Intra-host evolutionary dynamics of hepatitis C virus E2 in treated patients

Victoria Alfonso, Viviana Andrea Mbayed, Silvia Sookoian and Rodolfo Héctor Campos

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

Hepatitis C virus (HCV) displays high genetic diversity. Inter-host sequence variability may mainly reflect a neutral drift evolution. In contrast, intra-host evolution may be driven by an adaptive selection to host responses to infection. Here, HCV E2 intra-host evolution in two patients during the course and follow-up of successive treatments with IFN-α and IFN-α/ribavirin was investigated. Phylogenetic analyses suggested that adaptive pressures prompt a continuous selection of viral variants derived from the previous ones (intra-lineage evolution) and/or a swapping of viral lineages during the course of the infection (inter-lineage evolution). Selection would act not only on the phenotypic features of hypervariable region 1 (HVR1) but also on those of the flanking regions. The pressures operate mainly at the amino acid level, but they also appeared to act on nucleotide sequences. Moreover, HVR1 heterogeneity seemed to be strongly constrained. This work contributes to the knowledge of HCV intra-host evolution during chronicity.

Hepatitis C virus (HCV) is a major cause of chronic hepatitis worldwide (Choo et al., 1989; Alter & Seeff, 2000) and can lead to liver cirrhosis and hepatocellular carcinoma (Kiyosawa et al., 1990; Saito et al., 1990; Muller, 1996). This RNA virus displays high genetic diversity. Six genetic clades and various subtypes have been described (Simmonds, 1999) and are associated either with a specific geographical distribution or with different biological features (Pawlotsky, 2003; Zeuzem, 2004). In spite of the predominance of the neutral sequence drift process, part of this divergence could result from adaptation to individual immune responses (Simmonds, 2004). One target for a variety of selective pressures is the viral envelope glycoprotein E2 (McAllister et al., 1998; Sobolev et al., 2000). This protein contains not only the most variable region of the virus, known as hypervariable region 1 (HVR1) (Hijikata et al., 1991; Weiner et al., 1991; Kurosaki et al., 1996), which is thought to be a target for neutralizing antibodies (Farci et al., 1996; Shimizu et al., 1996), but also conserved regions. In spite of the marked inter-host diversity of HVR1, there is a strong constraint on the replacement of some amino acids, as most are conservative for their physico-chemical properties (McAllister et al., 1998; Penin et al., 2001). The quasispecies theory and population genetics provide theoretical frameworks for the study of the evolutionary dynamics of most RNA viruses (Domingo et al., 1998; Stumpf & Pybus, 2002).

The study of evolution of the HCV populations within hosts must consider the dynamics and complexity of the quasispecies in the patient during the course of the infectious process. There have been many cross-sectional studies of HVR1 quasispecies composition, but only a few have analysed virus evolution by longitudinal studies (Farci et al., 2000; Abate et al., 2004).

In this work, we investigated the composition and molecular evolution of the HCV E2 quasispecies in two patients during the course and follow-up of two successive treatments with alpha interferon (IFN-α) and IFN-α/ribavirin. The analysed fragment (nt 1425–1972) was composed of HVR1 (nt 1491–1571) and 467 nt flanking this region. This genomic fragment thus contained both variable and conserved regions and was amplified by RT-PCR using Pfu DNA polymerase (Promega) to diminish the introduction of artefactual mutations (primer sequences are available in Supplementary Table S1, available in JGV Online). PCR products were cloned and independent clones were bidirectionally sequenced (ALFexpress II; Amersham Biosciences). Sequences were aligned using the CLUSTAL X program (Thompson et al., 1997). Nucleotide analyses were performed with PAUP* version 4.0b10 (Swofford, 2002). The best-fitting nucleotide substitution models for the different data analysed in each patient were chosen with MODELTEST version 3.06 (Posada & Crandall, 1998) (see Supplementary Table S2, available in JGV Online). Phylogenetic trees were constructed according to the maximum-likelihood method.

To assess diversity, genetic distances were calculated for all the GenBank/EMBL/DBJ accession numbers of the sequences determined in this work are AY876391–AY876493.

Supplementary material is available in JGV Online.
pairs of sequences. Phylogenetic trees constructed on the basis of amino acid sequences were obtained using the neighbour-joining method in the MEGA2 program (Kumar et al., 2001). The reliability of the phylogenetic trees was assessed by bootstrap resampling (1000 datasets). Patient A, infected with a genotype 1b strain, was treated for 6 months with IFN-α 2b and with IFN-α 2b plus ribavirin for another 6 months. He did not respond to either therapy. Serum samples were taken: (I) just before the IFN-α treatment; (II) 6 months after the end of the monotherapy; (III) 4 months later, at the beginning of the combined treatment; and (IV) 3 months and (V) 6 months after the end of this therapy. Patient B, infected with HCV genotype 1a, was treated for 6 months with IFN-α 2b but did not respond. He received a second treatment for 12 months of IFN-α 2b plus ribavirin. This patient initially responded to the combined therapy, but relapsed at the end of the treatment. Serum samples were taken: (I) just before the first treatment; (II) 7 months after the end of this IFN-α treatment; (III) 3 months later, before the beginning of the IFN-α/ribavirin treatment; (IV) at the end of the combined treatment; and (V) 6 months later.

Phylogenetic analysis of the nucleotide sequences from all clones from patient A showed a shift pattern of evolution, with selection of clearly distinct lineages of sequences at different times of the infectious process, supported by high bootstrap values (Fig. 1a). In the basal sample, we found two different clusters of sequences: Ia (seven clones) and Ib (three clones). In sample II, the viral sequences composed a single group that was different from both lineages found in sample I. Three months later, there was a single lineage of sequences, different from the previous ones. We found a single clone in sample II that grouped with lineage III and a single clone in sample III that was associated with lineage II. Clone sequences from samples IV and V formed a new group genetically distant from the previous lineages. In this patient, the divergence between lineage sequences was remarkably higher than within them (see Supplementary Table S3, available in JGV Online). This inter-lineage evolution has been previously observed (Pawlotsky et al., 1999; Alfonso et al., 2004). In contrast, analysis of all clones sequenced for patient B revealed a phylogenetic structure that reflected the continual selection of genomes within lineages, a process that could be defined as intra-lineage evolution.

![Phylogenetic trees](https://example.com/phylogenetic-trees.png)

**Fig. 1.** Phylogenetic trees corresponding to the entire sequenced fragment of all clones from the two patients: sample I, green; sample II, red; sample III, blue; sample IV, yellow; sample V, pink. (a) Nucleotide sequences from patient A. (b, c) Nucleotide (b) and amino acid (c) sequences from patient B. Lineages are indicated. Bootstrap values higher than 70% are shown for each genetic group. Asterisk (*), clone III*, as indicated in the text.
evolution (Fig. 1b). Clones from sample I grouped into four lineages of sequences (Fig. 1b: a, six clones; b, one clone; c, two clones; d, one clone). At the end of the therapy (samples II and III), there were only two lineages of sequences: lineage a, where populations Ila and IIIa appeared phylogenetically close to la clones, and lineage b composed of populations IIb and IIIB that clustered with clone lb. At the end of the combined therapy, all the viral sequences detected were included in or derived from lineage a. The sequences designated IVa were intermingled with those from lineage a, and the sequences designated IVa2 formed a different lineage. Sample V showed a group of viral sequences that clustered intermingled with sequences from population IVa2. Both populations constituted the so-called lineage a2. This lineage was considered to be derived from lineage a, since they shared a monophyletic origin in the tree. The viral sequences found at any one time seemed to derive from variants of the previous sample. This temporal evolutionary pattern has been described as the most possible representative for HCV intra-host dynamics (Grenfell et al., 2004). The genetic distances (see Supplementary Table S4, available in JGV Online) coincided with the clustering of the above-mentioned lineages.

In both patients, the use of nucleotide or amino acid sequences for the phylogenetic analysis did not modify the tree topologies. Nevertheless, in patient B, amino acid analysis was unable to discriminate the lineage a2, which was selected in the last time points (Fig. 1c). The reason for such inability was that the eight nucleotide substitutions that characterized the selected lineage were all synonymous. Both the appearance of these previously undetected substitutions in 50% of the viral population after the last treatment and the fixation of that lineage 6 months later suggested a shift mechanism rather than a consecutive accumulation of changes. These results could indicate a selective pressure exerted at the nucleotide level, although we cannot rule out a process of random drift.

To perform a more in-depth analysis, we studied HVR1 and the flanking region (FR) separately. HVR1 has been described as a target for positive selection by immune system-mediated neutralization (Weiner et al., 1992; Van Doorn et al., 1995; Farci et al., 2002) and for purifying selection (McAllister et al., 1998; Penin et al., 2001), suggesting an important role of this region in driving the evolutionary process. It has been widely demonstrated that HVR1 displays high diversity associated with immune escape when different samples from the same or distinct patients are compared (Hijikata et al., 1991; Kurosaki et al., 1993). HVR1 may be a target for neutralizing antibodies and HCV persistence may therefore require continuous virus amino acid changes (Taniguchi et al., 1993; Kanzanou et al., 2003). Phylogenetic analysis of HVR1 in patient A showed the same topology as for the entire fragment. This genomic region was highly homogeneous in the intra-lineage analysis, but was very heterogeneous in the inter-lineage analysis (Fig. 2a and b; see also Supplementary Table S3). The analysis in patient B also showed a similar topology to that presented for the entire fragment; the nucleotide sequences that belonged to lineage a2 were highly homogeneous in this region, as was observed for all lineages from patient A (Fig. 2c and d).

A high inter-patient variability has been widely reported for HVR1. In contrast, we observed nucleotide and amino acid homogeneity in different lineages from the two analysed patients, suggesting that the genomic structure of HVR1 populations in a defined environment is submitted to strong pressures exerted on this region as a whole. These results agree with our previous observations that HVR1 tends to be homogeneous several months after the end of IFN-α treatment (Alfonso et al., 2004). In this paper, we showed that, in some cases, not even a low degree of diversity is allowed. In patient B, it was evident that such pressure had driven the evolutionary process, since amino acid HVR1 sequences from populations IIIa, IV (with the exception of one clone) and V were identical (Fig. 2d). In addition, once the HVR1 amino acid sequence was selected (population IIIa), variants with defined synonymous substitutions increased their frequency progressively, tending to become fixed in the population (3/10 and 8/10 sequences were identical in samples IV and V, respectively). This fixation could be the result of a selective process exerted at the nucleotide level.

Many authors have shown the importance of HVR1 in driving the evolutionary process. However, there have been few longitudinal studies that have evaluated the quasispecies dynamics of the genomic region flanking HVR1 (McAllister et al., 1998; Sheridan et al., 2004). For patient A, the nucleotide FR sequence analysis [see Supplementary Fig. S1(a), available in JGV Online] showed the same topology as that of the entire fragment, but the heterogeneity within lineages was slightly higher than for HVR1 and the mean genetic distances between the lineages were lower than for HVR1 (see Supplementary Table S3). The FR phylogenetic analysis in patient B showed poor discriminative ability. Only lineages b and a2 presented significant bootstrap supports [see Supplementary Fig. S1(b)]. Six of eight synonymous substitutions that characterized lineage a2 belonged to the FR. The sequence of clone III* clustered with lineage b when we analysed HVR1, but did not group with this lineage in the FR. Although it is feasible that it may have been an artefact produced during the PCR or cloning process, the possibility that a recombination event has taken place could not be discarded, since inter- and intra-genotypic recombination in HCV has been reported by some authors in recent years (Kalina et al., 2002, 2004; Colina et al., 2004).

The amino acid substitutions outside HVR1 that define the main lineages in both patients were not evenly distributed along the studied fragment and most were concentrated between aa 435 and 475. McAllister et al. (1998) have described this region as hypervariable in subtype 1b isolates and mentioned that some substitutions tended to occur
Fig. 2. Phylogenetic trees corresponding to HVR1. (a, b) Nucleotide (a) and amino acid (b) sequences from patient A. (c, d) Nucleotide (c) and amino acid (d) sequences from patient B. Lineages are indicated. Bootstrap values higher than 70% are shown for each genetic group. Samples are indicated with the same colours as in Fig. 1. Asterisk (*), clone III*, as indicated in the text.
among restricted amino acids. For patient A (genotype 1b), we found amino acid substitutions that agreed with this previous report (A/S in aa 441, E/N/H/Y in aa 446, K/R in aa 447 and A/S in aa 450). In this patient, we also assessed a marked clustering of the substitutions at aa 445, 446 and 447. The substitutions that characterized the main lineages in patient B (genotype 1a) resulted in four modified amino acids at positions 439, 447, 467 and 475. Published data provide some evidence that at early time points after the beginning of the therapy, the non-HVR1 region of E2 evolves under purifying selection (Farci et al., 2002). In contrast, it has recently been suggested that this region may also be subject to immune pressure (Chambers et al., 2005). Both observations made in our study from patient A and, despite the different viral subtypes, the convergence of modifications at aa 439, 447 and 467 in both patients indicated the participation of this region during the evolutionary process as a target of selection (Table 1).

Finally, evolutionary quasispecies analysis, performed during the chronic course of HCV-infected patients undergoing antiviral treatment, confirmed that the evolution of this virus in each infection showed individual features, probably determined by different virus and host factors, as has been previously suggested (Manzin et al., 1998; Stumpf & Pybus, 2002). While the inter-host evolution of HCV reflects a process of neutral sequence drift, the intra-host evolution may be driven by a selection of changes associated with adaptation to different host responses to infection (Grenfell et al., 2004; Simmonds, 2004). Our results showed that such pressures prompt the continuous selection of variants derived from the previous ones and/or the swapping of viral lineages during the infectious process (Stumpf & Pybus, 2002). This selection would be exerted over phenotypic features of HVR1 and also over FRs. The pressures operate mainly at the amino acid sequence level, but in accordance with our results they would also act on nucleotide sequences, although a genetic process of random drift cannot be ruled out. In addition, HVR1 seems to be under constraints that strongly limit its heterogeneity at most time points of the infectious process. This work contributes to our knowledge of HCV intra-host evolution during the course of chronic infection of treated patients.

**Acknowledgements**

This work was supported by grants from the Universidad de Buenos Aires and Consejo Nacional de Investigaciones Científicas y Técnicas.

**References**


---

**Table 1.** Amino acid residues outside HVR1 that differentiate the lineages from patient A and/or patient B

<table>
<thead>
<tr>
<th>Patient Lineage</th>
<th>Amino acid position*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>435 439 441 445 446 447 450 455 458 467 475</td>
</tr>
<tr>
<td>Ia</td>
<td>Q F A H K A G S Y</td>
</tr>
<tr>
<td>Iib</td>
<td>H L V N E E A</td>
</tr>
<tr>
<td>II</td>
<td>L S Q A D</td>
</tr>
<tr>
<td>III</td>
<td>E S</td>
</tr>
<tr>
<td>IV/V</td>
<td>H L T N S D A</td>
</tr>
<tr>
<td>B</td>
<td>a-a2 A F G R N R N E A A Y</td>
</tr>
<tr>
<td>b</td>
<td>L K D H</td>
</tr>
</tbody>
</table>

*Based on the HCV-1 sequence (Choo et al., 1991).
glycoprotein of hepatitis C virus. Biochem Biophys Res Commun 175, 220–228.


