Hepatitis C virus population analysis of a single-source nosocomial outbreak reveals an inverse correlation between viral load and quasispecies complexity

Antonio Mas,1 Encarna Ulloa,1 Miguel Bruguera,2 Ivana Furčić,2† Damia Garriga,1‡ Silvia Fábregas,3 David Andreu,1 Juan Carlos Saiz4 and Juana Díez1

Correspondence
Juana Díez
juanadiez@upf.edu

1Departamento de Ciencias Experimentales y de la Salud, Universitat Pompeu Fabra, Dr. Aiguader 80, 08003 Barcelona, Spain
2Hepatología, Instituto de Enfermedades Digestivas, Hospital Clinic, 08036 Barcelona, Spain
3Hospital Figueres, 17600 Girona, Spain
4Departamento de Biotecnología, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28040 Madrid, Spain

Received 3 August 2004
Accepted 25 August 2004

Introduction

Hepatitis C virus (HCV) is an enveloped positive-strand RNA virus of the family Flaviviridae. Only about 15–30% of HCV infections are spontaneously cleared, the remaining result in virus persistence with subsequent development of chronic hepatitis, liver cirrhosis and hepatocellular carcinoma (Alter et al., 1999). Because worldwide over 170 million people are infected carriers, HCV is a major health problem and a key issue in antiviral research.

The mechanisms responsible for the high rate of viral persistence are thought to be the result of a complex host–virus interaction early after infection (Racanelli & Rehermann, 2003). However, little is known about these early virus and host determinants because the acute phase of infection is often asymptomatic and thus most diagnoses are made during the chronic stage, i.e. months or years after the events that determined the clinical course of the infection. While HCV induces strong humoral and cellular immune responses, their roles in virus clearance or persistence have not been fully elucidated. Studies in humans and chimpanzees indicate that a robust intrahepatic CD4+ and CD8+ T-cell response during the first weeks after infection was reflected in the structure of the quasispecies. A higher viral load correlated with the presence of a predominant HCV genome and a corresponding lower quasispecies complexity. The quasispecies complexity itself was not correlated with HCV clearance or persistence. Thus, the relationship between an intrapatient HCV quasispecies and the clinical outcome of an HCV infection is more complex than previously anticipated.
aspect is that HCV seems to have a wide cell tropism and can infect not only hepatocytes but also cells of the immune system (Bain et al., 2001; Sung et al., 2003). Another important feature is that HCV behaves in infected patients as a complex mixture of genetically distinct but closely related variants, termed quasispecies, which results in a high genetic variability and adaptability (Martell et al., 1992; Meller et al., 1995; Simmonds, 1995). The quasispecies distribution of RNA viruses may influence the transmission, the pathogenesis and the outcome of viral infections (Domingo et al., 2001; Pawlotsky, 2003). New variants are continuously generated during virus replication as a result of errors made by the viral RNA-dependent polymerase, which lacks proof-reading activity. This genetic variability could give the virus an advantage in adapting to a changing host environment including availability of permissive cells, the presence of innate and adaptive immune responses and antiviral treatment.

The quasispecies nature of HCV may help to establish viral persistence by escaping the host immune surveillance (Chang et al., 1997; Christie et al., 1999; Kao et al., 1995; Kato et al., 1994; Shimizu et al., 1994; Tsai et al., 1998; van Doorn et al., 1995; Weiner et al., 1992, 1995). However, due to the difficulty in obtaining samples at early time points after infection, very limited information is available on HCV quasispecies at that stage (Farcì et al., 2000; Manzin et al., 1998, 2000; Ray et al., 1999). Moreover, the above studies were conducted in patients usually infected with large quantities of virus from potentially heterogeneous sources, even belonging to different genotypes.

In the present study, we had the unique opportunity of analysing HCV quasispecies features as early as 11 weeks post-infection in samples from acutely infected patients that presented different disease outcomes (Bruguerà et al., 2002). All patients were infected from a single common donor during a nosocomial episode, allowing us to assess the evolution of a single viral strain in patients in whom the infection either became persistent or was spontaneously resolved. The study was done by extensive cloning and sequencing of an HCV E1/E2 segment that included the HVR1 region.

**METHODS**

**Patients.** A nosocomial epidemic outbreak of genotype 1b HCV infections at the Hospital Figueres (Spain) has been recently described. The incident was due to an accidental contamination of a multi-dose vial of heparin solution with blood from an unrecognised HCV carrier. The outbreak involved 10 patients from whom the demographical, clinical and biochemical features have been described elsewhere (Bruguerà et al., 2002). Early serum samples obtained within weeks 11–17 after exposure to HCV were available for the present study in 7 of 10 patients. In patients S7 and S9, the infection resolved spontaneously, while in the remaining five patients S1, S3, S8, S10 and S11, HCV viraemia persisted. One of the five chronically infected patients (S3) cleared HCV-RNA after treatment with interferon and ribavirin. The other four patients were not treated due to old age or contraindications for antiviral therapy and remained chronically infected. Serum samples for quasispecies analysis obtained within week 36 and 39 after HCV exposure were also available in two (S1 and S8) of the four patients that remained viraemic.

**PCR amplification.** HCV-RNA was extracted from serum by the Trizol method following the instructions of the manufacturer (Invitrogen). Specific amplification of HCV-RNA was carried out with primers described previously (Gimenez-Barcons et al., 2001) as follows: cDNA was synthesized from viral RNA using 1·5 μM reverse primer HV2 and 0·45 U AMV reverse transcriptase μl⁻¹ in 20 μl final volume. Ten microlitres of this cDNA was used as a template for PCR amplification using the forward primer HV1 and the reverse primer HV2 at 0·5 μM each, in a final volume of 50 μl. One tenth of the external PCR product was used in a second PCR amplification using the internal primers HV3 and HV4. The resulting fragment of 388 nt comprised an E1/E2 segment (position 1443–1784 in genotype 1b HCV consensus sequence; Los Alamos National Laboratory HCV database) that includes the hypervariable region 1 (HVR1). PCR conditions were as follows: first 5 min denaturation step at 95°C, then, 35 cycles consisting of denaturation for 30 s at 95°C, annealing for 30 s at 45°C and extension for 1 min at 72°C, followed by a final extension for 10 min at 72°C. All reactions were carried out using the Expand High Fidelity PCR system (Roche) with 1·8 mM MgCl₂ final concentration.

**HCV-quasispecies analysis.** At least two different PCR amplification reactions were done from each cDNA and combined to diminish molecular bottleneck during clonal analysis. These PCR products were purified from agarose gels and incubated for 30 min at 72°C with Taq polymerase (EcoTaq; EcoGen) to introduce single over-hanging A’s at the 3’ ends. The product of this reaction was cloned with the pGEM-T vector system (Promega) following manufacturer’s instructions. A mean of 16 recombinant clones from each sample (range from 10 to 24) were sequenced using standard dideoxy sequencing with fluorescence labelled nucleotides (Perkin-Elmer Applied Biosystems). Although the amount of RNA used for the amplification reactions could not be quantified before use, the observed higher genetic complexities in the serum from patients with low viral load strongly argues against a molecular bottleneck effect in sample preparation. Sequences were aligned using CLUSTAL W (Thompson et al., 1994). The features of intrapatient HCV populations were defined and calculated as follows. The genetic distance is defined as the number of nucleotide differences between two sequences and was calculated by Kimura two-parameters (Kimura, 1980). Distances between amino acid sequences are given as Hamming distances. The mean of genetic distances and Hamming distances is the mean of the values taken for all sequence pairs derived from a single sample. Both, the genetic distance and the number of synonymous and non-synonymous changes per synonymous and non-synonymous sites were calculated using the MEGA 2 program (Kumar et al., 2001). The mutation frequency is the number of different mutations found relative to the number of nucleotides sequenced; it is calculated by dividing the number of different mutations found in a set of genomes (compared to the consensus nucleotide sequence of the same set) by the total number of nucleotides sequenced. The Shannon entropy (Sₐ) is a measure of the proportion of identical sequences in a mutant distribution. The possible values of Sₐ range from zero (when all genomes are identical) to one (when all genomes differ from one another). Sₐ was calculated following the formula: Sₐ = Σ[i × ln pᵢ]/ln n, in which pᵢ is the frequency of each sequence in the mutant spectrum and n is the total number of sequences compared (Volkenstein, 1994). The statistical significance of comparisons among patient samples was analysed with a Student’s t-test. P values lower than 0·05 were considered significant.
**RESULTS**

**Clinical parameters of HCV-infected patients**

We have previously described a single-source nosocomial genotype 1b HCV outbreak (Bruguera et al., 2002). Serum samples were available from seven of 10 patients involved in the outbreak, and they were studied in detail. Five of seven patients, namely S1, S3, S8, S10 and S11 were persistently infected whereas the other two (S7 and S9) resolved the infection spontaneously. During the follow up of the infected patients it was observed that patient S10, who did not show evidence of ongoing HCV replication at the time of the previous report (Bruguera et al., 2002), presented an HCV load of 40 700 IU ml⁻¹ at 63 weeks post-infection. Consequently, this patient has been reclassified as chronically infected. In Table 1 the viral-load values and the levels of the liver enzymes alanine aminotransferase and aspartate aminotransferase (ALT/AST) are summarized for the two time points of serum sampling. Based on the viral load of the first sampling time, which was 11–17 weeks post-infection, the patients were divided into two groups, those with a viral load of \( \geq 76 \times 10^3 \) IU ml⁻¹ (patients S1, S3, S8 and S11) and those with a value of \( \leq 8 \times 10^3 \) IU ml⁻¹ (patients S7, S9 and S10). These two groups are named hereafter as high and low viral-load group, respectively. All patients included in the high viral-load group developed a chronic infection. Patient S3 resolved HCV infection after treatment with interferon plus ribavirin. Of the three patients of the low viral-load group, two resolved the HCV infection spontaneously, whereas in patient S10, the viral load increased from 5000 IU ml⁻¹ at week 17 to 40 700 IU ml⁻¹ at week 63.

**Homogeneity of viral isolates**

To verify the previously suggested epidemiological link amongst the patients HCV strains (Bruguera et al., 2002), we carried out RT-PCR amplification of a 388 nt E1/E2 segment that includes the hypervariable region 1 (HVR1). Amino acid sequences derived from the consensus nucleotide sequences were aligned and compared with nine non-related HCV genotype 1b sequences (CP1 to CP9) from the same geographical area (Fig. 1). The results confirmed that all patients were infected by a single-source and that little variation was generated during the first weeks of the new infection process. Compared with the number of mutations present in the nine control sequences (CP1 to CP9), very few mutations were detected among sequences from this study, even in the HVR1 region. Mean inter-patient distances on nucleotide and amino acid levels among HCV isolates from this study were 1·2 versus 2·4 %, while they were 18·5 versus 21·9 % for the control CP1 to CP9 samples. These values reflect the expected differences between linked and unlinked transmission events (Wang et al., 1998).

From patients S1 and S8, serum samples of a second time point post-infection were available. The HCV-RNA was amplified accordingly, and the consensus sequences were aligned with the previously obtained sequences (Fig. 1). Only one mutation (M456V) was found in the consensus sequence of sample S8w39 compared to S8w15. The sequence from sample S1w37 was identical to that from S1w16 (Fig. 1). This would indicate that sampling artefacts can be excluded and that the accumulation of HCV mutations in these patients proceeded in a longer time frame than a few months.

**Table 1.** Clinical parameters of the HCV seropositive patients infected from a common source and followed in this study

<table>
<thead>
<tr>
<th>Group</th>
<th>Patient</th>
<th>ALT/AST (IU ml⁻¹)*</th>
<th>Viral load ( \times 10^3 ) IU ml⁻¹*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1st sample (weeks p.i.)</td>
<td>2nd sample (weeks p.i.)</td>
</tr>
<tr>
<td>High viral load</td>
<td>S1</td>
<td>162/323 (16)</td>
<td>53/63 (37)</td>
</tr>
<tr>
<td></td>
<td>S3‡</td>
<td>2180/2670 (11)</td>
<td>&lt;40 (152)</td>
</tr>
<tr>
<td></td>
<td>S8</td>
<td>371/552 (15)</td>
<td>565/530 (39)</td>
</tr>
<tr>
<td></td>
<td>S11</td>
<td>836/1182 (17)</td>
<td>56/89 (51)</td>
</tr>
<tr>
<td>Low viral load</td>
<td>S7‖</td>
<td>59/62 (15)</td>
<td>&lt;40 (52)</td>
</tr>
<tr>
<td></td>
<td>S9‖</td>
<td>28/114 (15)</td>
<td>&lt;40 (76)</td>
</tr>
<tr>
<td></td>
<td>S10</td>
<td>2169/2320 (17)</td>
<td>52/105 (63)</td>
</tr>
</tbody>
</table>

*Liver enzymes aspartate aminotransferase (AST) and alanine aminotransferase (ALT) and viral-load values were determined by routine automated procedures.
†p.i., post-infection.
‡Negative after treatment with interferon and ribavirin.
§BL, Data below limit of detection.
‖Patient that spontaneously resolved the HCV infection.

**Nucleotide sequence accession numbers.** The nucleotide sequences reported here have been submitted to GenBank under the accession numbers AY683898–AY684043.
Quasispecies analysis reveals a correlation of high HCV load with the presence of a predominant viral form

The quasispecies nature of HCV has been related to important aspects of virus evolution and pathogenesis. To get an insight into the actual complexity of the intrapatient HCV quasispecies from this nosocomial outbreak and to analyse possible correlates with the clinical outcome of the infection, viral populations were determined for all seven patients early after infection. The above-mentioned amplified E1/E2 HCV-RNA segment was cloned into a bacterial vector. A mean of 16 clones (range from 10 to 24) per sample were sequenced and analysed. Fig. 2 shows a schematic representation of the quasispecies composition for each sample. In the high viral-load group, the major form ranged from 54 to 66 % of all nucleotide sequences while in the low viral-load group, it ranged from 23 to 25 %. Parallel observations...
were made from the analysis of the corresponding predicted amino acid sequence. The major deduced amino acid form ranged from 62 to 89 % and from 29 to 41 % in the high and low viral-load groups, respectively. These differences were statistically significant with $P$-values < 0.0007 and < 0.0004 for nucleotide and amino acid values, respectively. In patients S7 and S9, two major forms with an approximately equal frequency co-existed in the quasispecies. Taken together, these results indicated that grouping the patients according to viral load was reflected in the structure of the quasispecies.

From the obtained sequences we calculated some fundamental features of the viral populations such as the mutation and amino acid substitution frequencies, the distance between sequences at the nucleotide and amino acid level, the synonymous versus non-synonymous substitution rates and the normalized $S_n$ (Table 2). The mean values with the respective standard deviations for the mutation frequency, $S_n$ and distances for the isolates of each group of patients are shown in Fig. 3(a). Despite the tendency for higher mutation frequencies and higher distance values in the low viral-load group, the statistical analyses showed differences with statistical significance only for nucleotide and deduced amino acid $S_n$ comparisons ($P<0.001$ and $P<0.01$, respectively).

The high variability of the HVR1 region among HCV isolates has been proposed to play a pivotal role in the appearance of escape variants and, thus, in the establishment of a persistent infection (Booth et al., 1998; Kao et al., 1995; Kato et al., 1993, 1994; van Doorn et al., 1995; Weiner et al., 1992). To assess whether features of this region would better correlate with properties of the HCV infections than the whole E1/E2 subregion, we analysed the HVR1 and the rest of the sequence separately. The results expressed as mean values ± standard deviations are shown in Fig. 3(b). Considering the nucleotide sequences of the HVR1 region, only the difference in $S_n$ for the high and low viral-load group of patients was statistically significant. However, when the predicted amino acid data were taken into account, the low viral-load group of patients showed statistically significant higher values in all three parameters ($P<0.05$). The rest of the analysed HCV region also showed a tendency of higher

### Table 2. HCV population parameters

The different parameters are described in Methods.

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample code*</th>
<th>No. clones</th>
<th>Nucleotide</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mutation frequency</td>
<td>$dS/dN*$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(% distance)§</td>
<td>$S_n$</td>
</tr>
<tr>
<td>High viral load</td>
<td>S1w16</td>
<td>14</td>
<td>1·25 × 10⁻³</td>
<td>0·34</td>
</tr>
<tr>
<td></td>
<td>S1w37</td>
<td>14</td>
<td>2·29 × 10⁻³</td>
<td>0·53</td>
</tr>
<tr>
<td></td>
<td>S3w11</td>
<td>18</td>
<td>2·27 × 10⁻³</td>
<td>0·32</td>
</tr>
<tr>
<td></td>
<td>S8w15</td>
<td>24</td>
<td>3·53 × 10⁻³</td>
<td>0·42</td>
</tr>
<tr>
<td></td>
<td>S8w39</td>
<td>10</td>
<td>2·34 × 10⁻³</td>
<td>0·65</td>
</tr>
<tr>
<td></td>
<td>S11w17</td>
<td>20</td>
<td>2·48 × 10⁻³</td>
<td>0·49</td>
</tr>
<tr>
<td>Low viral load</td>
<td>S7w15</td>
<td>17</td>
<td>3·79 × 10⁻³</td>
<td>0·79</td>
</tr>
<tr>
<td></td>
<td>S9w15</td>
<td>12</td>
<td>4·63 × 10⁻³</td>
<td>0·81</td>
</tr>
<tr>
<td></td>
<td>S10w17</td>
<td>17</td>
<td>7·05 × 10⁻³</td>
<td>0·83</td>
</tr>
</tbody>
</table>

*Sample code represent patient number (Sx) and week of sampling post-infection (wx).
†Ratio synonymous/non-synonymous substitutions.
‡Number of substitutions per nucleotide or amino acid.
§Genetic distance for nucleotide and amino acid data.
||Normalized Shannon entropy.
¶For sample S1w16, one clone that showed a stop codon at position 420 was not included in the calculations. For sample S8w15, two clones that showed one stop codon at position 420 and two stop codons at positions 412 and 420, respectively, were not included in the calculations.
The infection history of our patients is particularly inter-

complicated within the HVR1 region. Three of four patients of the

high viral-load group became chronic carriers, while one

was interviewed by a needle bearing HCV-positive blood and

second, the low HCV load in both patients that were the

possible transmitters (Bruguera et al., 2002). This kind of

low dose HCV transmissions are expected to be representa-

tive of most new infections as blood screening programmes

have drastically reduced the risks of the high dose trans-

missions via blood transfusions.

It is interesting to note that from all the 10 HCV-infected

patients, two had cleared the infection. Although these are

very low numbers, this frequency is what one would expect

from any HCV infection: a chronic carrier state develops in

about 70–85 % of cases whereas a spontaneous virus

clearance is observed in 15–30 %. This, together with the

observation that the infecting virus strain was very homo-

geneous, would suggest that host factors contribute more to

the infection outcome than the particular virus variant.

Indeed this hypothesis fits well to studies in which chim-

panzees infected with the same clonal HCV sequence pre-

sented different disease outcomes (Major et al., 2004).

However to substantiate this point, studies with larger

cohorts and different HCV genotypes are warranted.

Previous reports have suggested a correlation between HCV

infection and either a higher quasispecies complexity or a

sequential increasing complexity early after infection

(Farci et al., 2000; Ray et al., 1999). These studies were

conducted with injecting drug users (Ray et al., 1999) or

post-transfused patients (Farci & Purcell, 2000) infected

with unrelated HCV strains, which even belonged to differ-

ent genotypes. Such a correlation was explained by a more

effective immune response in patients who resolved the

infection. In such a patient group, the immune pressure

was thought to progressively clear virus variants from the

quasispecies thus resulting in the decline of complexity

(Farci & Purcell, 2000). In contrast, in our present study, we

observed that the quasispecies with a lower complexity

were always related to persistent infection (patients S1, S3,

S8 and S11). In addition, the quasispecies with the higher

values for the mutation frequency, the distances and S\textsubscript{n} in

the low viral-load group, however only differences in S\textsubscript{n} for

the nucleotide sequences were statistically significant

(P<0.05). Thus, the differences observed between isolates

from both groups of patients were mainly due to variation

within the HVR1 region.

For the high viral-load patients S1 and S8, we obtained sera

from a later time point and were able to amplify the respective HCV region. The comparable quasispecies analy-

sis revealed a slight increase in S\textsubscript{n} over time while the

mutation frequencies, distances and ratio of synonymous/

non-synonymous substitutions did not show consistency

(Fig. 2 and Table 2). Importantly, despite the ongoing HCV

replication in both patients for about 20 weeks, the quais-

species characteristics of the high viral-load group were

maintained.

**DISCUSSION**

In the present study, features of the HCV quasispecies were

analysed for seven patients infected with a common highly

homogeneous HCV inoculum during a nosocomial out-

break episode. The major findings are the correlation of a

high viral load at an early time of sampling (11–17 weeks

post-infection) with the presence of a predominant HCV

form in the quasispecies, and a corresponding lower com-

plexity within the HVR1 region. Three of four patients of the

high viral-load group became chronic carriers, while one

was treated and cleared the infection. In the low viral-load

group, one patient became chronic and the other two sponta-
nously resolved the infection.

The infection history of our patients is particularly inter-

esting because (i) all 10 patients involved in the outbreak

were infected on the same day from a single-seropositive

person, (ii) samples were available at an early time point

after infection, and (iii) the infection dose was lower than in

previously studied transmissions via blood transfusion. This

latter aspect was inferred by two observations, first the

infection source was a vial of heparin presumably con-
taminated by a needle bearing HCV-positive blood and

second, the low HCV load in both patients that were the

possible transmitters (Bruguera et al., 2002). This kind of

low dose HCV transmissions are expected to be representa-
tive of most new infections as blood screening programmes

have drastically reduced the risks of the high dose trans-

missions via blood transfusions.

![Graphical representation of means intrapatient HCV population features. (a) Mutation frequency, genetic distance and normalized S\textsubscript{n} of the E1/E2 HCV-RNA segment (position 1443–1784 in genotype 1b HCV consensus sequence, Los Alamos National Laboratory HCV database) for high viral load (black bars) and low viral load (white bars) sample groups. Given is the mean and standard deviation. Differences were tested for significance using a Student's t-test. ** Indicate P<0.001 and * indicate P<0.01. (b) As in (a), but consid-

ering two separate regions: the HVR1 and the rest of the sequence. ** Indicate P<0.005 and * indicate P<0.05.](image)
complexity related either to persistent infection (patient S10) or to spontaneous HCV clearance (patients S7 and S9) demonstrating that the complexity per se is not a reliable predictor for the outcome of an HCV infection. A possible explanation for the observed differences between our observations and previous reports could be because of different genetic heterogeneity, genotypes and size of the inoculum. In any case, future studies of early post-infection samples should help to clarify this matter.

How might the correlation of the existence of a dominant HCV form within the quasispecies with the high viral load be rationalized? Suppose that a clonal virus transmission event will start an infection. Then, variants will accumulate inevitably because of the high error rate of the virus replication machinery. Since the mutations will spread across the genome, the created quasispecies will mainly consist of minor forms. However, mutants that might overcome intrapatients selective constraints will gain selective advantage and expand relative to the other variants. These mutants will subsequently become the major forms in the virus population. If this simplistic view is correct, then a successfully selected mutant will directly lead to a high viral load and a more homogeneous quasispecies. Indeed, this is what has been suggested in quasispecies models (Eigen & Biebricher, 1988). Furthermore, overcoming the initial intrapatients barrier will increase the chance to develop a chronic infection. In fact, this is the case in our limited cohort.

In conclusion, while this study on HCV quasispecies complexity did not result in a correlate with the clinical outcome of the infection, it revealed an interesting feature of basic HCV evolution. HCV populations in high viral-load patients seemed to harbour a major viral variant possibly as a result of overcoming intrapatients selective constraints. It would be interesting to understand the underlying process, be it immune-mediated selection or host-cell adaptation. The recently developed method to mimic the natural infection by HCV pseudotypes should help to shed light on this issue in the future (Hsu et al., 2003; Matsuura et al., 2001).

ACKNOWLEDGEMENTS

This work has been supported by a grant from the Spanish Ministerio de Ciencia y Tecnologia (BMC 2001-0834). I. F. was supported in part by the Asociacion Espanol de Cooperacion Internacional (AECI).

REFERENCES


