Evolutionary study of HVR1 of E2 in chronic hepatitis C virus infection

Victoria Alfonso,1 Diego Martín Flichman,1 Silvia Sookoian,2 Viviana Andrea Mbayed1 and Rodolfo Héctor Campos1

1Junín 956, 4to. piso, Buenos Aires (1113), Cátedra de Virología, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Argentina
2Unidad de Hepatología, Hospital Argerich, Argentina

INTRODUCTION

Hepatitis C virus (HCV), an enveloped RNA virus included in the Flaviviridae family (Choo et al., 1989; Miller & Purcell, 1990; Houghton et al., 1991), is the major causative agent of non-A, non-B hepatitis (Kuo et al., 1989). More than 80% of individuals infected with HCV become chronically infected and this persistent viral infection frequently develops into liver cirrhosis and hepatocellular carcinoma (Kiyosawa et al., 1990; Ohkoshi et al., 1990; Saito et al., 1990; Alter et al., 1992; Muller, 1996). HCV variability plays a crucial role in escaping the host immune surveillance and establishing persistent infection (Honda et al., 1994; Yamaguchi et al., 1994; Erickson et al., 2001). The HCV genome in single hosts is described as a dynamic population of different but closely related genomes designated quasispecies (Martell et al., 1992; Duarte et al., 1994; Domingo et al., 1998; Alfonso et al., 1999). This quasispecies composition undergoes extensive variations during the course of chronic infection (Okamoto et al., 1992; Kurosaki et al., 1993; Farci et al., 1997; Alfonso et al., 1999), which may result from the accumulation of random substitutions or from changes in the predominant quasispecies population under the selective pressure of the immune response (Lu et al., 2001). The HCV hypervariable region 1 (HVR1) is a 27 amino acid sequence located at the N terminus of glycoprotein E2 (Hijikata et al., 1991; Weiner et al., 1991; Kurosaki et al., 1993). HVR displays marked sequence variability, possibly corresponding to the emergence of immune escape mutants (Hijikata et al., 1991; Weiner et al., 1992; Kurosaki et al., 1993; Van Doorn et al., 1995; McAllister et al., 1998). In spite of the high variability of this region, there is a strong negative selection against some amino acid substitutions since at most codons there is selection for conservative amino acid replacement, pointing to a biological role in the virus life-cycle (McAllister et al., 1998; Sobolev et al., 2000; Penin et al., 2001).

In this study we present an HVR1 evolutionary quasispecies analysis from four chronically HCV-infected patients who were treated with interferon (IFN) for 6 months and followed up for 14–23 months.
METHODS

Patients. Thirteen patients with chronic hepatitis C, eligible for IFN therapy, were included in a clinical trial of IFN therapy at the Liver Unit of the Argerich Hospital, a tertiary university referral centre. The entry criteria for the present study included persistently raised alanine aminotransferase (ALT) for at least the previous 6 months, a positive anti-HCV test, detection of the HCV RNA 5’UT region of the viral genome in serum (Alfonso et al., 2001) by RT-PCR, and liver histology consistent with chronic hepatitis C, obtained within the last 18 months. All patients were negative for hepatitis B surface antigen and anti-human immunodeficiency virus. No patient had a previous history of autoimmune disease, alcohol intake, current intravenous drug use or any chronic liver disease. HCV genotypes were determined upon treatment initiation using the conventional methods of the line probe assay (Stuyver et al., 1996) and/or restriction fragment length polymorphism (Thiers et al., 1997). Four patients were infected with a genotype 1a strain, seven with a genotype 1b strain and two were classified as genotype 2a/c. All patients were treated with IFN-α 2b, 3 MU (Shering Plough) subcutaneously three times a week. Clinical and biochemical follow-up was performed at 1 month intervals. Written informed consent to participate in this study was obtained from all patients and the local ethical committee approved the study protocol in accordance with the 1975 Declaration of Helsinki.

Response to the treatment was defined according to the ALT level and HCV RNA detection during and following treatment as follows: (i) no response (NR), presenting abnormal ALT levels during treatment and/or detectable HCV-RNA; (ii) partial response (PR), with normalized ALT levels and HCV RNA undetectable during and following treatment; and (iii) sustained response (SR), with normalized ALT levels and HCV-RNA undetectable during and following treatment as follows: (i) sustained response (SR), with normalized HCV RNA detection during and following treatment as follows: (i) sustained response (SR), with normalized ALT levels and HCV-RNA undetectable during and following treatment. After undergoing the treatment, only one patient was SR (infected with genotype 1a), six were NR and six were PR.

cDNA synthesis and PCR amplification. Briefly, RNA was extracted from 100 μl serum with the Trizol system (Life Technologies), denatured at 80 °C for 5 min and primed with 0·4 μg random hexamers. The reverse transcription reaction was performed for 90 min at 37 °C, using M-MLV reverse transcriptase (Promega). After heat inactivation at 80 °C for 5 min and chilling on ice, the cDNA was amplified.

The amplification of HVR1 E2 was carried out using the following primers: E2OS (5’-GCCATATAACGGGTCACCGATGGC-3’, sense) and E2OA (5’-TCTCAGGACAGCCTGAAGMGTTGAA-3’, antisense) in the first round and E2NS (5’-GGATATGATGATGAACTGGTC-3’, sense) and E2NA (5’-GGTGTTGAGGCTATCGATTGTC-3’, antisense), in the nested reaction. The PCR reaction consisted of 40 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s and polymerization at 72 °C for 45 s. In order to diminish the introduction of artefactual mutations, reactions were carried out with pfu DNA polymerase (Promega), a high fidelity DNA polymerase. PCR products were purified with the Wizard PCR Prep DNA Purification System (Promega).

Cloning of PCR products. PCR products were cloned in the pGEM-T Easy Vector System (Promega) and transformed into Escherichia coli DH5α-competent cells. Plasmid DNA from transformants producing white colonies was extracted from minipreps cultured overnight and purified with the Wizard DNA Purification System (Promega). The sequences of inserted DNAs were determined in independent clones using a Thermo Sequenase Radio-labelled Terminator Cycle Sequencing Kit (USB Corporation) in both directions.

Phylogenetic analysis. The sequences were aligned according to their position homology using the CLUSTAL X program (Thompson et al., 1997). The mean genetic distance (gd) between all pairs of possible sequences and the standard error of the mean (SEM) were calculated using the MEGA2 program (Kumar et al., 2001) according to the Kimura two-parameter model. The phylogenetic trees were constructed from nucleotide sequences according to neighbour-joining and maximum-likelihood methods, using the MEGA2 program and PHYLIP package (Felsenstein, 1993), respectively. Their reliability was assessed by bootstrap resampling (1000 datasets).

As an index of quasispecies complexity, the normalized Shannon entropy (Sn) was calculated as follows: Sn = −Σi (pi ln pi)/ln N, where pi is the frequency of each sequence and N is the total number of sequences analysed in each sample. Sn varies from 0 (no diversity) to 1 (maximum diversity) (Wolinsky et al., 1996; Pawlotsky et al., 1998).

Antigenic analysis. Antigenic analysis was performed on the amino acid sequences translated from the nucleotide sequences of HVR1 using Parker’s method in the ANTHEPROT package 5 program (Deleage et al., 2001). Parker’s method uses a combination of three parameters for hydrophilicity, accessibility and flexibility to predict the antigenic profile of the protein. In order to analyse the entire HVR1 and because this method uses a calculation window of seven residues, 15 nucleotides were sequenced and translated at either side of HVR1.

RESULTS

The evolution of HCV genomes from 12 patients in whom sustained RNA clearance was not achieved after IFN treatment was studied by direct sequencing of HVR1 from serial samples. Serum samples were taken before (I), soon after (II) and 14–32 months (III) after a 6 month IFN treatment. The direct sequence analysis showed a wide range of mutation rates and a strong prevalence of non-synonymous modifications, which are evolutionary characteristics previously demonstrated for this region. A phylogenetic analysis performed with the sequences showed that, in spite of the high number of nucleotide modifications throughout the evolutionary process, all the isolates belonging to the same patient formed independent clades in the tree, thus confirming the absence of cross-contamination (data not shown; GenBank accession numbersAY390002–AY390035). We also observed a large number of ambiguities at several nucleotide positions as a consequence of the quasispecies nature of HCV viral genomes.

To perform further quasispecies analysis, we selected serum samples from two NR (Cb11 and Cb12) and two PR (Cb02 and Cb03) patients showing different evolutionary characteristics. Patients Cb12 and Cb03 were infected with genotype 1b, patient Cb11 with a 1a strain and patient Cb02 with a genotype 2a/c virus. In these four patients, the serial samples were analysed as follow: HVR1 was amplified, cloned and the clones were sequenced.

For each patient, a phylogenetic analysis of the sequences was performed and quasispecies complexity (Sn) and genetic distance (gd) were calculated. The translated amino acid sequences and their predicted antigenic profiles
were further analysed. Results from each patient are detailed below.

**Patient Cb12**

The phylogenetic analysis showed that the sequences from sample I (the basal sample, 2 months before the beginning of the treatment) grouped into two different clusters of sequences: lineage Ia (five clones) and lineage Ib (seven clones), with an intersample gd of 0·250. The sample II population, taken 1 month after the end of the therapy, presented high values of quasispecies complexity (Sn 0·77) and an intrasample gd of 0·051. Only one of the sample II sequences appeared phylogenetically close to those from lineage Ia. Sample III (25 months after the end of IFN treatment) was composed of a highly homogeneous group of sequences (Sn 0·00), which clustered with lineage Ib with a gd for Ib–III of 0·038 over 31 months (Fig. 1 and Table 1).

The amino acid sequences confirmed the phylogenetic analysis. In sample I there were two distinct lineages whose sequences differed in 12 out of 27 amino acids. Sequences from sample II were highly heterogeneous, though one clone presented only two amino acid substitutions compared with Ia sequences. The amino acid sequences from sample III clones were quite similar to those of lineage Ib (two amino acid positions modified out of 27) (Fig. 2a).

We compared the predicted HVR1 antigenic profiles. While lineages Ia and Ib presented clear differences, in sample II we found several profiles, agreeing with the genetic diversity. The population from sample III presented an antigenic profile closely related to that of lineage Ib (Fig. 2b).

**Patient Cb03**

Sample I from patient Cb03, taken 2 months before the beginning of the treatment, showed sequences with two different profiles: those of lineage Ia formed a heterogeneous group (Sn 0·41, gd 0·020), while those grouped in lineage Ib were highly homogeneous (Sn 0·00, gd 0·000). These groups presented a gd Ia–Ib value of 0·063. As observed in patient Cb12, sample II clones, taken 1 month after the end of therapy, formed a heterogeneous group of sequences (Sn 0·62, gd 0·024). The phylogenetic analysis showed substantial intermingling between viral sequences from lineage Ia and from sample II. The gd between both groups (gd Ia–II 0·021) was similar to that observed within them. On the other hand, sequences from sample III, taken 14 months after the end of treatment, had a very low quasispecies complexity (Sn 0·13) and were clearly associated with the minor lineage (Ib) found in the pre-treatment sample (gd Ib–III 0·001) (Fig. 1 and Table 1).

The amino acid sequences also showed the same two pre-treatment populations, Ia (eight and two clones) and Ib (three clones), which differed between them in four or five amino acids. Sequences from lineage Ia were closely related to the population from sample II but sequences from sample III were identical to those from the minor lineage Ib (Fig. 2a).

**Fig. 1.** Phylogenetic trees derived by the maximum-likelihood method corresponding to all clones analysed in each patient. Bootstrap values higher than 70 % of neighbour-joining congruent trees are indicated for each genetic group. The different symbols identify the sequences from each population. Numbers in parentheses represent the number of clones having the same sequence. The symbol * in patient Cb02 marks the III* clone as indicated in Results. GenBank accession numbers are indicated for each sequence.
Antigenic profiles showed differences between the pre-treatment lineages Ia and Ib and also a similarity of these with populations from samples II and III, respectively (Fig. 2b).

**Patient Cb02**

In contrast to the results from patients Cb12 and Cb03, sample I from patient Cb02, taken 2 months before treatment initiation, showed a single homogeneous population (Sn 0·00, gd 0·000).

Nucleotide sequences from early post-treatment clones (sample II), over a period of 10 months, were quite different from those belonging to sample I (gd I–II 0·195) and were phylogenetically close to those from sample III with a gd II–III value of 0·070. Sample II had a high complexity value (Sn 0·49) and the highest intrasample gd (0·036). Sample III, taken 23 months after the end of the therapy, showed a homogeneous population with low normalized entropy (Sn 0·13) and the highest intersample genetic distance (gd 0·070). While most of the clones (11 out of 12) had a sequence similar to those from sample II (Fig. 2a), a single clone (III*) appeared phylogenetically closer to the population from sample I with a gd I–III* value of 0·064 over a 31 month period (not shown).

The amino acid sequence analysis showed that populations from samples I and II differed in 11 or 12 out of 27 positions. It should be pointed out that, despite the quasispecies complexity of sample II, its clones appeared to be represented by only two amino acid sequences – corresponding to ten and three clones – with three amino acid substitutions, the population from sample III being close to the ten-clone group (Fig. 2a).

In these samples, antigenic analysis showed different profiles between clones from sample I and those from sample II. The antigenic profile from sample III was similar to the ten-clone group from sample II (Fig. 2b).

**Patient Cb11**

In this particular case, the direct sequencing of HVR1 did not present modifications throughout the follow-up. The cloning analysis of sample I (before the treatment) showed the presence of three nucleotide sequences (AY314963, three clones; AY314964, one clone; and AY314965, one clone). In sample II (4 months after the end of the treatment) we found four nucleotide sequences (AY314966, two clones; AY314967, one clone; AY314968 one clone; and AY314969, one clone). Pairwise comparison showed that all the sequences differed in up to four nucleotide positions. The nucleotide modifications were synonymous.

**DISCUSSION**

HVR1 was directly sequenced in samples from 12 chronically infected patients who had not responded to IFN treatment. Due to the quasispecies nature of HCV circulating genomes, in the direct sequence analysis of HVR1 several nucleotide positions remained undefined, thus impairing further studies. Therefore, a quasispecies analysis was performed in samples from four patients presenting different evolutionary characteristics. The analysis of evolving viral populations from patients Cb12, Cb03 and Cb02 showed that pre-treatment samples presented different genetic quasispecies compositions. Patient Cb12 presented two homogeneous lineages in sample I with clear differences between them. In patient Cb03, there were also two different lineages in sample I but one of them was homogeneous while the other one was heterogeneous. Finally, patient Cb02 showed only one homogeneous lineage detected in the pre-treatment sample. These dissimilar genetic

---

### Table 1. Normalized Shannon entropy and intra-/intersample genetic distances

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sample</th>
<th>No. of clones</th>
<th>Normalized Shannon entropy</th>
<th>Mean (SEM) intrasample genetic distance (gd)</th>
<th>Patients</th>
<th>Mean (SEM) intersample genetic distance (gd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cb12</td>
<td>Ia</td>
<td>5</td>
<td>0·00</td>
<td>0·000</td>
<td>Ia–Ib</td>
<td>0·250 (0·067)</td>
</tr>
<tr>
<td></td>
<td>Ib</td>
<td>7</td>
<td>0·00</td>
<td>0·000</td>
<td>Ia–II</td>
<td>0·114 (0·036)</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>15</td>
<td>0·77</td>
<td>0·051 (0·013)</td>
<td>Ia–III</td>
<td>0·267 (0·070)</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>10</td>
<td>0·00</td>
<td>0·000</td>
<td>Ib–II</td>
<td>0·232 (0·060)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ib–III</td>
<td>0·038 (0·022)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>II–III</td>
<td>0·249 (0·062)</td>
</tr>
<tr>
<td>Cb03</td>
<td>Ia</td>
<td>10</td>
<td>0·41</td>
<td>0·020 (0·008)</td>
<td>Ia–Ib</td>
<td>0·063 (0·027)</td>
</tr>
<tr>
<td></td>
<td>Ib</td>
<td>3</td>
<td>0·00</td>
<td>0·000</td>
<td>Ia–II</td>
<td>0·021 (0·009)</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>10</td>
<td>0·62</td>
<td>0·024 (0·012)</td>
<td>Ia–III</td>
<td>0·064 (0·027)</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>11</td>
<td>0·13</td>
<td>0·002 (0·002)</td>
<td>Ib–II</td>
<td>0·068 (0·028)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ib–III</td>
<td>0·001 (0·001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>II–III</td>
<td>0·070 (0·028)</td>
</tr>
<tr>
<td>Cb02</td>
<td>I</td>
<td>11</td>
<td>0·00</td>
<td>0·000</td>
<td>I–II</td>
<td>0·195 (0·052)</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>13</td>
<td>0·49</td>
<td>0·036 (0·014)</td>
<td>I–III</td>
<td>0·157 (0·048)</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>12</td>
<td>0·12</td>
<td>0·020 (0·007)</td>
<td>II–III</td>
<td>0·070 (0·025)</td>
</tr>
</tbody>
</table>

---

**V. Alfonso and others**
**Fig. 2.** Amino acid sequences (a) and antigenic profiles (b) obtained from patients Cb12, Cb03 and Cb02. Dots represent amino acid identity with the upper sequence and letters indicate amino acid substitutions. The number of clones is displayed in parentheses. The black bar indicates HVR1.
compositions before the treatment initiation may have resulted from different stages of the chronic infection in each patient (Stumpf & Pybus, 2002).

Samples taken soon after the end of the treatment (sample II) presented the highest intrasample gd and Sn values, which are related to a diversification process. It was suggested that the rise of new variants after treatment is consistent with the selection and expansion by IFN of variants present as minor populations in the pre-treatment specimen (Hassoba et al., 1999).

However, after 14–23 months of IFN withdrawal, the three samples III were highly homogeneous with the lowest Sn and intrasample gd values for each patient.

The evolution of the HVR1 region is mainly guided by the selection of non-synonymous substitutions. It may proceed by a point substitution pattern reflecting the fixation of random mutations or by the shift from one quasispecies variant population to another (Lu et al., 2001). Our results show that intermediate samples from PR or NR patients, taken during the lengthy follow-up period, allowed the visualization of the quasispecies shift pattern of evolution in addition to the point substitution mechanism.

In patient Cb12, the viral population from sample II may have resulted from the selection and further diversification of a minor population close to the Ia lineage. The population found in sample III could have resulted from the selection of a coexisting lineage in the pre-treatment sample (lineage Ib) after the incorporation of two amino acid modifications during the 31 months of the infectious process.

Patient Cb03 showed a different behaviour, since the population from sample II appeared to be very close to the pre-existing lineage Ia, i.e. not having any differentiating characteristics, but after 14 months of follow-up, clones from sample III presented the same amino acid composition as Ib clones. These clones could have derived from the selection of that pre-treatment population (detected 22 months earlier).

Patient Cb02 presented a different pattern of evolution. The population from sample II, after IFN treatment, presented 11 or 12 out of 27 amino acid differences with reference to clones from sample I and may have resulted from the selection of a previous undetected lineage. However, the low genetic distance between samples II and III and the few amino acid differences between them suggest that sequences from sample III evolved from those of sample II over 21 months through the accumulation of point mutations.

Had the study not comprised a cloning analysis and an intermediate sampling period (sample II), the results would have been very different. For example, if amino acid consensus sequences from samples I and III of patient Cb12 had been compared, we would have obtained a rate of substitution of two amino acids over 31 months. However, by including the intermediate sample II, we found ten amino acid substitutions during the first 9 months between samples I and II and twelve extra substitutions during the following 22 months between samples II and III. This highlights the importance of the follow-up of the evolutionary process, especially when the genomic region being analysed is highly sensitive to an applied selective force such as IFN treatment. Therefore, the existence of a shift mechanism, especially when applied to hypervariable regions, requires the use of this methodology to avoid a misinterpretation of the results.

Finally, patient Cb11 represented a case where different components of the quasispecies coexisted during long periods of chronic infection without replacement (McAllister et al., 1998; Gaud et al., 2003). Patients Cb12, Cb03 and Cb02 presented a fast evolution of viral consensus sequences, like many other RNA viruses. In contrast, Cb11 viral populations displayed long periods of replicative evolutionary stasis in which the consensus sequence remained unmodified (Domingo et al., 1995).

From a phenotypic framework of analysis, HVR1 has been considered as structurally flexible and antigenically variable. Variation is unequally distributed within the region, thus resulting in some invariant positions, with others highly variable but limited to discrete amino acid replacements. This phenomenon indicates that a positive selection is driving the mechanisms of replacement at most of the positions and that a negative pressure is also being exerted against the substitutions in other positions (McAllister et al., 1998; Afonso et al., 1999; Penin et al., 2001).

We compared the predicted HVR1 antigenic profiles and found that all the clone sequences displayed the same two antigenic regions previously described (Penin et al., 2001), but individual antigenic profiles differed markedly from one lineage to another, especially in the N-terminal region.

The amino acid analyses of our populations reproduced most of our genetic observations, such as the co-existence of different lineages (Cb12 Ia and Ib, Cb03 Ia and Ib), homogeneity at the end of the follow-up period (Cb12 III, Cb03 III, Cb03 II) and the selection of pre-treatment populations (Cb12 III from Cb12 Ib, Cb03 III from Cb03 Ib).

In conclusion, we have shown here that, regardless the complexity or diversity of the viral populations before treatment of our three patients, they evolved towards genetic diversification after a selective pressure. Once the environment became stable (at the end of the follow-up period, 14–23 months after the end of treatment), the populations tended towards homogeneity. We propose that the evolution of HVR1 E2 is most likely driven by selection of clearly different subpopulations (modification of quasispecies equilibrium), in agreement with previous reports (Pawlotsky et al., 1999), rather than by a continuous mechanism related to the successive accumulation of point mutations.
Despite the common characteristics noted above, this analysis has confirmed the individual features of HCV genetic evolution in the four subjects and shown that the failure of IFN treatment in vivo follows different routes. The results of this analysis probably reflect the molecular counterpart of viral fitness for individual environments (Manzin et al., 1998). The different patterns of viral molecular changes point to the existence of different virus and host factors contributing to the evolution of HCV quasispecies in vivo.

ACKNOWLEDGEMENTS

This study was supported by grants from the Universidad de Buenos Aires (SECyT–UBA, B076), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, PID 723), Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT, PICT 2000 05-08355) and Ministerio de Salud Pública de la Nación (Beca Carrillo–Octavía). We wish to thank Professor Rex Davis for language supervision.

REFERENCES


Ohkoshi, S., Kojima, H., Tawaraya, H. & 8 other authors (1990). Prevalence of antibody against non-A, non-B hepatitis virus in


