Contribution of insertions and deletions to the variability of hepatitis C virus populations

Manuela Torres-Puente,1 José M. Cuevas,1 Nuria Jiménez-Hernández,1 María A. Bracho,1 Immaculada García-Robles,1 Fernando Carnicer,2 Juan del Olmo,3 Enrique Ortega,4 Andrés Moya1 and Fernando González-Candelas1

1Instituto Cavanilles de Biodiversidad y Biología Evolutiva and Departamento de Genética, Universidad de Valencia, Spain
2Unidad de Hepatología, Hospital General de Alicante, Spain
3Servicio de Hepatología, Hospital Clínico de Valencia, Spain
4Unidad de Enfermedades Infecciosas, Hospital General de Valencia, Spain

Received 17 January 2007
Accepted 7 April 2007

Little is known about the potential effects of insertions and deletions (indels) on the evolutionary dynamics of hepatitis C virus (HCV). In fact, the consequences of indels on antiviral treatment response are a field of investigation completely unexplored. Here, an extensive sequencing project was undertaken by cloning and sequencing serum samples from 25 patients infected with HCV subtype 1a and 48 patients with subtype 1b. For 23 patients, samples obtained after treatment with alpha interferon plus ribavirin were also available. Two genome fragments containing the hypervariable regions in the envelope 2 glycoprotein and the PKR-BD domain in NS5A were sequenced, yielding almost 16 000 sequences. Our results show that insertions are quite rare, but they are often present in biologically relevant domains of the HCV genome. Moreover, their frequency distributions between different time samples reflect the quasispecies dynamics of HCV populations. Deletions seem to be subject to negative selection.

The hepatitis C virus (HCV) establishes persistent infection in more than two-thirds of infected individuals. The prevalence of chronic hepatitis ranges between 1 and 2 % in industrialized countries (Pawlotsky, 2006), affecting more than 170 million people worldwide (Wasley & Alter, 2000; Shepard et al., 2005). In approximately 20 % of infected people, the disease progresses to cirrhosis, liver failure or hepatocellular carcinoma (Afdhal, 2004). Antiviral therapies employing alpha interferon (IFN-α) were started before even knowing the infectious agent causing hepatitis C (Hoofnagle et al., 1986). Later on, the nucleotide analogue ribavirin was added in combination with IFN-α (Brillanti et al., 1994). More recently, the results of combination therapy were improved by pegylation of IFN-α molecules, which enhanced their efficacy (Manns et al., 2001; Fried et al., 2002). However, antiviral therapies involve a significant toxicity and are effective in only about 50 % of cases (Heathcote & Main, 2005; Dienstag & McHutchison, 2006).

HCV is a positive-strand RNA virus in the genus Hepacivirus that belongs to the family Flaviviridae. The HCV genome consists of approximately 9400 nt in length and encodes a polyprotein of about 3000 aa. This polyprotein is cleaved by both host and viral proteases to generate three structural (core, E1 and E2) and seven non-structural (p7, NS2–NS5B) proteins (reviewed in Lindenbach & Rice, 2005). The most remarkable feature of HCV is its considerable genetic heterogeneity, typical of RNA viruses, leading to populations composed of a spectrum of mutants usually termed quasispecies (Eigen & Biebricher, 1988).

In this study, we have sequenced two regions of the HCV genome. The E1–E2 region consisted of 169 nt from the envelope 1 (E1) glycoprotein C terminus and 303 nt from the envelope 2 (E2) glycoprotein N terminus (from nt 1322 to 1793 in the reference HCV genome sequence, GenBank accession no. AF009606, Kuiken et al., 2006). This region included hypervariable regions HVR1 (Weiner et al., 1991), HVR2 (Kato et al., 1992) and HVR3 (Torres-Puente, 2004; Troesch et al., 2006). The second region corresponded to 743 nt (6742–7484) from the NS5A protein coding portion of the HCV genome and included the PKR-BD and V3 domains. These two genome fragments were chosen because of their biological relevance of the regions included therein. On the one hand, three hypervariable regions are included in the E1–E2 region analysed: HVR1, which seems to be involved in target cell recognition and virus
attachment (Penin et al., 2001); HVR2, which could be involved in cell surface receptor binding (Yagnik et al., 2000); and HVR3, which could play a role in the process of binding with host cell receptors and virus entry into host cells (Troesch et al., 2006). On the other hand, two remarkable domains are included in the NS5A region: the V3 domain, seemingly involved in responsiveness to interferon (Duverlie et al., 1998; Durante Mangoni et al., 2003), and PKR-BD, which contains the putative interferon sensitivity determining region (ISDR) and seems to be involved in blocking the cellular antiviral response induced by interferon (Gale et al., 1997, 1998).

In this study, we used a cohort of 73 patients infected with genotype 1 HCV, 25 of them infected with subtype 1a and 48 with subtype 1b (see Supplementary Tables S1 and S2 available in JGV Online). Samples were provided by three hospitals from the Comunidad Valenciana, Spain. A sample from each patient was taken before a combined therapy of IFN-α plus ribavirin was started (T0 sample). Additionally, for 23 non-responder patients, we also obtained a sample after 6 (T1 sample) or 12 months of treatment (T2 sample), or both (Supplementary Tables S1 and S2). There were 41 patients that did not respond to treatment and 27 patients that did respond to treatment but this information was not available for five patients. Experimental procedures (RNA extraction, amplification, cloning and sequencing of viral populations) are detailed in Jimenez-Hernandez et al. (2007). Briefly, after viral RNA extraction, reverse transcription reactions were performed with random hexadecamers in order to prevent any bias during reactions because of unspecific oligonucleotides. Primers employed for subsequent PCR were previously reported (see Table 1 in Bracho et al., 2004). Amplified DNA products for each region were purified with High Pure PCR product Purification kit (Roche) and directly cloned into EcoRV-digested pBluescript II SK (+) phagemid (Stratagene). Plasmid DNA was purified with High Pure Plasmid Isolation kit (Roche). Cloned products for the E1–E2 region or the NS5A region were sequenced using vector-based primers KS and SK (Stratagene). Sequencing was carried out using ABI Prism BigDye Terminator v3.0 Ready Reaction Cycle Sequencing kit (Applied Biosystems) on an ABI 3700 automated sequencer. Sequences were verified and both strands assembled using the Staden package (Staden et al., 2000). Sequence alignments were obtained using CLUSTAL_X v1.81 (Thompson et al., 1997). For the E1–E2 region, we obtained about 100 clones from each patient, yielding a total of 9944 sequences. For the NS5A region, we obtained from 25 to 96 clones per patient and 5870 sequences in total were determined (Supplementary Tables S1 and S2).

A summary of the results is shown in Table 1, where the distribution of insertions and deletions (indels) for responder and non-responder patients is given separately for each viral subtype. In general terms, indels were not detected in most of the patients (61 of 73 patients) and were mainly found in non-responder patients, and more specifically in those with subtype 1b. In particular, eight patients showed insertions in the E1–E2 region, all of them located at two recurrent points (Fig. 1). Additional information for each patient, such as the position of the insertion detected, viral subtype, treatment response, available time samples, number of sequences obtained, amino acids present in each insertion, number of sequences containing the insertion, the relative frequency of the different types of insertion and total frequency is provided in Supplementary Table S3 (available in JGV Online). On the one hand, six patients showed insertions between positions 57 and 58, with an insertion size ranging between 1 and 4 aa (Fig. 1 and Supplementary Table S3). On the other hand, two patients showed insertions between positions 152 and 153, with an insertion size of 2 or 3 aa (Fig. 1 and Supplementary Table S3). In this region, position 57 coincides with the first position of HVR1, whereas positions 152–153 are within HVR2, which is located between positions 147 and 155 of the analysed fragment.

When more than one time point sample was available for a patient with an insertion detected in the T0 sample (patients C20, G06, G16, G22 and G28), sequences containing the insertions were present in all samples. For three of these patients (patients G16, G22 and G28), insertions were present in all sequences from all time point samples. In two cases, the frequency of the insertions varied between T0 and T1 samples. Whereas, patient C29 showed a remarkable increase from 9.1 (T0 sample) to 100 % (T1 sample), patient G06 presented the opposite tendency, decreasing the frequency of the insertion from 25 (T0

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Responders</th>
<th>Non-responders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Insertion</td>
<td>Deletion</td>
</tr>
<tr>
<td>1a</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1b</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
sample) to 9% (T1 sample). Therefore, in the first patient the insertion got fixed on the viral population, but in the second patient the insertion seemed to tend to extinction. Finally, for patients C11 and C35, where only T0 sample was available, the frequency of insertions was 9.1 and 30%, respectively.

Only one of the eight patients showing insertions was infected with subtype 1a of HCV. However, no significant differences were detected in the distribution of HCV subtypes between patients with or without insertions in the E1–E2 region (Fisher’s exact test, \(P=0.250\)). Therefore, the apparent excess of patients infected with subtype 1b HCV showing insertions could simply reflect the differential distribution of subtypes in the sample studied.

With regards to the response to antiviral treatment, seven of the eight patients showing insertions were non-responders. Again, no significant differences in the distribution of response or non-response to treatment among patients with or without insertions were observed (Fisher’s exact test, \(P=0.136\)). It is noticeable that all patients infected with subtype 1b HCV were non-responders, whereas the only patient infected with subtype 1a was a responder to the treatment. It is also remarkable that both points of insertion were present in more than one patient, although apparently there is not a common pattern with respect to the nature of the insertion (Supplementary Table S3). It is important to note that these infected patients were epidemiologically unrelated, because previous phylogenetic analyses showed that sequences from each patient grouped in independent clades (data not shown). Therefore, this suggests that a key factor in the dynamics of these

---

**Table 2. Summary of insertions (NS5A region) and deletions (E1–E2 and NS5A regions) detected in the HCV genome**

For each patient the position of the insertion/deletion, HCV subtype, treatment response and available time samples, are given. Abbreviations: Seqs, number of analysed sequences for each patient; aas_Ins, amino acids present in the insertion; Seqs_Ins/Del, number of sequences containing a particular insertion/deletion in a given patient; %Tot, percentage of sequences with insertion/deletion in a given patient; Resp, responder patient to therapy; NResp, non-responder patient to therapy.

**(a) Patients with insertions in the NS5A region**

<table>
<thead>
<tr>
<th>Patient</th>
<th>Insertion position</th>
<th>Subtype</th>
<th>Treatment response</th>
<th>Time sample</th>
<th>Seqs</th>
<th>aas_Ins</th>
<th>Seqs_Ins</th>
<th>%Tot</th>
</tr>
</thead>
<tbody>
<tr>
<td>A06</td>
<td>6–7</td>
<td>1b</td>
<td>NResp</td>
<td>T0</td>
<td>96</td>
<td>Y</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td>C13</td>
<td>94–95</td>
<td>1b</td>
<td>NResp</td>
<td>T0</td>
<td>96</td>
<td>W</td>
<td>2</td>
<td>2.1</td>
</tr>
<tr>
<td>C33</td>
<td>81–82</td>
<td>1a</td>
<td>Resp</td>
<td>T0</td>
<td>74</td>
<td>L</td>
<td>74</td>
<td>100</td>
</tr>
</tbody>
</table>

**(b) Deletions detected at the E1–E2 and NS5A regions**

<table>
<thead>
<tr>
<th>Patient</th>
<th>Region</th>
<th>Deletion position</th>
<th>Deletion size (aa)</th>
<th>Subtype</th>
<th>Treatment response</th>
<th>Time sample</th>
<th>Seqs</th>
<th>Seqs_Del</th>
<th>%Tot</th>
</tr>
</thead>
<tbody>
<tr>
<td>C29</td>
<td>E1–E2</td>
<td>8</td>
<td>1</td>
<td>1b</td>
<td>NResp</td>
<td>T0</td>
<td>100</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A21</td>
<td>NS5A</td>
<td>222</td>
<td>1</td>
<td>1b</td>
<td>NResp</td>
<td>T1</td>
<td>64</td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>A34</td>
<td>NS5A</td>
<td>236–239</td>
<td>4</td>
<td>1a</td>
<td>NResp</td>
<td>T0</td>
<td>29</td>
<td>1</td>
<td>3.5</td>
</tr>
<tr>
<td>G22</td>
<td>NS5A</td>
<td>98</td>
<td>1</td>
<td>1b</td>
<td>NResp</td>
<td>T0</td>
<td>36</td>
<td>1</td>
<td>2.8</td>
</tr>
</tbody>
</table>
mutations is not the type of insertion, but more probably its location.

Three patients showed insertions in the NS5A region, but no recurrence of the insertion points as observed in the E1–E2 region (Table 2a). Patient A06 showed two insertions of 1 aa between positions 6 and 7, and 94 and 95. Patient C33 also showed an insertion in the latter location, but in this case consisting of 3 aa. Finally, patient C13 presented an insertion of 1 aa between positions 81 and 82. Insertion points 81–82 and 94–95 are within the ISDR, which expands from amino acid position 71 to 110 in the analysed fragment. For these patients, only T0 samples were available. The frequency of insertions located at positions 6–7 and 81–82 was 100 %, whereas the insertions at position 94–95 showed a frequency of 39.1 % for patient C13 and of 2.1 % for patient A06.

The number of deletions detected in this study was lower than the number of insertions (Table 1). For the E1–E2 region, patient C29 showed the only deletion identified, which consisted of an amino acid at position 8 detected in its T0 sample (Table 2b). For the NS5A region, three patients presented deletions. In patient A21, a sequence showed an amino acid deleted at position 222 for the T1 sample. Patient A34 presented a sequence with 4 aa deleted at position 236 for the T0 sample. Finally, patient G22 showed a sequence with 1 aa deleted at position 98 for the T0 sample (Table 2b).

Several studies have shown that the quasispecies nature of HCV enables it to evade the immune response (Mänzini et al., 1998; Ray et al., 1999). In this sense, the regions under the strongest immune system pressures show the highest degree of genetic variability (Reed & Rice, 2000), as is the case for the HVR1 region. For this region, it has been shown that non-responder patients present an increase in genetic variability (Farci et al., 2000) and that the evolution of HVR1 in the early phases of antiviral treatment can be an indicator of its therapeutic outcome (Farci et al., 2002). Moreover, several studies have observed a correlation between genetic variability of HVR1 and a degree of resistance to antiviral treatments (Polyak et al., 1997) or severity of hepatic injury (Curran et al., 2002; Farci et al., 2006). Very recently, several studies have sought to correlate the outcome of antiviral treatments with the degree of genetic variability in different regions of HCV genome, such as the 5’UTR (Zekri et al., 2007), the p7 membrane protein (Castelain et al., 2007), the NS3 protein (Vallet et al., 2007) or the core region (Akuta et al., 2007).

Because of their low frequency compared with point mutations, only a few studies have described indels in HCV. Therefore, little is known about the consequences of these mutations on viral infective ability or response to antiviral treatment. Different studies have occasionally identified indels in the 5’UTR (Zhang et al., 1999; Revie et al., 2006). Interestingly, indels have been described between positions 384 and 385 for HVR1 (Kato et al., 1992; Abe et al., 1992; Hohne et al., 1994), which corresponds to position 57–58 in our analysed genome fragment, where six of eight insertions in the E1–E2 region were detected. Although insertions seem to appear repetitively in this region, HVR1 size is invariable for all genotypes, except for genotype 6, which suggests the existence of selective restrictions maintaining an adequate size for this region (Chamberlain et al., 1997).

A few studies have analysed the implications of size variability in HVR1. A statistically significant relationship between the presence of indels at position 384–385 and the development of mixed cryoglobulinemia type 2, a lymphoproliferative disorder associated with chronic HCV infection, has been described (Gerotto et al., 2001). According to a structural model predicted for the E2 protein (Yagnik et al., 2000), HVR1, apart from being a highly exposed antigenic site, is located very close to a CD81-binding region, and physical interactions could be established between them. The same happens with HVR2, which is not only a very exposed region but it is also located very close to a region involved in CD81 binding and/or dimerization (Yagnik et al., 2000). In our study, two of the eight insertions detected in the E1–E2 region were located in HVR2. Therefore, changes in HVR1 and HVR2 size could affect their antigenic properties and ability to bind to cellular receptors.

Indels have also been described for the NS5A protein, although at a lower frequency; such is the case of an 8 aa insertion in the ISDR (Enomoto et al., 1996). Size modifications in this protein seem to be genotype specific (Maertens, 1997). Similarly, the analysis of the coding region for 37 HCV whole genomes revealed 12 points with insertions or deletions, many of them strongly associated to a viral subtype (Casino et al., 1999). Therefore, these data suggest a certain influence of indel events at different levels of the HCV evolutionary dynamics, such as differentiation processes between viral subtypes or antiviral treatment response. Again, two of three insertions found in the NS5A region were located within a domain probably involved in immune response evasion, the ISDR, and this suggests that insertions could play a relevant role in this mechanism. In this sense, a recent study has shown that indel analyses permit the identification of regions in the NS5A protein that are not essential for viral genome replication, such as the ISDR or the C-terminal region (Liu et al., 2006).

Some deletions were also found in the analysed patients (Table 2b). Four patients showed deletions, one in the E1–E2 region and three in the NS5A region. For all patients, the deletion was present in a single sequence. This paucity of sequences with deletions could be due to selective pressures against these events because, for compact genomes such as those of RNA viruses, deletions are very likely to be deletious and to be removed by selection. In contrast, our results show that the frequency of insertions in viral populations can be very high (Supplementary Table S3 and Table 2a), thus suggesting that these kind of mutations promote almost neutral or even beneficial effects on viral
fitness. At this point, it is important to note that indels are very unlikely to revert and consequently their frequency in viral quasispecies would essentially depend on their effects on fitness.

In summary, our results indicate that the frequency distributions of insertions in the E1–E2 and NS5A regions between different time samples also reflect the naturally occurring quasispecies dynamics of HCV populations. In spite of the relatively low number of cases detected, our results also suggest, for subtype 1b patients, a potential association between the presence of insertions and lack of response to treatment (Table 1), although further studies are necessary to verify this hypothesis.

Acknowledgements

This work was supported by the Conselleria de Sanitat i Consum, Generalitat Valenciana (Spain), and project BFU2005-00503 from Ministerio de Educación y Ciencia (Spain).

References


