

## Significance of Pretreatment Analysis of Hepatitis C Virus Genotype 1b Hypervariable Region 1 Sequences To Predict Antiviral Outcome

Catherine Gaudy,<sup>1</sup> Alain Moreau,<sup>1</sup> Pascal Veillon,<sup>2</sup> Stephanie Temoin,<sup>1</sup> Francoise Lunel,<sup>2</sup> and Alain Goudeau<sup>1\*</sup>

*Département de Microbiologie Médicale et Moléculaire EA 3250, Université François Rabelais, F37044 Tours Cedex,<sup>1</sup> and Laboratoire de Virologie, CHU, F49033 Angers Cedex 01,<sup>2</sup> France*

Received 11 November 2002/Returned for modification 13 January 2003/Accepted 17 May 2003

**The heterogeneity of hypervariable region 1 (HVR1), located at the amino terminus of the E2 envelope, may be involved in resistance to alpha interferon (IFN- $\alpha$ ) treatment. We investigated whether peculiar HVR1 domain profiles before treatment were associated with the maintenance of sensitivity or the appearance of resistance to treatment. Fifteen patients infected with hepatitis C virus genotype 1b and treated with IFN with or without ribavirin were selected. Ten responded to treatment (groups R1 and R2) and five did not (group NR). The amino acid sequences of 150 naturally occurring HVR1 variants present in the serum before therapy were compared in relation to treatment outcome. HVR1 variants from the NR group contained a constant nonantigenic amino acid segment that was not found in HVR1 variants from the R groups.**

Hepatitis C virus (HCV) infection leads to viral persistence and chronic disease in 50 to 70% of cases. A significant proportion of chronically infected patients subsequently develop cirrhosis and hepatocellular carcinoma. The high prevalence of HCV infection in the general population, the absence of documented spontaneous recovery from chronic infection, and the potentially serious complications make treatment with alpha interferon (IFN- $\alpha$ ) or a combination of ribavirin and IFN- $\alpha$  necessary. However, the virological response is sustained in less than 20% of patients treated with IFN- $\alpha$  alone and in 40 to 45% of patients given combination therapy (30). The antiviral response is determined by the HCV genotype and the viral load: the frequency of the long-term response to IFN treatment is higher for patients infected with genotypes 2 and 3 than for those infected with genotype 1 (20). The covalent attachment of IFN- $\alpha$  to polyethylene glycol (pegylated interferon) increases its half-life and improves the overall effectiveness of treatment, although there is still a significant difference in the overall effectiveness of treatment for patients infected with different genotypes (44).

In infected individuals, HCV exists as pools of related genetic variants, referred to as quasispecies (3). The genes encoding the envelope glycoproteins (E1 and E2) are the most heterogeneous, especially the 81 nucleotides encoding hypervariable region 1 (HVR1) of E2 (24). Therefore, HVR1 can be used both to identify individual viral strains and to study HCV quasispecies. HVR1 is predicted to be flexible and well exposed (40). HVR1 contains linear B-cell epitopes and is thought to be the major immunogenic domain of E2 (39), although other B-cell epitopes outside of HVR1 have also been identified (16). There is considerable evidence suggesting

that HVR1 plays a functional role in the attachment and entry of HCV into target cells: (i) anti-HVR1 monoclonal antibodies inhibit the attachment of HCV to human T cells in vitro (45), (ii) rabbit hyperimmune serum directed against the C terminus of HVR1 neutralizes HCV infectivity in chimpanzees (5), and (iii) only small amounts of RNA are detected in chimpanzees after infection with an HCV variant in which the nucleotide sequence corresponding to HVR1 has been deleted (6).

HVR1 is mutated during the natural course of HCV infection in untreated immunocompetent individuals (9, 14). The mutation rate has been estimated to be 0.1 to 0.2 nucleotide substitutions per genome site per year, and several amino acid changes occur over a period of 1 year or more (13, 23). The treatment of chronic HCV carriers with IFN- $\alpha$  dramatically disrupts the quasispecies equilibrium. The number of variants (33) and the degree of sequence heterogeneity (43) decrease following IFN treatment among individuals showing a sustained or short-term response. Inconclusive results have been obtained for nonresponders. Most single-strand conformation polymorphism (SSCP) analysis studies have shown that HVR1 patterns change during treatment; thus, these patterns are considered a hallmark of the immunogenic shift associated with a lack of response or relapse (29). In contrast, an association was recently found between the apparent absence of changes in HVR1 quasispecies and the lack of response to treatment (27). This suggests that major pretreatment variants (i.e., the naturally occurring variants present in the serum before therapy) are intrinsically associated with resistance to IFN- $\alpha$ . Attempts to determine whether pretreatment genetic heterogeneity in HVR1 is correlated with the outcome of IFN- $\alpha$  treatment have also been inconclusive. The earliest studies describe a significant correlation between greater HVR1 heterogeneity in pretreatment serum and a lack of response to IFN- $\alpha$  (25, 38, 42). More recently, nonsynonymous pretreatment substitutions in HVR1 sequences were shown to

\* Corresponding author. Mailing address: Département de Microbiologie Médicale et Moléculaire EA 3250, Faculté de Médecine, 2, bis Boulevard Tonnellé, F37044 Tours Cedex, France. Phone: (33) 2 47 47 69 97. Fax: (33) 2 47 47 36 10. E-mail: goudeau@med.univ-tours.fr.

be more frequent than synonymous substitutions in responders, suggesting that the anti-HCV immune response is involved in HCV clearance during IFN- $\alpha$  treatment (34). These conflicting results suggest that it is not possible to use quantitative methods to explore fully the changes in HVR1 so that we can use these changes to guide and control antiviral therapy. Recent studies exploring the genomic and conformational variabilities of HVR1 in relation to its putative biological role show that amino acid replacements are constrained to yield particular chemico-physical properties and conformations, indicating that although it is highly variable, HVR1 plays an important biological role in HCV replication (10, 28). This functional analysis of HVR1 has not yet been confronted with IFN treatment outcome. We hypothesized that some of the variants generated during the natural progression of HCV infection acquire conformational and/or immunological properties that make them resistant to treatment. To test this hypothesis, we studied pretreatment serum samples from a panel of HCV genotype 1b (HCV-1b)-infected patients in an attempt to characterize the chemico-physical, conformational, or antigenic profiles of the E2 HVR1 domain that might make it possible to distinguish between patients who respond to treatment and those who do not. We analyzed 150 cloned sequences encompassing HVR1 and its flanking regions.

#### MATERIALS AND METHODS

**Patients.** Fifteen human immunodeficiency virus-negative patients infected with HCV-1b were selected. All subjects had volunteered to take part in a clinical trial comparing treatment with IFN alone ( $6 \times 10^6$  U three times a week for 12 months) to a regimen consisting of a combination of IFN and ribavirin (1.2 g/day for 9 months). The HCV variants in serum samples collected within the 8 weeks prior to the beginning of antiviral therapy were analyzed.

Ten patients (five who received IFN alone, called group R1, and five who received the combination therapy, called group R2) had sustained virological responses to antiviral therapy, as defined by normal alanine aminotransferase activities and negativity for HCV RNA by PCR 6 months after the end of the treatment; five patients (group NR) did not respond to the combination therapy. The patients in groups R1 and R2 were between 28 and 65 years of age (median age, 43 years), and those in group NR were between 34 and 67 years of age (median age, 49 years). The viral loads ranged from  $1.5 \times 10^4$  to  $6.9 \times 10^6$  IU/ml (median viral load,  $3 \times 10^5$  IU/ml) in groups R1 and R2 and from  $3 \times 10^5$  to  $7.9 \times 10^5$  IU/ml (median viral load,  $6 \times 10^5$  IU/ml) in group NR.

**Cloning and sequencing of the HVR1 domain and its flanking regions. (i) RNA extraction and nested RT-PCR.** We used a classical nested RT-PCR method that included an initial reverse transcription step followed by two amplification steps. Virus RNA was extracted from 200  $\mu$ l of serum by the guanidinium thiocyanate method. RNA extracts were reverse transcribed with a specific primer (5'-GGTGTGGAGGGAGTCATTGCAGTT-3') that annealed 75 nucleotides downstream of the HVR1 domain. Reverse transcription was carried out at 42°C for 30 min in the presence of 12.5 U of Moloney murine leukemia virus reverse transcriptase (Stratagene). For the nested PCR steps, an optimized genotype-independent set of primers specific for the amino-terminal region of the HCV E2 gene (HVR1) was used (17). The first round of PCR was carried out with the reverse transcription primer described above (20 pmol) as the outer antisense primer and with primer 5'-CGCATGGCGTGGGACATGATG-3' (20 pmol) as the outer sense primer. The first round of amplification consisted of 40 cycles consisting of denaturation at 94°C for 20 s, primer annealing at 60°C for 50 s, and primer extension at 72°C for 30 s, followed by a final prolonged extension at 72°C for 15 min. The second step was performed with the inner antisense primer 5'-TGCCACCTGCCATTGGTGT-3' (20 pmol) and the inner sense primer 5'-GCTTGATATGATGATGAAGTGGTC-3' (20 pmol). The conditions for the second round were as follows: 5 min of denaturation at 94°C, followed by 30 cycles of 94°C for 20 s, 57°C for 20 s, and 72°C for 40 s and a final extension at 72°C for 15 min. Negative controls containing distilled water instead of RNA or cDNA were included in each of the three steps to ensure that the reverse transcription and PCR mixes were not contaminated. HCV-negative serum samples were used as specificity controls. The amplified products were subjected to

electrophoresis in 1% agarose gels (Gibco BRL) and stained with ethidium bromide. A 308-bp PCR product, corresponding to nucleotides (nt) 955 to 1262 (195 nt upstream and 32 nt downstream from HCV HVR1), was generated.

**(ii) cDNA cloning.** HVR1 amplicons were recovered from liquid low-melting-point agarose and directly ligated into 50 ng of vector PCR 2.1 (Topo-TA Cloning kit; Invitrogen). Recombinant plasmids were used to transform competent *Escherichia coli* cells, and transformants were detected according to the protocol of the manufacturer. Fifteen positive clones from each patient were selected at random.

**(iii) Sequencing.** Plasmid DNA was isolated from 1.5 ml of a 16-h broth culture and purified with a Qiagen plasmid mini kit, according to the protocol of the manufacturer. The presence of the HVR1 insert was verified by restriction digestion with *EcoRI*. For each patient, both strands of 10 HVR1-containing plasmids were sequenced with the M13 universal and M13 reverse primers (Eurogentec) by the dideoxy chain termination method with the Thermo Sequenase II Dye Terminator cycle sequencing premix kit (Amersham Pharmacia Biotech) on an automatic DNA sequencer (ABI PRISM 377; Perkin-Elmer, Applied Biosystems). Thus, a set of 50 variants (10 clones from each of five patients) was analyzed for each of the three groups of patients (groups NR, R1, and R2). Sequence data were analyzed with the Sequence Navigator program (version 1.0.1; Applied Biosystems), and the nucleotide sequences were aligned by use of BIOEDIT software (Department of Microbiology, North Carolina State University [http://www.mbio.ncsu.edu/BioEdit/bioedit.html]).

**Sequence analysis of the HVR1 domain and its flanking regions. (i) Mutation frequency.** Sporadic mutations (i.e., mutations occurring once in a given set of amino acid sequences) can be due to PCR enzyme errors. The expected number of sporadic amino acid substitutions due to *Taq* polymerase can be calculated with a simple formula that takes into account the specific error rate for the enzyme and the number of PCR cycles (36). In our experiment, the number of artifacts due to *Taq* errors was estimated for each group of patients as follows: [error rate of *Taq* ( $0.2$  to  $2 \times 10^{-4}$ )  $\times$  length of HVR1 (81 nt)  $\times$  number of PCR cycles ( $n = 70$ )  $\times$  number of clones ( $n = 50$ )  $\times$  proportion of sporadic mutations expected to produce amino acid substitutions ( $2/3$ )]/number of samples (5). According to this formula, the expected number of sporadic amino acid substitutions induced by *Taq* errors ranged from 0.7 to 7.4 substitutions per group of patients.

The nucleotide substitution rate over sites within each set of 50 variants (nt 955 to 1262) was estimated by using DAMBE software (Department of Ecology and Biodiversity, University of Hong Kong [http://web.hku.hk/~xia/software/software.html]). The variability at each site was measured by calculating the entropy at site  $i$  ( $H_i$ ) by the following formula:  $H_i = -\sum_j^4 p_j \log_2 p_j$ , where  $j$  is equal to 1, 2, 3, and 4, corresponding to nucleotides A, C, G, and T, respectively, and  $p_j$  is the proportion of nucleotide  $j$  at site  $i$ . The types of mutational changes affecting the HVR1 domain (nt 1150 to 1230) were also determined by means of the MEGA program (version 2.1; Department of Biology, Arizona State University [http://www.megasoftware.net]). The numbers of synonymous substitutions ( $d_s$ ) and nonsynonymous substitutions ( $d_n$ ) per site were calculated by the Nei-Gobojori method by using the Jukes-Cantor correction to account for multiple substitutions at the same site (22). The ratio of nonsynonymous mutations to synonymous mutations (the  $d_n/d_s$  ratio) within each sample was used as an indicator of relative selective pressure.

**(ii) Amino acid sequence.** For each patient, the predicted amino acid sequences of 10 entire cloned domains (amino acids [aa] 319 to 420) were aligned. The properties of the predicted amino acids of HVR1 and its flanking regions were evaluated for each variant with DAMBE software. Polarity and hydropathy plots were prepared by the methods of Grantham (8) and Kyte and Doolittle (15), respectively. These methods calculate the average hydropathy or basicity index of overlapping segments of several amino acids (3 aa in our case). The whole sequence is scanned, and consecutive scores are plotted from the amino terminus to the carboxy terminus. The hydropathic and chemical properties of the residues were studied at each position of the HVR1 domain (aa 384 to 410). Three categories of residues were taken into account when the hydropathic pattern was determined: neutral, hydrophobic, and hydrophilic. Four categories of residues were considered when chemical properties were evaluated: basic, acid, nonpolar, and polar. The hydropathic and chemical patterns at individual sites were compared among the three groups (groups R1, R2, and NR) by Fisher's exact  $t$  test.  $P$  values were estimated by using the Benjamini and Hochberg correction, which is required for multiple comparisons.

**(iii) Prediction of conformational states and antigenic profiles.** The secondary structures of the entire cloned domains were further analyzed by the double prediction method of Deleage and Roux by using ANTHEPROT software (version 5.0; Institute of Biology and Chemistry of Proteins, University of Lyons [http://www.pbil.univ-lyon1.fr]). We determined the predicted conformational state (i.e., coil,  $\alpha$  helix,  $\beta$  sheet, and turn) at each amino acid position of the

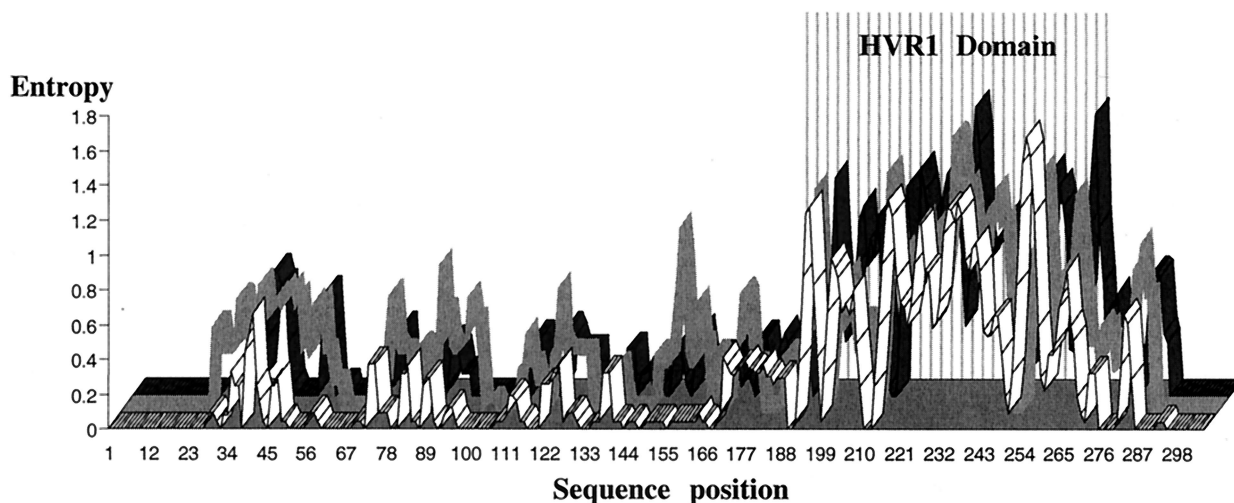


FIG. 1. Substitution rate at each site within the HVR1 domain and its flanking regions (nt 955 to 1262) for pretreatment isolates of HCV-1b-infected patients. For each alignment of 50 sequences, the nucleotide substitution rate at each site was estimated by calculating entropy with DAMBE software. Group R1 and R2 sequences are represented by a gray curve and a black curve, respectively, whereas group NR sequences are represented by a white curve. The hypervariable region (nt 1150 to 1230) is indicated by vertical lines.

HVR1 domain. The predicted secondary structures were aligned, and the conformational similarities were calculated by using the identity matrix. Antigenicity profiles were calculated with ANTHEPROT software (version 5.0) by the method of Parker et al. (26) on the basis of the atomic flexibilities, hydrophilicities, and experimental high-pressure liquid chromatography retention times of synthetic peptides to predict protein surfaces.

(iv) **Software.** The ANTHEPROT software package includes most of the methods required for protein sequence analysis. It can be used to search for functional sites, to study physicochemical profiles, and to predict secondary structures and antigenic scores.

The DAMBE software package can be used to retrieve, edit, and analyze molecular sequence data and can also be used for phylogenetic comparison. It allows visualization of the site-specific variabilities of orthologous sequences and can be used to plot amino acid properties along single or multiple sequences.

MEGA software is used for phylogenetic analyses with the maximum parsimony, distance matrix, and likelihood methods. It can also calculate basic statistical quantities such as nucleotide and codon frequencies, transition to transversion, and  $d_N/d_S$  ratios.

**Nucleotide sequence accession numbers.** The nucleotide sequences have been submitted to GenBank and have been assigned accession numbers AF473704, AF473705, AF473708 to AF473724, AF473727 to AF473735, AF473737 to AF473806, AF478360, and AF478361.

## RESULTS

**Analysis of mutation frequency.** Twelve sporadic amino acid substitutions were found among R1 variants, 14 were found among R2 variants, and 21 were found among NR variants. These rates are much higher than those expected from simple *Taq* polymerase errors (0.7 to 7.4). Therefore, some of these substitutions may correspond to segregating polymorphisms. None of the mutations were discounted in further analyses.

The analysis of substitution rates along HVR1 and flanking sequences did not reveal any major differences between the three groups (Fig. 1). The  $d_N/d_S$  ratios of the HVR1 sequences were similar for the three groups (0.278 to 3.307 for group NR, 0.200 to 3.850 for group R1, and 0.341 to 3.872 for group R2).

**Analysis of amino acids properties in the HVR1 domain and its flanking regions.** The frequency of particular amino acids at each HVR1 position was determined for each group (Fig. 2). Six residues (residues 2, 6, 7, 20, 23, and 26) were conserved among all 150 variants, whereas the residues at most positions

were variable. Some amino acid substitutions did not significantly change the hydrophobic or chemical properties (positions 16, 19, and 24). Amino acid substitutions at 12 positions triggered significant changes ( $P < 0.05$  calculated with the Benjamini and Hochberg correction) in both hydrophobic and chemical properties. Nine of these positions were located in the middle of the HVR1 domain, between residues 8 and 18. Positions 9 and 17 differed considerably between variants from responders and nonresponders. Variants from group NR tended to have hydrophilic and polar residues at position 9, whereas variants from groups R1 and R2 predominantly contained hydrophobic and nonpolar residues at this position. Furthermore, variants from group NR tended to have hydrophobic and nonpolar residues at position 17, whereas variants from groups R1 and R2 predominantly contained neutral and polar residues at this position. Polarity and hydrophobicity plots of the HVR1 and flanking regions did not show any significant differences according to treatment outcome (Fig. 3).

**Analysis of conformation and antigenic prediction.** A phylogenetic tree constructed on the basis of conformational similarities revealed no clustering of variants with respect to treatment response (i.e., variants from patients in groups R1, R2, and NR were scattered throughout the phylogenetic tree) (data not shown). The antigenic linear epitopes of the HVR1 domain were predicted from the sequence of the entire clone (308 nt). Two regions of HVR1 were predicted to be antigenic. They were located at positions 1 to 12 and 16 to 27 for variants from group R1, positions 1 to 13 and 16 to 27 for variants from group R2, and positions 1 to 12 and 18 to 27 for variant from group NR. Variants from nonresponders displayed a nonantigenic segment (between positions 16 and 17) that was not found among the variants from the two responder groups (Fig. 4).

## DISCUSSION

In the past few years, there has been growing interest in the study of the genetic heterogeneity of HCV isolates in relation to the clinical course of HCV-associated disease. In some stud-

**Responders Ifn : 50 clones**

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  | 21  | 22  | 23  | 24  | 25  | 26  | 27  |  |
| E20 | T50 | H21 | V40 | T50 | G50 | G50 | S20 | A40 | A39 | R21 | N23 | A29 | A20 | G23 | L32 | S21 | S43 | F26 | F50 | T28 | P18 | G49 | A25 | S30 | Q49 | N39 |  |
| S10 |     | Y20 | T10 |     |     |     | T10 | S10 | G10 | Q19 | T22 | T21 | Y19 | K10 | F18 | T19 | N7  | L23 |     | S13 | S10 | A1  | P21 | K19 | R1  | K10 |  |
| T10 |     | R9  |     |     |     |     | A10 |     | T1  | Y10 | S4  |     | R9  | R9  |     | A10 |     | I1  |     | A7  | L9  |     | S4  | R1  |     | S1  |  |
| N5  |     |     |     |     |     |     | L9  |     |     |     | Y1  |     | S1  | S8  |     |     |     |     |     | K1  | Y4  |     |     |     |     |     |  |
| D5  |     |     |     |     |     |     | Q1  |     |     |     |     |     | H1  |     |     |     |     |     |     | P1  | F4  |     |     |     |     |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | T3  |     |     |     |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | R2  |     |     |     |  |

**Responders Ifn / Riba : 50 clones**

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  | 21  | 22  | 23  | 24  | 25  | 26  | 27  |
| H28 | T50 | H21 | V28 | T20 | G49 | G50 | T18 | A31 | A36 | R31 | T23 | T32 | Y15 | G36 | L29 | T34 | S40 | L47 | F38 | T20 | R23 | G50 | P38 | Q18 | Q50 | K40 |
| S10 |     | Q18 | T21 | S20 | R1  |     | A17 | S10 | G10 | H7  | Q20 | A18 | A10 | S11 | F16 | S12 | G10 | F2  | L12 | S12 | L9  |     | A12 | S10 | Q10 |     |
| E10 |     | R10 | M1  | V9  |     |     | V11 | V9  | S4  | F6  | A7  |     | G9  | R3  | I5  | A4  |     | H1  |     | V7  | P7  |     |     | A10 |     |     |
| G1  |     | Y1  | M1  |     |     |     | S3  |     |     | Y6  |     |     | Q6  |     |     |     |     |     |     | F3  | S5  |     |     | T10 |     |     |
| R1  |     |     |     |     |     |     | M1  |     |     |     |     |     | R4  |     |     |     |     |     |     | I3  | Q4  |     |     | K2  |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     | L3  |     |     |     |     |     |     |     | L2  | F1  |     |     |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     | S2  |     |     |     |     |     |     |     | K1  | H1  |     |     |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     | C1  |     |     |     |     |     |     |     | Y1  |     |     |     |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | R1  |     |     |     |     |

**Non-responders : 50 clones**

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  | 21  | 22  | 23  | 24  | 25  | 26  | 27  |
| N11 | T49 | H17 | V38 | T34 | G50 | G50 | A23 | S22 | G33 | R25 | A13 | T38 | Y25 | G23 | F26 | A29 | S47 | L40 | F50 | S36 | V15 | G49 | A44 | S28 | Q50 | K50 |
| E10 | A1  | R11 | T11 | V10 |     |     | V11 | Q11 | A10 | H14 | N12 | A11 | L10 | R10 | L23 | T14 | A2  | F9  |     | A6  | S15 | E1  | P6  | A20 |     |     |
| Q10 |     | Y10 | E1  | A6  |     |     | S9  | E9  | S7  | Q10 | T12 | N1  | Q7  | A10 | R1  | S5  | F1  | H1  |     | N4  | P10 |     |     | Q1  |     |     |
| T10 |     | Q9  |     |     |     |     | T4  | A6  |     | S1  | K10 |     | S4  | S2  |     | V2  |     |     |     | R2  | M7  |     |     | H1  |     |     |
| G6  |     | V2  |     |     |     |     | E1  | M1  |     |     | D2  |     | K2  | K2  |     |     |     |     |     | D1  | F3  |     |     |     |     |     |
| D3  |     | F1  |     |     |     |     | M1  | V1  |     |     | G1  |     | N1  | T2  |     |     |     |     |     |     | L1  |     |     |     |     |     |
|     |     |     |     |     |     |     | P1  |     |     |     |     |     | R1  | E1  |     |     |     |     |     |     |     |     |     |     |     |     |

FIG. 2. Amino acid substitutions in the HVR1 domain in 50 variants obtained from pretreatment serum samples from subjects who responded to monotherapy (group R1) or combination therapy (group R2) and who failed to respond (group NR). The residues at each position (positions 1 to 27) are indicated with the single-letter amino acid code, together with the number of variants in which they were found. The most conserved residues are shaded.

ies, heterogeneity in the HVR1 domain has been found to be correlated with the duration of HCV infection (7), whereas other multivariate analyses failed to confirm these initial observations (12). SSCP analysis studies of the HVR1 domain have been extensively used to investigate whether patients whose isolates have multiple-band profiles have liver disease significantly more advanced than patients whose isolates have single-band profiles, generating conflicting results (21). Only

very limited information is available concerning the early evolution of viral quasispecies during the acute phase of hepatitis. HVR1 amino acid substitutions, which occur rapidly during the acute phase of infection, were assumed to help the virus escape from the immune system, thereby contributing to the development of persistent infection in many patients (41). Resolution of the hepatitis was found to be associated with the relative evolutionary stasis of the quasispecies (4).

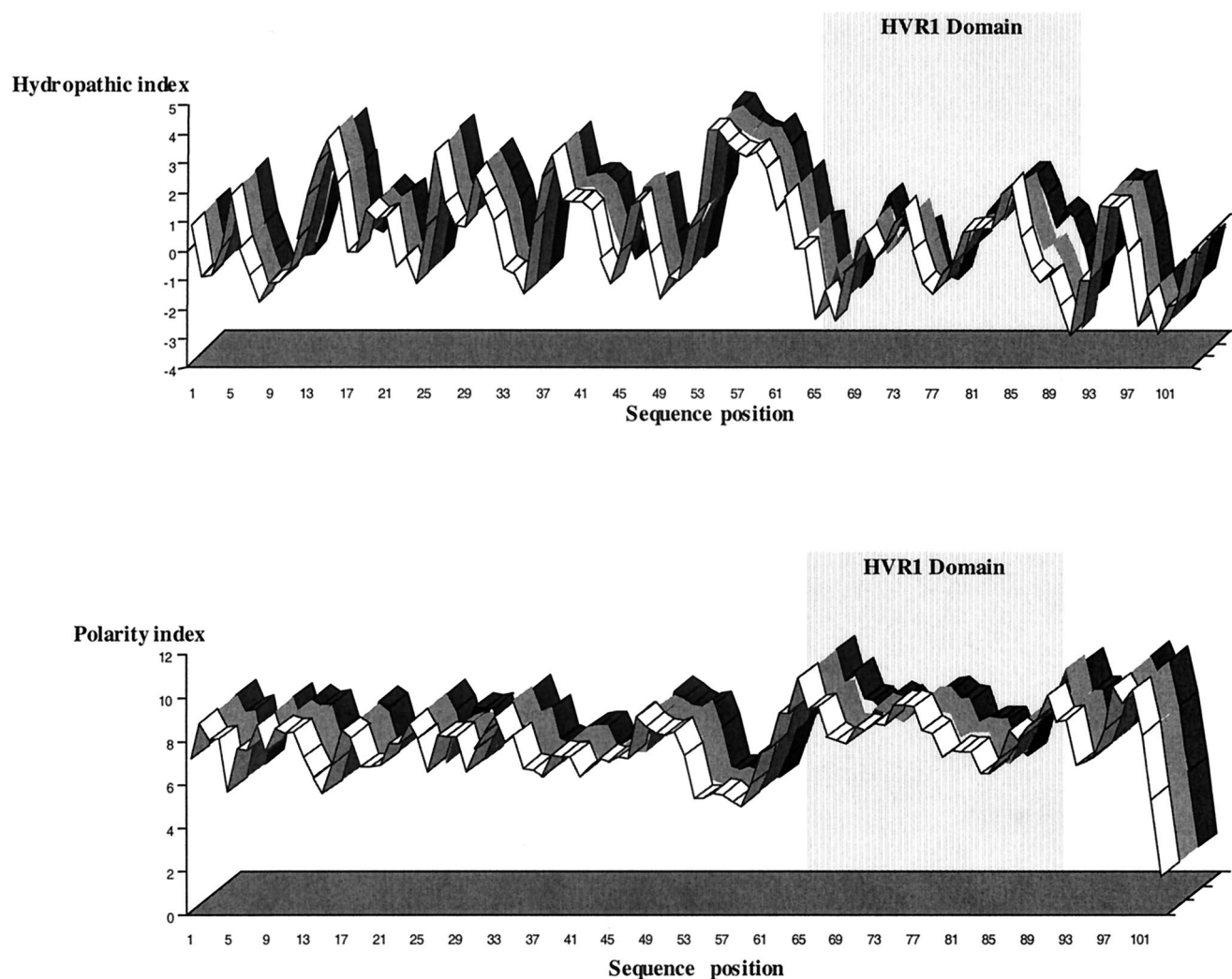


FIG. 3. Hydrophobicity and polarity plots for the HVR1 domain and its flanking regions (nt 955 to 1262) from pretreatment isolates of HCV-1b-infected patients. The average values (hydrophobic and polarity indices) were calculated at each position for the whole set of variants within a group by using DAMBE software. The plots for groups R1 and R2 are represented by a gray curve and a black curve, respectively, whereas the plots for group NR are represented by a white curve. The hypervariable region (residues 384 to 410) is indicated by vertical lines.

Much effort has been focused on evaluating the significance of quasispecies composition as a predictor of the response to IFN- $\alpha$  therapy in chronically infected patients. Several studies have suggested that patients with less complex viral populations before IFN- $\alpha$  therapy, as assessed by the number of HVR1 variants by SSCP analysis, are more likely to respond to treatment (11, 18), but conflicting results have also been published (19). Indeed, the results of many of these studies cannot be compared directly due to differences in the definitions of the biochemical and virological parameters used to evaluate therapeutic response, due to the use of methods that were not fully standardized, or due to the use of mixed populations infected with different viral genotypes. To avoid these problems, we only compared responders and nonresponders who were infected with the same HCV genotype (genotype 1b) and who were enrolled in a single controlled clinical trial.

We first attempted to identify specific mutations or conserved amino acids associated with the sensitivity or resistance of HCV-1b strains to IFN. Single amino acid substitutions in

this structurally flexible and antigenically variable domain may result in physicochemical and/or conformational changes. Single mutations can result in major changes in viral phenotype. For example, a single amino acid substitution within the V3 domain of gp120 of human immunodeficiency virus type 1 can alter the conformation of the loop, resulting in the loss of a neutralization epitope (2).

As described previously, we confirmed that HVR1 contains a central highly variable domain (residues 8 to 18), flanked by six conserved positions (positions 2, 6, 7, 20, 23, and 26) that are assumed to be important for the structure of the whole HVR1 domain (31). Amino acid substitutions at three other sites (positions 16, 19, and 24) did not affect hydrophobic or chemical properties. Overall, one-third of the amino acids in HVR1 appeared to retain properties that are highly conserved within a large pool of variants. Consistent with this, it could be argued that variability in HVR1 is not random, even though this region is tolerant to amino acid substitutions (37). As HVR1 may play a role in the entry of the virus through cell

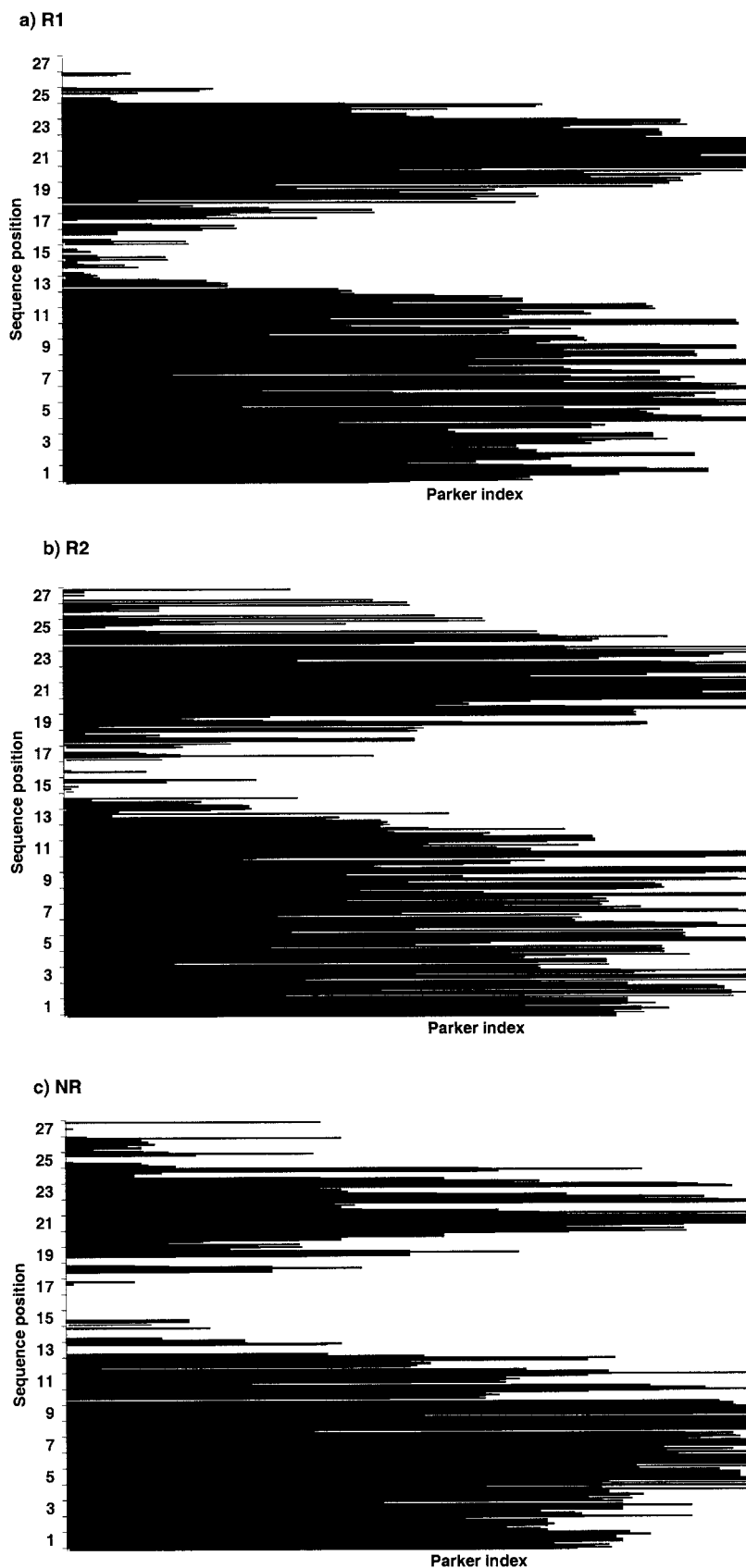


FIG. 4. Antigenicity histograms of the HVR1 domain from pretreatment isolates of HCV-1b-infected patients. The antigenicity at each residue for the 50 variants of each group (groups R1, R2, and NR) was predicted by the method of Parker et al. (26) with ANTHEPROT software.

attachment (28, 35), HVR1 variants must have limited conformational variability to ensure receptor recognition. Thus, although this constrained genomic variability suggests that HVR1 plays a role in HCV survival (10), the antigenic variations in E2, a major immunological target, may also enable HCV to adapt to the host environment. HVR1 was recently shown to be a decoy antigen that diverts the host immune response away from more conserved neutralizing epitopes (32).

Few groups have studied the structure of the whole HVR1 domain. Most studies focused on epitope mapping (39) or structural modeling (40) of HCV E2. The antigenicity profile of Parker et al. (26) has been used only once to study a large number of data bank sequences (28). The investigators predicted that HVR1 from HCV genotype 1 samples consisted of two antigenic domains separated by a less antigenic segment located between positions 14 and 18. An analysis of the antigenicity profiles of HVR1 sequences representative of the principal HCV subtypes of clades 1 to 6 provided the same results. The same group carried out a longitudinal study of HVR1 using serum samples from six patients who were infected with HCV genotype 1 but who did not clear HCV RNA after 6 months of treatment. The fact that the investigators did not look at specific clinical status or treatments in the population study and did not include any responders in the clinical study prevented them from drawing any conclusions about the differences in this basic structure between isolates from responders and nonresponders. Our study confirmed the existence of two antigenic regions in HVR1 separated by a non-antigenic segment of variable size. Presumably, variability in the two antigenic regions of E2 HVR1 accounts, at least in part, for the adaptive fitness of HCV to the host immune system. The prevalence of each amino acid at each position within HVR1 sequences was examined pretreatment, but no mutational patterns were found to be correlated with the treatment outcome. This suggests that simple sequencing of the HVR1 domain cannot be used to predict the response to treatment.

The study by Sandres et al. (34) with treatment-naive patients who were then treated with IFN- $\alpha$  alone suggested that differences in the dynamics of the host anti-HCV-specific immune response between responders and nonresponders (differences in the frequencies of nonsynonymous and synonymous substitutions) could be seen prior to treatment. However, the study population was heterogeneous (three different HCV genotypes were detected), and the physicochemical and immunological properties of the variants were not analyzed. We confirmed that HVR1 is subjected to strong selective pressure by the immune system, even before the onset of antiviral therapy. Two novel differences were also revealed in our homogeneous cohort of patients: the amino acid profiles at positions 17 and 19 differed between variants from responders and those from nonresponders, and major changes in the antigenicity profiles within the central part of HVR1 were detected. The variants in pretherapeutic serum samples from treatment-naive patients who were subsequently found to be nonresponders displayed a constant nonantigenic segment between positions 16 and 17. The exact functional meanings of these observations and their predictive values remain to be determined by further immunological studies of amino acid replacements in the central hypervariable region of HVR1,

especially the less antigenic sequence. Such studies could involve a panel of serum samples from responders and nonresponders and either peptides or HCV proteins produced after transfection of HCV cDNA (1).

#### ACKNOWLEDGMENT

This work was supported by a grant (ANRS grant 2001/016) from the Agence Nationale de Recherche sur le SIDA.

#### REFERENCES

- Blanchard, E., D. Brand, S. Trassard, A. Goudeau, and P. Roingard. 2002. Hepatitis C virus-like particle morphogenesis. *J. Virol.* **76**:4073–4079.
- Di Marzo Veronese, F., M. S. Reitz, G. Gupta, M. Robert-Gouffo, C. Boyer-Thompson, R. C. Gallo, and P. Lusso. 1993. Loss of a neutralizing epitope by a spontaneous point-mutation in the V3 loop of HIV-1 isolated from an infected laboratory worker. *J. Biol. Chem.* **268**:25894–25901.
- Duarte, E. A., I. S. Novella, S. C. Weaver, E. Domingo, S. Wain-Hobson, and D. K. Clarke. 1994. RNA virus quasispecies: significance for viral disease and epidemiology. *Infect. Agents Dis.* **3**:201–214.
- Farci, P., A. Shimoda, A. Coiana, G. Diaz, G. Peddis, J. C. Meldoper, A. Strazera, D. Y. Chien, S. J. Munoz, A. Balestrieri, R. H. Purcell, and H. J. Alter. 2000. The outcome of acute hepatitis C predicted by evolution of viral quasispecies. *Science* **288**:339–344.
- Farci, P., A. Shimoda, D. Wong, T. Cabezon, D. De Giovannis, A. Strazera, Y. Shimizu, M. Shapiro, H. J. Ater, and R. H. Purcell. 1996. Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. *Proc. Natl. Acad. Sci. USA* **93**:15394–15399.
- Forns, X., R. Thimme, S. Govindarajan, S. U. Emerson, R. H. Purcell, F. V. Chisari, and J. Bukh. 2000. Hepatitis C virus lacking the second envelope protein is infectious and causes resolving or persistent infection in chimpanzees. *Proc. Natl. Acad. Sci. USA* **97**:13318–13323.
- Gonzalez-Peralta, R. P., K. Qian, J. Y. She, G. L. Davis, T. Ohno, M. Mizokami, and J. Y. Lau. 1996. Clinical implications of viral quasispecies heterogeneity in chronic hepatitis C. *J. Med. Virol.* **49**:242–247.
- Grantham, R. 1974. Amino acid difference formula to help explain protein evolution. *Science* **185**:862–864.
- Higashi, Y., S. Kakuma, K. Yoshioka, T. Wakita, M. Mizokami, K. Ohba, Y. Ito, T. Ishikawa, M. Takayanagi, and Y. Nagai. 1993. Dynamics of genome change in the E2/NS1 region of hepatitis C virus *in vivo*. *Virology* **197**:659–668.
- Hino, K., M. Korenaga, E. Orito, Y. Katoh, Y. Yamaguchi, F. Ren, A. Kitase, Y. Satoh, D. Fujiwara, and K. Okita. 2002. Constrained genomic and conformational variability of the hypervariable region 1 of the hepatitis C virus in chronically infected patients. *J. Viral Hepat.* **9**:194–201.
- Hino, K., D. Yamaguchi, D. Fujiwara, Y. Katoh, M. Korenaga, M. Okazaki, M. Okuda, and K. Okita. 2000. Hepatitis C virus quasispecies and response to interferon therapy in patients with chronic hepatitis C: a prospective study. *J. Viral Hepat.* **7**:36–42.
- Koizumi, K., N. Enomoto, M. Kurosaki, T. Murakami, N. Izumi, F. Marumo, and C. Sato. 1995. Diversity of quasispecies in various disease stages of chronic hepatitis C virus infection and its significance in interferon treatment. *Hepatology* **22**:30–35.
- Kumar, U., J. Brown, J. Montjardino, and H. C. Thomas. 1993. Sequence variation in the large envelope glycoprotein (E2/NS1) of hepatitis C virus during chronic infection. *J. Infect. Dis.* **167**:726–730.
- Kurosaki, M., N. Enomoto, F. Marumo, and C. Sato. 1993. Rapid sequence variation of the hypervariable region of hepatitis C virus during the course of chronic infection. *Hepatology* **18**:1293–1299.
- Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**:105–132.
- Lechner, S., K. Rispeter, H. Meisel, W. Kraas, G. Jung, M. Roggendorf, and A. Zibert. 1998. Antibodies directed to envelope proteins of hepatitis C virus outside of hypervariable region 1. *Virology* **243**:313–321.
- Lee, J. H., T. Stripf, W. K. Roth, and S. Zeuzem. 1997. Non-isotopic detection of hepatitis C virus quasispecies by single-strand polymorphism. *J. Med. Virol.* **53**:245–251.
- Le Guen, B., G. Squadrito, B. Nalpas, P. Berthelot, S. Pol, and C. Brechot. 1997. Hepatitis C virus genome complexity correlates with response to interferon therapy: a study in French patients with chronic hepatitis C. *Hepatology* **25**:1250–1254.
- Lopez-Labrador, F. X., S. Ampurdanes, M. Gimenez-Barcons, M. Guilera, J. Costa, M. T. Jimenez de Anta, J. M. Sanchez-Tapias, J. Rodes, and J. C. Saiz. 1999. Relationship of genomic complexity of hepatitis C virus with liver disease severity and response to interferon in patients with chronic HCV genotype 1b infection. *Hepatology* **29**:897–903.
- Martinot-Peignoux, M., P. Marcellin, M. Pouteau, C. Castelnau, N. Boyer, M. Poliquin, C. Degott, I. Descombes, V. Le Breton, M. Milotova, J. P. Benhamou, and S. Erlinger. 1995. Pretreatment serum hepatitis C virus

- RNA levels and hepatitis C virus genotype are the main and independent prognostic factors of sustained response to interferon alpha therapy in chronic hepatitis C. *Hepatology* **22**:1050–1056.
21. Naito, M., N. Hayashi, T. Moribe, H. Hagiwara, E. Mita, Y. Kanazawa, A. Kasahara, H. Fusamoto, and T. Kamada. 1995. Hepatitis C viral quasispecies in hepatitis C virus carriers with normal liver enzymes and patients with type C chronic liver disease. *Hepatology* **22**:407–412.
  22. Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.
  23. Odeberg, J., Z. Yun, A. Sönnerborg, M. Uhlen, and J. Lundeberg. 1995. Dynamic analysis of heterogeneous hepatitis C virus populations by direct solid-phase sequencing. *J. Clin. Microbiol.* **33**:1870–1874.
  24. Ogata, N., H. J. Alter, R. H. Miller, and R. H. Purcell. 1991. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **88**:3392–3396.
  25. Okada, S. I., Y. Akahane, H. Suzuki, H. Okamoto, and S. Mishiro. 1992. The degree of variability in the amino terminal region of the E2/NS1 protein of the hepatitis C virus correlates with responsiveness to interferon therapy in viremic patients. *Hepatology* **16**:619–624.
  26. Parker, J. M., D. Guo, and R. S. Hodges. 1986. New hydrophobicity scale derived from high-performance liquid chromatography peptide retention data: correlation of the predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* **25**:5425–5432.
  27. Pawlotsky, J. M., G. Germanidis, P. O. Frainais, M. Bouvier, A. Soulier, M. Pellerin, and D. Dhumeaux. 1999. Evolution of the hepatitis C virus second envelope protein hypervariable region in chronically infected patients receiving alpha interferon therapy. *J. Virol.* **73**:6490–6499.
  28. Penin, F., C. Combet, G. Germanidis, P. O. Frainais, G. Deléage, and J. M. Pawlotsky. 2001. Conservation of the conformation and positive charges of hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment. *J. Virol.* **75**:5703–5710.
  29. Polyak, S. J., S. McArdle, S. L. Liu, D. G. Sullivan, M. Chung, W. T. Hofgärtner, R. L. Carithers, Jr., B. J. McMahon, J. I. Mullins, L. Corey, and D. R. Gretch. 1998. Evolution of the hepatitis C virus quasispecies in the hypervariable region 1 and the putative interferon sensitivity-determining region during interferon therapy and natural infection. *J. Virol.* **72**:4288–4296.
  30. Poynard, T., P. Marcellin, S. S. Lee, C. Niederau, G. S. Minuk, G. Ideo, V. Bain, J. Heathcote, S. Zeuzem, C. Trepo, and J. Albrecht. 1998. Randomised trial of interferon alpha2b plus ribavirin for 48 weeks or for 24 weeks versus interferon alpha2b plus placebo for 48 weeks for treatment of chronic infection with hepatitis C virus. *Lancet* **352**:1426–1432.
  31. Punterio, G., A. Meola, A. Lahm, S. Zuchelli, B. Bruni Ercole, R. Tafi, M. Pezzarena, M. U. Mondelli, R. Cortese, A. Tramontano, G. Galfre, and A. Nicosia. 1998. Towards a solution for hepatitis C virus hypervariability: mimotopes of the hypervariable region 1 can induce antibodies cross-reacting with a large number of viral variants. *EMBO J.* **17**:3521–3533.
  32. Ray, C., Y. M. Wang, O. Laeyendecker, J. R. Ticehurst, S. A. Villano, and D. L. Thomas. 1999. Acute hepatitis C virus structural gene sequences as predictors of persistent viremia: hypervariable region 1 as a decoy. *J. Virol.* **73**:2938–2946.
  33. Sakuma, I., N. Enomoto, M. Kurosaki, F. Marumo, and C. Sato. 1996. Selection of hepatitis C virus quasispecies during interferon treatment. *Arch. Virol.* **141**:1921–1932.
  34. Sandres, K., M. Dubois, C. Pasquier, J. L. Payen, L. Alric, M. Duffault, J. P. Vinel, J. P. Pascal, J. Puel, and J. Izopet. 2000. Genetic heterogeneity of the hypervariable region 1 of the hepatitis C virus genome and sensitivity of HCV to alpha interferon therapy. *J. Virol.* **74**:661–668.
  35. Scarselli, E., H. Ansuini, R. Cerino, R. M. Roccasecca, S. Acali, G. Filocamo, C. Traboni, A. Nicosia, R. Cortese, and A. Vitelli. 2002. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J.* **21**:5017–5025.
  36. Smith, D. B., J. McAllister, C. Casino, and P. Simmonds. 1997. Virus quasispecies: making a mountain out of a molehill? *J. Gen. Virol.* **78**:1511–1519.
  37. Sobolev, B. N., V. V. Poroikov, L. V. Olenina, E. F. Kolesanova, and A. I. Archakov. 2000. Comparative analysis of amino acid sequences from envelope proteins isolated from different hepatitis C virus variants: possible role of conservative and variable regions. *J. Viral Hepat.* **7**:368–374.
  38. Toyoda, H., T. Kumada, S. Nakano, I. Takeda, K. Sugiyama, T. Osada, S. Kiriya, Y. Sone, M. Kinoshita, and T. Hamada. 1997. Quasispecies nature of hepatitis C virus and response to alpha interferon: significance as a predictor of direct response to interferon. *J. Hepatol.* **26**:6–13.
  39. Weiner, A. J., H. M. Geysen, C. Christopherson, J. E. Hall, T. J. Mason, G. Saracco, F. Bonino, K. Crawford, C. D. Marion, K. A. Crawford, M. Brunetto, P. J. Barr, T. Miyamura, J. McHutchinson, and M. Houghton. 1992. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc. Natl. Acad. Sci. USA* **89**:3468–3472.
  40. Yagnik, A. T., A. Lahm, A. Meola, R. M. Roccasecca, B. B. Ercole, and A. Nicosia. 2000. A model for the hepatitis C virus envelope glycoprotein E2. *Proteins* **40**:355–366.
  41. Yamaguchi, K., E. Tanaka, K. Higashi, K. Kiyosawa, A. Matsumoto, S. Furuta, A. Hasegawa, S. Tanaka, and M. Kohara. 1994. Adaptation of hepatitis C virus for persistent infection in patients with acute hepatitis. *Gastroenterology* **106**:1344–1348.
  42. Yeh, B. I., K. H. Han, S. H. Oh, H. S. Kim, S. H. Hong, S. H. Oh, and Y. S. Kim. 1996. Nucleotide sequence variation in the hypervariable region of the hepatitis C virus in the sera of chronic hepatitis C patients undergoing controlled interferon- $\alpha$  therapy. *J. Med. Virol.* **49**:95–102.
  43. Yun, Z. B., J. Odeberg, J. Lundeberg, O. Weiland, M. Uhlen, and A. Sönnerborg. 1996. Restriction of hepatitis C virus heterogeneity during prolonged interferon- $\alpha$  therapy in relation to changes in viral load. *J. Infect. Dis.* **173**:992–996.
  44. Zeuzem, S., V. Feinman, J. Rasenack, J. Heathcote, M. Lai, E. Gane, J. O'Grady, J. Reichen, M. Diago, A. Lin, J. Hoffman, and M. J. Brunda. 2000. Peginterferon alpha-2a in patients with chronic hepatitis C. *N. Engl. J. Med.* **343**:1666–1672.
  45. Zhou, Y. H., Y. K. Shimizu, and M. Esumi. 2000. Monoclonal antibodies to the hypervariable region 1 of the hepatitis C virus capture virus and inhibit virus adsorption to susceptible cells *in vitro*. *Virology* **269**:276–283.