Complexity and Diversity of Hepatitis C Virus RNA in African Americans and Whites: Analysis of the Envelope-Coding Domain

Erica D. Keenan,1 Susan D. Rouster,1 Norah J. Shire,1 Paul S. Horn,2 and Kenneth E. Sherman1

1Division of Digestive Diseases, University of Cincinnati College of Medicine, and 2Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio

African Americans infected with hepatitis C virus (HCV) are less responsive than whites to interferon-based therapy. HCV quasi species have been implicated. Quasi-species complexity and diversity were evaluated in matched African American and white individuals. Complexity and diversity in the first hypervariable region were similar in the 2 groups. Both the frequency of nonsynonymous amino acid substitutions and the mean ratio of nonsynonymous mutations to synonymous mutations were greater in clones derived from white patients. Racial differences in amino acid usage at otherwise conserved positions were observed. Differences in amino acid representation at key positions are suggestive of immunologic and functional selection along racial lines.

Standard treatment for chronic hepatitis C virus (HCV) infection leads to a sustained virological response in ≥50% of treatment-naïve patients [1]. However, significant racial differences in treatment response have been described. African American patients are less responsive to antiviral therapy than are white patients [2, 3]. Various viral and host factors, including the quasi-species nature of HCV, may account for this difference in responsiveness to therapy. Greater complexity (total number of codominant variants within a population) and diversity (genetic heterogeneity of the population) of the viral population may allow HCV to establish and maintain chronic infection despite active immune response. Complexity and diversity in certain regions of the viral genome, particularly in the first hypervariable region (HVR1), have been associated with the outcome of treatment for HCV infection [4].

We postulated that greater complexity and diversity of HCV RNA might be observed in African American patients. We examined quasi-species complexity in the E2/NS1 HVR1 from matched African American and white patients, assuming that racial differences in both mutation and selection would be most apparent at this variable genome site. We conducted, in a subset of matched pairs of patients, a comparison of nucleotide and amino acid diversity in the HVR1 and flanking envelope domains.

Materials and methods. Blood was collected, under the auspices of an Institutional Review Board–approved protocol, from HCV-infected patients. Infection was confirmed by both HCV-antibody reactivity (EIA and/or RIBA) and detection of HCV RNA by polymerase chain reaction (PCR) or branched-chain–DNA assays. The serum samples were divided into aliquots and stored in a serum bank at −70°C ±3 hours of venipuncture. Sixty-four patients (32 African American and 32 white) of genotype 1 were matched on the basis of covariates including sex and age (matched to ±5 years), yielding 32 matched pairs of patients. Patients recorded racial identity on a demographic questionnaire. Exclusion criteria included previous interferon-based treatment, alcohol abuse, decompensated liver disease, and/or coinfection with HIV or hepatitis B virus. When they were available, outcomes of treatment of the patients were analyzed post hoc, to determine any detectable intergroup differences.

HCV RNA was isolated from 140 μL of patient serum by use of the QIAamp Viral RNA Mini Kit (QIAGEN). Nested reverse-transcription (RT) PCR of samples of the patients’ RNA was performed as described elsewhere [5]; both random-hexamer and specific external primers were utilized in the RT step. Paired sampling from several patients failed to demonstrate a measurable difference, in quasi-species complexity, between amplicons generated from random hexamers and those generated from a specific-primer set (data not shown). The primers flank a 136-bp region of the E2/NS1 envelope-coding domain that includes the HVR1.

Heteroduplex complexity analysis (HCA) was performed as described elsewhere [5, 6]. In brief, DNA polymerase present in PCR products was inactivated with EDTA. In an additional denaturation/renaturation step, the sample was heated to 95°C for 3 min and then was slowly cooled to room temperature. Electrophoresis was performed on a nondenaturing polyacryl-
amide mutation detection enhancement (MDE) gel (Cambrex BioScience). Samples were run at 1000 V for 5–6 h. The bands were stained with GelStar Nucleic Acid Gel Stain (Cambrex BioScience) and were visualized under UV light. Two independent investigators, blinded to race and other identifiers, determined band number by visual analysis. Discordant results were resolved by discussion and mutual agreement.

HVR1 amplicons from 10 randomly selected matched pairs of patients were cloned into pCR 2.1 vectors and were introduced into One Shot competent cells, according to the TA cloning method (Invitrogen). Selected clones were isolated and grown overnight, and the plasmids were isolated by use of the QIAprep Miniprep Kit (QIAGEN). Presence of the insert was confirmed by digestion with EcoRI, which yielded a 176-bp fragment.

Ten clones were obtained for each patient. Sequences were aligned manually, and gaps were created where any additions or deletions occurred. MEGA algorithms were used to calculate, for all clones, the mean Kimura 2-parameter pairwise distance and to generate, for each patient, a matrix of Kimura 2-parameter pairwise distances. The numbers of synonymous mutations ($d_S$) and nonsynonymous mutations ($d_N$) were determined by the Nei-Gojobori method with the Jukes-Cantor correction for multiple substitutions. The mean $d_S:d_N$ ratio was calculated for each patient. Phylogenetic trees of representative patients were generated by the neighbor-joining method, with the bootstrap test for reliability. Predicted amino acid sequences based on observed nucleotides were compared by group.

The $\chi^2$ test was used to perform demographic comparisons and to determine specific amino acid differences. Paired $t$ tests were used for band-number comparisons and to determine amino acid diversity. A 2-way mixed-model analysis of variance (ANOVA) was used to compare distance-matrix differences within pairs of patients. In all cases, statistical significance was set at a threshold of $\alpha = 0.05$, by use of a 2-tailed hypothesis. Nucleotide sequences used in this study are available in the GenBank database, under accession numbers AY194602–AY194802.

**Results.** Sixty-four patients of genotype 1 were included in this study. Thirty-two African American and 32 white patients were paired on the basis of age (matched to ± 5 years) and sex. The average age was 44.2 years for African American patients and 44.7 years for white patients ($P = NS$). Each group comprised 19 men and 13 women.

All patients were treatment naive when samples were collected for this study. Thirteen (40.6%) of 32 African American patients and 15 (46.9%) of 32 white patients were subsequently treated, either in the context of routine clinical care or after enrollment in a clinical-treatment trial ($P = .80$); 45% of treated African American patients received pegylated interferon with ribavirin, and 26.6% of treated white patients received this combination ($P = .42$). Of those receiving treatment, 46.7% of white patients and 15.4% of African American patients experienced viral clearance. This was a post hoc analysis, and the difference was not statistically significant.

Quasi-species complexity differences between matched pairs were assessed by nonisotopic HCA. This previously validated methodology has demonstrated strong and highly statistically significant correlation with sequence analysis ($r = 0.99; P < .005$) [5]. For each patient, band number on an MDE gel was quantified, as an indirect measure of complexity. The mean ± SE band number was 2.63 ± 0.27 for African American patients and 2.59 ± 0.22 for white patients ($P = .90$). To determine nucleic acid diversity, 10 pairs of patients (5 men of each race and 5 women of each race) were randomly selected for cloning and sequence analysis of 176-bp amplicons containing the HVR1. The mean ± SE band number was 3.40 ± 0.43 for white patients and 3.20 ± 0.63 for African American patients ($P = .74$). On average, 10 clones per patient were obtained. The mean ± SE Kimura 2-parameter distance was 0.28 ± 0.036 for clones derived from African American patients and 0.43 ± 0.046 for

**Figure 1.** A, Phylogenetic tree of sequence clones from patient 1246, a 42-year-old white woman with a quasi-species band count of 1. For this patient, the mean ratio of nonsynonymous substitutions to synonymous substitutions was 1 (0.1:0.1), and the mean ± SE pairwise distance of clones was 0.14 ± 0.028. B, Phylogenetic tree of sequence clones from patient 1167, a 42-year-old African American woman, with a quasi-species band count of 1. For this patient, the mean ratio of nonsynonymous substitutions to synonymous substitutions was 1 (0.2:0.2), and the mean ± SE pairwise distance of clones was 0.027 ± 0.023.
clones derived from white patients. A matrix of Kimura 2-parameter pairwise distances was generated for each patient. Comparison of these distance measurements, by a 2-way mixed-model ANOVA, did not demonstrate a statistically significant racial difference in nucleic acid diversity ($P = .19$). Analysis of the 81-bp HVR1 sequence, located within the amplified E2/NS1 envelope-coding domain, also failed to demonstrate a difference between the 2 groups ($P = .22$).

Phylogenetic trees of representative patients from each group were generated by the neighbor-joining method. The phylogenetic tree derived from all clones from white patient 1246 (figure 1A) shows a heterogeneous clonal population, whereas the tree from her African American match, patient 1167 (figure 1B), shows less diversity, consistent with the observed trend in this African American population.

The $d_S:d_I$ ratio was calculated for each clone; a ratio $>1$ indicates a rate of structurally or functionally significant mutations that is higher than the rate of random mutations that are also occurring at synonymous sites. The mean $\pm$ SE $d_S:d_I$ ratio of all clones derived from white patients ($1.35 \pm 0.7132$) was statistically significantly greater than that derived from African American patients ($0.681 \pm 0.3593$) ($P < .001$, $t$ test with unequal variance).

Predicted HVR1 amino acid sequences (aa 384–410) were grouped by race and were aligned to determine the number of unique amino acids observed at each position. From these data, a profile of amino acid variation in the HVR1, in African Americans and in whites, was created (figure 2). The mean $\pm$ SE substitutions per site was $4.2 \pm 0.47$ for clones derived from African American patients and $5.1 \pm 0.52$ for clones derived from white patients ($P = .01$).

Despite the hypervariability of this region, certain residues were highly conserved, occurring in $\geq 98.9\%$ clones, irrespective of race: in all but one clone, a Thr residue was found at position 2 of the amplified sequence, a Gly residue was highly conserved at positions 6 and 23, Gln occupies position 26 in 100% and 98.8% of clones derived from African Americans and from whites, respectively, and either a Phe or an Leu residue occupies position 20. However, racial differences were observed in amino acid usage at other conserved positions: a Thr residue occupies position 5 in 86 (93.5%) of 92 clones derived from African American patients but in only 57 (64.0%) of 89 clones derived from white patients ($P < .0001$).

**Discussion.** HCV quasi-species heterogeneity has been implicated as a factor affecting response to interferon-based therapy. Some investigators have detected an inverse association between quasi-species complexity in the HVR1 and treatment response [7–9]. Cohort differences have also been described in patients coinfected with HIV (although not all reports detect this association [10, 11]) and in those with alcoholic liver disease and elevated levels of alanine aminotransferase [12]. Preliminary data from our group had suggested that HCV quasi-species complexity in the HVR1 is greater in African Americans than in whites [13]; however, that analysis had included patients with potential confounding factors, such as prior treatment and alcohol abuse.
We evaluated quasi-species complexity and diversity of the HVR1 in a cohort of treatment-naive African American and white patients matched on the basis of potential confounding factors such as sex, age, and HCV genotype. A retrospective analysis of treatment responses in the study cohort revealed that the rate of sustained viral response in African Americans was lower than that in whites (15.4% vs. 46.7%), a finding that is consistent with the results of other published studies [2]. Quasi-species complexity was similar in African American and white patients \( (P = .90) \). Although the mean Kimura 2-parameter distance was greater in clones derived from whites than in clones derived from African Americans, ANOVA failed to detect a difference in nucleic acid diversity within matched pairs of patients. This may be due to the relatively small numbers of matched pairs and therefore may reflect a type II statistical error. We assessed intergroup differences in the \( d_{S_0}:d_\beta \) ratio of the HVR1; this ratio serves as a measure of relative selective pressure, because nonsynonymous mutations are a method of evasion of host immune response. A greater proportion of nonsynonymous substitutions was observed in clones derived from white patients, suggesting that immune selection in this group was greater than that in African American patients.

A profile of amino acid substitutions in the HVR1 was generated. Clones derived from white patients demonstrated a greater number of amino acid substitutions per site than did those derived from African American patients \( (P = .01) \); similar observations have been reported in comparisons of interferon responders and nonresponders [14, 15]. We observed amino acid differences at key conserved positions. It is possible that these conserved positions are involved in interactions between the HVR1 and molecules located on the cell surface; altered amino acid placement at these positions may reflect possible group-specific phenotypic differences in infected cells.

There are limitations to this study that should be acknowledged. This was a retrospective study that was unable to control for potentially confounding variables such as fibrosis score and baseline HCV viral loads. Treatment intervention was not an element of this study’s design. Not all patients in this study initiated or completed treatment, and not all patients were not treated similarly; therefore, our treatment-response analysis was post hoc and not powered to detect a difference. Although the heteroduplex assay methodology described above has been validated at our site, external validation has not been published. Future studies to compare heteroduplex methods and single-strand conformational–polymorphism assays to sequence analysis would be helpful.

In conclusion, we find that, when the entire HVR1 nucleotide sequence and flanking E2/NS1 sequences are evaluated, there is very little evidence for significant racial differences in HCV RNA quasi-species complexity or diversity. However, African American and white patients in this cohort demonstrated notably different response rates to interferon-based therapies. Greater amino acid variability occurs in whites, suggesting the presence of a more vigorous immune response driving selection. Race is associated with substitutions at key amino acid sites, which could have both functional (replicative) and immunologic significance and may warrant further investigation.

References