Available online at www.sciencedirect.com
ScienceDirect
Virology 352 (2006) 357–367
VIROLOGY
www.elsevier.com/locate/yviro

Study of a novel hypervariable region in hepatitis C virus (HCV) E2 envelope glycoprotein

Myriam Troesch [a,d], Isabelle Meunier [a,d,1], Pascal Lapierre [b,e], Normand Lapointe [c,f], Fernando Alvarez [b,f], Marc Boucher [c,g], Hugo Soudeyns [a,d,f,*]

[a] *Unité d'immunopathologie virale, Centre de recherche du CHU mère-enfant Sainte-Justine, Montreal, Quebec, Canada H3T 1C5*
[b] *Service de gastroentérologie, hépatologie et nutrition, Centre de recherche du CHU mère-enfant Sainte-Justine, Montreal, Quebec, Canada H3T 1C5*
[c] *Centre maternel et infantile sur le SIDA, Centre de recherche du CHU mère-enfant Sainte-Justine, Montreal, Quebec, Canada H3T 1C5*
[d] *Department of Microbiology and Immunology, Université de Montréal, Montreal, Quebec, Canada H3C 3J7*
[e] *Program of Biomedical Sciences, Université de Montréal, Montreal, Quebec, Canada H3C 3J7*
[f] *Department of Pediatrics, Université de Montréal, Montreal, Quebec, Canada H3C 3J7*
[g] *Department of Gynecology and Obstetrics, Université de Montréal, Montreal, Quebec, Canada H3C 3J7*

Received 3 February 2006; returned to author for revision 30 April 2006; accepted 10 May 2006
Available online 16 June 2006

Abstract

A large share of hepatitis C virus amino acid sequence variation is concentrated within two hypervariable regions located at the N-terminus of the E2 envelope glycoprotein (HVR1 and HVR2). Interhost and intrahost comparison of 391 E2 sequences derived from 17 subjects infected with HCV using amino acid entropy revealed clustering of amino acid variability at a third site (residues 431–466), which was termed HVR3. Genetic distance analysis supported the division of HVR3 into three subdomains (HVR3a, HVR3b, and HVR3c). Study of synonymous and nonsynonymous nucleic acid substitutions confirmed that HVR3a was subjected to strong intrahost-selective pressure. Physicochemical and antigenicity predictions, conservation of key residues, and molecular modeling were concordant with one another and further validated the proposed organization of HVR3. Taken together, these results are suggestive of a role for HVR3 in cell surface receptor binding and viral entry akin to that proposed for HVR1 and HVR2.

Keywords: Hepatitis C virus; Quasispecies; Envelope; Positive selection; Molecular modeling

Introduction

Hepatitis C virus (HCV) is an enveloped RNA virus belonging to the Flaviviridae family (Choo et al., 1989, 1991). Despite the fact that some infected individuals are able to clear the virus, more than 80% become chronically infected and may develop liver cirrhosis and hepatocellular carcinoma (Saito et al., 1990; Alter et al., 1992; Bach et al., 1992). Hepatitis C is a leading cause of liver transplantation and represents a continuing public health concern. There is currently no vaccine to prevent HCV infection.

As a result of common modes of transmission, the global prevalence of HCV infection among human immunodeficiency virus type 1 (HIV-1)-infected subjects ranges between 30% and 50%, with rates of coinfection as high as 90% in injection drug users and almost 100% in patients with hemophilia (Dieterich, 1999; Dodig and Tavill, 2001). Several studies have shown that coinfection with HCV and HIV-1 adversely affects liver fibrosis, HCV viral load, and progression of HCV disease (Soriano et al., 1999; Bonacini and Puoti, 2000; Benhamou et al., 2001).

The high genetic variability of HCV contributes to the chronicity of hepatitis C. Based on nucleotide sequence analysis, HCV was classified into 6 genotypes and a series of subtypes (Simmonds et al., 1993, 2005). Moreover, in individual subjects,

* Corresponding author. Unité d'immunopathologie virale, Centre de recherche du CHU mère-enfant Sainte-Justine, 3175 Côte Sainte-Catherine, room 6735, Montreal, Quebec, Canada H3T 1C5. Fax: +1 514 345 4794.
E-mail address: hugo.soudeyns@recherche-ste-justine.qc.ca (H. Soudeyns).
[1] Present address: INRS-Institut Armand-Frappier, Laval, Quebec, Canada H7V 1B7.

0042-6822/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2006.05.015

HCV is present as a dynamic population of distinct but closely related variants designated quasispecies (Martell et al., 1992; Bukh et al., 1995). HCV quasispecies evolution is influenced mainly by (a) the high production rate of viral particles; (b) the lack of error correction mechanisms by viral RNA-dependent RNA polymerase; and (c) selective pressure exerted by host immune responses. The HCV E2 viral envelope glycoprotein is a preferred target for humoral (Weiner et al., 1992; Kato et al., 1993) and cell-mediated immune responses (Shirai et al., 1999). Not surprisingly, a large share of HCV sequence variation is concentrated within the hypervariable regions of E2, including hypervariable region 1 (HVR1), a sequence of 27 amino acids located at the N-terminus of the protein (Weiner et al., 1991; Kato et al., 1992). The conservation of overall conformation and positively charged amino acid residues at specific positions of HVR1 are consistent with a role in target cell recognition and virus attachment (Penin et al., 2001). HVR1 exhibits the highest degree of intersequence variability in the HCV genome and constitute a practical model for the study of host-selective pressure and quasispeciation; that is, the process of viral genetic diversification leading to quasispecies formation following clonal infection (Sala and Wain-Hobson, 2000). Quasispecies dynamics based on HVR1 are indicative of outcomes such as spontaneous viral clearance (Farci et al., 2000; Chen and Wang, 2005), response to interferon treatment (Farci et al., 2002; Gaudy et al., 2003; Abbate et al., 2004), and HCV-associated liver histopathology (Honda et al., 1994; Gonzalez-Peralta et al., 1996). The second hypervariable region of E2, HVR2, consists of 9 amino acids located downstream of HVR1 (Kato et al., 1992). In this case, structure predictions are consistent with a potential involvement in cell surface receptor binding (Yagnik et al., 2000). Here, we report on the identification of an additional highly variable region in HCV glycoprotein E2, which we termed HVR3. Quasispecies diversity was studied in order to determine whether HVR3 was potentially subjected to selective pressure exerted by host HCV specific immunity, and the physicochemical and structural characteristics of this region were evaluated to gain insights into its possible functional roles.

Results

Identification and characterization of HVR3

HCV E1 and E2 envelope gene sequences were amplified starting with serial serum samples (pregestational; 1st, 2nd, 3rd trimesters; postpartum) obtained from 17 pregnant women infected with HCV, including 13 who were coinfected with HCV and HIV-1. Four women were studied throughout two conscutive pregnancies. Nine patients were infected with HCV genotype 1a, 5 with HCV genotype 1b, and 3 with HCV genotype 3a. A total of 1337 nucleic acid sequences (570–576 nucleotides) were obtained, including 1101 that were nonredundant, representing a mean of 78.7 sequences per patient (range = 52–147). Nucleotide sequences were edited and translated, spanning amino acid positions 319–508 of the HCV-1a reference sequence (GenBank accession no. M62321), and corresponding to the 65 C-terminal amino acid residues of E1 and the 125 N-terminal amino acid residues of E2 (E1/128–E2/125).

First, for the purposes of estimating amino acid sequence diversity, a data set comprised of all 391 E1–E2 sequences derived from third trimester samples from all patients was assembled, of which 331 (84.7%) were nonredundant at the nucleic acid level, representing a mean of 23 sequences per patient (range = 17–38). Hence, this data set simultaneously sampled roughly equivalent contributions from both intrahost and interhost E1–E2 amino acid sequence diversity. Amino acid sequences were aligned, and the entropy or variability at each position of the alignment was determined (Fig. 1). The entropy

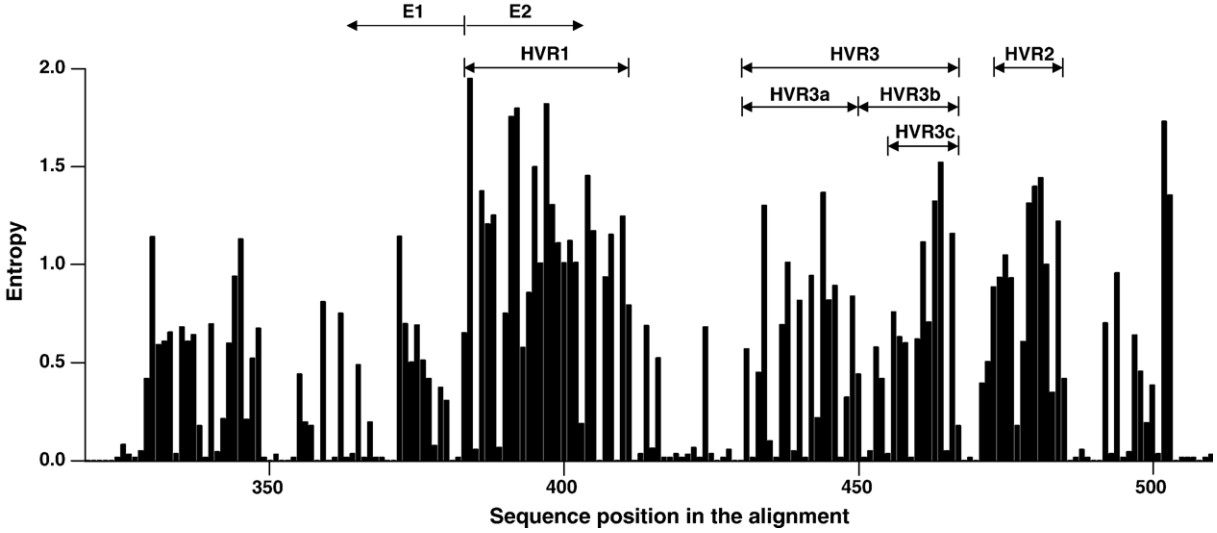

Fig. 1. Amino acid sequence diversity in HCV E1–E2 envelope glycoproteins. Variability at each amino acid position was computed using a data set comprised of 391 E1–E2 sequences derived from third trimester samples from all study subjects and the Entropy-ONE Web tool (Korber et al., 1994), as described under Materials and methods. Amino acid positions correspond to the M62321 reference sequence (Choo et al., 1989). Boundaries of E1, E2, HVR1, and HVR2 were set in accordance with previous reports (Weiner et al., 1991; Kato et al., 1992).

profile revealed that amino acid variability was not uniformly distributed across the amplified E1–E2 segments, with visually discernible clusters of amino acid sequence diversity. The first cluster (amino acid positions 384–410) was well defined and corresponded to the 27 amino acids of HVR1 (Weiner et al., 1991), whereas a second cluster, further downstream (amino acid positions 474–482), corresponded to HVR2 (Kato et al., 1992). Interestingly, distribution of amino acid entropy in the region between HVR1 and HVR2 (amino acid positions 411–473) was clearly discontinuous, and a third cluster of amino acid variability was readily delineated, spanning amino acid positions 431–466, approximately halfway between HVR1 and HVR2 (Fig. 1). For the purpose of the present study, we termed this region HVR3. HVR3 was empirically subdivided into two regions of roughly equal size, which were termed HVR3a (amino acid positions 431–449) and HVR3b (amino acid positions 450–466). Additional sequence variation was observed at the N terminus and C terminus of the amplified segment (Fig. 1). Analysis was focused on HVR3 because (a) the first 65 N-terminal amino acids are in fact localized within the C-terminal portion of the E1 envelope protein; and (b) amino acids at the C terminus are localized at the extreme end of our sequence reads, where base miscalls are more prevalent. The overall entropy profile was not affected by removal from the alignment of all HCV-3a sequences (n = 58; subjects TV453, TVC45, and TVC73) and/or all sequences from patient TV531 (n = 17), who respectively exhibited a 1 and 2 amino acid insertion between positions 476 and 477 (data not shown). Likewise, comparable amino acid entropy profiles were obtained when analysis was performed with (a) additional subsamples from the complete 1337 sequence data set (Troesch et al., unpublished); or (b) a set of 187 full-length genomic HCV sequences comprising all 6 major HCV subtypes (http://hcv.lanl.gov/content/hcv-db/; Kuiken et al., 2005) (Supplemental Fig. 1; data not shown).

To test relative amino acid sequence variability, genetic distances were estimated within each of these domains and subdomains (Table 1). As expected, highest variability was observed at the positions corresponding to HVR1 and HVR2. Mean genetic distance was lower in HVR3 (431–466) than in HVR1 or HVR2 but higher than that obtained when the complete E1/128–E2/125 segment was analyzed (Table 1). Within HVR3, the HVR3b subdomain was comparatively more variable than HVR3a or than the whole of HVR3. Additional truncation of the HVR3b subdomain into HVR3c (amino acid positions 456–466) circumscribed a region of high sequence diversity, with a mean genetic distance of 0.430 ± 0.064 (Fig. 1). These values represented 84.5% and 79.5% of the mean genetic distances measured within HVR1 and HVR2, respectively, and 170% of the mean variability observed in the complete E1/128–E2/125 region (Table 1). Again, none of these estimates was adversely affected by exclusion of all HCV-3a sequences and/or sequences from patient TV531 (see above), and all were comparable to those obtained using a set of 187 full-length genomic HCV sequences retrieved from the Los Alamos database (Kuiken et al., 2005) (Supplemental Table 1; data not shown).

Table 1
Amino acid sequence variability in selected subregions of HCV E1 and E2 envelope glycoproteins

Sequence identification	Amino acid position [a]	Mean genetic distance [b]	Standard error [c]
E1/128–E2/125	319–508	0.253	0.020
HVR1	384–410	0.509	0.049
HVR2	474–482	0.541	0.044
HVR3	431–466	0.315	0.042
HVR3a	431–449	0.305	0.056
HVR3b	450–466	0.326	0.061
HVR3c	456–466	0.430	0.064

Nucleic acid sequences were obtained, edited, translated, and analyzed as described under Materials and methods.

[a] Amino acid positions correspond to the M62321 reference sequence (Choo et al., 1989).

[b] Mean genetic distance was calculated as mean of pairwise p distances.

[c] Standard error of the mean was computed by bootstrapping (500 replicates) (Kumar et al., 2001).

Intrahost selective pressure

Next, the presence and extent of intrahost selective pressure exerted on HCV E2 sequences were examined within gene segments encoding the HVR1, HVR2, HVR3, HVR3a, HVR3b, and HVR3c subdomains, which showed the highest amino acid variability when initial interhost comparisons were performed (Table 1). To do this, we assembled data sets comprised of nucleic acid sequences derived from longitudinal samples obtained from each patient throughout pregnancy (pregestational; 1st, 2nd, 3rd trimesters; postpartum). The ratio of the mean number of nonsynonymous substitutions per nonsynonymous site over the mean number of synonymous substitutions per synonymous site (dN/dS) was then computed within the above-mentioned E1–E2 subdomains, with a dN/dS ratio >1 taken as an index of positive selection for specific amino acid residues (Nei and Gojobori, 1986). As HVR3 and HVR3c represented broadly overlapping and/or nested subregions, comparisons of dN/dS ratios were not drawn that involved either of these regions. Finally, it should be noted that the Los Alamos data set essentially samples interhost not intrahost variability (i.e., all of these sequences are presumably unrelated). The extent of intrahost-selective pressure can therefore not be computed using such a data set.

Overall, statistically significant differences in median dN/dS ratio were observed between subregions in multiple group analysis (P = 0.0002, Kruskal–Wallis test). As expected, median dN/dS ratio for HVR1 (2.391; interquartile range (IQR) = 0.816–3.092) was significantly higher than that observed in HVR2 (0.227; IQR = 0.089–0.858), HVR3b (0.204; IQR = 0.129–0.330), but not HVR3a (0.719; IQR = 0.280–1.120) ($P < 0.01$, $P < 0.001$, and $P > 0.05$, respectively, Dunn's multiple comparison test) (Table 2). Thirteen of 17 HCV-infected subjects tested (76.5%) had a dN/dS ratio >1 in HVR1, as compared with 6 of 17 (35.3%) in HVR3a, 2 of 17 (11.8%) in HVR2, and 1 of 17 (5.88%) in HVR3b (Table 2). In addition, 10 of 17 HCV-infected subjects (58.8%) exhibited statistically significant positive selection in HVR1 in analysis of variance ($P < 0.10$, Z test), as compared with 2 of 17 (11.8%) in HVR3a, 1 of 17 (5.88%) in HVR2, and 0 of 17 (0%) in HVR3b (Table 2). These results are

Table 2
Estimate of selective pressure exerted on various subregions of the HCV E2 envelope glycoprotein

Subject	HIV coinfection	HCV genotype	d*N*/d*S*					
			HVR1 [a] (384–410)	HVR2 (474–482)	HVR3 (431–466)	HVR3a (431–449)	HVR3b (450–466)	HVR3c (456–466)
TVC57	−	1a	2.614 [b, c]	0.090	0.812	1.177 [b]	0.424	2.812 [b, c]
TVC67	−	1a	1.482 [b]	0.163	0.607	0.634	0.498	0.246
TVC75	−	1b	1.181 [b]	0.227	0.282	0.395	0.163	0.151
TVC79	−	1a	3.826 [b, c]	0.000	0.513	1.542 [b]	0.174	0.119
TV075 [d]	+	1a	0.081	0.024	0.110	0.041	0.340	0.766
TV179 [d]	+	1b	0.136	7.317 [b, c]	0.202	0.164	0.204	0.163
TV233 [d]	+	1b	2.743 [b, c]	0.175	0.685	3.825 [b, c]	0.096	0.180
TV289	+	1b	2.391 [b, c]	nc	0.461	0.431	1.549 [b]	1.283 [b]
TV453	+	3a	0.428	0.166	0.037	0.023	0.052	0.037
TV519	+	1b	4.263 [b, c]	nc	0.552	1.062 [b]	0.044	0.204
TV531	+	1a	1.860 [b, c]	0.253	0.391	0.955	0.233	0.521
TVC17	+	1a	3.975 [b, c]	2.191 [b]	1.182 [b]	4.674 [b, c]	0.320	1.592 [b]
TVC33	+	1a	3.054 [b, c]	0.515	0.558	1.036 [b]	0.289	0.210
TVC45	+	3a	1.040 [b]	0.858	0.626	0.904	0.227	0.345
TVC55 [d]	+	1a	2.724 [b, c]	0.965	0.363	0.693	0.173	0.283
TVC59	+	1a	0.592	0.248	0.165	0.139	0.185	0.220
TVC73	+	3a	3.131 [b, c]	0.065	0.383	0.719	0.088	0.314

Nucleic acid sequencing, HCV genotyping, and calculation of the d*N*/d*S* ratio were performed as described under Materials and methods.
[a] Amino acid positions correspond to the M62321 reference sequence (Choo et al., 1989).
[b] Denotes d*N* > d*S*.
[c] Denotes significant positive selection as determined by analysis of variance ($P < 0.10$, Z test) (Kumar et al., 2001). nc: not computed due to absence of synonymous substitutions in that particular sequence subset.
[d] Analysis performed on two consecutive pregnancies.

consistent with the established role of HVR1 as one of the principal foci of intrahost immune-selective pressure (Weiner et al., 1992; Kato et al., 1993; Farci et al., 2000, 2002; Canobio et al., 2004). These results also indicate that HVR3a (residues 431–449) was the second most highly selected subdomain after HVR1 within the E2 region examined, and HVR3b was the least selected. Overall, this observed hierarchy in positive subdomain selection provides experimental support for the amino acid entropy and genetic distance-based clustering approach initially used to define HVR3, HVR3a, HVR3b, and HVR3c (Fig. 1; Table 1; Supplemental Fig. 1; Supplemental Table 1). Interestingly, linear regression analysis revealed a statistically significant relationship between the d*N*/d*S* ratios computed in HVR1 and HVR3a ($r^2 = 0.360$, $P = 0.0105$). However, there were no such significant relationships observed that involved either HVR2 or HVR3b. These results suggest that the selective forces that shape the amino acid sequences of HVR1 and HVR3a act in a similar and/or synchronous manner on both subregions, whereas genetic diversification in HVR2 and HVR3b would appear to be driven by distinct selective forces.

Influence of HCV subtype and coinfection with HIV-1

Three subjects in our study group (TV453, TVC45, and TCV73) were infected with HCV-3a, a subtype that not only displays marked amino acid sequence dissimilarity as compared with HCV-1a and HCV-1b, but also exhibits variant biological properties in terms of pathogenesis and resistance to interferon treatment (Pawlotsky, 2000; Rubbia-Brandt et al., 2001). For these reasons, intrahost-selective pressure was examined separately in these subjects. Median d*N*/d*S* ratio was lower in patients infected with HCV-3a than HCV-1a or HCV-1b for all subdomains analyzed (Table 2). Within HVR1, 2 of 3 subjects (66.6%) with HCV-3a infection displayed a d*N*/d*S* ratio >1, and 1 of 3 (33.3%) exhibited significant positive selection as determined by analysis of variance (Table 2). These proportions were similar to those observed in subjects infected with HCV-1a or HCV-1b with respect to HVR1 and all other E2 subdomains, but the statistical significance of these observations was not tested due to the limited number of HCV-3a-infected subjects enrolled in our study group.

In addition, a number of studies have reported that infection with human immunodeficiency virus type 1 (HIV-1) influenced HCV quasispecies distribution and reduced host-selective pressure exerted on HVR1 in coinfected subjects (Sherman et al., 1996; Mao et al., 2001; Canobio et al., 2004). Because of this, intrahost-selective pressure on HVR1, HVR2, HVR3a, and HVR3b was compared between HCV/HIV-1 coinfected patients ($n = 14$) and patients infected solely with HCV ($n = 4$). Overall, the median d*N*/d*S* ratio was not significantly higher in subjects infected only with HCV than in coinfected patients ($P > 0.117$, Mann–Whitney U test), and there was no significant difference in the proportion of coinfected subjects exhibiting either a d*N*/d*S* ratio >1 or significant positive selection for any of the E2 subdomains analyzed ($P > 0.235$, Fisher's exact test).

Physicochemical properties and predicted antigenicity of HVR3

Because HVR3, and in particular HVR3a, stood out as HCV E2 envelope glycoprotein subregions subjected to relatively high intrahost selective pressure, details of their physicochemical

properties and predicted antigenicity profiles were further investigated. This analysis was based on an HVR3 consensus sequence of 36 amino acid residues (DSLNTGFLAGLFYYHKFNSSGCPERLASCRPITDFA) derived from the 391 third trimester E1–E2 sequences (see above and Table 1). According to the Kyte and Doolittle (1982) method, HVR3 displays a rather hydrophilic central region located within the C-terminal portion of HVR3a, whereas its flanks (i.e., N-terminal HVR3a, HVR3b) are generally hydrophobic (Fig. 2A). The Parker index, which integrates hydrophilicity, atomic flexibility, and HPLC retention components, remains one of the most extensively validated in silico predictors of protein antigenicity (Parker et al., 1986). Computing of this index over the HVR3 consensus sequence revealed two salient peaks of antigenicity, both located within the C-terminal part of HVR3a and both overlapping regions of relatively high hydrophilicity highlighted in the Kyte–Doolittle plot (Figs. 2A and B). These regions also overlapped with segments predicted to be highly surface-exposed by the Janin accessibility method (Fig. 2C) (Janin, 1979). Overall, combination of these prediction algorithms suggests that the main antigenic portion of HVR3 lies within HVR3a, and positions HVR3a at the surface of E2, where it should be accessible and recognizable by antibodies. These observations further validate the HVR3a, HVR3b, and HVR3c subdomain organization initially proposed on the basis of amino acid sequence entropy and genetic distance.

Detailed examination of the amino acid sequence composition of HVR3a revealed a domain comprised of relatively conserved hydrophobic residues surrounded by relatively hydrophilic or neutral amino acids (Fig. 3). Thus, in spite of the relative amino acid sequence variability observed within this region, its overall hydropathic character appears highly conserved, both in interhost and intrahost comparisons (Fig. 3). In particular, amino acid positions 433, 437, 438, 439, 441, 442, 443, and 447 were almost exclusively held by hydrophobic residues. Positions 432, 435, and 436 were occupied by neutral residues, with a fully invariant glycine at position 436. The remaining positions were held by hydrophilic (431, 434, 445, 446, and 448) or variable amino acids (440, 444, and 449) (Fig. 3). Overall, positions of highest amino acid sequence variability within HVR3a overlapped those of predicted antigenic sites (Fig. 2B). Positively charged residues were mainly found C-terminal from the hydrophobic center (positions 445 and 446 in 81.0% and 93.3% of sequences, respectively) and to a lesser extent at position 444 (7.44%). Position 434 exhibited both positively and negatively charged residues, with a preference for basic amino acids (28.0% versus 10.0%). Interestingly, the first position of HVR3a (residue 431) was almost always occupied by a basic residue. Fully consistent profiles were observed when analysis was performed using a set of 187 full-length genomic HCV sequences from the Los Alamos HCV sequence database (data not shown) (Kuiken et al., 2005). The conservation of hydrophobic amino acids and charged residues suggests that the overall conformation of HVR3a is likewise conserved among HCV variants comprised within the quasispecies.

Molecular modeling of HVR3

The tick-borne encephalitis virus (TBEV) E protein, which is distantly related to HCV envelope glycoproteins, remains the closest available crystal structure amenable to three-dimensional modeling of E2 (Rey et al., 1995). Work using this model has led to the formulation of several testable hypotheses regarding E1–E2 structure, higher order organization, immunogenicity, and function (Yagnik et al., 2000; Slater-Handshy et al., 2004; Heo et al., 2004). This model was reconstructed by incorporating the consensus amino acid sequence of HVR3 derived from alignment of the 391 third trimester E1–E2 sequences described

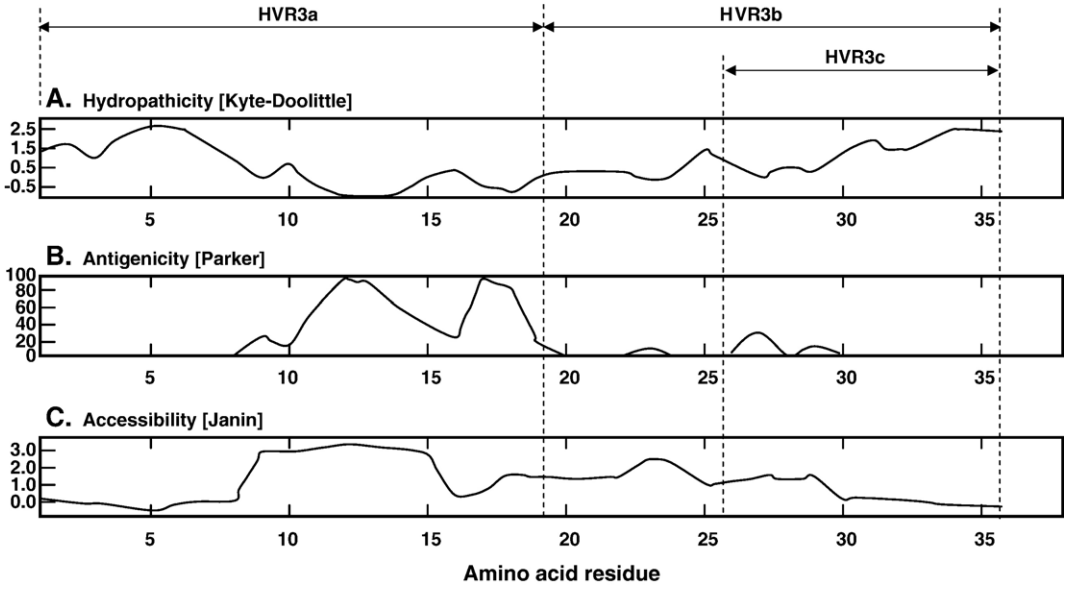

Fig. 2. Physicochemical properties and antigenic characterization of HVR3. Physicochemical profiles of the consensus HVR3 amino acid sequence derived from third trimester samples were computed using the NPS@ webware (http://npsa-pbil.ibcp.fr). (A) Hydropathicity plot (Kyte and Doolittle, 1982). (B) Antigenicity index (Parker et al., 1986). (C) Accessibility profile (Janin, 1979).

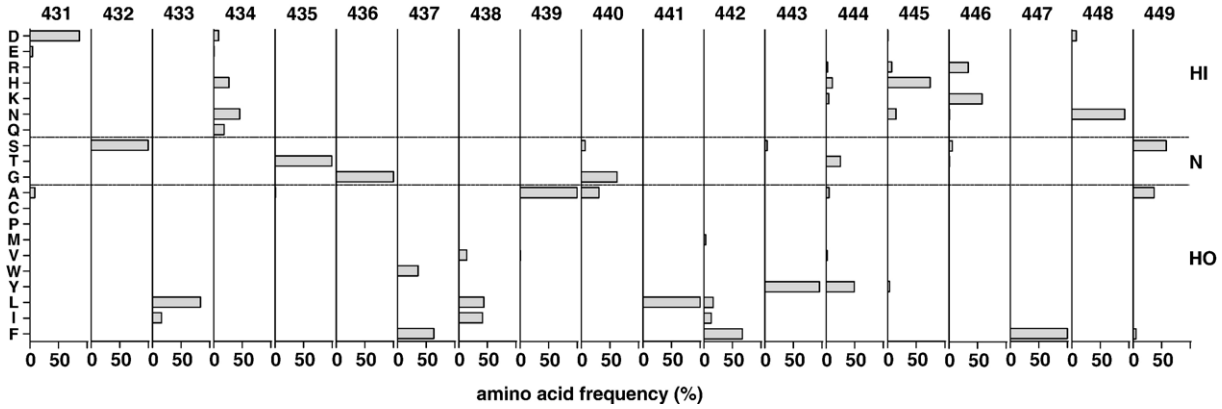

Fig. 3. Amino acid usage in HVR3a. The range of amino acid residues that occupy each position of HVR3a was determined from a data set comprised of 391 E1–E2 sequences derived from third trimester samples from all study subjects. HI: hydrophilic; N: neutral; HO: hydrophobic. Amino acid positions correspond to the M62321 reference sequence (Choo et al., 1989).

above. Confirming antigenicity and accessibility predictions, HVR3 was found to be very largely exposed at the surface of E2, with a potential contact surface of approximately 800Å^2 and 50% of HVR3 amino acids having more than 30% of their total surface exposed (Fig. 4). In fact, the only amino acid residues that appear to be entirely buried within the structure are phenylalanine 437 and leucine 438 (0% exposure), which are primarily hidden underneath cysteine 503 and glycine 504, whereas alanine 439 and glycine 440 have <5% exposure. In this three-dimensional model, HVR3 lies distal to the proposed dimerization interface (Rey et al., 1995). HVR3 is also strikingly distant from HVR1, which is situated at the opposite end of E2 and largely positioned on the other face of the protein. This structural arrangement would make direct interactions between these two domains improbable, both in the context of monomeric and dimeric envelope structures. In contrast, and most interestingly, HVR3c is in close spatial proximity with HVR2: the apex of HVR2 (GSG tripeptide) and the center of HVR3c (RPL tripeptide) are separated by an estimated distance of 27Å, whereas the closest point-to-point distance between HVR2 and HVR3 is only 9Å.

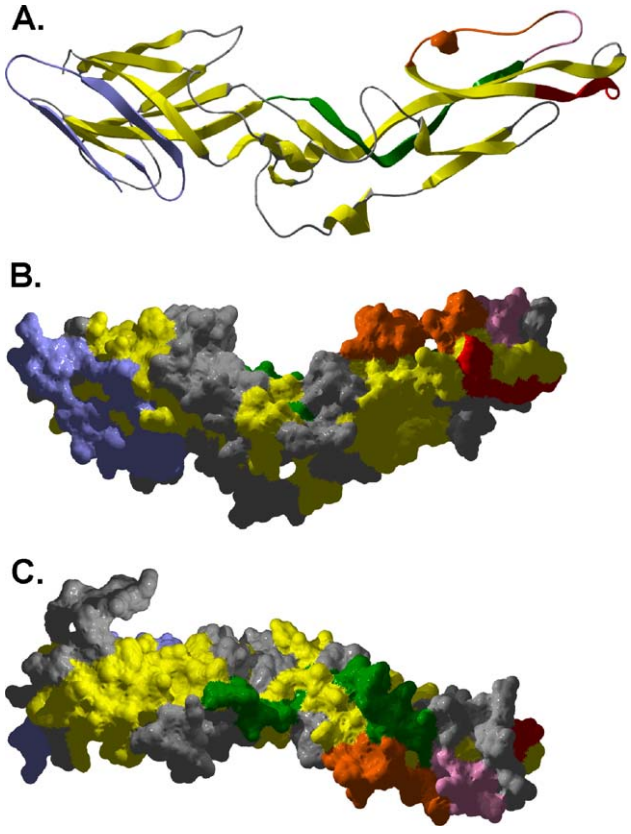

Fig. 4. Molecular modeling of hepatitis C virus E2 envelope glycoprotein. (A) Schematic representation of the E2 protein of HCV with hypervariable regions highlighted. (B and C) Surface of the E2 protein based on tick-borne encephalitis virus E protein crystal structure (front and reverse side shown) (Rey et al., 1995; Yagnik et al., 2000). Blue: hypervariable region 1 (HVR1); red: hypervariable region 2 (HVR2); orange: hypervariable region 3c (HVR3c); orange and pink: hypervariable region 3b (HVR3b); orange, pink, and green: hypervariable region 3 (HVR3).

Discussion

HCV E1 and E2 envelope glycoproteins are of particular interest because of their central implication in the initial steps of viral infection and because they constitute important targets for the host's HCV-specific immune responses. Herein we report that significant clustering of interhost genetic diversity takes place outside of already-described E2 hypervariable regions HVR1 and HVR2. In addition to HVR1 and HVR2, amino acid sequence variability was particularly striking within HVR3 as a whole (amino acid positions 431–466) and HVR3c (amino acid positions 456–466). Evidence was also provided that these newly described envelope subdomains were subjected to differential levels of intrahost selective pressure. Indeed, positive selection, as expressed by the mean d*N*/d*S* ratio, was highest in HVR1 > HVR3a > HVR2 > HVR3b. In fact, the d*N*/d*S* ratio was higher in HVR3a than HVR2 in 13 of 17 patients (76.5%) examined in the present study (Table 2). Interhost genetic diversity and intrahost selection pressure were not strictly dependent on HCV genotype: these parameters were indistinguishable in subjects infected with HCV-1a, HCV-1b, or HCV-3a, even though the limited number of subjects examined per stratum may have obscured the detection of subtle differences in

this regard. Interestingly, interhost genetic variability and intrahost selective pressure were not focalized on identical E2 subdomains, with the exception of HVR1, which was both highly diversified and subjected to robust positive selection. Furthermore, significant correlation between the d*N*/d*S* ratios observed in HVR1 and HVR3a, but not HVR3b and HVR2, is potentially reflective of functional differences between respective subdomains (see below).

Intrahost selective pressure can be exerted at least in part by host HCV-specific immune responses, and humoral immunity is an important component of this response. Indeed, neutralizing monoclonal antibodies 1/39, 2/69a, 7/16b, and 11/20, that block the binding of HCV-like particles to human CD81 and inhibit infection of Huh-7.5 hepatoma cells by HIV pseudotyped particles bearing native HCV E1 and E2, bind within a 15 amino acids region of glycoprotein E2 (amino acid residues 432–447) that maps precisely to HVR3a (Flint et al., 1999; Owsianka et al., 2001; Hsu et al., 2003). In addition, we cannot exclude that cell-mediated immune responses, in particular class I MHC-restricted cytotoxic T lymphocytes (CTL), also play a role in exerting selective pressure on particular HVR3 segments. Multiple CTL epitopes were described in E2, including some located within hypervariable regions (Shirai et al., 1999; Frasca et al., 1999; Tester et al., 2005). In particular, the HLA-B53-restricted epitope CRPLTDFDQGV maps to the C-terminal part of HVR3c (Wong et al., 1998), whereas the E2–21 peptide (DFAQGWGPISYANGS), that straddles the C-terminal part of HVR3c and the N-terminal part of HVR2, is recognized by $CD8^+$ T cells in HCV-infected subjects expressing HLA-A2, B15, and B51 (Dutoit et al., 2005). Moreover, positive selection associated with CTL escape was observed in acute and chronic HCV infection (Timm et al., 2004; Cox et al., 2005) and could serve to explain interhost and intrahost amino acid sequence variability at positions occupied by hydrophobic residues (i.e., amino acid positions 437 and 438) that were predicted to be largely buried within E2 based on molecular modeling (Fig. 4).

HIV infection induces a decline in $CD4^+$ T cell numbers and, as a consequence, can result in suppression of HCV-specific immune responses in coinfected hosts (Lauer et al., 2002). Consistent with this principle, several groups of investigators have shown that coinfection of HCV^+ subjects with HIV-1 led to a modulation of HCV quasispecies evolution and to reduced intrahost selective pressure on HVR1 (Mao et al., 2001; Canobio et al., 2004). However, in the present study, a significant reduction in host selective pressure exerted on E2 was not observed in coinfected subjects, as compared with patients infected with HCV alone. This can be explained by the fact that mean $CD4^+$ T cell counts in coinfected study participants were >500 cells/mm^3, with only 2 subjects below the 200 cells/mm^3 AIDS-defining threshold. These values are not consistent with substantial HIV-associated immunosuppression. Hence, a marked reduction in selective pressure would not be anticipated. As well, 11 of 13 coinfected subjects were treated with antiretroviral therapy during the course of the study, including 10 who were treated with combinations of antiretroviral agents. This commonly leads to a rapid restoration of $CD4^+$ T cell counts and of functional immunocompetence (Autran et al., 1997). In fact, the two lowermost d*N*/d*S* ratios observed in HVR1 during the course of the study were observed in TV075 and TV179, two coinfected subjects who were not initially treated with antiretrovirals.

In summary, data were presented that supports the existence in HCV E2 of a third hypervariable region located between HVR1 and HVR2. Concordance between results of the amino acid entropy-based clustering approach (Fig. 1), interhost genetic distance comparisons (Table 1), intrahost selective pressure data (Table 2), secondary structure, antigenicity, and accessibility predictions (Fig. 2), and three-dimensional molecular modeling (Fig. 4) provides compelling evidence in support of the proposed HVR3 subdomain organization. Taken together, these data also suggest that (a) at least sizable portions of HVR3, including the C-terminal part of HVR3a, are being exposed at the surface of E2; (b) consistent with monoclonal antibody binding, HVR3, including HVR3a, can be targeted by host humoral immune responses; and (c) consistent with antibody neutralization experiments, HVR3 plays a role in the process of binding with host cell receptors and entry of the virus into host cells. In this regard, whereas amino acid sequence variation observed in HVR3a was extensive in some cases, the basic organization and physicochemical characteristics of residues at key positions in HVR3a were generally well-conserved (Fig. 3). In particular, further analysis showed that the hydropathic profile at the center of HVR3a was conserved among genotypes (data not shown), suggestive of functional constraints on variation potentially related to conformational conservation of the E2 glycoprotein, viral fitness, and/or specific subdomain function. Likewise, the conserved positive charges localized within the rather hydrophilic environment at the edge of the hydrophobic center of HVR3a represents a feature conserved across genotypes that suggests a functional role of this region, for instance in protein–protein interaction. Amino acid residues proximal to HVR2 have been implicated in binding of E2 to the CD81 cell surface molecule, an interaction that has strong potential relevance for viral entry of HCV in the host cell and the pathogenesis of hepatitis C (Pileri et al., 1998; Hsu et al., 2003). Molecular modeling (Fig. 4) revealed that HVR3c and, to a certain extent, HVR3a and HVR3b, were closely clustered in the E2 envelope glycoprotein structure, suggestive of a potential role in binding of putative HCV cell surface receptors, including CD81, scavenger receptor class B type I (SR-BI) (Scarselli et al., 2002), and/or heparan sulfate (Barth et al., 2003). Alternatively, a region corresponding to HVR3a was recently reported to contain a fusion peptide-like domain (amino acid positions 436–443; GWLAGLFY) homologous to that found in TBEV, dengue virus, and other flaviviruses, which would have bipartite functions in binding CD81 and mediating fusion of the viral and host cell membranes (Drummer et al., 2005). It is tempting to speculate that neighboring HVR3b, which shows marked amino acid sequence diversity in interhost comparisons but appears subjected to little if any intrahost selection, could be functionally involved in pH-dependent conformational changes in E2, leading to exposure of this putative fusion peptide. However, resolution of these issues will need to await direct experimentation.

In view of the results presented herein, we strongly feel that future investigations into HCV envelope function and HCV envelope-specific immune responses should take into account the full extent of amino acid sequence variability of the E2 glycoprotein, which clearly extends beyond HVR1 and HVR2. In addition, because of the high levels of interhost sequence variability observed outside of HVR1, cross-sectional studies based on comparisons of selective pressure in HVR1 and so-called invariant flanking regions should be interpreted with caution. A fuller understanding of the structure–function relationships in HCV envelope glycoproteins will contribute to the design of HCV entry inhibitors and may lead to the development of immunogens to prevent and/or treat hepatitis C.

Materials and methods

Patients and clinical parameters

Study subjects (n = 17) were selected among participants to the Centre Maternel et Infantile sur le SIDA mother–child cohort (CHU mère-enfant Sainte-Justine, Montreal, Canada) and were previously enrolled in a study of immune responses and HCV quasispecies evolution during pregnancy (Troesch et al., unpublished). This study was conducted according to the Declaration of Helsinki, and the research protocol was approved by the Ethics Review Board of CHU mère-enfant Sainte-Justine, Montreal, Canada. All patients and their children were provided with medical care and counseling required by their condition. Mean age at study entry was 27.9 years (range = 22.2–40.7 years). Four of the subjects were infected with HCV only and 13 were coinfected with HCV and HIV-1, as determined using standard diagnostic algorithms routinely applied in the Province of Quebec. At the time of the study, all coinfected patients were treated with combination antiretroviral therapy for the purpose of preventing mother-to-child transmission of HIV-1, with the exception of subjects TV075 and TV179, who were only treated during the second of two consecutive pregnancies analyzed. In coinfected subjects, mean $CD4^+$ and $CD8^+$ T lymphocyte counts, measured by flow cytometry, were 558 cells/mm^3 (range = 144–1287 cells/mm^3) and 731 cells/mm^3 (range = 431–1122 cells/mm^3), respectively. Serum was extracted from whole blood by centrifugation and kept at −80 °C until used. HCV genotyping was performed by sequence analysis of the 5′ noncoding region and NS5B, as described (Murphy et al., 1994).

RNA extraction, RT-PCR, cDNA cloning and sequencing

RNA was extracted from 280 µl serum according to the QIAamp Viral RNA protocol and cDNA was synthesized using the OneStep RT-PCR method (QIAGEN, Mississauga, Canada). A portion of the E1 and E2 genes from HCV genotype 1 (nucleotide positions 1278–1889) was amplified by RT-PCR, with the nucleotide sequence of HCV-1a used as reference (GenBank accession no. M62321). Amplification was carried out using primers E2/NS1a (5′-ATA ACG GGT CAC CGC ATG GCA TGG GAT AT-3′) and E2/NS1b (5′-CAC CAC CAC GGG GCT GGG AGT GAA GCA AT-3′) (Farci et al., 2000). In the case of patient TVC55, primers E2/NS5aBIS (5′-GGC ATG GGA CAT GAT GAT GA-3′) and E2/NS1b were used (nucleotide positions 1295–1889). In the case of subjects infected with HCV genotype 3, primers E2/NS3a (5′-TGT ACC CAG GCC ATC TTT CA-3′) and E2/NS1b or E2/NS3b (5′-ATC AGT AGT GCC TAC GAC CA-3′) were used (nucleotide positions 1262–1890 or 1262–1902, respectively), with the nucleotide sequence of HCV-3a serving as reference (GenBank accession no. D17763). Reverse transcription was performed at 50 °C for 30 min. The PCR reaction consisted of 40 cycles of denaturing at 94 °C for 50 s, annealing at 55 °C for 1 min, and extension at 72 °C for 1 min. The fact that the enzyme used was not a proofreading polymerase would have for effect to increase the observed mutation rate. However, this increase will be largely unbiased with respect to the localization of mutations within the amplified segments, and comparisons between various subloci located within a single amplicons should therefore remain valid (Polyak et al., 2005). PCR products were purified from agarose gels using the GFX system (Amersham Biosciences, Piscataway, NJ). Amplicons were cloned between the twin *Eco*RI sites of pCR2.1-TOPO and were transformed into *Escherichia coli* TOP10-competent cells (Invitrogen, Carlsbad, CA). Plasmid DNA from transformants was extracted and purified with the Perfectprep Plasmid 96 Vac Direct Bind System (Eppendorf, Hamburg, Germany). Unidirectional sequencing of recombinants was performed on a Genetic Analyser 3100 (Applied Biosystems, Foster City, CA) using dye terminator chemistry.

Sequence analysis

Chromatograms were edited manually (Chromas version 1.45, Technelysium, Southport, Australia). All nucleotide sequences were submitted to GenBank (accession nos. DQ650805–DQ652141). Sequence alignments were visualized, compared, and edited using ClustalX version 1.81 (Thompson et al., 1997). Variability at each amino acid position was calculated as the entropy, defined as $-\sum P(s_i)\log P(s_i)$, where $P(s_i)$ is the probability of a given amino acid (s) appearing at a given position (i). Entropy was computed and consensus amino acid sequence was determined using the Entropy-ONE Web tool (http://hcv.lanl.gov/content/hcv-db/ENTROPY) (Korber et al., 1994). Mean genetic distance, defined as the mean of all pairwise p distance comparisons (number of amino acid differences divided by the total number of amino acid sites compared), and its standard deviation were calculated using MEGA version 2.1 (Kumar et al., 2001). The number of nonsynonymous (amino acid replacement) nucleotide substitutions per nonsynonymous site (dN), the number of synonymous (silent) nucleotide substitution per synonymous site (dS), and the dN/dS ratio were computed according to the Nei–Gojobori method with the Jukes–Cantor correction (Nei and Gojobori, 1986). dN/dS was expressed as median and interquartile range (IQR), and the significance of positive selection was determined using analysis of variance (Z test), as implemented in MEGA 2.1 (Kumar et al., 2001). As dN/dS does not follow the Gaussian distribution (Kolmogorov–Smirnov test), the Kruskal–Wallis test for multiple groups was selected for comparisons between medians.

Comparisons between regions were examined using Dunn's post test, which corrects for multiple comparisons in nonparametric tests.

Molecular modeling of HCV E2 glycoprotein

The model of E2 was generated essentially as described by Yagnik and coworkers (2000). Briefly, secondary structure predictions were used to align HCV E2 with the sequence of the E envelope glycoprotein of TBEV. The alignment generated was used to model the E2 protein using the structure of the TBEV E protein (PDB code *1SVB*; Rey et al., 1995) and the MODELLER program (Sali and Blundell, 1993). Energy minimization of the resulting model was then performed using 100 steps of the steepest descents algorithm with the consistent valence force field (CVFF) (Dauber-Osguthorpe et al., 1988). The protein structure was analyzed with the Swiss PDB Viewer 3.7 program (Guex and Peitsch, 1997), and the final images were rendered using POVRAY version 3.5.

Acknowledgments

The authors wish to thank Martine Caty, Silvie Valois, and Ampha Khammy for expert technical assistance, and Marc-André Rodrigue and Julie Lacaille for automated DNA sequencing. Supported by grants from the CIHR-Health Canada Research Initiative on Hepatitis C (grant no. EOP-41537), CANFAR-Canadian Foundation for AIDS Research (grant no. 013515), and The Elizabeth Glaser Pediatric AIDS Foundation (grant no. 28-PG-51355). M.T. was the recipient of a scholarship from Fondation de l'Hôpital Sainte-Justine. H.S. is a Junior-II Scientist of le Fonds de la recherche en santé du Québec (FRSQ).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2006.05.015.

References

Abbate, I., Lo Iacono, O., Di Stefano, R., Cappiello, G., Girardi, E., Longo, R., Ferraro, D., Antonucci, G., Di Marco, V., Solmone, M., Craxi, A., Ippolito, G., Capobianchi, M.R., 2004. HVR-1 quasispecies modifications occur early and are correlated to initial but not sustained response in HCV-infected patients treated with pegylated- or standard-interferon and ribavirin. J. Hepatol. 40, 831–836.

Alter, M.J., Margolis, H.S., Krawczynski, K., Judson, F.N., Mares, A., Alexander, W.J., Hu, P.Y., Gerber, M.A., Sampliner, R.E., Meeks, E.L., Beach, M.J., 1992. The natural history of community-acquired hepatitis C in the United States. The Sentinel Counties Chronic non-A, non-B Hepatitis Study Team. N. Engl. J. Med. 327, 1899–1905.

Autran, B., Carcelain, G., Li, T.S., Blanc, C., Mathez, D., Tubiana, R., Katlama, C., Debre, P., Leibowitch, J., 1997. Positive effects of combined antiretroviral therapy on $CD4^+$ T cell homeostasis and function in advanced HIV disease. Science 277, 112–116.

Bach, N., Thung, S.N., Schaffner, F., 1992. The histological features of chronic hepatitis C and autoimmune chronic hepatitis: a comparative analysis. Hepatology 15, 572–577.

Barth, H., Schafer, C., Adah, M.I., Zhang, F., Linhardt, R.J., Toyoda, H., Kinoshita-Toyoda, A., Toida, T., Van Kuppevelt, T.H., Depla, E., Von Weizsacker, F., Blum, H.E., Baumert, T.F., 2003. Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate. J. Biol. Chem. 278, 41003–41012.

Benhamou, Y., Di Martino, V., Bochet, M., Colombet, G., Thibault, V., Liou, A., Katlama, C., Poynard, T., 2001. Factors affecting liver fibrosis in human immunodeficiency virus- and hepatitis C virus-coinfected patients: impact of protease inhibitor therapy. Hepatology 34, 283–287.

Bonacini, M., Puoti, M., 2000. Hepatitis C in patients with human immunodeficiency virus infection: diagnosis, natural history, meta-analysis of sexual and vertical transmission, and therapeutic issues. Arch. Intern. Med. 160, 3365–3373.

Bukh, J., Miller, R.H., Purcell, R.H., 1995. Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. Semin. Liver Dis. 15, 41–63.

Canobio, S., Guilbert, C., Troesch, M., Samson, J., Lemay, M., Pelletier, V.A., Bernard-Bonnin, A.C., Kozielski, R., Lapointe, N., Martin, S.R., Soudeyns, H., 2004. Differing patterns of liver disease progression and hepatitis C virus (HCV) quasispecies evolution in children vertically coinfected with HCV and human immunodeficiency virus type 1. J. Clin. Microbiol. 42, 4365–4369.

Chen, S., Wang, Y.M., 2005. Multigene tracking of quasispecies in viral persistence and clearance of hepatitis C virus. World J. Gastroenterol. 11, 2874–2884.

Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., Houghton, M., 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244, 359–362.

Choo, Q.L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R., Barr, P.J., Weiner, A.J., Bradley, D.W., Kuo, G., Houghton, M., 1991. Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. U.S.A. 88, 2451–2455.

Cox, A.L., Mosbruger, T., Mao, Q., Liu, Z., Wang, X.H., Yang, H.C., Sidney, J., Sette, A., Pardoll, D., Thomas, D.L., Ray, S.C., 2005. Cellular immune selection with hepatitis C virus persistence in humans. J. Exp. Med. 201, 1741–1752.

Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M., Hagler, A.T., 1988. Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase-trimethoprim, a drug-receptor system. Proteins 4, 31–47.

Dieterich, D.T., 1999. Hepatitis C virus and human immunodeficiency virus: clinical issues in coinfection. Am. J. Med. 107, 79S–84S.

Dodig, M., Tavill, A.S., 2001. Hepatitis C and human immunodeficiency virus coinfections. J. Clin. Gastroenterol. 33, 367–374.

Drummer, H., Maerz, A., Poumbourios, P., 2005. Structural and functional role of a fusion-peptide like sequence in hepatitis C virus glycoprotein E2. Abstracts of the 12th International Meeting on HCV and Related Viruses, Montreal, Canada, p. 26.

Dutoit, V., Ciuffreda, D., Comte, D., Gonvers, J.J., Pantaleo, G., 2005. Differences in HCV-specific T cell responses between chronic HCV infection and HIV/HCV co-infection. Eur. J. Immunol. 35, 3493–3504.

Farci, P., Shimoda, A., Coiana, A., Diaz, G., Peddis, G., Melpolder, J.C., Strazzera, A., Chien, D.Y., Munoz, S.J., Balestrieri, A., Purcell, R.H., Alter, H.J., 2000. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. Science 288, 339–344.

Farci, P., Strazzera, R., Alter, H.J., Farci, S., Degioannis, D., Coiana, A., Peddis, G., Usai, F., Serra, G., Chessa, L., Diaz, G., Balestrieri, A., Purcell, R.H., 2002. Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. Proc. Natl. Acad. Sci. U.S.A. 99, 3081–3086.

Flint, M., Maidens, C., Loomis-Price, L.D., Shotton, C., Dubuisson, J., Monk, P., Higginbottom, A., Levy, S., McKeating, J.A., 1999. Characterization of hepatitis C virus E2 glycoprotein interaction with a putative cellular receptor, CD81. J. Virol. 73, 6235–6244.

Frasca, L., Del Porto, P., Tuosto, L., Marinari, B., Scotta, C., Carbonari, M., Nicosia, A., Piccoletta, E., 1999. Hypervariable region 1 variants act as TCR antagonists for hepatitis C virus-specific $CD4^+$ T cells. J. Immunol. 163, 650–658.

Gaudy, C., Moreau, A., Veillon, P., Temoin, S., Lunel, F., Goudeau, A., 2003. Significance of pretreatment analysis of hepatitis C virus genotype 1b

hypervariable region 1 sequences to predict antiviral outcome. J. Clin. Microbiol. 41, 3615–3622.
Gonzalez-Peralta, R.P., Qian, K., She, J.Y., Davis, G.L., Ohno, T., Mizokami, M., Lau, J.Y., 1996. Clinical implications of viral quasispecies heterogeneity in chronic hepatitis C. J. Med. Virol. 49, 242–247.
Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723.
Heo, T.H., Chang, J.H., Lee, J.W., Foung, S.K.H., Dubuisson, J., Kang, C.Y., 2004. Incomplete humoral immunity against hepatitis C virus is linked with distinct recognition of putative multiple receptors by E2 envelope glycoproteins. J. Immunol. 173, 446–455.
Honda, M., Kaneko, S., Sakai, A., Unoura, M., Murakami, S., Kobayashi, K., 1994. Degree of diversity of hepatitis C virus quasispecies and progression of liver disease. Hepatology 20, 1144–1151.
Hsu, M., Zhang, J., Flint, M., Logvinoff, C., Cheng-Mayer, C., Rice, C.M., McKeating, J.A., 2003. Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped viral particles. Proc. Natl. Acad. Sci. U.S.A. 100, 7271–7276.
Janin, J., 1979. Surface and inside volumes in globular proteins. Nature 277, 491–492.
Kato, N., Ootsuyama, Y., Tanaka, T., Nakagawa, M., Nakazawa, T., Muraiso, K., Ohkoshi, S., Hijikata, M., Shimotohno, K., 1992. Marked sequence diversity in the putative envelope proteins of hepatitis C viruses. Virus Res. 22, 107–123.
Kato, N., Sekiya, H., Ootsuyama, Y., Nakazawa, T., Hijikata, M., Ohkoshi, S., Shimotohno, K., 1993. Humoral immune response to hypervariable region 1 of the putative envelope glycoprotein (gp70) of hepatitis C virus. J. Virol. 67, 3923–3930.
Korber, B.T.M., Kunstman, K.J., Patterson, B.K., Furtado, M., McEvilly, M.M., Levy, R., Wolinsky, S.M., 1994. Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. J. Virol. 68, 7467–7481.
Kuiken, C., Yusim, K., Boykin, L., Richardson, R., 2005. The Los Alamos HCV sequence database. Bioinformatics 21, 379–384.
Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.
Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.
Lauer, G.M., Nguyen, T.N., Day, C.L., Robbins, G.K., Flynn, T., McGowan, K., Rosenberg, E.S., Lucas, M., Klenerman, P., Chung, R.T., Walker, B.D., 2002. Human immunodeficiency virus type 1-hepatitis C virus coinfection: intraindividual comparison of cellular immune responses against two persistent viruses. J. Virol. 76, 2817–2826.
Mao, Q., Ray, S.C., Laeyendecker, O., Ticehurst, J.R., Strathdee, S.A., Vlahov, D., Thomas, D.L., 2001. Human immunodeficiency virus seroconversion and evolution of the hepatitis C virus quasispecies. J. Virol. 75, 3259–3267.
Martell, M., Esteban, J.I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J., Gomez, J., 1992. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. J. Virol. 66, 3225–3229.
Murphy, D., Willems, B., Delage, G., 1994. Use of the 5′ noncoding region for genotyping hepatitis C virus. J. Infect. Dis. 169, 473–475.
Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.
Owsianka, A., Clayton, R.F., Loomis-Price, L.D., McKeating, J.A., Patel, A.H., 2001. Functional analysis of hepatitis C virus E2 glycoproteins and virus-like particles reveals structural dissimilarities between different forms of E2. J. Virol. 82, 1877–1883.
Parker, J.M., Guo, D., Hodges, R.S., 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry 25, 5425–5432.
Pawlotsky, J.M., 2000. Hepatitis C virus resistance to antiviral therapy. Hepatology 32, 889–896.
Penin, F., Combet, C., Germanidis, G., Frainais, P.O., Deleage, G., Pawlotsky, J.M., 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.
Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A.J., Houghton, M., Rosa, D., Grandi, G., Abrignani, S., 1998. Binding of hepatitis C virus to CD81. Science 282, 938–941.
Polyak, S.J., Sullivan, D.G., Austin, M.A., Dai, J.Y., Shunhart, M.C., Lindsay, K.L., Bonkovsky, H.L., Di Bisceglie, A.M., Lee, W.M., Morishima, C., Gretch, D.R., 2005. Comparison of amplification enzymes for hepatitis C virus quasispecies analysis. Virol J 2, 41.
Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. Nature 375, 291–298.
Rubbia-Brandt, L., Leandro, G., Spahr, L., Giostra, E., Quadri, R., Male, P.J., Negro, F., 2001. Liver steatosis in chronic hepatitis C: a morphological sign suggesting infection with HCV genotype 3. Histopathology 39, 119–124.
Saito, I., Miyamura, T., Ohbayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, Y., Koi, S., Onji, M., Ohta, Y., Choo, Q.L., Houghton, M., Kuo, G., 1990. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. Proc. Natl. Acad. Sci. U.S.A. 87, 6547–6549.
Sala, M., Wain-Hobson, S., 2000. Are RNA viruses adapting or merely changing? J. Mol. Evol. 51, 12–20.
Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815.
Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R.M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R., Vitelli, A., 2002. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. EMBO J. 21, 5017–5025.
Sherman, K.E., Andreatta, C., O'Brien, J., Gutierrez, A., Harris, R., 1996. Hepatitis C in human immunodeficiency virus-coinfected patients: increased variability in the hypervariable envelope coding domain. Hepatology 23, 688–694.
Shirai, M., Arichi, T., Chen, M., Masaki, T., Nishioka, M., Ikeda, K., Takahashi, H., Enomoto, N., Saito, T., Major, M.E., Nakazawa, T., Akatsuka, T., Feinstone, S.M., Berzofsky, J.A., 1999. T cell recognition of hypervariable region-1 from hepatitis C virus envelope protein with multiple class II MHC molecules in mice and humans: preferential help for induction of antibodies to the hypervariable region. J. Immunol. 162, 568–576.
Simmonds, P., Holmes, E.C., Cha, T.A., Chan, S.W., McOmish, F., Irvine, B., Beall, E., Yap, P.L., Kolberg, J., Urdea, M.S., 1993. Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J. Gen. Virol. 74, 2391–2399.
Simmonds, P., Bukh, J., Combet, C., Deleage, G., Enomoto, N., Feistone, S., Halfon, P., Inchauspe, G., Kuiken, C., Maertens, G., Mizokami, M., Murphy, D.G., Okamoto, H., Pawlotsky, J.M., Penin, F., Sablon, E., Shin-I, T., Stuyver, L.J., Thiel, H.J., Viazov, S., Weiner, A.J., Widell, A., 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology 42, 962–973.
Slater-Handshy, T., Droll, D.A., Di Bisceglie, A.M., Chambers, T.J., 2004. HCV E2 glycoprotein: mutagenesis of N-linked glycosylation sites and its effects on E2 expression and processing. Virology 319, 36–48.
Soriano, V., Rodriguez-Rosado, R., Garcia-Samaniego, J., 1999. Management of chronic hepatitis C in HIV-infected patients. AIDS 13, 539–546.
Tester, I., Smyk-Pearson, S., Wang, P., Wertheimer, A., Yao, E., Lewinsohn, D.M., Tavis, J.E., Rosen, H.R., 2005. Immune evasion versus recovery after acute hepatitis C virus infection from a shared source. J. Exp. Med. 201, 1725–1731.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
Timm, J., Lauer, G.M., Kavanagh, D.G., Sheridan, I., Kim, A.Y., Lucas, M., Pillay, T., Ouchi, K., Reyor, L.L., Zur Wiesch, J.S., Gandhi, R.T., Chung, R.T., Bhardwaj, N., Klenerman, P., Walker, B.D., Allen, T.M., 2004. CD8

epitope escape and reversion in acute HCV infection. J. Exp. Med. 200, 1593–1604.

Weiner, A.J., Brauer, M.J., Rosenblatt, J., Richman, K.H., Tung, J., Crawford, K., Bonino, F., Saracco, G., Choo, Q.L., Houghton, M., 1991. Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology 180, 842–848.

Weiner, A.J., Geysen, H.M., Christopherson, C., Hall, J.E., Mason, T.J., Saracco, G., Bonino, F., Crawford, K., Marion, C.D., Crawford, K.A., Brunetto, M., Barr, P.J., Miyamura, T., McHutchinson, J., Houghton, M., 1992. Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. Proc. Natl. Acad. Sci. U.S.A. 89, 3468–3472.

Wong, D.K., Dudley, D.D., Afdhal, N.H., Dienstag, J., Rice, C.M., Wang, L., Houghton, M., Walter, B.D., Koziel, M.J., 1998. Liver-derived CTL in hepatitis C virus infection: breadth and specificity of responses in a cohort of persons with chronic infection. J. Immunol. 160, 1479–1488.

Yagnik, A.T., Lahm, A., Meola, A., Roccasecca, R.M., Ercole, B.B., Nicosia, A., Tramontano, A., 2000. A model for the hepatitis C virus envelope glycoprotein E2. Proteins 40, 355–366.