available immunoassays for anti-HEV detection were inadequate for the serological identification of HEV-US1 acute infection (Kwo et al., 1997), we developed strain specific ELISAs based on HEV-US1 amino acid sequences (Schlauder et al., 1998). These immunoassays aided in the identification of an additional isolate of the US strain (HEV-US2). This isolate was found in a US patient (USP-2) suffering from acute viral hepatitis of unknown aetiology. In order to determine the genetic variability within this strain and the HEV genera, the near full-length genomic sequences of HEV-US1 and HEV-US2 were determined. Cynomolgus macaques were inoculated with HEV-US RNA positive sera to characterize HEV-US infection and establish an animal model.

**Methods**

**Patient history.** The clinical presentation of USP-1 has been described previously (Kwo et al., 1997; Schlauder et al., 1998). USP-2 was a 62-year-old male admitted to a US hospital with jaundice, elevated serum alanine transaminase (ALT; 1270 IU/L) and fatigue. Serological markers for hepatitis A virus, hepatitis B virus and hepatitis C virus were absent, though the patient had travelled to Cancun, Mexico. USP-2 serum was strongly reactive to HEV-US1 based ELISAs for anti-HEV-IgG and anti-HEV-IgM (Schlauder et al., 1998).

**Genomic extension.** The HEV-US1 and HEV-US2 genomes have been extended to near genome length by RT–PCR (primer sequences can be obtained upon request). Briefly, nucleic acids extracted from human sera using a Total Nucleic Acid Extraction kit (United States Biochemical), QIAmp RNA kit (Qiagen) or Ultraspec RNA isolation system (Biotec) were reverse transcribed using random hexamers and the GeneAmp RT–PCR kit according to the manufacturer’s protocol (Perkin Elmer). These cDNAs were used as the template for the following PCR methodologies. Amplifications utilized the GeneAmp RT–PCR kit with final concentrations of 2 mM MgCl₂ and 1 µM of each primer. In some cases, modified PCR reactions utilized the 10× Buffer and Q solution as recommended (Qiagen). For HEV-US2, amplifications involved (Fig. 1): (i) ‘touchdown’ PCR (Roux, 1995) using HEV-US1 specific primers, (ii) anchored PCR utilizing a gene specific primer and an HEV specific primer designed within regions of identity between Burmese-like and Mexican isolates, with the assumption that sequences conserved between these HEV strains are also conserved in the HEV-US strain and (iii) conventional amplification with HEV-US2 specific primers. The 3’ most sequences were obtained using specifically primed cDNA as recommended in the 3’ RACE Kit ( Gibco-BRL). The 5’ most sequences were isolated by 5’ inverse PCR (Zeiner et al., 1994) using RNA extracted from human sera as a 10% macaque faecal suspension. Briefly, cDNA synthesis was performed using the cDNA synthesis kit from Boehringer Mannheim and an HEV-US specific primer, 575a (5’ GCCGGGTTGATGGCAGACCTC 3’). Following second strand synthesis and blunting, cDNAs were purified over glass milk (GeneClean II, Bio101). Eluates were ethanol precipitated in the presence of 20 µg glycogen and 0.1 M sodium acetate. The cDNA pellets were dissolved in 88.5 µL H₂O and 10 µL 10× ligation buffer and 600 units T4 DNA ligase (Promega) were added. Following overnight ligation at 15 °C, 2 µL of each ligation was utilized in a PCR reaction with primers 426s (5’ CGTTGTGCTTTTGCTGAGACC 3’) and 84a (5’ GAAACGGCCGAAACACCCACCA CGC). Nested PCR was performed with primers 484s (5’ CAGCTTAGTTGAGGCTATGG 3’) and 78a (5’ GCCGAACCCACACGATCACC 3’). Cycling conditions involved 94 °C for 1 min followed by 35 replicates of 94 °C, 20 s; 55 °C, 30 s; and 72 °C, 1 min. Reactions were held at 72 °C for 10 min and then chilled. For the isolation of HEV-US1 sequences between nucleotides 1349 and 3850, amplifications involved (Fig. 1): (i) anchored PCR utilizing a gene specific primer and an HEV-US2 specific primer, (ii) ‘touchdown’ PCR using HEV-US2 specific primers and (iii) conventional amplification with HEV-US1 specific primers. PCR products were cloned into pT7Blue T-Vector (Novagen) and multiple clones sequenced using an Applied Biosystems model 373 DNA sequencer.

**Sequence analysis.** Consensus sequences were generated and analysed using the programs of the Wisconsin Sequence Analysis Package (Genetics Computer Group, Madison, WI, USA, version 9), the Phylip package (Felsenstein, 1993) and TreeView (Page, 1996) as previously described (Schlauder et al., 1998). The sequences used were: Mexico (M1, GenBank accession no. M74506), Burma (B1, M73218; B2, D10330), Pakistan (P1, M80581), China (C1, D11092; C2, L25547; C3, M94177; C4, D11093) and India (I1, X98292; I2, X99441; I3, AF076239). Partial sequences from the hypervariable region included Morocco (MOR12, AF010429; MOR23, AF010428), Tashkent, Uzbekistan (TASH, AF010427) and Osh, Kirgizia (OSH, M84803). The US sequences reported here are: HEV-US1 [AF006066 (previously AF035436 and AF035437)] and HEV-US2 (AF006069).

**Animal transmission studies.** Cynomolgus macaques (Macaca fascicularis) were obtained through the Southwest Foundation for Biomedical Research (SFBR) in San Antonio, Texas, USA. These animals were maintained and monitored in accordance with guidelines established by SFBR to ensure humane care and the ethical use of primates. Sera were obtained twice weekly for at least 4 weeks prior to inoculation in order
to establish the baseline levels for serum ALT. Cut-off (CO) values were determined based on the mean of the baseline plus 3.75 times the standard deviation. Two macaques were intravenously inoculated with 0.4–0.625 ml of HEV positive USP-1 serum and one macaque was inoculated with 2.0 ml of HEV positive USP-2 serum. Serum and faecal samples were collected twice weekly for up to 16 weeks post-inoculation (p.i.). Sera were tested for changes in ALT and values greater than the CO were considered positive and suggestive of liver damage. Antibodies to HEV were assessed via ELISA (Schlauder et al., 1998). Sera and faecal samples were tested for HEV RNA by RT–PCR. Macaque sera or 10% faecal suspension (25–100 μl) were extracted and reverse transcribed as above. PCR reactions contained 2 mM MgCl₂ and 0.5 μM of each oligonucleotide primer as per the manufacturer (Perkin-Elmer). The PCR primers were designed within regions of identity between Burmese, Mexican and HEV-US isolates. These degenerate primers are located within ORF2 between nucleotides 6348 and 6544 of HEV-US2. First round amplification utilized HEVORF2con-s1 (GACAGAATTRATTTCGTCGGCTGG) and HEVORF2con-a1 (CTTGTTCRTGYTGGTTRTCTATAATC) to produce a product of 197 nucleotides. Nested reactions used HEVORF2con-s2 (GTYTCTCRGCACATGGCGAGC) and HEVORF2con-a2 (GTTCTCAGTGGTTRTCTATAATC) to produce a 145 nucleotide product. Amplified products were separated on a 1.5% agarose gel and analysed for the presence of PCR products of the appropriate size. Such products were sequenced directly using the methods described above.

### Results

#### Genomic analysis

The near full-length genomes of HEV-US1 and HEV-US2 have been sequenced. The HEV-US1 genome [excluding the poly(A) tail] was 7186 nucleotides in length starting at position 55 of the original Burmese isolate (B1). HEV-US2 was 7251 nucleotides in length, the longest of the reported HEV genomes. Overall, HEV-US1 possessed 92% nucleotide identity to HEV-US2. These sequences were only 73.5–74.5% identical to the full-length sequences from Burma, India, China, Pakistan and Mexico. This was close to the 75–76% nucleotide identity observed between the Mexican isolate and any one of the Burmese-like isolates (Table 1).

Although primer extension experiments were used to verify the length of the B1 5’ NTR (Tam et al., 1991), similar experiments on HEV-US material were unsuccessful (data not shown). However, inverse RT–PCR on USP-2 serum as well as macaque faecal material revealed the presence of 8 nucleotides at the extreme 5’ end of the genome not previously identified for other isolates. Thus, the 5’ NTR of HEV-US2 was 35 nucleotides in length. The overlapping region of the 5’ NTR was highly conserved with only two variable nucleotide positions between HEV-US2 and the ten reported 5’ NTR sequences from Burmese-like isolates. The 5’ NTR has not been reported from the Mexican strain. The 3’ NTR sequences were less conserved than the 5’ NTR. There was an insertion of 6 nucleotides within both of the HEV-US 3’ NTR sequences when aligned to any of the Burmese isolates. The Mexican isolate (M1) had an additional 9 nucleotides when compared to B1, at the same relative position. Across this region, approximately 87% nucleotide identity was observed between the HEV-US isolates; however, identity of the HEV-US isolates to the Burmese-like and M1 isolates was near 74%. While greater than 95% identity was seen between Burmese-like isolates across this same region, there was only 77% identity between the Burmese-like isolates and M1.

Several distinct features of the HEV-US isolates have emerged from this work. The initiating methionine of the HEV-US1 ORF1 has not been isolated; however, similarity to HEV-
Table 1. Pairwise comparisons and genetic distances of the full-length nucleotide sequences

Isolates of the three HEV genotypes are separated by lines. Shading indicates the percentage identities and genetic distances within potential subtypes of the Burma genotype.

<table>
<thead>
<tr>
<th>US1</th>
<th>US2</th>
<th>B1</th>
<th>B2</th>
<th>I2</th>
<th>I3</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>P1</th>
<th>C4</th>
<th>I1</th>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>97.5</td>
<td>98.5</td>
<td>98.5</td>
<td>98.5</td>
<td>98.5</td>
<td>98.5</td>
<td>98.5</td>
<td>98.5</td>
<td>98.5</td>
<td>98.5</td>
<td>98.5</td>
<td>98.5</td>
<td>98.5</td>
</tr>
</tbody>
</table>

Table 2. Pairwise comparison of amino acid sequences encoded by ORF1, ORF2 and ORF3, and the nucleotide sequences of the hypervariable region

US2 suggested the absence of 9 amino acids at the amino terminus. The complete HEV-US2 ORF1 polyprotein was 1708 amino acids. The HEV-US1 and HEV-US2 polyproteins were 97.8% identical to each other and nearly 85% identical to the Burmese-like and M1 sequences (Table 2). These polyproteins were 14–17 amino acids longer than the Burmese and
Mexican ORF1 polyproteins. This variability was due to nucleotide insertions (Fig. 2) within the hypervariable/proline-rich hinge region (nucleotide positions 2154–2384 of HEV-US2). Within this region, HEV-US isolates contained a unique poly(C) tract. Modified RT–PCR protocols using USP-1, USP-2, or macaque serum or macaque faecal material consistently

<table>
<thead>
<tr>
<th>US1</th>
<th>US2</th>
<th>MOR12</th>
<th>MOR23</th>
<th>B1</th>
<th>B2</th>
<th>I2</th>
<th>I3</th>
<th>OSH</th>
<th>TASH</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>P1</th>
<th>I1</th>
<th>M1</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>acatc</td>
<td>acctc</td>
<td>a</td>
<td>a</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>t</td>
<td>a</td>
<td>g</td>
<td>g</td>
<td>a</td>
<td>c</td>
<td>acac</td>
<td>cttaa</td>
</tr>
<tr>
<td>tttc</td>
<td>tttc</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>g</td>
<td>t</td>
<td>c</td>
<td>g</td>
<td>t</td>
<td>c</td>
<td>t</td>
<td>g</td>
<td>t</td>
<td>t</td>
<td>tt</td>
<td>c</td>
</tr>
<tr>
<td>cc</td>
<td>cc</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>c</td>
<td>t</td>
<td>c</td>
<td>t</td>
<td>tt</td>
<td>t</td>
</tr>
<tr>
<td>gcagc</td>
<td>gcagc</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>c</td>
<td>t</td>
<td>c</td>
<td>t</td>
<td>ct</td>
<td>c</td>
</tr>
<tr>
<td>gacgc</td>
<td>gacgc</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>g</td>
<td>c</td>
<td>c</td>
<td>t</td>
<td>c</td>
<td>t</td>
<td>ct</td>
<td>c</td>
</tr>
<tr>
<td>GC</td>
<td>GC</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>GACTG</td>
<td>GACTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>CTTG</td>
<td>CTTG</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

Fig. 2. Nucleotide sequence alignments of the hypervariable region from thirteen full-length HEV sequences and four partial sequences. The nucleotide positions are based on the HEV-US2 sequence. The consensus sequence (Cons) was determined based on the nucleotide which occurs most frequently in a given position. Dashes (–) indicate agreement with the consensus while gaps in the alignment are shown as a period (.). Bases are shown only at those positions in the alignment which differ from the consensus. The shaded area denotes the poly(C) tract of HEV-US1 and HEV-US2.
amplified products with poly(C) tracts of various lengths. Regardless of this observation, there was very low nucleotide or amino acid identity between geographically distributed isolates within the hypervariable region (Table 2). Across this region, the HEV-US isolates were 85% identical to each other at the nucleotide level and only 31–42% identical to the Burmese-like and M1 isolates. Similarly, most Burmese-like isolates possessed 88–99% identity to each other and only 33–38% identity to M1. The Burmese-like isolates from Morocco were slightly more divergent, exhibiting approximately 75% identity to other Burmese-like isolates. Analysis of the nucleotide alignment across the hypervariable region (Fig. 2) demonstrated the genetic grouping of geographically distributed isolates. The consensus sequence was essentially derived from the 14 Burmese-like isolates. The two HEV-US isolates had many common nucleotide substitutions and insertions differing from the consensus. The single Mexican isolate also possessed many nucleotide substitutions which were unique from the Burmese-like and HEV-US sequences.

Within the HEV-US ORF1 polyproteins, motifs have been identified suggesting protein function as has been seen in other HEV isolates, and many other positive-strand RNA viruses (Fry et al., 1992; Gorbalenya et al., 1991; Koonin et al., 1992). For example, within HEV-US2 a conserved methyltransferase (amino acids 1–240), helicase (975–1207), RNA-dependent RNA polymerase (1222–1708), Y domain (216–433) and X domain (799–957) have been identified. The presence of a putative papain-like protease was identified between amino acids 1–240, helicase (975–1207), RNA-dependent RNA polymerase (1222–1708), Y domain (216–433) and X domain (799–957) have been identified. The presence of a putative papain-like protease was identified between amino acids 433 and 706 (Koonin et al., 1992). The cysteine proposed as the protease active site (amino acid 483), based on limited identity to the putative rubella virus papain-like protease (Koonin et al., 1992), was conserved in all full-length HEV sequences. In comparison, the proposed histidine residue at position 590 was not conserved. Both of the US isolates, M1 and several Burmese-like isolates possessed a tyrosine or leucine at this position. Therefore, it is likely that there is an alternate catalytic histidine residue, such as that conserved at position 671, or that each strain or isolate of HEV may utilize different histidine residues.

The ORF2 or capsid protein was the most highly conserved ORF protein (Table 2). This protein was 660 amino acids in length for the HEV-US and Burmese-like isolates, while the M1 sequence was 659 amino acids. The HEV-US isolates shared 98·3% amino acid identity across this region. The HEV-US ORF2 proteins were approximately 93% identical to the Burmese-like ORF2 proteins and 91·5% identical to the M1 ORF2 protein. These were similar to the observed identities between Burmese-like ORF2 proteins (98·2–99·7%) and between the Burmese-like and M1 isolates (94–95%). A hydrophobic signal peptide sequence was identified at the extreme amino terminus of the protein, with potential cleavage between positions 18 (proline) and 19 (alanine) as predicted by the method of Nielsen et al. (1997). Also, as noted for other HEV isolates (Huang et al., 1992; Tam et al., 1991), the isoelectric point of the amino half of the protein was quite high for both HEV-US1 (11·03) and HEV-US2 (10·98). The immunodominant epitope located at the carboxyl end of the protein (designated 3-2) was highly conserved between the US, Burmese-like and Mexican isolates as was the overall hydrophobicity profile (data not shown).

The ORF3 protein was the least conserved ORF product between the two HEV-US isolates, with 95·9% amino acid identity (Table 2). This was slightly more variable than the amino acid identity observed between Burmese-like products (96·8–100%). The HEV-US ORF3 proteins were approximately 84% and 80% identical to the Burmese-like and M1 proteins, respectively. The Burmese-like ORF3 products were more similar to the M1 ORF3, with 85·4–87% identity. Both the HEV-US isolates had a unique 3 nucleotide deletion just downstream of the ORF3 ATG producing a protein of 122 amino acids versus the 123 amino acid product identified in the Burmese-like and Mexican isolates. The hydrophobicity profiles were very similar between the HEV-US, Burmese-like and Mexican isolates with two hydrophobic domains at the amino half of the protein (data not shown). Within the 4·2 epitope located at the carboxyl terminus of ORF3, only 1 of 33 positions differed between the two HEV-US isolates and 2 of 33 positions differed between ten Burmese-like isolates. However, there was a significant degree of amino acid sequence variability between the HEV-US and Burmese-like isolates (10 of 33), HEV-US and M1 (12 of 33) and the Burmese-like and M1 (8 of 33). Most of the variability was localized within the amino half of the epitope. This apparent geographically distributed sequence variation may affect the ability to identify HEV infection or exposure with current immunoassays.

Phylogenetic analysis

To clarify the relationship between the HEV isolates, phylogenetic analysis was performed. Full-length HEV nucleotide sequences were aligned and phylogenetic distances determined using the region of overlap to the shortest sequence (Table 1). Comparison of HEV-US1 and HEV-US2 revealed a genetic distance of 0·0805 nucleotide substitutions per position. This was similar to the distances between the most divergent of the closely related Burmese-like isolates (0·0120–0·0855). The US isolates had distances greater than 0·3295 to either Burmese-like or Mexican isolates, which is similar to those between the Burmese-like and Mexican isolates (greater than 0·3053).

Genetic distances were used to generate the unrooted phylogenetic tree shown in Fig. 3. HEV-US1 and HEV-US2 grouped closely on one major branch. The ten Burmese-like isolates formed a cluster near the end of a second major branch. The Mexican isolate grouped independently on a third major branch. Bootstrap values produced from 100 random resamplings of the data demonstrate strong support for these groups. The only groupings not supported in 100% of the
trees were those internal to the Burmese-like cluster; however, many of these values were similar to or greater than the 70% significance CO value. Similar groupings were identified by Schlauder et al. (1998) using sequences of a concatemer encompassing 65% of the HEV-US1 genome as well as portions of ORF1, ORF2 or ORF3 alone. Similar groups were also observed using 231 nucleotides across the hypervariable region (Fig. 2) and the 4–2 epitope region. These isolates, therefore, form three distinct geographically distributed genotypes (Fig. 3): the Burmese-like isolates from Asia and Africa are genotype 1, the Mexican isolate is genotype 2 and the US isolates form genotype 3. Also of interest were the observed genetic distances and groupings of the Burmese-like isolates. Analysis of full-length or partial nucleotide sequence alignments consistently revealed three subgroups: isolates B1, B2, I2 and I3 formed one group, the Chinese and Pakistani isolates another, and I1 grouped independently. Genetic distances within these groups were less than 0.0476 substitutions per position and distances between subgroups were greater than 0.0577 (Table 1). This suggests that the Burmese-like isolates can be further classified into three distinct subtypes: a, b and c (Fig. 3).

Transmission study

Although intravenous inoculation of 0.4–0.625 ml of USP-1 sera into two cynomolgus macaques failed to produce infection (data not shown), inoculation of 2.0 ml USP-2 sera resulted in viraemia and elevations of liver enzyme levels in the serum (Fig. 4). HEV RNA was first detected in faecal material on day 15 p.i. and remained positive through 64 days p.i. Serum specimens collected between days 28–56 p.i. were HEV RNA positive. Elevated ALT values were noted on days 15, 44–58, 72 and 93 p.i., with the peak ALT value (116 IU/l) on day 51 p.i.

Six strain specific ELISAs were utilized to assess antibody response. Measurable response was found only to the US 3-2 peptide assay with no noted cross-reactivity to the Burmese or Mexican peptides. IgM class antibody directed against HEV was detectable between 28 and 58 days p.i. This was followed by a strong anti-HEV-IgG response at day 44 p.i.

Discussion

The near full-length genomic sequences of two HEV isolates of a novel US strain have been characterized. The genomic organization is similar to that of the Burmese and Mexican strains of HEV with the 5’ two-thirds of the genome encoding the putative nonstructural proteins and the 3’ third encoding a capsid protein and a small ORF of unknown function. The individual nonstructural proteins are cleaved from a large polyprotein possessing recognizable motifs for a methy/transferase, helicase and RNA-dependent RNA polymerase. Comparisons with alpha- and rubivirus sequences suggest the presence of an X domain possibly involved in polyprotein processing together with a putative papain-like protease. While the HEV isolates show 73–99% nucleotide identity and greater than 80% amino acid identity to each other, within the hypervariable region of ORF1 there is as
much as 67% nucleotide variability and very low amino acid sequence identity. The significance of this high degree of variability within a nonstructural region is unclear. Interestingly, the two HEV-US isolates are more variable across the hypervariable region, ORF3 and the 3' NTR than are the most divergent Burmese-like isolates. Whether this is an artifact of the number of sequences available for analysis or an actual difference between HEV isolates from developed regions versus those from developing countries awaits further studies.

The function of the ORF3 protein is uncertain; however, several characteristics have been investigated. A recent report by Zafrullah et al. (1997) has suggested that the Burmese ORF3 product is a phosphoprotein modified at serine-80. While the position of the modified serine is conserved (serine-79 in the HEV-US ORF3 proteins), the consensus sequence for MAPK, the implicated kinase (proline-X-serine/threonine-proline or proline-X-X-serine/threonine-proline), is not completely conserved (proline-methionine-serine-phenylalanine). Neither the serine residue nor the MAPK recognition motif is conserved in the Mexican isolate. Whether modification is not required for protein function or the HEV-US and Mexican proteins are modified through a different mechanism awaits further studies. Also, utilizing the computer program described by Nielsen et al. (1997), we were unable to distinguish the hydrophobic amino terminus either as a eukaryotic signal peptide or as an anchor sequence. Since no significant processing of expressed ORF3 proteins has been observed (Jameel et al., 1996) and an association with the cytoskeleton has been identified (Zafrullah et al., 1997), it is likely that this is a membrane anchor region.

Nucleotide and amino acid sequence alignments, and phylogenetic analysis, suggest the presence of three distinct genetic groups of HEV isolates. The first group (genotype 1) represents isolates closely related to the original HEV isolate from Burma. These genomic sequences show greater than 92% nucleotide identity to each other with genetic distances less than 0.0850 nucleotide substitutions per position. Interestingly, this group may be further divided into three subtypes, a, b and c, distinguished by sequences with greater than 95.4% nucleotide identity with distances less than 0.0476 substitutions per position and those with less than 94.5% identity with distances greater than 0.0577 substitutions per position (Table 1). Characterization of small fragments of additional isolates from Northern Africa and Central Asia suggest the presence of at least two additional subtypes, d and e, respectively, within genotype 1 (Chatterjee et al., 1997; van Cuyck-Grande et al., 1997). The second group (genotype 2) is represented by the single Mexican isolate. This isolate shares less than 76.1% nucleotide sequence identity and genetic distances greater than 0.3053 substitutions to the Burmese-like isolates. The third group, genotype 3, is represented by HEV-US1 and HEV-US2. These isolates share less than 74.3% nucleotide sequence identity and genetic distances greater than 0.3295 substitution to the Burmese-like and Mexican isolates. The nucleotide identity and genetic distance observed between the two HEV-US isolates are similar to those between the subtypes of Burmese-like isolates, suggesting that HEV-US1 and HEV-US2 each represent unique subtypes of genotype 3, a and b, respectively.

The development of an animal model for the HEV-US strain is useful for the virological, serological and biological characterization of this strain relative to other HEV strains. Burmese-like and Mexican isolates have been inoculated into several species of non-human primate, typically by intravenous injection of a faecal suspension (Ticehurst et al., 1992; Tsarev et al., 1992). Faecal material from USP-1 and USP-2 was not available for study; therefore sera were used as the inocula. The USP-1 inoculum did not produce an active infection, potentially due to low titre (Schlauder et al., 1998). The USP-2 inoculated macaque showed patterns of virus production and shedding, elevated liver enzymes characteristic of liver damage, and an immune response very similar to those of animals inoculated with other HEV strains. The infection of an inoculated macaque with HEV RNA positive serum identifies an alternate mode of transmission other than the known faecal/oral route.

Recently, the partial sequence of a domestic US swine isolate of HEV was reported (Meng et al., 1997). This sequence was most closely related to HEV-US1, sharing 92% nucleotide identity, and grouped closely on a phylogenetic tree (Schlauder et al., 1998). HEV-US2 was also 92% identical to the swine isolate. All three isolates encode an ORF3 product of 122 amino acids, each possessing a common deletion when compared to B1. The ORF2 product is 660 amino acids, similar to the HEV-US and Burmese-like isolates. The similarity of the swine sequences to HEV-US suggests the ability of HEV to infect multiple species with a potential for zoonotic infections. A number of animals including rats, swine, lambs and wild-caught non-human primates have been shown to be susceptible to or exposed to HEV by various methods (Balayan et al., 1990; Bradley et al., 1987; Clayson et al., 1996; Maneerat et al., 1996). These data suggest that animals other than humans could serve as a worldwide reservoir for HEV.

The significance of the amino acid variability of the epitope region 4-2 and 3-2 to diagnostic immunoassay design needs further investigation. There is little amino acid variability within the 3-2 epitope; however, the 4-2 epitope contains several strain specific amino acid substitutions. In correlation with this sequence variation, there is a preferred reactivity of HEV-US infected patient sera and macaque sera to the US peptides relative to the Burmese and Mexican based peptide ELISAs (Schlauder et al., 1998). Immunoassays utilizing US specific recombinant proteins are currently being evaluated to determine sensitivity and cross-reactivity, as these assays may be more specific than peptide based ELISAs (Mast et al., 1998). While antibody titre may affect the ability to detect cross-reactivity, one cannot ignore the potential for multiple
serotypes of HEV. Thus, the number of acute hepatitis cases due to HEV infection, particularly in developed countries, may be underestimated. Serological studies in the US and Europe have demonstrated an unexplainable 2–10% of IgG positive, IgM negative individuals (Mushahwar et al., 1996). The IgG response was markedly lower than that of recently infected individuals, suggesting previous exposure to HEV or infection with an unknown HEV strain.

We would like to thank Mark F. Knigge for his excellent technical assistance and Dr A. Scott Muerhoff for his critical review of the manuscript.

References


Received 24 September 1998; Accepted 17 November 1998