Longitudinal Evaluation of the Structure of Replicating and Circulating Hepatitis C Virus Quasispecies in Nonprogressive Chronic Hepatitis C Patients

BEATRIZ CABOT, MARÍA MARTELL, JUAN I. ESTEBAN,* MARIA PIRON, TERESA OTERO, RAFAEL ESTEBAN, JAIME GUARDIA, AND JORDI GÓMEZ

Liver Unit, Department of Internal Medicine, Hospital General Universitari Vall d’Hebron, Universitat Autònoma de Barcelona, 08035 Barcelona, Spain

Received 23 May 2001/Accepted 5 September 2001

Hepatitis C virus (HCV) is one of the leading causes of chronic liver disease worldwide (22). The quasispecies nature of the single-stranded RNA genome of HCV is thought to play a central role in maintaining and modulating viral replication (10, 29). In general, the natural history of HCV infection does not follow a defined pattern, but the rates of progression of the disease (from minimal changes of the liver to cirrhosis and hepatocarcinoma) vary greatly in different individuals (22). The cytopathic potential of the virus and the characteristics of the host’s immune response have both been postulated to explain the observed differences in disease progression (6, 17, 18, 45).

In the absence of an appropriate animal model or culture system and in the context of the extreme heterogeneity of HCV, it has not been possible to associate particular sequences with distinct cytopathic potential. Similarly, attempts to correlate the process of quasispecies diversification with disease progression have yielded controversial results. While in cross-sectional studies some authors found a correlation between quasispecies complexity and extent of liver damage (20, 20, 21, 25, 48), others did not (19, 27, 33, 41). In longitudinal studies, the nucleotide complexity of the hypervariable domain of the HCV E2 region (HVR1) did not increase cumulatively over time but fluctuated in consecutive serum samples (2). Similarly, at more conserved regions of the genome, an oscillatory pattern of quasispecies complexity over time has been observed in long-term studies with frequent serum sampling (7).

The factors that drive these fluctuations and their relative contributions are not defined, and, to further complicate the interpretation of the biological meaning of circulating quasispecies complexity, it has been proven that within an infected patient the composition of the circulating viral population does not necessarily reflect that of the hepatic population (4, 15, 28, 34, 38).

Recently, we observed a significant correlation between quasispecies complexity at the amino acid level, in both serum and liver quasispecies, and the extent of liver fibrosis, albeit only in those patients with a similar levels of complexity in both compartments (5). It is currently unknown whether the ratio of quasispecies complexity in the liver to that in serum (liver/serum complexity ratio) is a stable parameter, so that patients may be characterized by their ratios, or if it fluctuates over time.

We have investigated the behavior of viral population parameters over time in consecutive liver biopsy and serum samples obtained from four patients with nonprogressive chronic hepatitis C.

MATERIALS AND METHODS

Viral isolates. HCV RNA was isolated from consecutive paired serum and liver biopsy samples obtained from four HCV-infected patients (A to D) at intervals of 3 to 6 years (Table 1). In one of the patients (patient A), four additional serum samples obtained during the 3 years between the biopsies were also studied. All patients had persistently normal alanine transaminase (ALT) levels and histologically nonprogressive liver disease (Table 1). All patients were
infected with HCV genotype 1b, were negative for other hepatitis viruses and human immunodeficiency virus, and had not been treated with antiviral therapy. Informed consent was obtained before they underwent liver biopsies. In all cases blood was drawn in Vacutainer tubes and centrifuged within 2 h and the serum was stored at -80°C. Three-millimeter-long fragments of liver biopsy samples were frozen in liquid nitrogen.

**RNA extraction, reverse transcription-PCR, cloning, and sequencing.** Viral RNA was extracted from both serum (0.05-g) and liver (0.05-g) samples using QIAamp viral RNA binding columns (Qiagen). Isolated HCV RNA was reverse transcribed into cDNA with a genotype 1b-specific primer from the E2(p7)-NS2 region (MJJ3, 5'-CTGAGCGCTTGAGGCGGAG-3', at positions 2925 to 2942). Nested PCR was performed with specific oligonucleotides to amplify a 212-bp fragment (outer set, MJJ3 and MJJ4 [5'-TCTGAGCGCTTGAGGCGGAG-3', at positions 2534 to 2551]; inner set, MJJS [5'-CTAGAATTCACCAAAAATATTGTATGACCA-3', at positions 2870 to 2888] and MJJS [5'-ACAGGATCCACCTGGTTGTTCTC-3', at positions 2639 to 2677]). Amplified products were purified using the QIAquick PCR purification kit (Qiagen) and cloned in Escherichia coli DH5α. Individual clones were sequenced by the dideoxy chain terminator method using the DNA dRhodamine sequencing kit (Applied Biosystems) and the ABI Prism 310 genetic analyzer (Applied Biosystems).

**Viral RNA quantitation at E2(p7)-NS2.** To ensure that the number of template RNA molecules that enter the amplification reaction did not limit the level of resolution of the quasispecies analysis, we quantified the HCV RNA in the E2(p7)-NS2 region (the locus used for the quasispecies analysis) in at least one liver-serum sample pair from three of the patients, as already described (30). Specific oligonucleotides and the fluorogenic probe for genotype 1b were used (sense primer: C-FT-2782, 5'-CTAGAATTCACCAAAAATATTGTATGACCA-3', at positions 2639 to 2888; antisense primer: CR-2799, 5'-CTCCTGCACGATGGA-3', at positions 2799 to 2785; probe: C-FT-2782, 5'-ATTACCCGTTGCATCCGACGGTGTTCTG-3', at positions 2762 to 2782). An HCV RNA standard used as reference for quantitation, representing the E2(p7)-NS2 region (at positions 2639 to 2888), was synthesized in vitro and purified by CF11 cellulose chromatography and polyacrylamide gel electrophoresis, in accordance with the protocol published by Martel et al. (30). Transcripts were quantitated by isotopic tracing (24).

**HCV quasispecies parameters.** Complexity of the quasispecies was estimated in all tissue and serum samples according to two parameters: mutation frequency (Mf) and Shannon entropy (S; frequencies of different sequences). Mf was calculated as the total number of mutations (nucleotide or amino acid) relative to the consensus sequence of each sample divided by the total number of nucleotides or amino acids sequenced; heterogeneity of the quasispecies increases as Mf increases (8). Shannon entropy has been defined in terms of the probabilities of the different sequences or clusters of sequences than can be present at a given time point (40, 46). This measure was calculated as $S = -\sum(p_i \ln p_i)$, where $p_i$ is the frequency of each sequence in the viral quasispecies. The resulting number was normalized as a function of the number of clones analyzed, thus allowing comparison of complexities among different isolates. The normalized entropy, $S_{np}$, was calculated as $S_{np} = S/N \times N$, where $N$ is the total number of sequences analyzed. $S_{np}$ varies from 0 (no diversity) to 1 (maximum diversity) (40). Intraclassic sample distances were calculated by the Kimura two-parameter modification method. To track the evolution of particular sequences over time, neighbor-joining phylogenetic trees were constructed with Phylip, version 3.572c (Phylogeny Inference Package [16]).

**Quantitation of nucleotide misincorporation with Taq Gold DNA polymerase.** To ensure that the observed heterogeneity was not due to nucleotide misincorporations introduced by the Taq Gold DNA polymerase (Applied Biosystems), one of the E2(p7)-NS2 clones of known sequence was amplified by PCR and subcloned. Among 47 such clones (totaling 9,964 nucleotides), only two nucleotide changes were detected in one of the clones. This low level of background noise coincides with our previous estimate of the misincorporation rate of Taq DNA polymerase (Perkin-Elmer) (one nucleotide change in 5,451 bases sequenced) (29) and confirms that most of the observed heterogeneity was independent of artifacts during the amplification procedure.


**RESULTS**

On average, 19 sequences (range: 7 to 24) for each sample were obtained (Table 2). We estimated the mutation frequency, normalized Shannon entropy, and genetic distance at both the nucleotide and amino acid levels in all serum and liver samples analyzed.

**Comparison of liver and serum quasispecies in all patients.** Table 2 summarizes the results obtained for patients A to D at the amino acid level. For patients A, B, and D amino acid consensus sequences in the four samples analyzed (L0, S0, L1, and S1) were identical. In patient C, three of the four samples analyzed (L0, S0, and L1) had the same amino acid consensus sequence but this sequence differed at two residues from the consensus sequence of sample S1. Nevertheless this difference was due to the existence of a double population in samples L1 and S1, with a different proportion of each population in the samples (the consensus residue at a given position is defined when the residue is present in 60% or more of the sequences [8]).

Phylogenetic analysis showed that, in all patients, amino acid sequences from the first liver-serum sample pair did not segregate from those of the second pair (see Fig. 2B; data not shown).

Regarding the quasispecies structure at the amino acid level, we observed that the liver-serum complexity ratios (Mf, Shannon entropy, and genetic distance) in the paired samples analyzed (L0-S0 and L1-S1) from patients B and D were similar (Table 2; patient B had the same level of complexity in the liver and serum in both paired samples, and patient D had more complex serum quasispecies in both paired samples). In contrast, in patients A and C we found differences in this ratio between the two paired samples. In patient A, whereas the serum population was, on average, twofold less complex than that in liver in the first pair of samples (L0-S0), in the second pair (L5-S5) the serum population was fourfold more complex than the liver population (Table 2 and Fig. 1B). In patient C, for the first pair of liver and serum samples (L0-S0) the quasispecies in serum was twofold more complex than that in liver while in the second pair of samples (L1-S1) the complexities of the quasispecies in the two compartments were similar (Table 2). No correlation between the degree of complexity at the amino acid level and the stage of liver disease in the first pair of samples (L0-S0) was found in patient C. However, in agreement with our previous results, the liver-serum complexity ra-
Phylogenetic analysis of sequences of serum and liver samples was performed to establish the relationship between samples during the infection (Fig. 2A) and specifically to monitor the persistence of minor sequences and the changes of master sequences. The majority of sequence groups in all serum samples had sequences identical or similar (0 to 3 single mutations) to those observed in the first liver biopsy sample (L0), even though sometimes at a very low frequency. Surprisingly, S5 and L5 major populations seemed to come from a minor group of the S2 sample, undetected in the first liver sample.

The master sequence in the first serum sample (S0) was completely replaced after 171 days by a complex population of sequences in S2; these were subsequently replaced by a mutant sequence (S3) in 210 days, which in turn evolved after 252 days into a mixture of two subpopulations (S5 and 15%), one of which contained sequences identical to the S0 master. Both populations were replaced by two other new ones in the last serum sample studied (S5): a minor subpopulation derived from sequences present in the original liver population and a master sequence which clustered with the major population found in the second liver specimen.

### Patient A: amino acid fluctuations in HCV quasispecies over 3 years

The signature mutation pattern and phylogenetic analysis of deduced amino acid sequences from serum and liver viral populations are presented in Fig. 1B and 2B, respectively. All but one of the amino acid consensus sequences (S5) were identical; after 3 years no amino acid substitution became fixed. Synonymous substitutions dominated overall variation both in serum and liver populations, except in serum samples.

### Table 2: Sequence complexity at the amino acid level of the E2(p7)-NS2 genomic region in the liver and serum quasispecies of four chronic hepatitis patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Date (day.month)</th>
<th>Sample (day.month)</th>
<th>Sample source</th>
<th>No. of clones sequenced</th>
<th>Quasispecies complexity</th>
<th>Genetic distance</th>
<th>% Synonymous mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>28.1.95</td>
<td>L0 Liver 28.1.95</td>
<td>Liver</td>
<td>24</td>
<td>1/98 0.6 0.03 ± 0.3</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.6.95</td>
<td>L0 Liver 5.6.95</td>
<td>Liver</td>
<td>7</td>
<td>1/56 0.6 0.004 ± 0.001</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.12.97</td>
<td>L0 Liver 4.12.97</td>
<td>Liver</td>
<td>8</td>
<td>1/490 0.4 0.004 ± 0.001</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.3.00</td>
<td>L0 Liver 9.3.00</td>
<td>Liver</td>
<td>23</td>
<td>1/544 0.11 0.004 ± 0.02</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25.10.95</td>
<td>L0 Liver 25.10.95</td>
<td>Liver</td>
<td>10</td>
<td>1/710 0.13 0.003 ± 0.007</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2.12.94</td>
<td>L0 Liver 2.12.94</td>
<td>Liver</td>
<td>10</td>
<td>1/140 0.7 0.01 ± 0.004</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.3.00</td>
<td>L0 Liver 9.3.00</td>
<td>Liver</td>
<td>10</td>
<td>1/350 0.1 0.006 ± 0.003</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>25.10.95</td>
<td>L0 Liver 25.10.95</td>
<td>Liver</td>
<td>10</td>
<td>1/168 0.2 0.02 ± 0.04</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.3.00</td>
<td>L0 Liver 3.3.00</td>
<td>Liver</td>
<td>21</td>
<td>1/186 0.61 0.03 ± 0.1</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.6.95</td>
<td>L0 Liver 5.6.95</td>
<td>Liver</td>
<td>21</td>
<td>1/249 0.49 0.03 ± 0.08</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.12.94</td>
<td>L0 Liver 2.12.94</td>
<td>Liver</td>
<td>10</td>
<td>1/140 0.7 0.01 ± 0.004</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.3.00</td>
<td>L0 Liver 9.3.00</td>
<td>Liver</td>
<td>10</td>
<td>1/350 0.1 0.006 ± 0.003</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>9.3.00</td>
<td>L0 Liver 9.3.00</td>
<td>Liver</td>
<td>10</td>
<td>1/168 0.2 0.02 ± 0.04</td>
<td>76</td>
<td></td>
</tr>
</tbody>
</table>

* Seventy amino acids were analyzed for each clone.
S0 and S3; thus amino acid sequence complexity of the viral populations was considerably narrower than the complexity of the nucleotide sequences (Table 2).

Viral RNA quantitation at the E2(p7)-NS2 region. In all samples investigated (see Materials and Methods), the number of E2(p7)-NS2 RNA molecules subject to amplification was at least 10^4, as confirmed by both RNA E2(p7)-NS2 quantitation and the positive result of amplification of the 1:10 dilution of the cDNA transcribed from the template with the same primer set used for quasispecies analysis in all serum and liver samples evaluated. This suggests that neither the input RNA molecules nor the efficiency of the reverse transcriptase to transcribe the RNA of this region led to underestimation of the quasispecies complexity.

DISCUSSION

For Darwin and most evolutionists, the individual organism is the object of selection (31). Nevertheless, according to Eigen's theory of quasispecies, the individual RNA replicon, as an entity for natural selection, is absent (12). Instead, selection acts on an ensemble of variant replicons: the quasispecies as a whole. In fact, what appears to be a revolutionary concept in theoretical biology (11), arising from Eigen's theory, was found throughout the world of RNA viruses as a way of organizing their genetic information, and HCV is a good example. The uncertainty of the genetic identity of the viral agent is maintained, both in space (serum and liver) and in time, throughout the infection. Our understanding of the viral life cycle relies on the knowledge of the behavior and relationships among components of the population of individuals, which, in turn, might help in understanding viral persistence and failure of treatment (1, 3, 13, 14, 39). In an attempt to define the continuity of the genetic constitution of the viral quasispecies in space and time, we analyzed series of sequential liver and serum samples in four patients with chronic hepatitis C.

In a previous work (4) we described two patients in whom the replicating HCV quasispecies was twice as complex as that circulating in the serum at the E2(p7)-NS2 region. We then investigated the liver-serum complexity ratio in a larger group of HCV-infected patients and identified a subset of patients (40%) with similar levels of complexity in both compartments, in whom the degree of complexity at the amino acid level correlated significantly with disease stage (5) and did not correlate with other clinical parameters (duration of infection, age, necroinflammation degree, etc.). In this study, we expanded original cross-sectional findings for four patients with nonprogressive hepatitis C by analyzing viral population parameters in repeat liver-serum sample pairs obtained 3 to 6 years apart.

The results of the present study confirm and extend our previous findings regarding differences between replicating and circulating viral population in all four patients (samples L0 and S0; Table 2) (4). As our results demonstrate, these differences were not due to limitations in the input RNA molecules, efficiency of the reverse transcriptase in the region analyzed, or misincorporation errors during the amplification procedure.

In previous cross-sectional studies dissimilarities between liver and serum quasispecies raised the question of the potential origin of the circulating viruses. We proposed that the higher complexity in the liver could be due to the existence of different functional compartments with distinct replication kinetics in the infected liver or by an excess contribution of sequences unable to become incorporated into mature virions and to be released to the circulation (4). For the opposite finding, that is, higher complexity in the circulating pool in addition to the minor contribution of variants replicating in extrahepatic sites (26, 34, 38), the possibility of differences in the clearance rates of major variants leading to an overrepresentation of the mutant repertoire has been suggested (5, 15).

The results of the present study provide additional information to clarify this issue. First, in all patients most circulating sequences found in the consecutive samples could be traced back to sequences present in the mutant spectrum of the initial liver specimen (L0) (Fig. 2A for patient A; data not shown for patients B to D). Second, the lack of complete nucleotide identity in some patients between the replicating and circulating major species (for instance, in the second liver-serum pair from patient A [L5-S5]) could not be easily explained given the different rates of turnover of virus in the two compartments and the greater stability of virions in the liver (half-life \( t_{1/2} \) = 2.7 h for free serum virions; \( t_{1/2} = 1.7 \) to 70 days for infected cells) (35). All this reinforces our original suggestions that there is either a differential effect of selective forces or hepatocytes with different kinetics of viral replication in the two compartments and that preferential release from one type of cell or another fluctuates over time (4). Also, the reemergence of minority components from the first replicating viral quasispecies in consecutive serum samples in patient A could be interpreted as a molecular record of a past quasispecies composition, as if a molecular memory was acting (43). Molecular memory may be glimpsed from reinoculation studies of chronically HCV-infected chimpanzees in which preexistent, slower-replicating quasispecies outlive faster-replicating variants present in the second inoculation (47). Hence, no extrahepatic origin must be invoked to explain discrepancies in quasispecies

---

**FIG. 1.** Signature mutation pattern: diversity of viral forms and shift in the HCV quasispecies in serum and liver samples over time. Nucleotide (A) and amino acid (B) sequences of the E2(p7)-NS2 junction quasispecies in serum (S0 to S5) and liver (L0 to L5) were grouped as clusters and represented as histograms. Groups were made according to identity at some points so that nucleotide sequences in any subgroup shared three or more specific mutations which were not present together in any other sequence subset in the same or in other samples; subsequent intergroup differences ranged from three to seven mutations and intragroup nucleotide differences were single point mutations (A). Similarly, amino acid sequences in any subgroup share one or more amino acid replacements, so intergroup amino acid differences consisted of one or two substitutions and intragroup amino acid differences were single point substitutions (B). Same colors and patterns identify identical subgroups. Same colors but different patterns identify related groups. Mutations in the consensus sequence of the studied viral quasispecies over time are shown at the bottom of each histogram (each symbol represents one mutation). Mf, Shannon entropy, and genetic distance for each sample are shown below the respective histogram.
FIG. 2. HCV phylogenetic reconstruction of long-term evolutionary relationship from a nonprogressive chronic hepatitis C patient. The phylogenetic analysis shown consists of unrooted neighbor-joining trees of nucleotide (A) and deduced amino acid (B) sequences from two liver (S0 and S5) and six serum (L0 to L5) samples. Numbers in brackets, numbers of identical sequences. Numbers indicate the bootstrap support (500 replicates); values of 50% or more are indicated.
composition, and most, if not all, sequences found in the replicating compartment at our level of resolution (44) are viable and able to gain access to the circulation (Fig. 1A and 2A).

Also, the observed replacement of the replicating pool by a homogeneous population of sequences in the last biopsy sample of patient A is more apparent than real, since the minor population of sequences in the coincident circulating pool was more related to sequences in the first biopsy sample, implying that at higher resolution the parental sequence should have been detected.

It is well known that the evolutionary potential of the HCV genome varies markedly at different loci depending on structural and functional constraints (i.e., those of the 5′-noncoding region) and positive selective forces such as that of the immune system on the hypervariable region of E2 (14, 36, 37). It is possible that the quasispecies fluctuation of the genomic region here investigated is a consequence of selective forces acting at other functional or antigenic regions of the encoded HCV proteins, favoring replication of a specific sequence subset from those present in the liver. Whatever the specific nature of the forces driving the observed fluctuations in composition and complexity of the viral population in both compartments, it is noteworthy that they have been found to occur in patients with persistently normal ALT levels and non-
progressive liver damage over a 3- to 6-year period (Table 1). Hence, in such a stable clinical situation it is tempting to speculate that the observed nucleotide fluctuation is not the result of a substantial change in the environmental conditions but rather that continuous reshaping of the viral quasispecies is necessary for the virus to persist and maintain steady-state replication. In fact, fitness variations in constant environments have been described in other viral systems (9, 42).

The high rate of nucleotide changes within and between samples contrasts with the overall low rate of amino acid changes (32) so that the most heterogeneous populations had no more than five groups of sequences differing at one or two amino acid positions (Fig. 1B). All but one of the consensus amino acid sequences were identical in both serum and liver specimens. This finding implies some kind of negative selective force driven by functional or structural constraints of the protein product encoded at this level.

With regard to our previous observation of a significant correlation between liver/serum complexity ratio and disease stage, the present study demonstrates that this parameter is not stable but fluctuates over relatively short periods of time in the absence of apparent disease progression. Hence, the liver/serum complexity ratio does not identify a particular group of patients but rather a particular state of the infecting quasispecies, which may change over time. Thus, one should be cautious when trying to correlate liver damage and quasispecies complexity. As our results have demonstrated, such correlation is found only in those coincident-time point samples, obtained from individual patients, in which the liver and serum quasispecies have the same or similar levels of complexity (patients were considered to have the same level of complexity when the ratio between liver and serum values for a given parameter was between 0.5 and 2). The issue of whether other methods of complexity analysis or investigation of quasispecies behavior at other genomic regions may provide more practical information warrants further studies.

ACKNOWLEDGMENTS

We are grateful to Mariona Parera for her excellent technical assistance.

This investigation was supported in part by grants 99-0108 and 2000-0347 from the Comisión Interministerial de Ciencia y Tecnología (CICYT) and by the Fundació per a la Recerca Biomèdica i la Do
cència de l’ Hospital Vall d’Hebron.

REFERENCES


