Hepatitis C virus genotype 4 is highly prevalent in central Africa (Gabon)

Li-Zhe Xu,1 Daniel Larzul,2 Eric Delaporte,3 Christian Bréchot1,4,5 and Dina Kremsdorf4*

1Hybridotest, and 2 Texcell, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, 3 INSERM U13, Claude Bernard Hospital, 190 Boulevard MacDonald, 75944 Paris Cedex 19, 4 INSERM U370, and 5 Liver Unit, CHU Necker, 156 rue de Vaugirard, 75730 Paris Cedex 15, France

Following a survey of hepatitis C virus (HCV) infection recently carried in central Africa (Gabon), we cloned and sequenced PCR products of the 5' non-coding and capsid-encoding regions of HCV RNA from three randomly selected HCV RNA-positive Gabonese subjects. In the capsid-encoding region, the identity between the three Gabonese isolates was 91 to 98%. The three Gabonese sequences showed a divergence of 11 to 17% from published HCV genotypes I to IV (la, lb, 2a and 2b) isolates and of 6 to 11% from HCV genotype 4 isolates. Thus the Gabonese isolates, termed HC-G, belong to HCV genotype 4. Based on the sequences of the three isolates, a specific probe (cpsG) was designed to detect the HC-G genotype in 30 randomly selected anti-HCV-positive Gabonese subjects, 14 of whom were HCV RNA-positive. Analysis with cpsG showed that 10 of 14 of the HCV RNA-positive subjects were infected by the HC-G genotype. HC-G is therefore highly prevalent in the HCV RNA-positive Gabonese population. The availability of these Gabonese sequences should facilitate the design of specific serological tests for African HCV isolates.

Since the discovery of the hepatitis C virus (HCV) by Choo et al. (1989), several HCV isolates have been characterized worldwide (Enomoto et al., 1990; Kato et al., 1990; Kremsdorf et al., 1991; Li et al., 1991; Xu et al., 1994). Nucleotide sequence data have shown that HCV is related to the flaviviruses and pestiviruses (Miller & Purcell, 1990), by the presence of a long open reading frame coding for a single polyprotein of 3010 to 3033 amino acids, and flanked by two non-coding regions (NCRs) located at the 5' and 3' ends of the genome. The polyprotein is subsequently cleaved into several proteins, as follows (N to C terminus): the capsid (C) and two putative envelope proteins (E1 and E2) (structural proteins), and NS2, NS3, NS4, NS5 proteins (non-structural region) (Han et al., 1991; Miller & Purcell, 1990).

HCV isolates have been separated into distinct genotypes (Cha et al., 1992; Chan et al., 1992; Enomoto et al., 1990; McOmish et al., 1993; Okamoto et al., 1992c). Several authors have proposed a genotyping system based on various criteria: PCR with genotype-specific primers in the C region (Okamoto et al., 1992c), analysis of amplification products of the NS5 region by slot blotting (Enomoto et al., 1990), or of the 5' NCR (McOmish et al., 1993) and NS5 region (Nakao et al., 1991), by restriction fragment length polymorphism, and sequence analysis of the 5' NCR, E1 or NS5 genes (Bukh et al., 1992, 1993; Okamoto et al., 1993; Simmonds et al., 1993a, b).

Some authors have suggested an association between certain genotypes and disease severity, i.e. cirrhosis and primary liver cancer (Pozzato et al., 1992; Takada et al., 1993), and the response to interferon treatment (Féray et al., 1993; Nousbaum et al., 1993; Pozzato et al., 1992; Yoshioka et al., 1992). This underlines the need for further studies on the natural variation of HCV isolates and their geographical distribution. However few HCV sequence data are available for Africa, particularly central Africa.

We have recently surveyed HCV infection in Gabon (Delaporte et al., 1993) and found a prevalence of up to 6.5% in a general population of 1172 subjects, as determined by a second-generation ELISA method. Fourteen of 30 randomly selected anti-HCV-positive subjects were PCR-positive with primers derived from the 5' NCR of the HCV genome (Delaporte et al., 1993; Féray et al., 1992). To investigate the HCV genotype of the Gabonese isolates, three HCV RNA-positive subjects were randomly selected and various PCR products were cloned for sequence analysis. The PCR primers (Table 1) were selected from the 5' NCR and C region to amplify fragments encompassing the genotyping motif sequences published for different HCV isolates (Cha et al., 1992; McOmish et al., 1993; Okamoto et al., 1992c).

A fragment of 656 bp was amplified and cloned from subjects 6 and 11 by means of the single-tube nested PCR method (Féray et al., 1992; Xu, 1993) using primers (CPS12/CPA14 as outer primers and CPS13/CPA15 as...
Table 1. Sequences of primers used for the amplification of the HCV genome and for HCV genotyping

<table>
<thead>
<tr>
<th>Name</th>
<th>Polarity</th>
<th>Sequence (5' to 3')</th>
<th>Position</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPS12</td>
<td>+</td>
<td>GCGCGATCCCTGCGTTAGTATGAGTGTGC</td>
<td>-256 to -236</td>
<td>This work</td>
</tr>
<tr>
<td>CPS13</td>
<td>+</td>
<td>CTGAGATCTGCTGTCGCCGAGCTGAGT</td>
<td>-197 to -178</td>
<td>This work</td>
</tr>
<tr>
<td>CPS14</td>
<td>+</td>
<td>GCGGAATTCTGTTTCTCGTGCTGATGGTTG</td>
<td>-59 to -39</td>
<td>This work</td>
</tr>
<tr>
<td>CPS15</td>
<td>+</td>
<td>GGCAGATCTCTCTGAGCCGGTACCATG</td>
<td>-18 to 3</td>
<td>This work</td>
</tr>
<tr>
<td>CPA14</td>
<td>-</td>
<td>GGCAGATCCGGGAAAGATAGAAGAGACACC</td>
<td>532 to 511</td>
<td>This work</td>
</tr>
<tr>
<td>CPA15</td>
<td>-</td>
<td>GCTAGATCTTCTTCTGTCGCATAGTGCACGCG</td>
<td>502 to 480</td>
<td>This work</td>
</tr>
<tr>
<td>186</td>
<td></td>
<td>ATGTAACCCATGAGGTCGGC</td>
<td>391 to 410</td>
<td>Okamoto et al. (1992c)</td>
</tr>
<tr>
<td>256</td>
<td>+</td>
<td>GCAGACGACGGAAAGACCTTC</td>
<td>139 to 158</td>
<td>Okamoto et al. (1992c)</td>
</tr>
<tr>
<td>104</td>
<td>+</td>
<td>AGGAAGACTTCGCCAGGCTG</td>
<td>148 to 167</td>
<td>Okamoto et al. (1992c)</td>
</tr>
<tr>
<td>cpsG</td>
<td>+</td>
<td>GTCTGGGCGCCCAATGAT</td>
<td>315 to 333</td>
<td>This work</td>
</tr>
</tbody>
</table>

inner primers) derived from the 5' NCR and C region (Table 1). At least three different clones from each subject were sequenced. The sequences thus represented the 3' end of the 5' NCR and the 5' end of the C region. For the 5' NCR, all the clones from the two subjects 6 and 11 had the same nucleotide sequence, with the exception of one clone from subject 11 which had a mutation of A to C at position -177 (Fig. 1). Only a product of 476 bp in the C region (amplified by CPS14/CPA14 as outer primers and CPS15/CPA15 as inner primers; Table 1) could be cloned for subject 29 (Fig. 1). In the C region, the isolate from subject 29 showed a nucleotide divergence of 9% from the other two Gabonese isolates (6 and 11) which differed by only 2% (Table 2). At the amino acid level the three Gabonese isolates showed a divergence of 1 to 4%. These results indicated that the three Gabonese isolates, termed HC-G, were closely related.

Eleven published full-length HCV RNA sequences corresponding to HCV genotypes I to IV (1a, 1b, 2a and 2b) (Chen et al., 1992; Choo et al., 1989; Inchauspe et al., 1991; Kato et al., 1990; Okamoto et al., 1991, 1992a, b; Takamizawa et al., 1991; Tanaka et al., 1992) were aligned with the Gabonese sequences (data not shown). In the 5' NCR the HC-G isolates showed a divergence of 3 to 8% from these sequence isolates (Table 2). Using Okamoto's classification, the C region of the Gabonese sequences showed divergences of 10 to 14% from types I and II, 15 to 17% from type III and 15% from type IV isolates (Okamoto et al., 1992c). Types I to IV correspond to genotypes 1a, 1b, 2a and 2b of Simmonds' classification (Chan et al., 1992; McOmish et al., 1993) (Table 2). Thus the Gabonese sequences differed considerably from previous isolates and probably represent an HCV genotype distinct from types I to IV of the Okamoto classification.

Recent studies on analysis of sequence variability in the 5' NCR, C and E1 regions have led to the identification of new HCV genotypes when compared with Okamoto's classification (Bukh et al., 1993; Okamoto et al., 1993; Simmonds et al., 1993a, b). These new isolates have been classified as types V/3a, 4, 5a and 6a (Bukh et al., 1993; Okamoto et al., 1993; Simmonds et al., 1993a, b). When the HC-G 5' NCR consensus sequence was compared with 5' NCR sequences of genotype 4, an identical sequence was observed with two samples from Zaire and one sample from Denmark (Bukh et al., 1992, 1993). A maximum divergence of 3% was determined in this region between HC-G isolates and the other isolates belonging to genotype 4 (Table 2). In addition, sequence alignment showed the existence, in almost all the HC-G and genotype 4 isolates, of conserved point mutations when compared with the HCV-1 sequence (these mutations are underlined in Fig. 1). All these data indicate that the three HC-G isolates belong to HCV genotype 4. To analyse this point further, the HC-G C sequences were compared with those of three Egyptian isolates classified as HCV genotype 4 (Simmonds et al., 1993b) (Fig. 1). A nucleotide sequence divergence of 5 to 12%, and an amino acid divergence of 0 to 4% were estimated between Gabonese and Egyptian isolates (Table 2). This further supported the typing of HC-G isolates as genotype 4, since divergences of 13 to 20% for nucleic acids and of 5 to 11% for amino acids were observed when compared with the other genotypes (Table 2).

A fragment of 20 bp in the C region, termed cpsG (Table 1 and Fig. 1), was highly conserved between the three Gabonese isolates (only one mismatch in isolate 29), whereas three to eight substitutions were observed relative to the 11 published complete HCV sequences (types I to IV). CpsG was therefore used as a probe in Southern blot analysis to detect specifically the HC-G isolate in the 30 randomly selected Gabonese subjects (Table 2).

After amplification with the universal typing primers
Fig. 1. Alignment of the nucleotide sequences, part of the 5' NCR and C regions, from the three Gabonese subjects (HC-G6, -G11 and -G29) with the HCV type 1 prototype (HCV-1; Choo et al., 1989; Han et al., 1991) and with C sequences of three HCV type 4 Egyptian isolates (EG-21, -29 and -33) (Simmonds et al., 1993b). The nucleotide variations between different clones from the same subject are shown as lowercase letters. Dots (.) indicate sequence identity with HCV-1. Type 4 conserved nucleotide variations in the 5' NCR, as compared to the HCV-1 sequence, are underlined. The area from which the cpsG probe specific for HC-G has been derived is underlined in the HCV-1 sequence.
Table 2. Nucleotide and amino acid sequence divergence between HC-G isolates and published HCV genotypes

<table>
<thead>
<tr>
<th>Region†</th>
<th>Type‡</th>
<th>HC-G6</th>
<th>HC-G11</th>
<th>HC-G29</th>
<th>I/la</th>
<th>II/1b</th>
<th>III/2a</th>
<th>IV/2b</th>
<th>V/3a</th>
<th>4</th>
<th>5a</th>
<th>6a</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' NCR</td>
<td></td>
<td>0</td>
<td>0-1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>0-3</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td></td>
<td>1 (1)</td>
<td>1-2 (1-2)</td>
<td>8-9 (3-4)</td>
<td>12-14 (6-8)</td>
<td>12-13 (6-9)</td>
<td>15 (8-9)</td>
<td>15 (9-10)</td>
<td>15-16 (6-7)</td>
<td>5-6 (0-1)</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>HC-G6</td>
<td>0</td>
<td>0</td>
<td>0-1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>0-3</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>HC-G11</td>
<td>0</td>
<td>0</td>
<td>0-1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>0-3</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>I/la</td>
<td>0</td>
<td>0-1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>0-3</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>II/1b</td>
<td>0</td>
<td>0-1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>0-3</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>III/2a</td>
<td>0</td>
<td>0-1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>0-3</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>IV/2b</td>
<td>0</td>
<td>0-1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>0-3</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Core</td>
<td>V/3a</td>
<td>0</td>
<td>0-1</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>0-3</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

* The nucleotide and amino acid sequence divergences are given in percentages.
† The 5' NCR and the core region used in the sequence comparisons correspond to nucleotides -1 to -177 and 1 to 479, respectively.
‡ Type I/la sequences correspond to HCV-1 (Choo et al., 1989; Hart et al., 1991) and HCV-H (Inchauspe et al., 1991). Type II/1b sequences correspond to HCV-BK (Takamizawa et al., 1991), HCV-J (Kato et al., 1990), HCV-JT/JT' (Tanaka et al., 1992), HC-J4-83/91 (Okamoto et al., 1992a) and HCV-T3 (Chen et al., 1992). Type III/2a sequences correspond to HC-J6 (Okamoto et al., 1991). Type IV/2b sequences correspond to HC-J8 (Okamoto et al., 1992b). Type V/3a, 4, 5a and 6a sequences are from Bukh et al. (1992, 1993), Okamoto et al. (1993) and Simmonds et al. (1993b).
§ ND, Not determined.

Fig. 2. Detection of the HC-G genotype by Southern blot hybridization. The upper panel shows the amplification products on an ethidium bromide-stained 2% agarose gel. The 263 bp products of the C region were obtained by semi-nested PCR with a universal primer set (186/256 and 186/104) (Table 1) using the conditions described by Okamoto et al. (1992c). The lower panel is the corresponding Southern blot hybridized with 32P-labelled cpsG at 45 °C for 2 h in the presence of 10% formaldehyde, exposed at -80 °C for 24 h. Lanes 1 to 10 correspond to subjects 6, 25, 18, 24, 26, 5, 28, 10, 27 and 3, respectively (Delaporte et al., 1993). Lanes I, II, III and IV correspond to HCV genotypes I to IV of Okamoto's classification (supplied by Dr S. Mishiro, Institute of Immunology, Koraku 1-1-10, Bunkyo-ku, Tokyo, Japan). Lane T is a control (predetermined as genotype II) extracted in the same way as the Gabonese isolates. Lanes M contain 100 Base-Pair Ladder markers (Pharmacia).

of the C region (Table 1 and Fig. 2), none of the samples from the 16 non-viraemic Gabonese subjects (negative in 5' NCR PCR) reacted with the cpsG probe. In contrast, 10 of the samples from the 14 viraemic subjects yielded specific hybridization with cpsG (Table 2 and Fig. 2). In the four cpsG-negative subjects, the HCV RNA was genotyped as type I (type Ia) in one case, and could not be classified in the remaining three cases. (Further studies of these latter isolates are in progress.) The suitability of the cpsG probe to detect HC-G isolates specifically was assessed in stringent hybridization conditions and by the absence of hybridization with samples predetermined as genotypes I to IV using Okamoto's classification (Fig. 2). Specific semi-nested PCR using the cpsG oligonucleotide as the primer was also performed on samples from the 14 viraemic subjects. The first round of PCR was carried out with two universal primers (186 and 256), and was followed by amplification with primers 186 and cpsG (Table 1). The 10 samples testing positive with the cpsG probe also yielded a specific PCR product of the expected size (95 bp). In contrast, the other four samples and certain samples predetermined as genotypes I to IV could not be amplified with the specific primer (data not shown), suggesting that Southern blot analysis with the cpsG probe reliably detects HC-G isolates. The reactivity of the cpsG probe with PCR products of genotypes 3, 5 and 6 remains to be tested.

In summary, we have analysed the sequences of the 5' NCR and C regions of three HCV isolates from Gabonese subjects. Sequence comparisons showed that Gabonese isolates are closely related (91 to 99% identity in the C region), and could be classified as genotype 4.
using Simmond’s classification (97 to 100% and 88 to
95% identity in the 5’ NCR and C regions, respectively, with HCV genotype 4 isolates) (Bukh et al., 1992, 1993; Simmonds et al. 1993b). Based on their 5’ NCR and E1 sequences Bukh et al. (1993) have identified subtypes in genotype 4. From our 5’ NCR sequence data HC-G sequences Bukh means of the cpsG probe (L.-Z. Xu, unpublished results). Preliminary analysis of 13 HCV RNA-positive subjects agreement with previous studies where the majority of Gambia has failed to identify the HC-G genotype by HCV RNA-positive Gabonese subjects. This is in
Simmonds It thus appears that the HC-G genotype is restricted to certain areas of central Africa.

The successful detection of the HC-G genotype by means of the cpsG probe again confirms the importance of such short sequences for rapid and simple HCV genotyping (Cha et al., 1992; Chan et al., 1992; Enomoto et al., 1990; Okamoto et al., 1992c; Simmonds et al., 1993b; Stuyver et al., 1993). The different genotyping techniques have so far been based on only one region of the HCV genome (Enomoto et al., 1990; McOmish et al., 1993; Okamoto et al., 1992c; Simmonds et al., 1993b; Stuyver et al., 1993). Stuyver et al. (1993) suggested that, because of possible nucleotide mutations, at least two specific probes were necessary for genotype determination; thus, the cpsG probe might be used in combination with the available genotyping assays based on the 5’ NCR (Simmonds et al., 1993b; Stuyver et al., 1993) to detect this new HC-G genotype. Finally, the availability of these new sequences might facilitate the design of specific serological tests for African HCV isolates.

The authors thank Bernard Larouzé for initiating the first study on these samples (Delaporte et al., 1993). Valerie Thiers for technical assistance and David Young for editing the manuscript. Work was supported by CNAM, EEC, INSERM, Diagnostic Pasteur-SANOFI and Institut Pasteur.

References


(Received 15 February 1994; Accepted 15 April 1994)