New Natural Intergenotypic (2/5) Recombinant of Hepatitis C Virus

Florence Legrand-Abravel, Julie Claudinon, Florence Nicot, Martine Dubois, Sabine Chapuy-Regaud, Karine Sandres-Saune, Christophe Pasquier, and Jacques Izopet*

Laboratoire de Virologie, Institut Fédératif de Biologie, INSERM U563-IFR30, 330 Avenue de Grande-Bretagne, TSA 40031, 31059 Toulouse Cédez 9, France

Received 29 November 2006/Accepted 19 January 2007

A 9.2-kb sequence from a hepatitis C virus (HCV) strain found in southwest France was compared to sequences from reference strains in HCV sequence databases. We found a recombinant virus with genotype 2 at the 5′ end and genotype 5 at the 3′ end. The crossover point was located between genes NS2 and NS3. Recombination between HCV genotypes must now be considered in studies on HCV epidemiology and evolution and in predictions of the virus response to antiviral therapy. Knowing the location of the recombination point may also be useful for constructing infectious chimeric viruses.

Studies of the nucleotide sequences of hepatitis C virus (HCV) variants from different individuals and geographical regions have revealed at least six genotypes and many subtypes (15, 16). The genotype is presently a major factor in the choice of treatment and the prognosis (2, 7). Knowledge of the distributions of genotypes in different parts of the world may help us to understand the epidemiology and evolution of HCV. Until 2001, the evolution of HCV was thought to proceed in a clonal manner, with diversity generated through the accumulation of mutations. However, homologous recombinations have been demonstrated in places as far apart as St. Petersburg (5), Vietnam (10), the Philippines (4), and Peru (1). We have now identified a new natural intergenotypic recombinant. Full-genome sequencing indicated that it was produced by recombination between genotypes 2 and 5.

Identification of a potential recombinant strain of HCV. We previously analyzed the responses to treatment of 26 patients infected with HCV genotype 5 (HCV-5) in southwest France and the associated epidemiology (6). Phylogenetic analysis of the NS5B genes of the 26 corresponding strains pointed out a particular strain, R1. Twenty-five local strains formed a cluster with the genotype 5a reference sequences, while the R1 strain formed a distinct branch in the tree (Fig. 1A). The NS5B sequence with accession number AY373486 corresponds to genotype 2, while the 3′ part of the R1 sequence, from 5′ untranslated region (UTR; n = 19) clustered with genotype 2 and all the clones from the NS5B region (n = 35) clustered with genotype 5 (data not shown). This result excluded a double infection.

Determination of the entire sequence of the R1 strain. cDNA was synthesized with reverse transcriptase (Superscript II; Invitrogen) and an RNase inhibitor (RNaseOUT; Invitrogen). A 7,400-nucleotide (nt) fragment was amplified and cloned using the Expand long-template PCR system (Roche Diagnostics, France) with sense primer KS3-2a and antisense primer ENO2 (Table 1). The other parts of the genome were directly sequenced. Almost the whole genome was sequenced (9,246 nt) with the Applied Biosystems Taq polymerase kit. The structure of the HCV genome was checked using the sequence locator tool (http://hcv.lanl.gov/content/hcv-db/LOCATE/locate.html). The various regions of the HCV genome were identified, and no stop codon was identified. The R1 sequences downstream (fragment of 3,430 nt) and upstream (fragment of 5,390 nt) of the putative breakpoint were compared with reference sequences (Fig. 2). The 5′ end of R1, from the 5′ UTR to NS2, corresponded to genotype 2, while the 3′ part of the R1 sequence, from NS3 to NS5B, was closer to that of genotype 5.

Identification of the recombination point. Recombination analysis was performed with the Recombinant Detection Program version 2 (http://darwin.uvigo.es/rdp/rdp.html) (8). This tool provides statistical evidence for the breakpoint site by using six methods (Recombination Detection Program, Geneconv, BOOTSCAN, maximum chi-square, Chimera, and sister scanning analyses). Recombination analyses of the R1 sequence and genome 2k (AB0331663), genome 2c (D50409), genome 2a (AB047639 and AX057395), genome 2b (AB030907 and AF238486), and genome 5a (AF064490 and Y13184) reference sequences indicated a crossover point at the beginning of the NS3 region, between nucleotides 3420 and 3440, depending on the method (P value, <0.01 for each method) (Fig. 3). Thus, the analysis by the Recombination Detection Program version 2 corroborated the phylogenetic tree results and showed that the 5′ part of the R1 strain is related to genotype 2k and the 3′ part is related to genotype 5. Further analyses did not reveal recombination with other genotypes.

* Corresponding author. Mailing address: Laboratoire de Virologie, Institut Fédératif de Biologie, INSERM U563-IFR30, 330 Ave. de Grande-Bretagne, TSA 40031, 31059 Toulouse Cédez 9, France. Phone: (33) 5 67 69 04 24. Fax: (33) 5 67 69 04 25. E-mail: izopetj@chu-toulouse.fr

† Published ahead of print on 31 January 2007.
0.1 substitutions by site

FIG. 1. (A) Phylogenetic analysis of the R1 strain and 25 sequences from HCV-5 from southwestern France, performed with 400 bp within the NS5B region. (B) Phylogenetic analysis of the R1 strain and 137 sequences from HCV-2 from southwestern France, performed with 200 bp within the E2 region. Reference sequences from HCV strains from the Los Alamos HCV sequence database are listed by their accession numbers.

TABLE 1. Primers used to amplify overlapping PCR fragments

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
<th>5' position</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW11</td>
<td>GCCTTATAGTATAGTGTCATCGTACAGCCCTCAG</td>
<td>1</td>
<td>Sense</td>
</tr>
<tr>
<td>KY78</td>
<td>CTGCAAGGACCCTATCAGGCACTG</td>
<td>288</td>
<td>Antisense</td>
</tr>
<tr>
<td>ENO1</td>
<td>TGGGATCCCTGATACCCGCTGCTTTGA</td>
<td>8,245</td>
<td>Sense</td>
</tr>
<tr>
<td>ENO2</td>
<td>GCGGAAATCCCTTGATAGCCTCCGTGAA</td>
<td>8,645</td>
<td>Antisense</td>
</tr>
<tr>
<td>KS3-2a</td>
<td>GCTGGGATATGATACCCGCTTCAGTC</td>
<td>1,296</td>
<td>Sense</td>
</tr>
<tr>
<td>LP2-4a</td>
<td>CGGGGTCGRGCAYGRGACATGCTGTGATAAA</td>
<td>9,246</td>
<td>Antisense</td>
</tr>
<tr>
<td>S1-SW-G2</td>
<td>ACGGCTGCGCGCCGTGATGCGATCCATCTGGG</td>
<td>2,951</td>
<td>Sense</td>
</tr>
<tr>
<td>AS-SW-G2</td>
<td>CCTGCGCGCGGAGTAGATCAGACCTGCTGTAAG</td>
<td>3,642</td>
<td>Antisense</td>
</tr>
<tr>
<td>S1-SW-G5</td>
<td>TCATACTGGCTGCGCTATGATACCTGATCAGCC</td>
<td>2,806</td>
<td>Sense</td>
</tr>
<tr>
<td>AS1-SW-G5</td>
<td>CCCACCAAGAGTTGGTGTGCTGCATTAGGATCATT</td>
<td>3,579</td>
<td>Antisense</td>
</tr>
</tbody>
</table>

a The nucleotide positions are numbered according to the sequence locator tool numbering program from Los Alamos for the strain M67463 subtype 1a.
FIG. 1—Continued.
FIG. 2. Neighbor-joining tree depicting the phylogenetic relationships among HCV reference sequences and sequences from the R1 strain. Trees were constructed for a 3,430-nucleotide fragment within the 5’ UTR to the NS2 gene (A) and a 5,390-nucleotide fragment within the NS3 region to the NS5B gene (B) by using the MEGA program with the Tamura-Nei model with gamma-corrected rates. Confidence values calculated by bootstrap analysis (100 replicates) are indicated at the major branching points. Reference sequences from HCV strains from the Los Alamos HCV sequence database are listed by their accession numbers.
Confirmation of the recombination point. We ruled out the possibility of artifactual intra-PCR recombination by amplifying by PCR fragments covering part of the E2 regions to the beginning of the NS3 regions of the R1, genotype 2, and genotype 5 strains. The reverse transcription-PCR kit (QIAGEN) was used with 140 μl of serum HCV RNA from the patient infected with the R1 strain. The sets of primers used were as follows: (i) a genotype 2-specific sense primer and a genotype 5-specific antisense primer (S1-SW-G2 and AS1-SW-G5), (ii) genotype 2-specific sense and antisense primers (S1-SW-G2 and AS1-SW-G2), and (iii) genotype 5-specific sense and antisense primers (S1-SW-G5 and AS1-SW-G5) (Table 1). The R1 strain sequence was successfully amplified only when the reverse transcription-PCR used a type 2-specific sense primer and a type 5-specific antisense primer, while the genotype 2 strain sequence was successfully amplified with the type 2-specific

FIG. 3. Determination of the recombination point with the Recombinant Detection Program method (A) or the BOOTSCAN method (B) using the Recombinant Detection Program version 2. Plots for comparing the R1 and 2k strains are shown in purple, and plots for comparing the R1 strain and the 5a reference strains are in yellow. An arrow indicates the crossover point. Reference sequences from HCV strains from the Los Alamos HCV sequence database are listed by their accession numbers.
primers and the genotype 5 strain sequence with the type 5-specific primers (data not shown). The PCR product obtained with the type 2-specific sense primer and the type 5-specific antisense primer was sequenced; this analysis confirmed the same recombination point.

We have characterized a new HCV-2/5 recombinant genome that was identified because of a difference between the results from E2 and NS5B genotyping. Neither genotype 2 nor genotype 5 strains were found in the serum of the patient infected with R1, suggesting that she was contaminated directly by the R1 strain or that the recombinant strain had replaced the parental strains. The genotype 5-related sequence of the 3′ region from the R1 strain diverges from the other genotype 5 sequences, clustering as subtype 5a. This finding suggests that other HCV-5 subtypes remain to be characterized. Similarly, our analyses of the E2 region or of the sequence from the 5′ UTR to NS2 indicate that the R1 strain belongs to genotype 2, but the subtype is unknown. Our recent work with genotype 2 in our area has shown that this genotype varies greatly. Eight subtypes have been described, and more than 15% of the reported strains are uncharacterized (17). Salemi and Vandamme have demonstrated lower phylogenetic analysis values for the 5′ part of the HCV genome than for the NS5B region (11).

Although recombination may not appear to be common among natural populations of HCV, it must be taken into account as a cause of HCV genetic diversity, which may influence epidemiology and diagnosis. The recombinant strain type 2k/1b is spreading in Europe; 10 strains are described in the HCV database. No data are yet available on the response of HCV genotype 5 strains were found in the serum of the patient infected with R1, suggesting that she was contaminated directly by the R1 strain or that the recombinant strain had replaced the parental strains. The genotype 5-related sequence of the 3′ region from the R1 strain diverges from the other genotype 5 sequences, clustering as subtype 5a. This finding suggests that other HCV-5 subtypes remain to be characterized. Similarly, our analyses of the E2 region or of the sequence from the 5′ UTR to NS2 indicate that the R1 strain belongs to genotype 2, but the subtype is unknown. Our recent work with genotype 2 in our area has shown that this genotype varies greatly. Eight subtypes have been described, and more than 15% of the reported strains are uncharacterized (17). Salemi and Vandamme have demonstrated lower phylogenetic analysis values for the 5′ part of the HCV genome than for the NS5B region (11).

Although recombination may not appear to be common among natural populations of HCV, it must be taken into account as a cause of HCV genetic diversity, which may influence epidemiology and diagnosis. The recombinant strain type 2k/1b is spreading in Europe; 10 strains are described in the HCV database. No data are yet available on the response of this recombinant virus to therapy (5, 9). Genotyping is generally done with small fragments of the HCV genome. The HCV genotype determined directly influences the therapeutic strategy. The widespread existence of recombinant forms would considerably limit the use of genotyping assays that are based on single genome regions, such as the 5′ UTR or the core gene. The genotype may be determined by analyzing regions of interest, like the targets of new inhibitors of the HCV NS5B or NS3 gene.

The entire genome was sequenced, and the recombination site was identified at the beginning of the NS3 region. Several RNA viruses, like the Picornaviridae and the dengue virus, have recombination sites in the region of the genome that codes for the first nonstructural protein (3, 13, 14, 18). The recombination sites of the previously described recombinant HCV strains are also at the NS2/NS3 junction. The NS2/NS3 junction could be a favorable site for generating recombination events, especially with genotype 2 (4, 5, 10).

In conclusion, these data could be most important for vaccine development, diagnosis, and therapeutic strategy.

**Nucleotide sequence accession number.** The recombinant sequence determined in this study has been deposited in the EMBL database with accession number AM408911.

**REFERENCES**