Nucleotide and Amino Acid Complexity of Hepatitis C Virus Quasispecies in Serum and Liver

BEATRIZ CABOT, MARÍA MARTELL, JUAN I. ESTEBAN,* SILVIA SAULEDA, TERESA OTERO, RAFAEL ESTEBAN, JAIME GUÀRDIA, AND JORDI GÓMEZ

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

Received 15 July 1999/Accepted 19 October 1999

The quasispecies nature of the hepatitis C virus (HCV) is thought to play a central role in maintaining and modulating viral replication. Several studies have tried to unravel, through the parameters that characterize HCV circulating quasispecies, prognostic markers of the disease. In a previous work we demonstrated that the parameters of circulating viral quasispecies do not always reflect those of the intrahepatic virus. Here, we have analyzed paired serum and liver quasispecies from 39 genotype 1b-infected patients with different degrees of liver damage, ranging from minimal changes to cirrhosis. Viral level was quantified by real-time reverse transcription-PCR, and viral heterogeneity was characterized through the cloning and sequencing of 540 HCV variants of a genomic fragment encompassing the E2-NS2 junction. Although in 95% of patients, serum and liver consensus HCV amino acid sequences were identical, quasispecies complexity varied considerably between the viruses isolated from each compartment. Patients with HCV quasispecies in serum more complex (26%) than, less complex (28%) than, or similarly complex (41%) to those in liver were found. Among the last, a significant correlation between fibrosis and all the parameters that measure the viral amino acid complexity was found. Correlation between fibrosis and serum viral load was found as well ($R = 0.7$). With regard to the origin of the differences in quasispecies complexity between serum and liver populations, sequence analysis argued against extrahepatic replication as a quantitatively important contributing factor and supported the idea of a differential effect or different selective forces on the virus depending on whether it is circulating in serum or replicating in the liver.

Hepatitis C virus (HCV) is an enveloped virus classified in the family Flaviviridae (7, 29, 45). Its genome consists of a single-stranded RNA, with plus polarity, of 9,600 nucleotides, which does not integrate in the host genome, yet persistence is the rule. The damage caused during infection ranges from minimal changes to cirrhosis of the liver and hepatocarcinoma, but little is known about the mechanism of hepatocyte injury due to chronic infection. It seems very likely that the pathogenesis of HCV infection is directly related to a strong interplay between the host defense mechanisms and the virus’s ability to evade them efficiently. Moreover, in order to persist, HCV must regulate its lytic potential and avoid elimination by the host immune system. Due to the quasispecies structure of the HCV viral population infecting single patients (39), the virus may use a variety of strategies to fulfill both requirements (11, 17).

The dynamic component of the quasispecies structure is responsible for the rapid virus evolution (12). It works through a complex mixture of genomic sequences (quasispecies) which behaves as a single unit when facing changes in the environment. The genetic interaction within the viral population allows the system to distinguish the best possibility at any given time and, therefore, to avoid replicative efforts in unrewarding directions (12–16). It has been experimentally proven for RNA phage (14) and viruses (27) that the use of this formula is intended to produce successful adaptation to environmental changes (12). This mechanism has been invoked in the HCV virus model to explain both the high frequency of viral persistence and the wide range of disease (3, 4, 10, 18, 21, 39). Phylogenetic methods may be used for understanding virus evolution.

The static component of the quasispecies structure of the viral population can be analyzed through the total number of viral particles (viral load), the proportion of different viral genomes present in that total (normalized Shannon entropy $P_n$) among the variants (21, 40). These parameters can be measured at single time points and are of analytical interest because they may fluctuate over a wide range of values and may be used to categorize quasispecies in relation to the clinical state of the patient. Many previous studies have examined the significance of both parameters independently. The wide range of clinico-pathological correlations between serum and intrahepatic RNA levels (20, 23, 24, 31, 32, 35, 38, 42, 47, 50, 51), together with discrepancies in the literature among authors who find correlation between quasispecies complexity and liver damage (25, 28, 33, 61) and those who do not (22, 36, 48, 56), suggests that the relation between viral load and liver injury is more complex than expected.

Recently, we have proven that, within an infected patient, the composition of the circulating viral population does not necessarily reflect the composition of the hepatic population (5), although the causes for this difference remain obscure. Heterogeneous quasispecies in peripheral blood mononuclear cells (PBMC) of humans and in chimpanzees have been described (34, 37, 49, 54, 57), and it has been proposed that replication in this tissue might contribute to HCV serum quasispecies complexity. In the present study we have evaluated the implications of serum and liver quasispecies complexity in the natural course of the disease. To do this, we have per-
formed an analysis of the viral population parameters of a genomic region encompassing the envelope 2-nonstructural region 2 (E2-NS2) junction in paired serum and liver samples from 39 patients with chronic hepatitis C.

**MATERIALS AND METHODS**

**Viral isolates.** HCV was isolated from paired serum and liver samples from 39 HCV-infected patients. The degree of liver damage was semiquantified according to the scoring system of Ishak et al. (30). This showed that 7 patients had mild chronic hepatitis, 17 had moderate hepatitis, 11 had severe hepatitis, and 4 had established cirrhosis (Table 1). A parenteral risk factor with a known date of infection was present for 24 patients. The estimated mean duration of infection of these patients was 28 ± 14 years. Demographic variables are summarized in Table 1. All patients were infected with genotype 1b and had detectable HCV RNA, but none was positive for other hepatitis viruses or human immunodeficiency virus (HIV).

Six patients had received a 6- to 12-month course of interferon treatment 2 to 4 years before the samples were obtained. Written informed consent was obtained from all patients before they underwent liver biopsy. 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 biopsies were frozen in liquid nitrogen.

**TABLE 1. Demographic, clinical, biochemical, and viral quasispecies parameters of paired liver and serum samples from a cohort of HCV-positive patients**

<table>
<thead>
<tr>
<th>Patient no. and liver damage</th>
<th>Sex</th>
<th>Age</th>
<th>D.I. (yr)</th>
<th>ALT level</th>
<th>HCV RNA load</th>
<th>Quasispecies complexity</th>
<th>ds/dn ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Serum</td>
<td>Liver</td>
<td>Serum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shannon entropy</td>
<td>Pn</td>
<td>ds/dn ratio</td>
</tr>
</tbody>
</table>

a M, male; F, female.

b Duration of infection (D.I.) was deduced from the date of transfusion.

c For serum, values are log10 (copies/milliliter); for liver, values are copies per microgram of total RNA.

d Shannon entropy and Pn are shown at the amino acid level.
e Numbers of sequenced clones (serum, liver).
**TABLE 2. Shannon entropy and \( P_n \)**

| Parameter and level | Serum | | Liver |
|---------------------|-------|-------|
| Shannon entropy     |       |       |
| Nucleotide          | 0.8 ± 0.2 | 0–1 | 0.8 ± 0.2 | 0.1–1 |
| Amino acid          | 0.4 ± 0.3 | 0–1 | 0.4 ± 0.3 | 0–1 |
| \( P_n \) Nucleotide| 1/152 ± 1/235 | 1/53–1/1,272 | 1/141 ± 1/238 | 1/56–1/2,544 |
| Amino acid          | 1/178 ± 1/263 | 1/58–1/700 | 1/178 ± 1/232 | 1/40–1/840 |

**RESULTS**

Table 1 summarizes viral population parameters of complexity in serum-liver pairs in the 39 patients studied. Viral levels ranged from \( 2.7 \times 10^7 \) to \( 2.9 \times 10^9 \) copies/ml of serum (mean, 2.6 \( \times 10^8 \) ± 4.8 \( \times 10^6 \) copies/ml) and from 0.021 to 1.9 \( \times 10^6 \) copies/\( \mu g \) of total liver RNA (mean, 24 ± 42 copies/\( \mu g \) of total liver RNA). The characteristic quasispecies structure was found in all samples, which were subsequently analyzed by using population parameters. On average, 7 E2-NS2 sequences (4 to 12) were obtained from each sample (Table 1). Overall, as shown in Table 2, mean polymorphism values and proportions of variants present in both serum and liver samples were high, although they varied widely from patient to patient. This variability indicates that in this region both parameters were of analytical interest and might be used to characterize HCV quasispecies. Serum and liver viral sequences appeared to be strongly selected for synonymous replacements (mean percentage of synonymous mutations in serum and liver, 70% ± 23% and 71% ± 25%, respectively; mean ratio of synonymous to nonsynonymous \([s/dn] \) substitutions in serum and liver, 1.6 ± 1.2 and 1.6 ± 0.9, respectively) (Table 1). Samples from four patients contained sequences differing in more than 10 residues (5% of the total fragment length) from the other sequences from the same compartment. In these cases, patients are said to have double populations.

**Comparison of serum and liver quasispecies.** The same consensus nucleotide sequence was found in 46% of patients. In 33% there were one to four ambiguities (when the consensus residue at a given position is not defined by 60% or more of the sequences, the consensus is not assigned to any residue [10]) in serum or liver consensus sequences. Twenty-one percent of patients had different nucleotide consensus sequences. At the amino acid level, 82% of the patients had identical sequences; in 13% there were one or two ambiguities, and 5% had different sequences in each compartment. Except for patients 9 and 36, phylogenetic analysis (Fig. 1) showed that sequences obtained from each individual did not segregate according to their tissues of origin.

Mean nucleotide and amino acid intrapatient distances for serum, liver, and serum/liver sequences were similar, while such distances were found to be seven or eight times higher in the interpatient analysis (Table 3). Nevertheless, as previously reported, HCV quasispecies structure (in terms of complexity) in serum did not always reflect the quasispecies structure in the liver (Table 4) at both the nucleotide and amino acid levels.

Patients were initially classified into three groups according to their tissues of origin.
to the similarity between HCV quasispecies complexities of serum and liver for each parameter, although the percentage of patients that were included in each group differed according to the parameter chosen (Table 3). 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 (group A), to have less-complex serum HCV quasispecies when the ratio was 2 or higher (group B), and to have more-complex serum HCV quasispecies when the ratio was 0.5 or less (group C).

**FIG. 1.** HCV phylogenetic reconstructions of evolutionary relationships among viruses from patients. The phylogenetic analysis shown consists of unrooted neighbor-joining trees. (A) Serum and liver nucleotide consensus sequences of HCV from each patient. (B and C) Serum and liver HCV nucleotide sequences from the two patients with manifest tissue segregation. (D) Representative tree for the 37 patients without tissue segregation.

**TABLE 3.** Intra- and interpatient genetic distances at the nucleotide and amino acid levels

<table>
<thead>
<tr>
<th>Type of comparison</th>
<th>Mean sequence distance* ± SD for:</th>
<th>Serum</th>
<th>Liver</th>
<th>Serum/liver</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nt</td>
<td>aa</td>
<td>nt</td>
<td>aa</td>
</tr>
<tr>
<td>Intrapatient</td>
<td>0.016 ± 0.01</td>
<td>0.012 ± 0.009</td>
<td>0.017 ± 0.01</td>
<td>0.012 ± 0.01</td>
</tr>
<tr>
<td>Intepatient</td>
<td>0.13 ± 9</td>
<td>0.1 ± 14</td>
<td>0.13 ± 8</td>
<td>0.1 ± 14</td>
</tr>
</tbody>
</table>

* nt, nucleotide; aa, amino acid.
TABLE 4. Differences in quasispecies complexity between serum and liver HCV (twofold or higher) as determined by Shannon entropy and polymorphism degree at the nucleotide and amino acid levels

<table>
<thead>
<tr>
<th>Group</th>
<th>% Patients at indicated level by:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shannon entropy</td>
</tr>
<tr>
<td></td>
<td>Nucleotide</td>
</tr>
<tr>
<td>A</td>
<td>87</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
</tr>
</tbody>
</table>

* According to amino acid parameters, for further correlations patients were grouped as follows: group A (n = 16), patients 2, 3, 5, 6, 9–12, 14, 21, 26, 31–33, 37, and 38; group B (n = 11), patients 1, 4, 8, 13, 15, 16, 18, 19, 25, 30, and 34; group C (n = 10), patients 7, 17, 20, 22–24, 27, 28, 35, and 39.

Since a phenotypic criterion, i.e., liver damage, had been used to select the study patients, the phenotypic parameters (those referring to amino acid sequence) were used to further study the correlations between quasispecies complexity for each compartment and clinicopathological variables. In those cases in which there was disagreement between the two amino acid parameters, final classification was based on nucleotide parameters. Accordingly, 16 patients (41%) were classified into group A, 11 (28%) were classified into group B, and 10 (26%) were classified into group C. Two patients (5%) were not included in the classification because they had different serum and liver HCV consensus amino acid sequences.

Correlation between the values of quasispecies parameters in serum and in liver. Statistical analysis of the results showed that there was a correlation between serum and liver viral RNA concentrations (R = 0.4; P < 0.02). There was a strong correlation among the six population parameters within each compartment (0.4 < R < 0.96; P < 0.02) but not between compartments, with the exception of the degree of nucleotide polymorphism and the intrapatient nucleotide distance (R = 0.66; P < 0.01). Total heterogeneity, which integrates viral load and normalized Shannon entropy, was correlated at both nucleotide and amino acid levels within and between compartments (0.4 < R < 0.9; P < 0.02).

Correlation of viral quasispecies parameters and liver damage. Viral population parameters for the 39 patients were classified according to the degree of liver damage. Significant correlation between serum viral load and fibrosis, necroinflammation, and ALT level was found (R = 0.7, P < 0.01; R = 0.5, P < 0.01; and R = 0.4, P = 0.02, respectively). Fibrosis correlated with total nucleotide heterogeneity of liver quasispecies (R = 0.4, P = 0.03) and total amino acid heterogeneity of both circulating and hepatic viral populations (R = 0.4, P = 0.01; and R = 0.3, P = 0.038, respectively). Necroinflammation correlated with total nucleotide heterogeneity of hepatic virus (R = 0.4, P = 0.03) and total amino acid heterogeneity of circulating virus (R = 0.4, P = 0.03).

Correlations between clinical and viral parameters according to liver/serum complexity ratio. Within group A patients, viral load and amino acid complexities of serum and liver quasispecies correlated with the degree of fibrosis (Table 5). In contrast, quasispecies complexity at the nucleotide level did not correlate with fibrosis. Among these patients, viral HCV RNA and amino acid complexities of circulating and hepatic viruses were higher in those with severe liver damage than in those with mild or moderate disease (data not shown). In contrast, among patients from groups B and C no correlation between viral parameters and the degree of fibrosis was found.

DISCUSSION

In a previous work (5), we found that the structure of replicating HCV quasispecies in the liver does not always reflect that of circulating HCV virions. We observed that two of four patients had a twofold-more-complex HCV quasispecies in liver than in serum. Subsequently, others have reported finding more-complex circulating quasispecies (49). The present study, involving a large number of genotype 1b-infected patients with a wide range of liver lesions, confirms and expands these observations. Most patients (95%) had HCV quasispecies with the same consensus amino acid sequence in serum and liver at the E2(p7)-NS2 junction; sequences from the majority of patients (95%) (Fig. 1) did not cluster separately between the two compartments, and in the same line, the intrapatient nucleotide and amino acid distances were seven and eight times lower than the interpatient distances. However, the amino acid complexities of the quasispecies in this region showed a twofold or higher difference between the two compartments in 54% of the patients. Accordingly, patients could be classified into three groups as a function of the degree of similarity in the complexities of viral quasispecies in both compartments. The origin and the clinical implications of this finding are unknown. In our previous work, we tried to explain the higher complexity in the liver by suggesting the existence of distinct functional compartments with different replication kinetics (5). Alternatively, since the final fate of sequences found in the replicating pool is unknown, the finding of highest complexity in the liver quasispecies might be explained by an excess contribution of sequences that will not be incorporated into mature virions and

TABLE 5. Pearson’s correlation between virological, clinical, and biochemical parameters for 16 patients with hepatitis C infection and similar quasispecies parameters in serum and liver (group A)*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Serum Shannon entropy</th>
<th>Liver Shannon entropy</th>
<th>Total heterogeneity Serum</th>
<th>Liver Total heterogeneity</th>
<th>HCV RNA Serum</th>
<th>Liver HCV RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCV RNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum</td>
<td>0.738†</td>
<td>0.567*</td>
<td>0.614*</td>
<td>0.596*</td>
<td>0.703†</td>
<td>0.737†</td>
</tr>
<tr>
<td>Liver</td>
<td>0.578*</td>
<td>0.570*</td>
<td>0.632†</td>
<td>0.986†</td>
<td>0.851†</td>
<td>0.622*</td>
</tr>
<tr>
<td>Fibrosis</td>
<td>0.846†</td>
<td>0.727†</td>
<td>0.552†</td>
<td>0.549*</td>
<td>0.537*</td>
<td>0.721†</td>
</tr>
<tr>
<td>Necroinflammation</td>
<td>0.524*</td>
<td>0.534*</td>
<td>0.567†</td>
<td>0.535†</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* nt, nucleotide; aa, amino acid; *, P < 0.05; †, P < 0.01.
released to the circulation. The opposite finding, that is, higher complexity in the circulating pool, does not have a readily obvious explanation. The possible contribution of variants replicating in extrahepatic sites has been proposed. In fact, several studies have reported the presence of distinct viral quasispecies in PBMC of infected patients (34, 37, 49, 54, 57). Any extrahepatic contribution to the circulating pool should lead to the presence of readily obvious mixed populations in the serum. This should be more apparent in long-standing infections (6, 26, 46), in which virus in the two replicating compartments (i.e., PBMC and liver) would evolve separately from a common ancestor. However, in our study a double virus population in the serum was found in only 1 of the 14 patients with long-standing infection. In addition, in two of the four patients who had a double population of sequences in the serum, the corresponding sequences were also present in the liver. All these data, the phylogenetic clustering of serum and liver sequences (in 37 of 39 patients; Fig. 1) and the finding that virus level was correlated between both compartments, argue against a significant contribution (in quantitative terms) of extrahepatic HCV replication to the serum (9, 43).

Alternatively, differences in the clearance rates of some variants might be responsible for the observed differences (19). Rapid elimination of a major variant by circulating antibodies could lead to an overrepresentation of the mutant repertoire. In that situation, the observed differences between the circulating and the hepatic virus would be more apparent than real.

Several studies have tried to correlate the complexity of the circulating quasispecies and degree of liver damage (22, 25, 28, 33, 36, 48, 56, 61). In the present study we found no correlation between quasispecies complexity at the nucleotide level and liver damage. In contrast, a significant correlation between quasispecies complexity at the nucleotide level and liver damage might have pathogenic relevance since the interaction between virus and the host immune system occurs at the phenotypic level. This observation would fit theoretical models of HIV diversification, in which antigenic variants are not completely replaced by emerging ones, so that the continuous accumulation of variants could reflect the history of immune evasion and cell destruction (52, 53). In these patients (group A), the good correlation between serum viral load and degree of fibrosis would allow the monitoring of disease progression in individual patients by HCV RNA quantitation (1, 8, 41, 58, 60). Nevertheless, the potential use of amino acid parameters of complexity and/or viral load as an indirect measure of ongoing liver damage is limited for two reasons. First, correlation is restricted to those patients with similar levels of quasispecies complexity in both compartments, and these cannot be differentiated from the other two groups by any clinical or readily accessible parameter. Second, it is currently unknown whether the liver/serum complexity ratio is a stable parameter or fluctuates over time. Longitudinal studies of sequential serum and liver pairs would be required to clarify this issue. It is possible that coincident quasispecies complexities represent a steady-state level, in which more complexity implies more damage. Such an equilibrium may transiently be lost when viral or immune factors influence the complexity of the circulating or replicating pool. Further investigation of the dynamic behavior of viral quasispecies in both compartments would increase our understanding of the influence of quasispecies complexity in liver damage.

ACKNOWLEDGMENTS

This investigation was supported in part by grant 94/1682 from the Fondo de Investigaciones Sanitarias (FIS), by grant 97-0148 from the Comisión Interministerial de Ciencia y Tecnología (CICYT), and by the Fundació per a la Recerca Biomèdica i la Docència de l’Hospital Vall d’Hebron.

REFERENCES


and liver damage in patients with chronic hepatitis C virus infection. Hepatology 25:697–701.


