

Hepatitis C virus

Brief Review

P. G. W. Plagemann

Department of Microbiology, University of Minnesota Medical School, Minneapolis,
Minnesota, U.S.A.

Accepted June 4, 1991

Summary. HepCV is the major cause of NANB PT hepatitis and is also implicated as the cause in a large proportion of sporadic cases of NANBH. Chronic infection with HepCV has also been linked to the development of hepatocellular carcinoma. Chimpanzees and marmosets are the only animals found to be experimentally infectable and the virus has not been propagated in any cell culture system.

HepCV is an enveloped virus with a diameter of 30–60 nm and a 10-kb positive-stranded RNA genome. Its genome organization resembles that of the flaviviruses and pestiviruses. A 5'-untranslated segment of 341 nucleotides precedes a continuous ORF of 9030/9033 nucleotides which is followed by a 54 nucleotides long 3'-non-coding segment. Further work is required to resolve the question of whether the genomic RNA possesses a 3'-poly(U) or poly(A) tail. The genome also carries an internal poly(A) segment towards the 5'-end of its ORF. Genomic RNA is probably translated into a single polyprotein of 3010/3011 amino acids which is processed into functional proteins. The viral proteins have not been identified, but on the basis of the predicted amino acid sequences, hydrophobicity plots, location of potential glycosylation sites and similarities of these properties to those of pesti- and flaviviruses, the following genome organization has been predicted. The predicted viral structural proteins, a nucleocapsid protein and two envelope glycoproteins are located at the amino-terminal end of the polyprotein. They are followed by a highly hydrophobic protein and proteins that exhibit proteinase, helicase and replicase domains and thus are probably involved in RNA replication and protein processing. The replicase domain is located close to the carboxy terminus of the polyprotein.

Although the overall nucleotide and amino acid homologies between HepCV and pestiviruses are low, a number of similarities exist that point to a closer ancestral relationship to the latter than the flaviviruses. First, the 5'-untranslated

segment of the HepCV genome resembles that of the pestivirus genomes in size and presence of several short ORFs and it contains several segments with high nucleotide homology. Second, the two putative envelope glycoproteins of HepCV resemble two of the three putative envelope glycoproteins of the pestiviruses. Because its genome organization and predicted virion structure closely resemble those of the flaviviruses and pestiviruses, HepCV has been proposed to be placed in the family *Flaviviridae*. It has been suggested to be classified as a new third genus in this family because it is only remotely related to the pestiviruses and flaviviruses in nucleotide sequence of its genome and the amino acid sequences of the predicted viral proteins.

On the basis of genomic sequence information, an immunoprobe has been devised for screening blood supplies and donors for anti-HepCV antibodies to a non-structural protein of HepCV. The immuno-assay exhibits a high efficiency in detecting infected donors, though one caveat is that antibodies to the test antigen develop in infected individuals only between 2 and 8 months post infection. On the other hand, viral RNA can be detected in plasma by reverse transcription and amplification of the cDNA by PCR within a few days post infection. Thus the latter technique may become more important in the detection of HepCV in blood and tissues once the technique becomes more widely established as a diagnostic tool.

The untranslated 5'-segment has been found to be highly conserved in the genomes of different HepCV isolates from various parts of the world. The replicase domain is also highly conserved, but considerable amino acid and nucleotide differences exist in other segments of the long ORFs of various HepCV isolates. Divergence among different isolates is particularly great (up to 30%) in the segment encoding the two putative envelope glycoproteins and the upstream hydrophobic protein. The variability in envelope glycoproteins needs to be considered in the development of immuno-probes and of vaccines for HepCV.

Introduction

Hepatitis C virus (HepCV) is a new pesti-flavi-like virus which has been discovered only relatively recently as the main cause of non A, non B hepatitis (NANBH). The abbreviation HepCV has been suggested for hepatitis C virus [6] instead of HCV used previously in most communications in order to distinguish it from hog cholera virus and human cytomegalovirus which also have been generally abbreviated HCV. The abbreviation HoCV has been suggested for hog cholera virus [35].

The present review will only highlight main aspects of HepCV pathogenesis and molecular properties, especially as they relate to pestiviruses [for review of pestiviruses, see 35]. For more detailed information the reader is referred to recent reviews by Choo et al. [14], Bradley et al. [9] and Choo [11].

Post-transfusion (PT) hepatitis is still a relatively common consequence of blood transfusion [14]. From studies before 1980 the incidence in the U.S.A.

was estimated as 7–12% [17, 18]. The application of diagnostic tests for hepatitis B virus (HBV) and hepatitis A virus (HAV) during the 70's indicated that most of PT hepatitis was not caused by these viruses but rather by another virus(es) termed NANBH virus. Recent estimates are that 60 to > 90% of PT-NANBH presently observed in the U.S.A., Europe and Japan is caused by HepCV [14], which amounts to 100,000 to 300,000 cases per annum in the U.S.A. alone [15, 50]. In addition, intravenous drug addicts and hemophiliacs are at high risk of acquiring HepCV infections and HepCV has also been implicated as the cause of 30–60% of sporadic cases of NANBH [4, 14]. Another recently discovered hepatitis virus, hepatitis E virus (HEV), is a calicivirus-like virus mainly associated with enterically transmitted hepatitis [9].

Acute infections with HepCV seem to be less severe than those caused by HBV or HAV but a greater proportion of HepCV than HBV infections (about 50%) are likely to progress to a persistent chronic state and it is estimated that about 20% of the chronic carriers develop cirrhosis of the liver [2, 3, 14]. Furthermore, chronic infections with HepCV have been linked to the development of hepatocellular carcinoma in Europe, Japan and the U.S.A. [14, 40] (see below).

Transmission studies

No definitive information is presently available on the mechanism of non-parenteral transmission of HepCV. No vector or alternate hosts have been identified. Intrafamilial and sexual transmission have been implicated [2, 14, 30]. Little is also presently known about the pathogenesis of HepCV. Attempts to infect various cell culture systems or small laboratory animals have been unsuccessful. However, HepCV can be readily transmitted by parental injection to chimpanzees which have served for years as primary model system for NANBH [see 7, 9, 14]. Chimpanzees also developed hepatitis after injection of Factor VIII concentrate and Factor IX complex that had been implicated in the transmission of NANBH to human recipients and NANBH was serially transmitted by parental injection from chimpanzee to chimpanzee [see 8, 9]. The clinical disease in chimpanzees mimics that observed in humans, but may vary greatly between individual humans or chimpanzees. Infections often remain asymptomatic. When acute illness develops, it generally occurs 7–14 weeks p.i. and consists of a range of symptoms including fever, chills, headache, weakness and jaundice. After parenteral or non-parenteral transmission HepCV infections seem to be mainly localized in the liver since tissue damage seems limited to this tissue [7].

Tissue injury

Liver injury is associated with characteristic alterations in the cytoplasm of infected or affected hepatocytes in both humans and chimpanzees [reviewed in 9]. These include dense reticular inclusion bodies, convoluted membranes de-

rived from proliferated smooth endoplasmic reticulum, bundles of granular microtubules and characteristic tubular structures, which initially earned HepCV the name tubule-forming agent [9]. Also many endothelial cells were found to contain peculiar bundles of hexagonal-packed microtubules, which, when sectioned, resembled arrays of 25 nm particles. Some of these ultrastructural changes, however, are not unique to HepCV infections; similar changes have been observed in cells infected with other RNA viruses [reviewed in 9] and may represent general responses of cells to cytotoxic effects of these viruses. Monoclonal antibodies have been generated that specifically interact with tubular aggregates in liver biopsies of NANBH and hepatitis delta virus-infected chimpanzees [41, 42]. It seems likely that the ultrastructural changes and antigen formation are mediated by interferon formed in the course of host cell-specific responses to infection by these viruses [9, 41].

Immunological studies and other potential diagnostic methods

The mechanism of liver injury is not fully understood and factors other than cytocidal replication of HepCV in hepatocytes may be involved. Liver injury is usually associated with an inflammatory reaction and marked increases in alanine aminotransferase (ALT) activity in the circulation, which are general features reflecting liver injury. Elevated plasma ALT activity serves as general indicator of liver injury and is measured by blood banks in a "surrogate" assay to screen blood donors for potential infection by NANBH viruses [4, 9, 14]. The recent development of assays for antibodies to a HepCV polypeptide (C 100-3) that is part of the viral non-structural proteins (see below) has allowed studies on the humoral antibody response of HepCV-infected patients and chimpanzees. A recent study of stored blood samples from prospective studies of 15 U.S.A. patients with transfusion-associated chronic hepatitis documented by liver biopsies indicated that the time course of changes in plasma ALT activity and of the formation of anti-HepCV antibodies can vary considerably between patients [4]. A number of common features, however, were observed. In agreement with earlier studies, the onset of acute infection generally occurred between 5 and 10 weeks after transfusion and was associated with increases in plasma ALT activity of 10 fold or more as well as with increased serum bilirubin levels. All of the 15 patients developed anti-C 100-3 HepCV antibodies but only at an average of about 22 weeks after transfusion. However, high antibody levels persisted in all but one patient for at least 7–12 years. Serum ALT activity levels varied greatly between patients during the course of the chronic infection and a significant elevation was not always prevalent. Similarly, 88% of 16 donor sera tested contained significant anti-HepCV antibodies, but of the anti-HepCV antibody positive sera only 33% exhibited elevated ALT activity.

In another prospective study in Taiwan, 59% of chronic PT-NANBH patients developed anti-C 100-3 HepCV antibodies 6–32 weeks after transfusion [48]. Comparative results have been reported from studies in the U.S.A., Japan, Italy, Germany, Greece and Holland [summarized in 14, 40]. Of a total of 156

patients with chronic PT-NANBH about 80% were seropositive for HepCV antibodies. The prevalence of serum anti-HepCV antibodies in patients with sporadic NANBH, hemophiliacs and intravenous drug abusers was only slightly lower, whereas the prevalence among blood donors with normal serum ALT activity ranged from 0.4 to 1.4%.

In two experimentally infected chimpanzees anti-C 100-3 HepCV antibodies became detectable only 13 and 32 weeks p.i., respectively, whereas ultrastructural changes in the liver became apparent during the first week p.i. and infection-related increases in serum ALT activity occurred between 5 and 10 weeks p.i. [43]. Ultrastructural abnormalities in the liver and serum anti-C 100-3 antibodies persisted in one of the chimpanzees for at least 11 months p.i. but disappeared in the other one, and serum ALT activity returned to near normal in both.

The overall results have suggested that screening of donors for anti-HepCV antibodies could prevent the majority of cases of PT-NANBH, including many that might be missed by the assay of surrogate markers [4, 14, 34]. The long latent period between HepCV infection and development of anti-C 100-3 antibodies in both humans and chimpanzees (4–32 weeks), however, indicates that a considerable time period exists when HepCV can be transmitted but not detected by the immunological test. Of interest is the question of whether the long latent period may reflect a late antibody response to the non-structural proteins measured by the anti-C 100-3 immune assay. Perhaps antibodies to the structural proteins of HepCV (see below) are formed more rapidly. Indeed, it has been reported that antibodies to the HepCV core protein were detected by Western blotting in a higher proportion of sera from patients with chronic NANBH (81%) than antibodies to the C 100-3 antigen (61%) [36]. In another recent study, 89% and 82% of sera from 65 chronic NANBH patients were positive for antibodies to HepC virus core protein and C 100 antigen, respectively [25]. Furthermore, in one NANBH PT patient, anti-core protein antibodies became first detectable in serum 8 weeks after transfusion, whereas antibodies to the C 100 antigen were first detected 13 weeks after transfusion [25].

It also has become apparent that HepCV RNA becomes detectable by polymerase chain reaction (PCR) procedures much sooner after infection than anti-HepCV antibodies and that HepCV RNA might be detectable in the absence of detectable levels of serum antibodies. The procedure involves centrifugation of plasma at 50,000 rpm in a Beckmann ultracentrifuge for 2 h [7], extraction of RNA from the pelleted material, reverse transcription of the RNA using an appropriate “antisense” oligonucleotide primer to a HepCV genomic sequence and amplification of the product after addition of an appropriate “sense” oligonucleotide primer [45, 49]. The PCR product is then detected by Southern blot analysis with the use of a labeled HepCV cDNA that lies between the two primers used [49]. Or, if so desired, the PCR product is cloned and sequenced [28, 32, 45, 46]. Using primers to published sequences (see below) HepCV RNA was detectable in serum of two experimentally infected chimpanzees 3 days p.i. [43]. It was also detected in serum of one patient with acute

PT NANBH, who was seronegative for anti-C 100-3 antibodies, in serum from one chronically infected chimpanzee [49], and in livers or sera of patients with chronic NANBH [28, 32, 36, 45, 49].

In this context, it is of interest that only 20–30% of patients with acute and resolving HepCV infections develop anti-C 100-3 antibodies. The question arises as to what factors are involved in the resolution of the infection and elimination of the virus. Perhaps the latter are mediated by antibodies to viral proteins not represented by the C 100-3 antigen or by cellular immune responses. Since the only HepCV titration assay available at present is an expensive endpoint dilution assay in chimpanzees or marmosets [see 9, 19], only limited information is available on HepCV viremia in infected humans and chimpanzees over the course of the disease. Nevertheless, all data indicate that serum HepCV titers are generally low; i.e., $< 10^3$ chimpanzee infectious doses (CID)/ml. However, titers of up to 10^6 CID/ml of serum have been observed in plasma of selected chimpanzees during periods of exacerbated chronic phase disease [9]. Also, liver tissue from one of the animals contained about 10^7 CID/g of tissue [9] and an unusual titer of at least 10^6 CID/ml has been observed in serum from one patient with acute NANBH (strain H) [19].

Association with hepatocellular carcinoma

In a recent study in Japan, 69% of 156 patients with hepatocellular carcinoma who lacked serum antibodies to HBV were found to be seropositive for the HepCV C 100-3 antigen compared to an incidence of 4–10% in control groups [40]. A similar incidence of serum anti-C 100-3 HepCV antibodies (64%) has been reported for another group of 33 patients with hepatocellular carcinoma in Japan [37]. Although these and similar results reported from the U.S.A. and Europe implicate chronic HepCV infections as the cause of hepatocellular carcinoma, potential mechanisms of this effect remain to be elucidated. Since viruses, like HepCV, lack oncogenes and do not replicate via DNA intermediates, indirect effects are indicated. Perhaps extensive stimulation of hepatocyte proliferation in the course of chronic liver injury plays a role.

Interferon α treatment

Recent randomized, controlled trials have indicated that treatment with 1–3 10^6 units of recombinant interferon α 2 b (Intron A, Schering-Plough, Kenilworth, NJ) thrice weekly is effective in controlling NANBH in many patients [15, 16]. Both decreases in serum ALT activity and improved histological features of the liver were observed. However, interferon treatment does not seem to result in termination of the infection since most patients relapsed when treatment was discontinued after 6 months.

Virion structure and properties

Only limited information is presently available on the structure and biochemical properties of HepCV. Filtration studies have shown that the diameter of infectious

virions is < 80 nm [7] or between 30 and 60 nm [26]. Electron microscopic examinations revealed togavirus-like particles with diameters of 36–62 nm in sera of chimpanzees and humans with chronic NANBH and similar particles were observed in cytoplasmic cisternae of hepatocytes of infected chimpanzees but the origin and identity of these particles has not been ascertained [1]. The virus-like particles detected by electron microscopy exhibited a buoyant density in sucrose density gradients of 1.14 – 1.18 g/cm³ [1]. In contrast, a buoyant density of 1.09 – 1.11 g/cm³ was observed in sucrose density gradient centrifugations in which the gradient fractions were assayed for infectious HepCV by titration in chimpanzees (D. W. Bradley, pers. comm.). From other sedimentation analyses it has been concluded that the sedimentation coefficient of HepCV is < 150 – 200 S [7, 9]

HepCV is an enveloped virus since its infectivity is destroyed by extraction with chloroform [10, 20]. Infectivity is also abolished by heating at 60°C for 30 min [39].

Genome organization and expression

Initial identification and cDNA cloning of HepCV RNA and development of immunoassays

The sequence of the HepCV RNA genome has been determined via sequencing of cDNAs [13, 27, 29, 44]. Initially a random-primed expression cDNA library in λ gt11 was prepared from RNA extracted from infectious material concentrated from a large volume of high titer plasma from a chimpanzee (see above) with a chronic NANBH infection. The animal represented the second chimpanzee passage of the virus present in a human Factor VIII concentrate [8, 9]. The cDNA library was screened with serum from a chronic NANBH patient as a potential source of antiviral antibodies [12]. A single positive clone (5-1-1; see Fig. 1) was identified among a total of about 10^6 clones. Using the 5-1-1 clone, a larger overlapping cDNA clone (81) was isolated. Clone 81 cDNA did not hybridize to human or chimpanzee cell DNA in Southern hybridization analyses, but it hybridized to RNA extracted from the liver of NANBH virus-

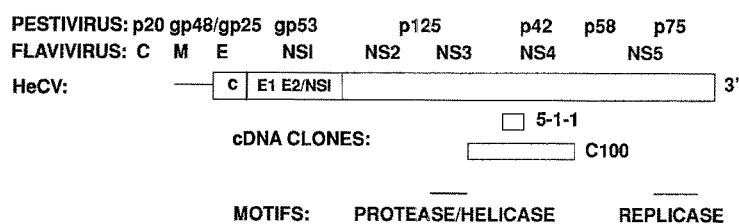


Fig. 1. Tentative genome organization of HepCV in relation to those of flaviviruses and pestiviruses and approximate location of 5-1-1 and C 100 cDNA sequences and of functional motifs of the putative encoded proteins. C Capsid protein, M matrix protein, E envelope glycoprotein. Modified from Choo et al. [14]

infected chimpanzees [12]. In Northern hybridization analyses, clone 81 cDNA hybridized to oligo(dT)-selected RNA in the 5 to 10 kb range. The results indicated that the putative positive strand viral RNA isolated from infectious plasma had a size of at least 10 kb. The results also suggested that HepCV RNA contains a poly(A) sequence(s) but it is unclear whether its binding to oligo(dT) is due to an internal or a 3'-terminal poly(A) sequence [6, 9, 13] (see below).

Three additional cDNA clones were isolated from the original cDNA library using clone 5-1-1 cDNA as a probe and found to encode a continuous ORF [31]. A reconstructed clone (C 100) encoding this ORF (see Fig. 1) was fused to a human superoxide dismutase (SOD) cDNA (C 100-3), which facilitates the efficient expression of foreign proteins in yeast, and expressed as a fusion protein containing 363 amino acids in yeast. The SOD/HepCV polypeptide produced in yeast has been used to develop a radioimmune assay (RIA) and an ELISA for the assay of anti-HepCV antibodies [31]. Commercial ELISA kits are available from Ortho Diagnostic Systems (Raritan, NY) and Abbott Diagnostics (Chicago, IL). As discussed already, the assays have been used extensively already in prospective studies and in screening blood donors for serum anti-C 100 HepCV antibodies.

Genome sequence analyses

By primer extension analyses a sequence of a total of 7310 of the HepCV genome was initially determined which encodes a continuous ORF of 2436 amino acids [27]. This sequence is located towards the 3'-end of the HepCV genome (see Fig. 1) and exhibits certain similarities to those of flaviviruses and pestiviruses (see below). Recently, sequence analyses of the 5'-region, presumably encoding the structural proteins, as well as of the 3'-end of the original chimpanzee HepCV isolate (variously referred to as prototype HepCV or HepCV-pt, HepCV-1 or HepCV-UC) have been completed [13, 24, 47]. A sequence of a total of 9407 nucleotides has been reported for the HepCV-1 genome which encodes a continuous ORF of 3011 amino acids (Table 1). The translation initiation codon starts 342 nucleotides from the reported 5'-terminus.

Recently, consensus sequences have also been reported for two HepCV genomes isolated from pooled plasma of nine Japanese NANBH patients (referred to as HepCV-J) [29] and of fifty Japanese blood donors with elevated plasma ALT activity (referred to as HepCV-KB) [44]. In both studies virus particles were collected from plasma by ultracentrifugation, RNA was extracted from pelleted material and reverse transcribed and cDNA libraries prepared in λ gt11. In one study, the viral RNA was reverse transcribed using appropriate antisense oligonucleotides as primers or the viral RNA was polyadenylated and reverse transcribed using oligo(dT) as primer [29]. In the other study, viral RNA was reverse transcribed using random hexanucleotide primers [44]. The 9413 nucleotides long sequence of HepCV-J RNA contains 5'- and 3'-terminal untranslated segments of 329 and 54 nucleotides in length, respectively (Table 1).

Table 1. Comparison of genomic sequences of three independent isolates of HepCV

HepCV isolate	Length of isolated genome (nucleotides)	Non-coding segment (nucleotides)		ORF (amino acids)	Ref.	Genbank data base accession number
		5'-term.	3'-term.			
1 (UC)	9401	341	27	3011	13	M 62321
					24	M 58406
						M 58407
J	9413	329	41+12U+C	3010	29	D 90208
KB	9416	332	41+C+12U	3010	44	M 58335

The genome of HepCV-KB is 9416 nucleotides long and possesses 332 and 54 nucleotides long 5'- and 3'-terminal untranslated segments (Table 1).

The 5'-untranslated segment of HepCV-1 is 12 and 9 nucleotides longer than those reported for HepCV-J and -KB, respectively. It may represent the true 5'-terminus of the HepCV genome since several independent clones had the same terminal sequence [24]. The 5'-terminus can potentially form a hairpin structure with a calculated ΔG value of -14.4 kcal [24]. The 5'-untranslated segment resembles those of pestiviruses, rather than those of flaviviruses, in its greater length and presence of short ORFs [13]. The HepCV 5'-untranslated segment encodes three ORFs of 12–15 amino acids [24] but no functions have been identified for these ORFs. In addition, it exhibits considerable nucleotide homology (47–49%) with the 5'-untranslated segments of pestivirus genomes [13, 24, 47]. In fact, four blocks of 8, 22, 30, and 37 nucleotides show sequence identity of $>70\%$ [24]. Whether the 5'-ends of these genomes are capped is unknown.

The 3'-untranslated segments of HepCV-J and -KB RNA are 28 nucleotides longer than that reported for HepCV-1 and contain a 5'-AGGCCA-3' repeat separated by AU as well as a terminal string of 12 Us (Table 1). This sequence probably represents the actual 3'-terminus of the viral genome since the same 3'-terminal sequence, except for the displacement of one C has been reported for HepCV-KB and -J RNA (see Table 1). However, some uncertainties still exist concerning the true 3'-terminus since for both reported genomes the 3'-terminal sequence has been derived from a single cDNA.

These results may suggest that the genomic RNA of HepCV does not possess a 3'-poly(A) tail, but rather a poly(U) tail. If this conclusion is correct, it follows that the complementary negative strand will possess a 5'-poly(A) tail. Interestingly, the genomic RNA also possesses an A-rich segment in the most 5'-end of its ORF (nucleotides 23–36) [13]. This segment is 5'-AAAAAAAAAACA-3' for HepCV-1 and 5'-AAAGAAAAACCA-3' for HepCV-J and -KB.

Whether these A-rich segments have a function in HepCV replication is unclear at present. Nevertheless, they could account for a weak binding of both positive and negative strands of HepCV RNA to oligo(dT) as has been observed for the genomic RNA [12].

On the other hand, Han et al. [24] suggested that at least some HepCV RNA molecules may be polyadenylated because cDNAs to the 3'-end of the viral genome could be generated by reverse transcription of RNA extracted from high titer plasma of an infected chimpanzee using an unrelated oligo-nucleotide with a 3'-oligo(dT) tail as primer. Amplification of the product by PCR yielded three types of cDNAs, all of which contained 3'-poly(A) tails. One cDNA encompassed the 27 nucleotide 3'-non-coding segment of HepCV-1 RNA plus a poly(A) tail.

In addition to the sequences of the HepCV-1, -J, and -KB genomes, the 5'-terminal end 1413 nucleotides of a human HepCV isolate derived from another batch of plasma of healthy Japanese carriers have been determined [32, 45–47]. A non-coding segment of 90 nucleotides is followed by an ORF of 441 amino acids. It overlaps with a 308 amino acid-long sequence (amino acids 129–437) that has been determined for six different HepCV isolates, four from the U.S.A. and two from Italy [50].

Potential gene products

The gene products of HepCV have not been identified, but on the basis of the amino acid sequences of the long ORFs of the various HepCV isolates, the hydrophobicity profiles of the predicted proteins, the location of potential glycosylation sites and similarities in these properties to the proteins encoded by the pestivirus and flavivirus genomes, certain predictions have been made. The first 190 amino acids of the N-terminal end are thought to represent the nucleocapsid protein (C; Fig. 1). The formation of a 22 kDa (p 22) protein has been demonstrated in mammalian cells transfected with this cDNA sequence under the control of a foreign promoter [25, 45, 46] and in transfected bacteria [36]. Furthermore, the putative nucleocapsid protein of HepCV exhibits a similar high content of basic amino acids as the nucleocapsid proteins of flaviviruses and pestiviruses and lacks N-glycosylation sites [13, 29, 45, 47].

The next 190 amino acids in the HepCV genome are thought to represent an envelope glycoprotein (E 1; Fig. 1) because the potential protein is mostly hydrophobic and contains six possible N-glycosylation sites. The hydrophobic plot of this domain resembles that of gp 25 of the pestiviruses [13, 47]. The HepCV genome seems to encode a second adjacent glycoprotein, about 380 amino acids in length, which seems to correspond in location to the major structural glycoprotein of pestiviruses (gp 53) and the non-structural NS 1 glycoprotein of the flaviviruses (see Fig. 1). The two predicted HepCV glycoprotein ORFs are preceded by putative signal peptide sequences [29, 50]. Other preliminary studies indicate that HepCV may possess at least three structural proteins, the 18–22 kDa basic nucleocapsid protein (C) and two glycoproteins,

a hydrophobic 33 kDa protein (E 1) and a 70 kDa glycoprotein (E 2/NS 1) [9, 11]. This arrangement basically resembles that of the pestiviruses rather than that of the flaviviruses, except that the HepCV genome does not encode a protein that corresponds to gp48 of the pestiviruses and therefore is shorter [for review of pestiviruses, see 35].

Another interpretation of the reported HepCV genomic sequences is that only the first 114 amino acids make up the nucleocapsid protein, that the next approximately 80 amino acids represent a matrix (M) protein, that E 1 is the single envelope glycoprotein of HepCV and that E 2/NS 1 is a non-structural protein [44]. This genome organization would be equivalent to that of the flaviviruses, except that the pre-M segment would be deleted. The question of which of these interpretations is correct can only be resolved by isolation and characterization of the structural and non-structural proteins of the virus.

Considerable similarities also exist between HepCV and flaviviruses and pestivirus genomes in the location of specific motifs generally associated with the genomes of positive-strand RNA viruses (see Fig. 1). The replicase motif characterized by the consensus GDD sequence is located close to the 3'-end of the HepCV ORF comparable in its location to the NS 5 and p 75 proteins of flaviviruses and pestiviruses, respectively [5, 6, 23, 38] (see Fig. 1). A serine proteinase motif (GxSGxP) [21] is found in the HepCV genome at the same relative position as in the genomes of flaviviruses (NS 3) and pestiviruses (p 125; see Fig. 1) [6]. Immediately downstream from the putative proteinase sequence are consensus sequences in the genomes of the three viruses that are typical for the proposed superfamily 2 of helicase proteins [22, 23]. The HepCV genome also encodes a highly hydrophobic protein sequence equivalent to NS 2 of the flaviviruses and the amino terminal end of p 125 of the pestiviruses, respectively [6, 9, 13] (see Fig. 1). Overall, however, the amino acid sequence of HepCV-1 is only distantly related to those of the flaviviruses and pestiviruses [13, 27, 33, 44]. Because of these differences but otherwise great similarities in genome organization and viron properties, HepCV has been suggested to be classified as a new genus within the family *Flaviviridae* [9, 13, 44].

Sequence comparisons of various HepCV isolates

The 5'-untranslated segment is highly conserved. For example, the 329-nucleotides 5'-segment upstream of the continuous ORF of HepCV-1, -KB, and -J differ from each other by only one or two nucleotides [24, 29, 44]. Similarly, a maximum divergence of 1-5 nucleotides was observed between the 5'-untranslated segment of HepCV-1 and those of 10 other isolates from the U.S.A., Australia, Italy, Japan, Korea, Argentina, South Africa, and Taiwan [24]. The high degree of conservation of the 5'-segment suggests that it provides important regulatory functions. The 3'-terminal segment seems somewhat less conserved. Although only one nucleotide differs in the 3'-untranslated segments of HepCV-KB and -J [29, 44], the 27 nucleotides 3'-segment of HepCV-1 differs from those of the Japanese isolates by at least 30% [13, 24].

Considerable sequence differences also exist between various HepCV isolates in the long ORF, but not all parts vary equally. For example, the ORFs of the two Japanese isolates HepCV-KB and -J exhibit 95.0% amino acid identity, but the amino acid identity is only 84.5 and 85.4% for the chimpanzee isolate HepCV-1 and HepCV-KB and HepCV-J, respectively. (Sequences were analyzed using the Molecular Biology Information Resource suite of programs (MBIR; Baylor College of Medicine, Houston, TX). Multiple alignments were produced with implementation of the Klotz and Blanten method [30 a]. Protein sequences were further analyzed using "homology" [32 a] with the Dayhoff scoring matrix.) However, even though the overall percentage of amino acid differences between the HepCV-1 ORF and those of the two Japanese isolates are about the same, a considerable number of different amino acids of the latter are involved indicating a rather complex relationship between the reported genomic sequences. Amino acid differences between the three strains are particularly prevalent in a segment encompassing amino acids 240–1140, that is the segment representing E1/E2/NS1, and the amino terminal portion of the adjacent downstream protein (see Fig. 1), which seem to contain areas with higher than average variability (see below). In this segment amino acid divergence amounts to 8.3, 20.8, and 27.2% for HepCV-KB and -J, HepCV-1 and -KB, and HepCV-1 and -J, respectively. In contrast a replicase domain of 48 amino acids is the same for HepCV-KB and -J, and that of HepCV-1 differs only by two amino acids (see Fig. 8 in [38]). The overall nucleotide sequence homology of the coding segment is 91.5% between HepCV-KB and -J, 78.3% between HepCV-1 and -J and 77.9% between HepCV-1 and -KB.

In addition, specific sequences of various HepCV isolates have been amplified by PCR, cloned and sequenced. The putative nucleocapsid protein sequence derived from PCR products of RNA extracted from healthy Japanese carrier plasma exhibited a 97.4% amino acid identity with the sequence determined for the original chimpanzee HepCV isolate (HepCV-1) but only a 90.5% identity at the nucleotide level [45–47]. Furthermore, only 70–75% amino acid and nucleotide identity was observed for a 351 nucleotides sequence encoding part of the putative envelope proteins (Fig. 1) of the two HepCV isolates [45–47]. By comparing the nucleotide and amino acid sequences of four HepCV isolates from the U.S.A. and two from Italy (amino acids 129–437), a moderately variable domain of approximately 40 amino acids in the E1 region and a hypervariable domain (region V) of approximately 28 amino acids in the junction between E1 and E2/NS1 (see Fig. 1) have been identified [50]. Although the overall divergence between the six isolates was only 7% for nucleotides and 6% for amino acids, the hypervariable region accounted for up to 60% of the observed amino acid differences between the isolates. The significance of the apparent variabilities in the glycoprotein coding sequences is unclear. It could be a consequence of immune selection and needs to be considered in the future development of HepCV vaccines.

Considerable differences were also reported for the non-structural protein

sequences of various other HepCV isolates in some studies but not in others. cDNA clones (282 bp) generated by PCR from RNA extracted from Japanese patients with chronic NANBH exhibited only 76–77% nucleotide identity with HepCV-1 [28] which resembles the difference between HepCV-1 and -KB/J. Furthermore, the HepCV RNA extracted from these patients was not detectable by PCR using certain primers based on the sequence of HepCV-1 [28]. Other cDNA clones to the non-structural region generated by PCR from RNA extracted from Japanese blood donors, however, exhibited less divergence from the prototype HepCV-1.

Sequencing of selective genome segments by the cDNA/PCR method has shown that the predicted amino acid sequences of most Japanese HepCV isolates diverge from those of HepCV-J by < 10%, but that some Japanese isolates are more homologous to HepCV-1 than HepCV-J [29, 32]. Combined the results indicate that numerous variants of HepCV are present in various parts of the world and that some of these variants, such as HepCV-J, might represent subtypes of HepCV-1 [29, 50]. Further work is required to establish the prevalence of various variants in different countries. Since amino acid sequence divergence is most prevalent in the putative envelope glycoproteins and the segment equivalent to the amino terminus of p 125 of the pestiviruses and NS 2 of the flaviviruses (up to 30%), another interesting question is whether the variants become selected in the course of chronic infections due to immunological pressures.

Acknowledgements

I thank Dr. Daniel W. Bradley for providing preprints of publications from his laboratory and helpful discussion and comments, and Joy Laiti for competent secretarial assistance.

References

1. Abe K, Kurata T, Shikata T (1989) Non-A, non-B hepatitis: visualization of virus-like particles from chimpanzee and human sera. *Arch Virol* 104: 351–355
2. Alter HJ (1989) Chronic consequences of non-A, non-B hepatitis. In: Seef LB, Lewis JH (eds) *Current perspectives in hepatology*. Plenum, New York, pp 83–97
3. Alter HJ, Coleman PJ, Alexander WJ, Kramer E, Miller JK, Mandel E, Hadler SC, Margolis HS (1989) Importance of heterosexual activity in the transmission of hepatitis B and non-A, non-B hepatitis. *JAMA* 262: 1201–1205
4. Alter HJ, Purcell RH, Shih JW, Melpolder JC, Houghton M, Choo Q-L, Kuo G (1990) Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. *N Engl J Med* 321: 1494–1500
5. Argos P (1988) A sequence motif in many polymerases. *Nucleic Acids Res* 16: 9909–9916
6. Beach M, Bradley D (1991) Analysis of the putative nonstructural gene region of hepatitis C virus. In: Hollinger FB, Lemon S, Margolis H (eds) *Viral hepatitis and liver disease*. Williams and Wilkins, Baltimore (in press)
7. Bradley DW (1985) The agents of non-A, non-B viral hepatitis. *J Virol Methods* 10: 307–319
8. Bradley DW, Cook EH, Maynard JE, McCaustland KA, Ebert JW, Dolana GH, Petzel

- RA, Kanator RJ, Heilbrunn A, Fields HA, Murphy BL (1979) Experimental infection of chimpanzees with antihemophilic (factor VIII) materials: recovery of virus-like particles associated with non-A, non-B hepatitis. *J Med Virol* 3: 253–269
9. Bradley DW, Krawczynski K, Beach MJ, Purdy MA (1991) Non-A, non-B hepatitis: toward the discovery of hepatitis C and E viruses. *Sem Liver Dis* (in press)
 10. Bradley DW, Maynard JE, Popper H, Cook EH, Ebert JW, McCaustland KA, Schable CA, Fields HA (1983) Posttransfusion non-A, non-B hepatitis: physicochemical properties of two distinct agents. *J Infect Dis* 48: 254–264
 11. Choo Q-L (1991) HCV encoded proteins. In: Hollinger FB, Lemon S, Margolis H (eds) *Viral hepatitis and liver disease*. Williams and Wilkins, Baltimore (in press)
 12. Choo Q-L, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 244: 359–362
 13. Choo Q-L, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina-Selby A, Barr PJ, Weiner AJ, Bradley DW, Kuo G, Houghton M (1991) Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci USA* 88: 2451–2455
 14. Choo Q-L, Weiner AJ, Overby LR, Kuo G, Houghton M, Bradley DW (1990) Hepatitis C virus: the major causative agent of viral non-A, non-B hepatitis. *Br Med Bull* 46: 423–441
 15. Davis GL, Balart LA, Schiff ER, Lindsay K, Bodenheimer HC Jr, Perrillo RP, Carey W, Jacobson IM, Payne J, Dienstag JL, van Thiel DH, Tamburra C, Lefkowitz J, Albrecht J, Meschivitz C, Ortego TJ, Gibas A (1990) Treatment of chronic hepatitis C with recombinant interferon α : a multicenter randomized, controlled trial. *N Engl J Med* 321: 1501–1506
 16. Di Bisceglie AM, Martin P, Kassianides C, Lisker-Melman M, Murray L, Waggoner J, Goodman Z, Banks ASM, Hoofnagle JH (1990) Recombinant interferon alfa therapy for chronic hepatitis C: a randomized, double-blind, placebo-controlled trial. *N Engl J Med* 321: 1506–1510
 17. Dienstag JL (1983) Non-A, non-B hepatitis. I. Recognition, epidemiology, and clinical features. *Gastroenterology* 85: 439–462
 18. Dienstag J (1985) Non-A, non-B hepatitis. II. Experimental transmission, putative virus agents and markers, and prevention. *Gastroenterology* 85: 743–768
 19. Feinstone SM, Alter HJ, Dienes HP, Shimizu Y, Popper H, Blackmore D, Sly D, London WT, Purcell RH (1981) Non-A, non-B hepatitis in chimpanzees and marmosets. *J Infect Dis* 144: 588–598
 20. Feinstone SM, Mihalik KB, Kaminira T, Alter HJ, London WT, Purcell RH (1983) Inactivation of hepatitis B virus and non-A, non-B hepatitis by chloroform. *Infect Immun* 41: 816–823
 21. Gorbalenya AE, Donchenko AP, Koonin EV, Blinov VM (1989) N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. *Nucleic Acids Res* 17: 3889–3397
 22. Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM (1989) Two related super-families of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res* 17: 4713–4730
 23. Gorbalenya AE, Koonin EV, Donchenko AP, Binov VM (1989) Coronavirus genome: prediction of putative functional domains in the structural glycoproteins by comparative amino acid analysis. *Nucleic Acids Res* 17: 4847–4859
 24. Han JH, Shyamala V, Richman KH, Brauer MJ, Irvine B, Urdea M, Tekamp-Olson P, Kuo G, Choo Q-L, Houghton M (1991) Characterization of the terminal regions of hepatitis C virus RNA: identification of conserved sequences in the 5' untranslated region and poly (A) tails at the 3' end. *Proc Natl Acad Sci USA* 88: 1711–1715

25. Harada S, Watanabe Y, Takeuchi K, Suzuki T, Katayama T, Takabe Y, Saito I, Miyamura T (1991) Expression of processed core protein of hepatitis C virus in mammalian cells. *J Virol* 65: 3015–3021
26. He L-F, Alling D, Popkin T, Shapiro M, Alter HJ, Purcell RH (1978) Determining the size of non-A, non-B hepatitis virus by filtration. *J Infect Dis* 156: 636–640
27. Houghton M, Choo Q-L, Kuo G (1988) Non-A, Non-B virus diagnostics and vaccines. *Eur Patent Appl* 88, 310, 922.5 and publication number 0318216
28. Kaneko S, Unoura M, Kobayashi K, Kuno K, Murakami S, Hattori N (1990) Detection of serum hepatitis C virus RNA. *Lancet* 335: 976
29. Kato N, Hijihata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, Shimotohno K (1990) Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc Natl Acad Sci USA* 87: 9524–9528
30. Kiyosawa K, Sodeyama T, Tanaka E, Shimizu S, Furuta S, Miyazaki Y, Akahane Y, Suzuki H (1991) Intrafamilial transmission of hepatitis C virus in Japan. *J Med Virol* 33: 114–116
- 30 a. Klotz LC, Blanken RL (1981) A practical method for calculating evolutionary trees from sequence data. *J Theoret Biol* 9: 261–272
31. Kuo G, Choo Q-L, Alter HJ, Gitnick GL, Redecker AG, Purcell RH, Miyamura T, Dienstag JL, Alter MJ, Stevens CE, Tegtmeier GE, Bonino F, Colombo M, Lee W-E, Kuo G, Berger K, Shuster JR, Overby LR, Bradley DW, Houghton M (1989) An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* 244: 362–364
32. Kubo T, Takeuchi YK, Boonmar S, Katayama T, Choo Q-L, Kuo G, Weiner AJ, Bradley DW, Houghton M, Saito I, Miyamura T (1989) A cDNA fragment of hepatitis C virus isolated from an implicated donor of post-transfusion non-A, non-B hepatitis in Japan. *Nucleic Acids Res* 17: 10367–10372
- 32 a. Lawrence CB, Goldman DA (1988) Definition and identification of homology domains. *Comp Appl Biol Sci* 4: 25–31
33. Miller RH, Purcell RH (1990) Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proc Natl Acad Sci USA* 87: 2057–2061
34. Miyamura T, Saito I, Katayama T, Kikuchi S, Tateda A, Houghton M, Choo Q-L, Kuo G (1990) Detection of antibody against antigen expressed by molecularly cloned hepatitis C virus cDNA: application to diagnosis and blood screening for posttransfusion hepatitis. *Proc Natl Acad Sci USA* 87: 983–987
35. Moening V, Plogemann PGW (1991) The pestiviruses. *Adv Virus Res* (in press)
36. Muraio K, Hijikata M, Ohkoshi S, Cho M-J, Kikuchi M, Kato N, Shimotohno K (1990) A structural protein of hepatitis C virus expressed in *E. coli* facilitates accurate detection of hepatitis C virus. *Biochem Biophys Res Commun* 172: 511–516
37. Ohkoshi S, Kato N, Kinoshita T, Hijikata M, Ohtsuyama Y, Okazaki M, Ohkura H, Hirohashi S, Honma A, Ozaki T, Yoshikawa A, Kojima H, Asakura H, Shimotohno K (1990) Detection of hepatitis C virus RNA in sera and liver tissues of non-A, non-B hepatitis patients using polymerase chain reaction. *Jap J Canc Res* 81: 862–865
38. Plogemann PGW, Moennig V (1991) Lactate dehydrogenase-elevating virus, equine arteritis virus and Simian hemorrhagic fever virus. *Adv Virus Res* (in press)
39. Purcell RH, Gerin JL, Pooper H, London WT, Cicmanec J, Eichberg JW, Newman J, Hindra ME (1985) Hepatitis B virus, hepatitis non-A, non-B virus and hepatitis delta virus in lyophilized anti-hemophilic factor: relative sensitivity to heat. *Hepatology* 6: 1091–1099
40. Saito I, Miyamura T, Ohbayashi A, Harada H, Katayama T, Kikuchi S, Watanabe TY, Koi S, Onji M, Ohta Y, Choo Q-L, Houghton M, Kuo G (1990) Hepatitis C virus

- infection is associated with the development of hepatocellular carcinoma. *Proc Natl Acad Sci USA* 87: 6547–6549
41. Shimizu YK, Purcell RH (1989) Cytoplasmic antigen in hepatocytes of chimpanzees infected with non-A, non-B hepatitis virus or hepatitis delta virus: relationship to interferon. *Hepatology* 10: 764–768
 42. Shimizu YK, Oomura M, Abe K, Uno M, Yamada E, Ono Y, Shikata T (1985) Production of antibody associated with non-A, non-B hepatitis in a chimpanzee lymphoblastoid cell line established by in vitro transformation with Epstein-Barr virus. *Proc Natl Acad Sci USA* 82: 2138–2142
 43. Shimizu YK, Weiner AJ, Rosenblatt J, Wong DC, Shapiro M, Popkin T, Houghton M, Alter HJ, Purcell RH (1990) Early events in hepatitis C virus infection of chimpanzees. *Proc Natl Acad Sci USA* 87: 6441–6444
 44. Takamizawa A, Mori E, Fuke I, Manabe S, Murakami S, Fujita J, Onishi E, Andoh T, Yoshida I, Okayama H (1991) Structure and organization of the hepatitis C virus genome isolated from human carriers. *J Virol* 65: 1105–1113
 45. Takeuchi K, Kubo Y, Boonmar S, Katayama T, Harada H, Ohbayashi A, Choo Q-L, Kuo G, Houghton M, Saito I, Miyamura T (1990) Hepatitis C viral cDNA clones isolated from a healthy carrier donor implicated in posttransfusion non-A, non-B hepatitis. *Gene* 91: 287–291
 46. Takeuchi K, Kubo Y, Boonmar S, Watanabe Y, Katayama T, Choo Q-L, Kuo G, Houghton M, Saito I, Miyamura T (1990) Nucleotide sequence of core and envelope genes of the hepatitis C virus genome derived directly from human healthy carriers. *Nucleic Acids Res* 18: 4626
 47. Takeuchi K, Kubo Y, Boonmar Y, Watanabe Y, Katayama T, Choo Q-L, Kuo G, Houghton M, Saito E, Miyamura T (1990) The putative nucleocapsid and envelope protein genes of hepatitis C virus determined by comparison of the nucleotide sequences of two isolates derived from an experimentally infected chimpanzee and healthy human carriers. *J Gen Virol* 71: 3027–3033
 48. Wang J-T, Wang T-H, Lin JT, Sheu J-C, Sung J-L, Chen D-S (1990) Hepatitis C virus in a prospective study of posttransfusion non-A, non-B hepatitis in Taiwan. *J Med Virol* 32: 83–86
 49. Weiner AJ, Kuo G, Bradley DW, Bonino F, Lee C, Rosenblatt J, Choo Q-L, Houghton M (1990) Detection of hepatitis C viral sequences in non-A, non-B hepatitis. *Lancet* 335: 1–3
 50. Weiner AJ, Brauer MJ, Rosenblatt J, Richman KH, Tung J, Crawford K, Bonino F, Saracco G, Choo Q-L, Houghton M, Han JH (1991) Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS 1 proteins and the pestivirus envelope glycoproteins. *Virology* 180: 842–848

Authors' address: Dr. P. Plagemann, Department of Microbiology, University of Minnesota Medical School, 420 Delaware Street S. E., Minneapolis, MN 55455-0312, U.S.A.

Received June 6, 1991