Molecular cloning and characterization of a GB virus C isolate from a patient with non-A–E hepatitis

James C. Erker, John N. Simons, A. Scott Muerhoff, Thomas P. Leary, Michelle L. Chalmers, Suresh M. Desai and Isa K. Mushahwar

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

Recently, the isolation of a novel virus, GB virus C (GBV-C), associated with cryptogenic hepatitis has been reported. Following the molecular cloning of this virus genome, it became apparent that the genomic sequence did not encode a protein resembling a nucleocapsid or core-like protein similar to those observed in other flaviviruses, pestiviruses, hepatitis C virus (HCV) and GB virus B. Similar findings were subsequently observed in the cloning of two viral genomes representing isolates of GBV-C, namely hepatitis G virus (HGV). To verify the presence or absence of a viral nucleocapsid protein, identify conserved protein motifs and determine the overall genomic variability, an additional virus isolate has been characterized. Here we report the full-length genomic sequence of GBV-C(EA), isolated from an East African suffering from acute non-A–E hepatitis. GBV-C(EA) was compared with the prototype West African isolate (GBV-C) and the two HGV isolates from the United States. The analyses demonstrate several characteristics of these novel viruses. (1) The degree of variability within the 5′ nontranslated region (NTR) approximates that observed between HCV isolates. (2) The nucleotide sequence of the coding region and the 3′ NTR is highly conserved between these isolates, in contrast to the extensive variability observed between HCV isolates from distinct geographical locations. (3) There is a high degree of amino acid conservation across the precursor polyproteins of these isolates; most striking is the lack of ‘hyper-variable’ regions within the envelope proteins. (4) There appears to be no nucleocapsid protein near the amino terminus of the GBV-C/HGV polyproteins.

Introduction

Utilizing oligonucleotide primers designed to amplify the NS3 genes of hepatitis C virus (HCV), GB virus A (GBV-A) and GB virus B (GBV-B), a novel virus was isolated from a West African individual (Simons et al., 1995). This virus, termed GB virus C (GBV-C), appears to be a member of the family Flaviviridae, most similar to GBV-A. It possesses an RNA genome 9377 nucleotides in length which contains a single long open reading frame (ORF). Structural and non-structural proteins are positioned at the amino terminus and carboxy terminus of the putative polyprotein, respectively (Leary et al., 1996 a). Subsequently, Linnen et al. (1996) reported the independent cloning of hepatitis G virus (HGV) from two United States (US) patients. Sequence analysis reveals that HGV is virtually identical to GBV-C (approximately 95% amino acid identity), and thus GBV-C and HGV are distinct isolates of the same virus (Zuckerman, 1996).

GBV-C/HGV has been detected in several human populations. Approximately 1% of US blood donors are GBV-C/HGV RNA positive. As expected for a transfusion transmitted agent, a higher percentage of positive specimens was found among intravenous drug users (33%) and multiply transfused patients (18%) (Linnen et al., 1996). Evidence for the association of GBV-C/HGV with hepatitis is manifold. GBV-C/HGV RNA is found at higher rates among chronic HBV and HCV carriers, as well as in cases of acute hepatitis, cryptogenic cirrhosis, hepatocellular carcinoma and fulminant hepatitis (Simons et al., 1995; Leary et al., 1996 b; Linnen et al., 1996; Yoshiha et al., 1995). Compelling evidence for hepatitis caused by GBV-C/HGV is found in patients who became GBV-
C/HGV RNA positive and developed hepatitis subsequent to being transfused with GBV-C/HGV RNA positive units of blood (Linnen et al., 1996).

One of the first GBV-C positive patients identified was an East African child diagnosed with acute non-A-E hepatitis (Simons et al., 1995). Here we report the genome length viral sequence from this patient [GBV-C(EA)]. Alignments of GBV-C(EA) with GBV-C and the two HGV sequences allow for characterization from serum collected at the time of presentation. The consensus nucleotide at each position was determined by the base that occurs most frequently at that position. The dashes (−) represent nucleotides identical to those determined in the consensus line. Base deletions are indicated by periods (.) For each isolate, nucleotides are shown only at those positions which differ from the consensus. Sequence encoding the first amino acid of E1 and the conserved Asn-Cys-Cys motif within the putative E1 protein are underlined and designated (a) and (b), respectively. Stop codons interfering with the extension of the ORF up-stream of E1 for each sequence are indicated as asterisks (**). Met residues located at the amino terminus of the encoded polyproteins are indicated (met). The position of the conserved Met is indicated as Met−→ below the consensus line.

### Methods

**Clinical history.** Serum was collected from an East African child with acute non-A-E hepatitis. Serial bleeds spanning 40 days post-presentation showed GBV-C genomic RNA presence despite the normalization of alanine amino transferase levels from 128 U/l to 30 U/l (Simons et al., 1995). The GBV-C(EA) genome was cloned and characterized from serum collected at the time of presentation.

**Genome extension.** Genome extension was performed as previously described (Leafy et al., 1996a) up- and downstream of the original 322 nucleotide NS3 sequence (GBV-C4, GenBank accession no. U25542; Simons et al., 1995). Nucleic acids were extracted from human serum using a total nucleic acid extraction method (DNA/RNA isolation kit; USB). RT-PCR was done with the GeneAmp RNA PCR kit (Perkin Elmer). RT reactions were primed with random hexamers. PCR reactions utilized 0.5 μM of each primer. Extension of the GBV-C(EA) genome utilized gene specific biotinylated primers and random flanking primers as described previously (Sorensen et al., 1993; Leafy et al., 1996a). Overlapping PCR products were also generated using primers derived from the prototype GBV-C sequence (GenBank accession no. U36380; Leafy et al., 1996a) using a ‘touchdown’ PCR protocol described previously (Roux, 1994). PCR products were cloned into pT7Blue T-Vector (Novagen) and at least three clones sequenced using the Sequenase version 2.0 sequencing kit (USB).

The 3′-most sequences were isolated by rapid amplification of cDNA ends (RACE). A polyadenosine tail was added to GBV-C RNA following nucleic acid extraction. Briefly, nucleic acid was dissolved in 5 μl of
DEPC-treated water and mixed with 5-5 μl 1 mM-dATP, 3 μl 5 x poly(A) polymerase buffer (USB) and 1193 units yeast poly(A) polymerase (USB). After incubation at 30 °C for 1 h, the reaction was extracted twice with equal volumes of phenol–chloroform and once with chloroform. Nucleic acids were ethanol precipitated and subjected to the 3’ RACE techniques as directed by the manufacturer (3’ RACE kit; Gibco BRL). Amplification conditions involved 35 cycles of PCR (94 °C, 30 s; 55 °C, 30 s; 72 °C, 60 s; with a final 10 min 72 °C extension) followed by semi-nested PCR with an internal GBV-C(EA) specific primer as above. Products were cloned into p7TBlue T-Vector and two clones sequenced on an ABI model 373 DNA Sequencer using the ABI Sequencing Ready Reaction kit (Perkin-Elmer).

Computer analysis. Sequences were analysed using programs of the Wisconsin Sequence Analysis Package (version 8; Genetics Computer Group). Sequences were compiled and edited in the Fragment Assembly programs. Associations to known viral sequences were identified using BLAST, GAP and PILEUP. Successive 100 amino acid windows of the GBV-C/HGV polyprotein multiple sequence alignment were analysed by PLOTSIMILARITY, with the identity setting in place, to search for overall and regional identity of amino acid sequences.

The following genomic sequences were used: GBV-C/E (GenBank accession no. U63715), updated GBV-C (U63800), HGV PNF2161 (U44402), HGV R10291 (U45966), updated GBV-A (U22303), GBV-B (U22304), HCV1 (M62321) and HCVJ8 (D10988).

### Results

#### Presence or absence of a nucleocapsid protein

The GBV-C/E NS3 sequence (Simons et al., 1995) was extended both up- and downstream to a total of 9367 nucleotides. This sequence, termed GBV-C(EA), contains a single long ORF encoding a 2842 amino acid polyprotein. Using analyses similar to those applied to GBV-C (Leary et al., 1996), GBV-C(EA) does not appear to encode a nucleocapsid or core-like protein at the amino terminus of the polyprotein. The sequence analysis of multiple clones derived from the East African sera by using RT–PCR procedures which reduce RNA secondary structure did not identify additional sequence upstream of the presumed E1 gene which would encode a nucleocapsid protein (data not shown). This suggests that the 5’ ends sequences of GBV-C(EA) are indeed representative of the viral genome and are not a result of RT, PCR or cloning artifacts. Additional analysis of the West African isolate revealed clones with two nucleotide insertions and one nucleotide deletion. The deletion, at nucleotide 489 of the previously reported sequence (Leary et al., 1996), introduced a frameshift which eliminated 31 amino acids from the amino terminus of the GBV-C polyprotein. In contrast to the African isolates, the US HGV isolates have been reported to encode core-like proteins of 71 and 34 amino acids (Linnen et al., 1996). In addition to the inconsistency of size, these proteins are considerably smaller than the core protein of HCV (191 amino acids) and GBV-B (156 amino acids). To explain the apparent anomalies at the 5’ end of the GBV-C/HGV isolates, alignments of the 5’ sequences from GBV-C(EA), GBV-C, HGV PNF2161 and HGV R10291 (Fig. 1) were examined. Single nucleotide insertions or deletions disrupt the ORF resulting in these non-conserved amino acid sequences. There is a single Met codon held in common among the GBV-C and HGV isolates immediately upstream of the presumptive E1 signal peptidase recognition sequence. This Met codon, at nucleotides 524–526 of GBV-C(EA), may represent the site of translation initiation and is used here as the first amino acid of the polyprotein. Thus, nucleotide sequences upstream of this position represent the 5’ nontranslated region (NTR).

#### Nucleotide sequence analysis

The 523 nucleotide 5’ NTR of GBV-C(EA) shows approximately 51% identity to that of GBV-A, with no significant identity to the homologous regions of HCV or GBV-B. Alignment of the 5’ end sequences of the GBV-C/HGV isolates (Fig. 1) shows a high degree of conservation with localized variability. Most variability is limited to short stretches, such as regions between positions 50–63, 159–194, 377–404 and 480–521. Several positions (63 and 159, 217 and 237, 242 and 254, 277 and 416) exhibit covariant substitutions.

### Table 1. Percentage nucleotide identity between GBV-C/HGV isolates

<table>
<thead>
<tr>
<th>Nucleotides</th>
<th>GBV-C</th>
<th>GBV-C(EA)</th>
<th>PNF2161</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) 5' NTR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C</td>
<td>57.5</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>GBV-C(EA)</td>
<td>45.5</td>
<td>88.5</td>
<td>–</td>
</tr>
<tr>
<td>PNF2161</td>
<td>74.5</td>
<td>88.5</td>
<td>93.1</td>
</tr>
<tr>
<td>R10291</td>
<td>1.479</td>
<td>88.5</td>
<td>92.7</td>
</tr>
<tr>
<td>95.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) Coding region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C(EA)</td>
<td>53.4</td>
<td>86.4</td>
<td>–</td>
</tr>
<tr>
<td>PNF2161</td>
<td>552.6</td>
<td>86.1</td>
<td>88.5</td>
</tr>
<tr>
<td>R10291</td>
<td>480.6</td>
<td>86.2</td>
<td>88.6</td>
</tr>
<tr>
<td>90.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) 3' NTR*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBV-C(EA)</td>
<td>90.66</td>
<td>89.4</td>
<td>94.6</td>
</tr>
<tr>
<td>PNF2161</td>
<td>908.1</td>
<td>94.6</td>
<td>98.1</td>
</tr>
</tbody>
</table>

* The 3’ NTR of R10291 has not been reported.

### Table 2. Percentage amino acid identity/similarity of the GBV-C/HGV polyproteins

<table>
<thead>
<tr>
<th></th>
<th>GBV-C</th>
<th>GBV-C(EA)</th>
<th>PNF2161</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBV-C(EA)</td>
<td>90.4</td>
<td>96.2</td>
<td>98.1</td>
</tr>
<tr>
<td>PNF2161</td>
<td>95.5</td>
<td>97.7</td>
<td>98.3</td>
</tr>
<tr>
<td>R10291</td>
<td>99.6</td>
<td>97.7</td>
<td>98.6</td>
</tr>
</tbody>
</table>

All polyproteins are 2842 amino acids except GBV-C which is 2843 due to the insertion of a serine residue within NS5B.

Nucleotide sequence analysis

The 523 nucleotide 5’ NTR of GBV-C(EA) shows approximately 51% identity to that of GBV-A, with no significant identity to the homologous regions of HCV or GBV-B. Alignment of the 5’ end sequences of the GBV-C/HGV isolates (Fig. 1) shows a high degree of conservation with localized variability. Most variability is limited to short stretches, such as regions between positions 50–63, 159–194, 377–404 and 480–521. Several positions (63 and 159, 217 and 237, 242 and 254, 277 and 416) exhibit covariant substitutions.
which may be vital in conserving secondary structure of this region. At positions 63, 128, 159, 185, 194, 382, 480 and 521 there are nucleotide substitutions, insertions or deletions held in common between isolates from the same relative geographical location. However, sequence identity within the 5′ NTR (Table 1a) demonstrates that GBV-C(EA) is more closely related to the HGV US isolates than to the West African GBV-C isolate. This relationship is also demonstrated in comparisons of the coding region (Table 1b). GBV-C is the most divergent with 86% identity to any of the other isolates, while the HGV isolates are 90% identical. Interestingly, most of this nucleotide variability within the coding region results in silent or conserved amino acid changes.

The 315 nucleotide 3′ NTR is the most highly conserved region of these viruses, with identities between 94.6% to 98.1% (Table 1c). As in the 5′ NTR, there are covariant substitutions at positions 9079–9085 and 9093–9099 of GBV-C(EA), possibly conserving a secondary structure within this region. A poly(A) or poly(U) tail, similar to those seen in HCV and GBV-B (Han et al., 1991; Muerhoff et al., 1995) has not been identified in GBV-C(EA) or GBV-C, the HGV isolates or GBV-A (Leary et al., 1996a; Linnen et al., 1996; Muerhoff et al., 1995).

Analysis of the polyprotein
Amino acid comparisons of the GBV-C/HGV isolates demonstrate extensive conservation with identities of more than 95% and similarities greater than 98% (Table 2). A more detailed analysis to identify overall and regional sequence variability between these isolates was performed by aligning the four polyproteins and plotting the amino acid identity within 100 amino acid windows of comparison along the
alignment. Fig. 2(a) shows identity greater than the mean of approximately 97% within the NS3, NS4B and NS5B proteins. Areas within the E1, NS2, NS4A and NS5A proteins show conservation slightly lower than the mean. The putative E2 protein exhibits the highest degree of diversity (approximately 8%) across all four sequences.

There are many conserved protein motifs between the GBV-C/HGV envelope proteins and those of HCV, GBV-A and GBV-B. An Asn-Cys-Cys motif is conserved near the amino terminus of E1 between the GBV-C/HGV sequences as well as GBV-A and GBV-B. This is similar to the conserved Asn-Asp-Cys motif of the HCV E1 protein (Muerhoff et al., 1995). There is one conserved potential N-linked glycosylation site in E1 and three in E2 (Fig 3). The N-glycosylation site within E1 appears to be spatially conserved between the GBV-C and HGV isolates, GBV-A, GBV-B and HCV, while those in
E2 appear to be spatially conserved only between the GBV-C/HGV isolates and HCV (Fig. 3). There are 11 Cys residues within E1; four of these have been shown to be conserved between GBV-A, GBV-B, HCV, GBV-C and HGV isolates (Muerhoff et al., 1995; Leary et al., 1996a; Linnen et al., 1996). There is also a fifth Cys conserved between GBV-A, HCV and the GBV-C/HGV sequences at amino acid 129 of GBV-C(EA). Five of the remaining six Cys residues are conserved with GBV-A. Within E2, there are 17 Cys residues conserved between the GBV-C/HGV sequences, all of which are conserved with GBV-A. Besides Cys residues, there are also conserved Pro and Gly residues which may be important to the formation of secondary and tertiary structures of the GBV-C/HGV envelope proteins. Many of these residues are also conserved with the GBV-A envelope proteins (Fig. 3).

The polyprotein of GBV-C/HGV is thought to be processed into individual proteins (Fig. 2b) by both host and viral proteases (Leary et al., 1996a), similar to other members of the Flaviviridae (Dawson et al., 1996a, and references therein). Sequence motifs possibly recognized by host cell signal peptidase are similar to those in HCV (Bukh et al., 1993). The alignment of E2 sequences presented here suggests an alternative E2/p7 signal sequence immediately upstream of the sequence previously reported for GBV-A and GBV-C (Fig. 3). The sequence at amino acids 548–561 of GBV-C(EA) aligns well with HCV and GBV-A, conserving a Leu-Ala-Asp/Glu-Ala-Arg motif at the P4-P1' positions. In addition, the amino acid composition of this sequence corresponds more closely with the eukaryotic signal sequences identified by von Heijne (1986).

Potential cleavage sites for the putative viral serine protease have a conserved Thr or Ala at position P1 and an Asp or Ser at the P1' position (Leary et al., 1996a). Using this motif, an NS4B/NS5A cleavage site has not been identified in these viruses. However, site-directed mutagenesis of the serine protease substrates of HCV has suggested that strict conservation of the P6 through P1' positions may not be necessary for proteolytic cleavage (Kolykhalov et al., 1994; Bartenschlager et al., 1995). Thus, a potential NS4B/NS5A cleavage site which aligns well with HCV and GBV-A sequences is located between amino acids 1864 (Val) and 1865 (Glu) of GBV-C(EA). The sequence surrounding this site possesses a similar hydropathy profile to that of HCV (data not shown), supporting the existence of this cleavage site.

Discussion

We have reported the isolation and genomic sequence of a new GBV-C isolate, GBV-C(EA). This East African isolate possesses many of the attributes of the prototype GBV-C isolate (Leary et al., 1996a). The near full-length genome is 9367 nucleotides with a 5' NTR of 523 nucleotides and a 3' NTR of 315 nucleotides. The putative precursor polyprotein of 2842 amino acids contains many of the protein motifs conserved in other positive-strand RNA viruses. Potential cleavage sites recognized by host and viral proteases divide the polyprotein into distinct structural and nonstructural proteins (Fig. 2b). Conserved protein motifs suggest the presence of two envelope proteins, a p7-like protein, a zinc protease, a serine protease, a supergroup II helicase and a supergroup II RNA-dependent RNA polymerase (Simons et al., 1995; Leary et al., 1996a; Linnen et al., 1996).

The high degree of conservation seen across the entire GBV-C/HGV genome is surprising in view of the fact that GBV-C/HGV is phylogenetically related to HCV. One would reasonably expect to see variability across the genome similar to that seen between HCV isolates. Typically, HCV isolates from geographically distinct regions possess 31–35% nucleotide divergence across the entire genome. In contrast, HCV isolates from a single geographical location vary by less than 10% across the genome (Miyakawa et al., 1995; Simmonds, 1995). These four GBV-C/HGV isolates from three distinct geographical locations bear no greater than 14% nucleotide variability. Analysis of multiple clones as well as clones from multiple serum samples from the same patient do not indicate any significant degree of sequence variability (data not shown). Why is the GBV-C/HGV genome so well conserved?

The low amount of variability between the GBV-C/HGV isolates within the 5' and 3' NTRs and the existence of covariant substitutions suggests that maintenance of the appropriate secondary structures may be critical for viral replication, packaging and translation initiation. The 5' NTR of GBV-C possess no significant identity to the 5' NTR of HCV; however, both viruses possess noted regional conservation with localized variability within this region (Fig. 1; Bukh et al., 1992). This pattern is also observed in the alignment of 5' NTR sequences from 35 GBV-C positive individuals (Muerhoff et al., 1996). Thus, most of the variability seen in this 5' NTR may be due to covariant nucleotide changes within a stem or nucleotide mutations, insertions or deletions within a loop structure. Similar mutations have been identified within the 5' NTR of geographically diverse isolates of HCV (Bukh et al., 1992). The degree of variability is also similar to that (up to 9%) observed between HCV isolates from distinct geographic locations (Simmonds, 1995). In contrast, the 3' NTR sequences of HCV upstream of the poly(U)/poly(A) tract vary greatly, up to 74% between isolates from distinct regions (Simmonds, 1995). However, recently identified sequences downstream of the poly(U)/poly(A) tract in HCV have been found to be highly conserved between geographically distinct isolates (approximately 98%) (Tanaka et al., 1995, 1996; Kolykhalov et al., 1996). This is similar to the 94% conservation observed between the GBV-C/HGV isolates. While these GBV-C/HGV sequences have no identity with the sequences downstream of the poly(U)/poly(A) tract in HCV, the high degree of conservation suggests that this is a functionally important part of the virus. Interestingly, the 3' NTR of GBV-C/HGV bears no identity to other known virus sequences except the 3'-most
sequences of GBV-A. Conservation within this region, as well
as within the 5' NTR, demonstrates the relatedness of GBV-C
to GBV-A and suggests a similar function for these regions.
The conservation within the 5’ and 3’ NTRs may therefore be
driven by strong evolutionary constraints to conserve sec-
ondary structure.

The degree of variability within the HCV and the GBV-
C/HGV polyproteins is similar to that seen across the entire
genome. HCV isolates possess up to 28% amino acid variability
across the entire polyprotein between geographi-
cally distinct isolates (Fig. 2c; Simmonds, 1995). However, the
degree of divergence across the GBV-C/HGV polyprotein is
less than 5%. This is similar to the variability seen between
HCV isolates from a single geographical location. The
envelope proteins of HCV possess up to 45% divergence (Fig.
2c; Simmonds, 1995). There are several 'hypervariable' regions
within the HCV envelope proteins which may allow the virus
to 'escape' neutralizing antibodies (Weiner et al., 1991; Farci et
al., 1992). Variability within the GBV-C/HGV envelope proteins
is at most 10%. There are no discernible GBV-
C/HGV E1 'hypervariable' regions, and while there is more
variability within the GBV-C/HGV E2 proteins, amino acid
substitutions do not cluster into 'hypervariable' regions as
seen in HCV. The relatively few amino acid substitutions
within the putative structural proteins of GBV-C/HGV along
with the persistence of infection noted in several individuals
(Leary et al., 1996b) may indicate a lack of immune selective
pressures exerted on these viruses. Supporting this hypothesis
is the lack of a consistently detectable immune response to any
of the GBV-C proteins expressed to date in infected individuals
(Pilot-Matias et al., 1996; Dawson et al., 1996b). Interestingly, a
serological immune response has not been detected in tamarins
infected with GBV-A (Schlauder et al., 1995; Pilot-Matias et al.,
1996). The lack of variability within the structural as well as the
nonstructural proteins of GBV-C/HGV is suggestive of a
potential proofreading function of the RNA-dependent RNA
polymerase. However, more likely is the possibility that GBV-
C/HGV has reached a point of evolutionary stasis (Holland et
al., 1992). This infers that GBV-C/HGV has become so highly
adapted that a master sequence dominates over the less fit
quasispecies. However, analysis of only four complete geno-
mes generated from analysis of relatively few clones may not
demonstrate the degree of variability actually present within
this group of viruses.

Protein motifs within the nonstructural proteins of GBV-
C/HGV demonstrate the relatedness of this novel virus to other
members of the Flaviviridae. Also, within the structural
proteins, conserved protein motifs can be identified between
the GBV-C/HGV isolates and GBV-A, GBV-B and HCV.
However, in contrast to HCV, flaviviruses, pestiviruses and
GBV-B (Collett et al., 1988; Han et al., 1991; Muerhoff et al.,
1995) the GBV-C/HGV and GBV-A genomes do not appear
to encode a nucleocapsid protein as part of the polyprotein.
GBV-C 5' sequences from 35 individuals are consistent with
this observation. All 5' NTR sequences analysed encode the
Met identified as position one of the GBV-C (EA) polyprotein
(Muerhoff et al., 1996). A Met residue, homologous to this
conserved Met in GBV-C/HGV, is present in GBV-A. The
ability of the 5' ends of GBV-C and GBV-A to direct translation in
a cell-free in vitro translation system (Simons et al., 1996)
supports polyprotein initiation from these Met residues. Thus,
GBV-C and GBV-A may be unique members of the Flaviviridae
lacking a nucleocapsid or core-like protein near the amino
terminus of the polyprotein.

Phylogenetic analysis of the helicase and replicase of GBV-
C/HGV has demonstrated that this novel virus is a member of the
Flaviviridae, most closely related to HCV (Leary et al.,
1996a). However, due to the lack of a nucleocapsid or core-like
protein near the amino terminus of the polyprotein, the
potentially low degree of sequence variability between isolates,
the extensive length of the 5' and 3' NTRs, differences in non-
human primate infectivity, as well as many other significant
differences, it is proposed that GBV-C (and GBV-A) be
grouped within a separate genus in the Flaviviridae. Based upon
similar comparisons, GBV-B should also be placed in a separate
genus. Thus, distinct genera for GBV-A/C, GBV-B and HCV
would be formed. Alternatively, following the nomenclature
and taxonomy recommended by Shukla et al. (1995) for HCV,
the new GBV-A/C and GBV-B genera could be placed within
the new family Hepciviridae.

References

Bartenschlager, R., Aihborn-Laake, L., Yasargil, K., Mous, J. &
trans by the hepatitis C virus NS3 proteinase. Journal of Virology 69,
198-205.

Bukh, J., Purcell, R. H. & Miller, R. H. (1992). Sequence analysis of the
5' noncoding region of hepatitis C virus. Proceedings of the National
Academy of Sciences, USA 89, 4942-4946.

Bukh, J., Purcell, R. H. & Miller, R. H. (1993). At least 12 genotypes of
hepatitis C virus predicted by sequence analysis of the putative E1 gene
of isolates collected world wide. Proceedings of the National Academy
of Sciences, USA 90, 8234-8238.

pestivirus bovine viral diarrhea virus with members of the Flaviviridae.


Dawson, G. J., Schlauder, G. S., Pilot-Matias, T. J., Thiele, D., Leary, T.
P., Murphy, P., Rosenblatt, J. E., Simons, J. N., Martinson, F. E. A.,
Gutiérrez, R. A., Lentino, J. R., Pachucki, C., Muerhoff, A. S., Widell, A.,
GB virus C infection using reverse-transcriptase-polymerase chain
reaction. Journal of Medical Virology 50, 97-103.

Farci, P., Alter, H. J., Govindarajan, S., Wong, D. C., Engle, R.,
Lesniewski, R. F., Mushahwar, I. K., Desai, S. M., Miller, R. H., Ogata,
N. & Purcell, R. H. (1992). Lack of protective immunity against
reinfection with hepatitis C virus. Science 258, 135-140.

Han, J. H., Shyamala, V., Richman, K. H., Brauer, M. J., Irvine, B.,
Urdea, M. S., Tekamp-Olson, P., Kuoh, G., Choo, Q.-L. & Houghton, M.


Received 20 May 1996; Accepted 16 July 1996