Immunoglobulin mimicry by Hepatitis C Virus envelope protein E2

Yu-Wen Hu\textsuperscript{a,b,},* Lynda Rocheleau\textsuperscript{b}, Bryce Larke\textsuperscript{a}, Linda Chui\textsuperscript{c}, Bonita Lee\textsuperscript{c}, Mang Ma\textsuperscript{c}, Song Liu\textsuperscript{a}, Teye Omlin\textsuperscript{a}, Martin Pelchat\textsuperscript{b}, Earl G. Brown\textsuperscript{b}

\textsuperscript{a}Canadian Blood Services, 1800 Alta Vista Drive, Ottawa, Ontario, Canada K1G 4J5

\textsuperscript{b}Department of Biochemistry Microbiology and Immunology, Faculty of Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, Canada K1H 8M5

\textsuperscript{c}Alberta Provincial Laboratory of Public Health, University of Alberta Hospital, 8440-112 Street, Edmonton, Alberta, Canada T6G 2J2

Received 28 September 2004; accepted 19 November 2004
Available online 12 January 2005

Abstract

Hepatitis C virus (HCV) establishes persistent infection in the majority of infected individuals. The currently accepted hypothesis of immune evasion by antigenic variation in hypervariable region 1 (HVR1) of glycoprotein E2 does not however, explain the lack of subsequent immune recognition. Here, we show that the N-terminal region of E2 is antigenically and structurally similar to human immunoglobulin (Ig) variable domains. E2 is recognized by anti-human IgG antibodies and also possesses common amino acid (aa) sequence features of the conserved v-gene framework regions of human Ig light chains in particular but also heavy chains and T cell receptors. Using a position specific scoring system, the degree of similarity of HVR1 to Ig types correlated with immune escape and persistence in humans and experimentally infected chimpanzees. We propose a unique role for threshold levels of Ig molecular mimicry in HCV biology that not only advances our concept of viral immune escape and persistent infection but also provides insight into host-dependent disease patterns.

D\textsuperscript{2}004 Elsevier Inc. All rights reserved.

Keywords: Immunoglobulin mimicry; Hepatitis C; Glycoprotein

Introduction

Hepatitis C virus (HCV) currently infects 170 million people constituting a major cause of chronic hepatitis, cirrhosis, and hepatocellular carcinoma (Alter and Seeff, 2004; Major et al., 2004). HCV is the only RNA virus infecting humans (excepting retroviruses) that persists in the majority of infected individuals. Viruses, especially those with RNA genomes, can undergo mutation at high frequencies, and under novel selective pressures, rapidly generate populations of viral variants. Such variability can provide a means of evading clearance by both T- and B-cell immunity. Accumulated data suggest that hypervariable region 1 (HVR1), located in a stretch of 27–31 residues at the amino terminus of the second envelope glycoprotein (E2) is the main target of the anti-HCV neutralizing response and therefore plays a significant role in the establishment of viral persistence (Farci et al., 2000; Kato, 2001). During HCV infection, amino acid substitutions in HVR1 generates populations of genetically related variants, termed quasispecies (Domingo et al., 1997; Ducoulombier et al., 2004; Hijikata et al., 1991), some of which are antibody escape mutants that are not recognized by the immune response and persist after seroconversion (Kato, 2001; Pavio and Lai, 2003). Although this is the most accepted hypothesis, the important question as to why the virus epitopes within HVR1 cannot be subsequently recognized by the immune system is unknown. The role of HVR1 in persistence has become controversial because recombinant HCV lacking HVR1, can persist in chimpanzees, but the resultant virus was not able to establish

---

Abbreviations: aa, amino acid(s); BLAST, basic local alignment search tool; CDR, complementarity determining region; E2, envelope protein 2; eIF2a, eukaryotic initiation factor 2a; FR, framework region; HCV, hepatitis C virus; Ig, immunoglobulin; HVR1, hypervariable region 1; IFN, interferon; IMGT, Immunogenetics database; pi, post infection; pd, post-diagnosis; PKR, protein kinase R; TCR, T cell receptor.

* Corresponding author. Canadian Blood Services, 1800 Alta Vista Drive, Ottawa, Ontario, Canada K1G 4J5. Fax: +1 613 739 2426.

E-mail address: yu-wen.hu@bloodservices.ca (Y.-W. Hu).

0042-6822/S - see front matter © 2004 Elsevier Inc. All rights reserved.
persistent infection on subsequent transmission suggesting a decreased ability in this regard (Forns et al., 2000).

Molecular mimicry, where viruses express proteins that are structurally similar to host defense proteins and immunomodulators, is an important immune-evasion strategy used to promote survival and persistence, especially in DNA viruses (Ploegh, 1998; Seet et al., 2003; Vossen et al., 2002). HCV is also known to employ molecular mimicry to resist type I IFN (Taylor et al., 2000). The virally induced antiviral cytokine, type I interferon (IFN), acts in part through the dsRNA-dependent protein kinase (PKR) to inhibit protein synthesis through phosphorylation of eukaryotic initiation factor 2α (eIF2α). The HCV envelope protein E2 contains a 12 aa sequence identical to phosphorylation domains of both eIF2α and the PKR kinase (Taylor et al., 1999). This domain operates to prevent PKR-dependent phosphorylation of eIF2α and inhibition of protein synthesis. The extent of PKR-eIF2α homology of this domain correlates with the ability of HCV to resist type I IFN treatment.

We reasoned that other instances of molecular mimicry, targeted to other human proteins, could also be contributing to HCV persistent infection. We found a second instance of molecular mimicry, where HCV encodes a sequence in E2 that is homologous to human immunoglobulins (Ig), and specifically possesses typical structural features of the variable region of Ig such that it cross-reacts with anti-human-IgG. Using a bioinformatic and evolutionary approach, sequence analysis of HCV variants that arise in the course of primary infection in humans and chimpanzees showed that the degree of similarity of HVR1 or its epitopes with Ig types was directly related to viral escape and persistence in a host specific manner, indicating a significant role for viral molecular mimicry.

**Results**

*E2 is structurally similar to IgG*

To elucidate the mechanism of immune escape by HCV, an investigation of E2 mimicry was performed. Surprisingly, full-length recombinant HCV genotype 3a E2 proteins expressed in *Escherichia coli* (*E. coli*) were repeatedly seen to bind antibody to human IgG Fab fragment (Fig. 1A, lanes 1 and 2), suggesting structural similarity between E2 and human Ig molecules. This binding was also obtained when using the first 113 or 123 aa of E2 (encompassing hypervariable regions 1 and 2 (HVR1, HVR2)) from different HCV variants of genotype 1a whether synthesized

![Fig. 1. Anti-human-IgG Binds HCV E2.](image-url)

(A) Full length HCV 3a E2 (aa1–322) or amino terminal regions from aa1–123 of HCV 1a proteins were expressed in *E. coli*. The left panel shows total protein staining as a sample loading control (Coomassie blue) and the right panel is stained with anti-human-IgG Fab fragment binding. Samples: U—uninduced HCV 3a clone B-d8-1 (aa1–322), 1—HCV 3a B-1 (aa1–322), 2—HCV 3a B-2 (aa1–322); 3—HCV 1a A-5 (aa1–123); 4—HCV 1a A-9 (aa1–123). In addition to the specific E2 proteins, two *E. coli* protein bands were observed to bind anti-human IgG (open arrows). (B) Amino terminal fragments of HCV1a were cloned and expressed in baculovirus infected insect cells. The left panel shows the loading controls that were detected by binding to anti-his6 antibody and the right panel is stained with anti-human-IgG Fab fragment. Sample C—cell lysate from empty baculovirus vector; 1—HCV 1a A19 (aa1–113); 2—HCV 1a A-5 (aa1–123).
in insect cells or *E. coli* when probed with antibodies against the Fab fragment (Fig. 1A lanes 3, 4; and 1B lanes 1, 2) or whole human IgG (not shown), indicating that the immunoglobulin antigenic cross-reactivity involves the protein component of the N-terminal portion of E2. Most importantly, these results satisfy the operational definition of mimicry (Cohen, 2004) and demonstrate the Ig-like nature of the N-terminal portion of E2, since both Ig and E2 are bound by common ligands (i.e., anti-IgG antibodies).

**Sequence alignment of E2 with immunoglobulin and T cell receptors**

To assess the nature of the structural similarities between E2 and Igs, an alignment between N-terminal portions of a variety of E2 sequences representing the six major HCV genotypes and sequences of various Ig types was executed. As a first step, computer-generated alignments (i.e., BLAST and ClustalW) were used to identify sequence similarities between E2 and conserved Ig aas. The regions of highest sequence similarity were found to be almost exclusively restricted to the variable region (v-gene) of Ig molecules and the first 104 aa of E2. Secondly, to better align these sequences, the Immunogenetics (IMGT) unique numbering system (Lefranc et al., 2003) was used because it allows comparison among members of the Ig superfamily (Barclay, 2003; Harpaz and Chothia, 1994). This numbering is derived from the sequence alignment and 3-dimensional structural comparisons of Ig superfamily members and provides a definition of the highly conserved framework regions (FR) that support the antibody binding site formed by complementarity determining regions (CDR). Therefore, in the IMGT numbering system, conserved amino acids always have the same number and position for immunoglobulins and Ig superfamily v-like genes. The numbering system accommodates the longest forms of CDR which are placed in justified alignments between the framework regions as shown for heavy, light and TCR in Fig. 2.

We were able to identify amino acids in each human Ig framework region that were completely conserved among E2 and Ig types (i.e., kappa light chain). These E2 positions (IMGT numbering: G16, L53, Y55, G70, R75, Y103, and C104) then served as references to manually complete the IMGT formatted alignment with the FR regions of immunoglobulins and TCR. It was thus possible to align FR1, FR2, and FR3 with corresponding regions in E2 which also resulted in alignment of intervening sequence regions that were the same or comparable lengths to the CDR1 and CDR2 of heavy and light immunoglobulin chains. In addition, the insertion of a single gap before aa position 103 was necessary to achieve the most parsimonious alignment because the E2 sequences were 6 aa longer than the corresponding Ig region (Fig. 2A).

The majority of aa positions in E2 matched identical aa found among Ig members (masked in Fig. 2A), and most of the highly conserved Ig aa are preserved in E2 (red aa in Fig. 2A). Significantly, a signature sequence comprised of 14 aa positions was highly maintained (>70% identity), among E2 sequences and all or individual Ig protein types including the heavy and the light (kappa and lambda) Ig chains as well as T cell receptor (TCR) α and β chains (see red bars in Fig. 2B). In the highly conserved sites with less than 100% conservation at aa21 and 80, alternative amino acids of similar chemistry were in common with IG and TCR groups (masked in blue Fig. 2). Amino acid identity and sequence spacing of FR and CDR regions indicated a closer relationship to kappa light chains (for example 24% identity of HCV 2a AAF59944.1 with kappa X59312, Fig. 2), than for the heavy and TCRα and TCRγ chains genes (*P* < 0.0001) suggesting specific rather than random relationships of E2 to Ig gene types. There were 20 sites that had >50% common identity and several other sites had highly conserved chemistry (i.e., aa26 (small hydroxyl); 50, 54, 76, 94 (hydrophobic); and 74 (acidic)). The sequence identity of E2 to kappa light chains was significantly higher than calculated for random sequences (*P* ≤ 4 × 10⁻⁷); and more importantly randomized Ig consensus sequences (*P* ≤ 1 × 10⁻⁹). Taken together these data indicate a conserved core structure among E2 and Ig-like genes rather than a chance association of amino acids in the variable CDR regions.

Such a level of sequence similarity was also reported between human cytomegalovirus UL18 and MHC-1 proteins (Beck and Barrell, 1988), where the viral protein acts as an MHC-1-like immunomodulator (Mocarski, 2002). The observation of shared antigenic and sequence structures between E2 and Ig suggests that this region of E2 has the ability to form the basic Ig fold structure (Barclay, 2003; Harpaz and Chothia, 1994) thus providing further support for an Ig-like nature of the N-terminal domain of E2 (Fig. 3). Indeed, three-dimensional modeling shows an overlap of FR regions of the kappa variable domain (in yellow in Fig. 3A) with E2 (superimposed in blue and red), including HVR1 (in red), where all the framework regions align within 2 Å of the kappa peptide chain. The corresponding regions of E2 possess other common features of immunoglobulin framework regions such as amino acid content (high S, T and A) as well as specific motifs such as 14SPG17, 33LFYRNN58, 85TDF87, 103YC104. Although the E2 structure does not maintain the conserved disulfide bond of Ig formed by C23 and C104 that is guarded by W41, the modeled E2 structure has C41 adjacent to C104 that may form an alternative disulfide bond (data not shown), as occurs for some members of the Ig superfamily (Barclay, 2003). The occurrence of aa sequence identities, in combination with shared antigenicity further supports the contention that E2 is a molecular mimic of immunoglobulins.

**Evolution of E2 is directional with respect to Ig**

The existence of Ig similarity in E2 implies that sequence variations in E2, in particular HVR1, could be directional in...
response to selective pressures from the immune system. If threshold levels of host similarity are required for HCV immune evasion, the patterns of E2 evolution are expected to reflect host-dependent similarities and requirements for change. The presence of quasispecies that differ in their ability to be recognized by the immune system and be cleared from the host has been observed in humans and chimpanzees (Korenaga et al., 1997; Wyatt et al., 1998). The only known HCV antibody neutralization sites are located in HVR1, where aa substitutions lead to life-long persistent infection due to escape from recognition by preexisting anti-HVR1 antibodies (Kato et al., 1994). Given that immunoglobulin variable regions are diverse but not random (Johnson and Wu, 2000), it is possible to compare aa usage relative to groups of immunoglobulin v-genes when formatted using the IMGT numbering system (Lefranc et al., 2003). Using a scoring system based on a position-specific scoring of aa similarity using the updated Dayhoff matrix of Jones (Jones et al., 1992), (as well as other matrices, see Methods) that quantify aa similarity, the similarities of variant neutralizing epitopes in HVR1 sequences with respect to large groups of human Ig chains were determined using HCV 2a variant sequences (Fig. 4A) that were previously shown to evolve to escape immunity (Kato et al., 1994). In these comparisons, the average sum of similarities for each amino acid in the epitope region (aa4–

Fig. 2. IMGT sequence alignment of the N-terminal domain of HCV E2 and the variable region of human immunoglobulins and T cell receptor genes. (A) Groups of 10 sequences representing the major genotypes of E2, as well as groups of 5 expressed and 5 germ line variable gene sequences are shown for each set of light kappa, light lambda, heavy, T cell receptor \( \gamma \), and T cell receptor \( \alpha \) v-genes; (genes are named with accession numbers). The numbering of the HCV polyprotein is shown at the top along with the location of HVR1 and HVR2. The location of the framework (FR) and complementarity determining regions (CDRs) are shown within the IMGT numbered region at the bottom of the alignment. Amino acid sites with \( \geq 70\% \) shared identity between both E2 and any or all Ig groups are masked in red with alternative common substitutions at these sites (of similar chemistry) shown in blue. All other instances of shared aa identity with E2 are masked in black. Average v-gene region identities of E2 relative to all sequences shown (i.e., 100 pair wise comparisons for each group) were significantly higher for the light chain genes (kappa 17.1 \( \pm \) 2.3%; lambda 16.3 \( \pm \) 2.1%) than for the heavy (11.8 \( \pm \) 1.8%) and TCR\( \alpha \) (9.9 \( \pm \) 1.3%) and TCR\( \beta \) chains (12.1 \( \pm \) 0.8%) genes ( \( P < 0.0001 \) ) indicating specific relationships of E2 Ig and TCR gene types rather than a random similarity that would not be expected to differ among groups. (B) The percent identity to E2 is shown in the bar graph with those aa sites that share \( \geq 70\% \) identity within E2 and any Ig group shown in red.
17 Fig. 4) was scored relative to the corresponding amino acids of a large group (see Methods) of Ig (heavy, kappa, lambda) and TCR (α, β); thus, providing a measure of sequence similarity to each group. Mutations that decreased epitope binding of patient sera collected at various times post infection, resulting in escape from immune recognition, were seen to incrementally and significantly increase Ig similarity scores with respect to Ig groups (P = 0.036 by paired t test at 11 months relative to prior samples) or as an average among all groups (comparison to 3496 sequences), consistent with an improvement in Ig mimicry (Fig. 4 B). Similar trends in the evolution of mimicry were obtained using alternative scoring matrices (data not shown for identity, PAM250 and BLOSUM62 matrices). This meant that each amino acid substitution that was selected in these epitopes was either identical to, or more like, the corresponding amino acids in the majority of individual Ig and TCR molecules. The increases in score were greatest for heavy and kappa chains that were equivalent to one amino acid identity substitution (increase of 1500) in the epitope region of HVR1. The 11-month immune escape epitope variant avoided subsequent recognition by the immune response, indicating that immune evasion and loss of immunogenicity was directional with respect to the extent of Ig molecular mimicry. This indicated that variation in HVR1 was not only able to escape neutralizing antibody but also to become non-immunogenic, coincident with the acquisition of mutations that increased similarity to Ig and TCR types consistent with a model of immune evasion through mimicry.

In another of the few characterizations of HVR1 evolution during immune evasion (Fig. 5A), similar increases

<table>
<thead>
<tr>
<th>Epitope #</th>
<th>Time (mo.PD)</th>
<th>Antibody binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 HSVGFTLSFS</td>
<td>0</td>
<td>++++</td>
</tr>
<tr>
<td>2 RGFTLSFSAGS</td>
<td>0</td>
<td>++++</td>
</tr>
<tr>
<td>1 . . . . L . . . .</td>
<td>6</td>
<td>+++</td>
</tr>
<tr>
<td>2 . . L . . . .</td>
<td>6</td>
<td>++</td>
</tr>
<tr>
<td>1 G . L . . . .</td>
<td>8</td>
<td>++</td>
</tr>
<tr>
<td>2 . . L . . . .</td>
<td>8</td>
<td>++</td>
</tr>
<tr>
<td>1 G . S . L . . .</td>
<td>11</td>
<td>–</td>
</tr>
<tr>
<td>2 S . L . . . .</td>
<td>11</td>
<td>–</td>
</tr>
</tbody>
</table>

17 Fig. 4) was scored relative to the corresponding amino acids of a large group (see Methods) of Ig (heavy, kappa, lambda) and TCR (α, β); thus, providing a measure of sequence similarity to each group. Mutations that decreased epitope binding of patient sera collected at various times post infection, resulting in escape from immune recognition, were seen to incrementally and significantly increase Ig similarity scores with respect to Ig groups (P = 0.036 by paired t test at 11 months relative to prior samples) or as an average among all groups (comparison to 3496 sequences), consistent with an improvement in Ig mimicry (Fig. 4B). Similar trends in the evolution of mimicry were obtained using alternative scoring matrices (data not shown for identity, PAM250 and BLOSUM62 matrices). This meant that each amino acid substitution that was selected in these epitopes was either identical to, or more like, the corresponding amino acids in the majority of individual Ig and TCR molecules. The increases in score were greatest for heavy and kappa chains that were equivalent to one amino acid identity substitution (increase of 1500) in the epitope region of HVR1. The 11-month immune escape epitope variant avoided subsequent recognition by the immune response, indicating that immune evasion and loss of immunogenicity was directional with respect to the extent of Ig molecular mimicry. This indicated that variation in HVR1 was not only able to escape neutralizing antibody but also to become non-immunogenic, coincident with the acquisition of mutations that increased similarity to Ig and TCR types consistent with a model of immune evasion through mimicry.

In another of the few characterizations of HVR1 evolution during immune evasion (Fig. 5A), similar increases
in Ig mimicry were observed during a 9.5-year follow-up study of experimentally HCV 1b infected chimpanzees (van Doorn et al., 1995). Among a group of six chimpanzees infected with a defined HCV 1b isolate, three patterns of evolution were observed where half the animals developed anti-HVR1 antibodies and resolved their infection without evolution of the HVR1 sequences (i.e., no change in Ig similarity). Among the persistent infections, one chimpanzee was not stimulated to produce antibody to HVR1 until 7 years of chronic infection followed by a single mutation in HVR1 that reduced the average Ig similarity score, thus constituting a pattern of structural stasis (Fig. 5B, Phil). Finally, for the two persistently infected animals that seroconverted early, their HVR1 sequences concomitantly began to evolve to continuously increase Ig mimicry and escape immunity \( (P = 0.000005) \) for similarity scores differences after seroconversion), consistent with selection of this E2 region on the basis of similarity to Ig proteins (Fig. 5B, Peggy and Hans). Convergent evolution was seen for Peggy and Hans that both independently acquired 3 common mutations that increased Ig similarities (Fig. 5A).

A second independent follow-up study of an experimentally HCV 1-infected chimp (Chimp #1) that developed a persistent infection (Lu et al., 2001) showed that the extent of mimicry stayed relatively constant through a 12-year period of monitoring (Figs. 5A and B). The consensus sequence incrementally acquired a group of 5 mutations that resulted in an oscillation of Ig mimicry around a relatively constant level suggesting that the virus was sufficiently similar to the host Ig structures and maintained this level of host mimicry.

Evolution of E2 mimicry during seroconversion and interferon therapy

The current therapy for HCV infection involves treatment with IFN and ribavirin (Kryczka and Zarebska-Michaluk, 2003). Since the majority of patients become IFN-resistant during treatment and this transformation is associated with changes in HVR1 (Farci et al., 2000, 2002; Taylor et al., 2000), the evolution of HVR1 relative to Ig was determined following natural infection, seroconversion and subsequent IFN treatment. HCV nucleic acid testing of approximately 5 million Canadian blood donors has detected the first, and to date only, blood donor to be HCV positive prior to seroconversion (Patient A), that was infected by a defined source patient (S). E2 molecular evolution was monitored before and during seroconversion and the establishment of persistent infection (Larke et al., 2002), followed by an unresponsive IFN treatment. E2 HVR1 sequence comparisons of molecular clones \( (n = 117, \text{Fig. 6A}) \) from samples collected over this time showed that multiple forms of variants with different Ig similarity scores appeared until seroconversion and persistent infection, when the most divergent viral variants
Fig. 6. Amino acid sequence analysis of cloned variants in HCV 1b populations in a nosocomial patient during the early phases of HCV primary infection and IFN treatment (Larke et al., 2002). (A) Mutations in the E2 region (E2 aa8–72) are indicated relative to the source virus sequence (S1, direct sequence, (S-D)) for sequences of clones in sequential patient samples from patient A. Dots indicate amino acid identity to the source virus, (S-D). The percentage prevalence of each genotype is shown for each sequence in each population. (B) Viral load decreased on seroconversion and the establishment of persistent infection. RT-PCR products of serially diluted viral RNA prepared as described in Methods are shown following gel electrophoresis; the location of the amplified DNA is indicated with an arrow. Seroconversion was monitored by EIA and indicated as negative (−), weak reactive (+), and positive (+). (C) Evolution of Ig similarity in HCV 1a patient A (Pat-A) infected by a defined source patient S (Pat-S). Each symbol represents the similarity score for HVR1 (aa1–20 IMGT) of individual molecular clones sampled at each time point. Ig similarity scores represent average scores among 5 groups of Ig and TCR molecules. (D) Evolution of Ig similarity in an immunocompromised HCV 3a infected lung transplant patient infected from patient D (Pat-D). Each symbol represents the average of Ig group similarity scores for HVR1 (aa1–20 IMGT) of individual molecular clones sampled at each time point. The patient letter and time of sampling is indicated for each viral population.
those with >5 mutations) disappeared to be replaced by viruses having Ig scores similar to the source virus, (Fig. 6C). The sequence encompassing the amino terminal (1–72 aa) region of E2 was quite homogeneous in the quasispecies variants of the source virus (S1) from the patient with chronic HCV infection, where 94% (16/17) of the clones analyzed were almost identical, with the remaining 6% of clones differing by more than six substitutions (aa8–72 shown in Fig. 6A). There was significantly more sequence variation seen in the variants from the nosocomially infected patient (patient A), prior to seroconversion, whereas 4 weeks after seroconversion, all the clones in sample A 2.4 m (29/29) were found to be identical or genetically similar in sequence to clones (≤5 mutations) from the source virus (S1). The composition of the quasispecies shifted significantly during the seroconversion where the most divergent viral variants, disappeared after the development of antibody (P<0.0004 by Chi-square analysis). The quasispecies shift suggests that two types of quasispecies variants co-existed during acute infection, one was restricted (i.e., non-persistent form), while the other became persistent after the selective pressure of seroconversion (i.e., persistent forms of variants with few or no mutations). The disappearance of the majority of the more divergent variants coincided with a >10-fold decrease in viral load (Fig. 6B) as HCV-specific antibodies were fully developed in sample A 2.4 m (Fig. 6B). Although the population of persistent variants selected after seroconversion in the A 2.4 m sample possessed HVR1 sequences (IMGT aa1-20) that had not increased in Ig similarity relative to the population of viruses present prior to seroconversion, the least variant-type viruses constituted a genetically distinct population that resists negative selection by the immune response. This suggests that variants that were not recognized and eliminated by the immune response in patient A, and were genetically similar to the source virus population, may have possessed currently undefined, qualitative, or quantitative properties of mimicry that mediated immune evasion. However, IFN treatment of patient A resulted in the selection of mutants with significantly increased average Ig similarity scores (P = 4 × 10^{-12}) suggesting a role for Ig mimicry in IFN resistance. Finally, as a control, in the absence of an immune response and IFN therapy, HVR1 evolution in an immunosuppressed lung transplant patient involved fewer mutations and maintained relatively constant average Ig similarity (P > 0.05) (Fig. 6D).

Discussion

HCV E2 protein possesses neutralization epitopes within HVR1. Although highly variable, HVR1 maintains its conformation and charge (Penin et al., 2001) and also demonstrates convergent evolution (Casino et al., 1999) which is evidence of positive selection and functional importance (Brown et al., 2001). We provide structural and sequence data to indicate that HCV mimics immunoglobulins and furthermore that evolutionary change in the extent of mimicry correlates with the host-dependent ability of HCV to escape immune recognition. We thus propose that HCV hides from the immune system by assuming the structures of the most variable aspect of the immune system itself (i.e., the variable region). Molecular mimicry of host antibodies and TCR by HCV constitutes a unique means of circumventing the immune response because it is focused on the variable region of antibodies, which are both the effectors as well as targets for humoral immunity and thus constitute a family of tolerated antigens. Host-like antigenic structures are non-immunogenic due to tolerance mechanisms that prevent the synthesis of self-reactive antibodies (Rajewsky, 1996; Starr et al., 2003). The structural diversity of the Ig molecules makes them ideal targets for a flexible pattern of mimicry by forms of E2 which may account for some of the genotypic and quasispecies diversity seen among HCV viruses. It also becomes possible to speculate that E2, as a viral homologue of immunoglobulins, may function as an immunomodulator and affect the nature of the immune response, including immune dysfunction of T, NK, and dendritic cells (reviewed (Eisen-Vanderwelde et al., 2004) to further benefit HCV survival or replication. This would extend the demonstration of structural mimicry by viruses for immune evasion beyond the DNA viruses that posses a large number of immunoglobulin superfamily homologues, to RNA viruses that must be much more conservative in their encoded molecular mimics. The evolution of Ig mimicry not only provides a mechanistic explanation of HVR1 variability and loss of immunogenicity, but also insight into host-dependent disease patterns.

Host-dependent evolution of Ig mimicry

Human Ig structure is dependent on both the expression of specific germ-line alleles as well as adaptive somatic mutations to result in multiple groups of v-gene variants that could be characterized as quasispecies groups. If E2 is adapting to a population of host molecular structures, it is possible that E2 is either well or poorly matched to the immune repertoire of a given host. If E2 meets or exceeds the threshold of similarity needed to avoid detection, it will not induce an immune response and thus not be driven by antibody selection and would be expected to stay relatively stable with respect to Ig mimicry. Alternatively, a poor match to a novel immune repertoire will result in either a clearance of virus if E2 cannot efficiently evolve to evade immune detection (as seen for 15–20% of human infections) or alternatively will adapt to more closely match the new host environment. We thus expect 2 patterns for the evolution of mimicry in persistence; either
stasis with little or no evolution of Ig mimicry or increased mimicry following seroconversion. Indeed, we found both patterns of HVR1 evolution during persistent infection in humans and experimentally infected chimpanzees.

Host-dependent correlations with clinical outcome have been documented for HCV patients possessing specific IgG (GM and KM) and HLA (Type I and II) allotypes (Pandey et al., 2004; Thio et al., 2001, 2002; Thursz et al., 1999). The relationship of persistent HCV infection with Ig alleles may be due to differences in extent of structural mimicry of HCV E2 proteins with specific heavy (GM) and light (KM) allotypes or to possible epistatic relationships with adjacent variable genes (Pandey et al., 2004). Similarly, the ability of patients with specific HLA type I and II allotypes to clear HCV infection may involve specific structural relationships to specific HLA molecules, which are also members of the immunoglobulin superfamily.

The direct relationship between quasispecies complexity and persistence is consistent with an increased probability of mimicry to a given host where the greater the E2 structural diversity, the greater the chance that effective levels of mimicry exist to evade immune detection (Abbate et al., 2003; Arenas et al., 2004).

Ig mimicry and IFN

Several studies have shown an association between variation in HVR1 (Taylor et al., 2000) and resistance to IFN indicating that functions of this region play a role in circumventing the inhibitory effects of IFN. IFN treatment of patient A resulted in the selection of mutants with increased average Ig similarity scores suggesting a role for Ig mimicry in IFN resistance. This observation is also consistent with the correlation between IFN resistance and a lack of immunogenicity (Del Porto et al., 2000; Nakamoto et al., 1996) that could be due to increased host mimicry. However, retrospective analysis of Ig mimicry in several IFN-treated HCV patients (Boulestin et al., 2002; Penin et al., 2001) showed both increased and decreased similarity scores in IFN resistant variants following multiple IFN treatments indicating that overall average levels of immune mimicry (averaged among Ig and TCR types) does not simply explain the variation in HVR1 of IFN resistant variants (unpublished data E.G. Brown, M. Pelchat and Y.-W. Hu).

Ig mimicry and autoimmunity

Although it is believed that autoimmunity is initiated when the host and infecting agent share genetic sequence similarities (Oldstone, 1998, 2000), there are few demonstrations of such homologies in RNA viruses. However, studies show a role for antigenic mimicry of the human immunodeficiency virus gp41 motif, GTDRV, with the HLA II, GTERV, motif found in immune cells, where cross-reactive antibodies promote immune dysfunction (Powell et al., 2000). The identification of E2 protein domains with homology to Ig provides the identification of shared antigenic structures that may, through chronic stimulation, induce autoimmune diseases such as mixed type II cryoglobulinemia and others that are associated with 75% of chronic HCV infections (Dammacco et al., 2000; Major et al., 2004). Additionally, there is a further association between HCV patients with mixed type II cryoglobulinemia and lymphoproliferative disease seen as non-Hodgkin’s B-cell lymphoma, indicating a progression of the former to the latter and that both thus implicate chronic HCV antigenic stimulation (Dammacco et al., 2000). Evidence of Ig mimicry now incriminates E2 as a candidate antigen for the induction of autoimmunity.

Ig mimicry as a model for HCV persistence

The unique discovery of Ig mimicry by E2 provides a model for virus-host co-evolution as a function of Ig mimicry, where HCV possessing threshold levels of mimicry can circumvent host reactions. In fact, Ig mimicry may represent a general pattern of evolution among viruses and significantly modify our understanding of viral immune escape and persistence with implications to viral pathogenesis and therapy. We present a hypothesis of Ig mimicry that not only accounts for many facts about HCV infection, but also provides a basis for further investigations of the role of Ig-mimicry in HCV biology. Future characterizations of the relationships between E2 and members of the immunoglobulin superfamily may illuminate avenues for diagnosis, treatment and prevention not the least of which is immunization for which the viral receptor, E2, is an important candidate.

Methods

Patient samples

Two patients with primary HCV infections transmitted from known virus sources were enrolled. Patient A had acquired an acute nosocomial HCV 1a infection from chronically infected patient S (Larke et al., 2002). Two samples were obtained before and soon after seroconversion (A0 and A 1.2 m at days 36 and 46 post infection (PI), respectively) as well as sample A 2.4 m that was collected 4 weeks after seroconversion (day 72 PI). On day 74 pi, patient A was treated with IFN and ribavirin for 6 months. Samples were also collected 1 and 2 months after 6 months of IFN treatment. Patient B was a lung transplant patient who had cleared a previous HCV infection and subsequently acquired HCV 3a from an organ donor (patient D). Serial samples were taken from patient B; on days 8, 50, 78, 109, and 171 after transplant/infection. HVR1 sequences in published follow-up studies in humans (Kato et al., 1994; Larke et
and chimpanzees (Lu et al., 2001; van Doorn et al., 1995) were analyzed.

**Anti-HCV antibody testing**

Anti-HCV was detected using an EIA (Ortho HCV version 3.0 ELISA, Ortho Diagnostic Systems Inc., Raritan, NJ or Abbott AxSym EIA) and confirmed using RIBA (Chiron RIBA HCV 3.0 SIA, CHIRON, Emeryville, CA).

**Cloning and sequencing**

Viral RNA was extracted from plasma using the QIAamp Viral RNA Kit (QIAGEN, Germany) for cDNA synthesis used random hexamer primer (pd(N)6) and the Amersham Pharmacia Biotech First-Strand cDNA Synthesis Kit. Nested RT-PCR was used to amplify the E2 region before cloning and sequencing as described previously (Larke et al., 2002).

**Computer analysis of E2 sequences**

Viral sequences were derived from patient samples, Genbank or the LANL HCV database (hev.lanl.gov). Ig genes were obtained from the NREF database. Sequences were locally aligned using ClustalW v1.82 and BLAST v2.2.6 followed by manual alignment according to the IMGT numbering scheme (imgt.cines.fr, (Lefranc et al., 2003)) which is approved by the Human Genome Organization (HUGO) Nomenclature Committee (HGNC). HVR1 sequences were then tested for Ig similarity (relative to IMGT alignment files of the variable domains of human heavy (n = 984); light kappa (n = 843); light lambda (n = 1212); TCRα (n = 111); and TCRβ (n = 346) chains) using a scoring system based on a position-specific scoring matrix that quantifies the physical and chemical properties of aas.

Specifically, for a given region of comparison, the similarities of each constituent aa to the corresponding position of reference sequences, was multiplied by a value found in the updated Dayhoff aa similarity matrix (Feng et al., 1984; Jones et al., 1992), derived from the observed frequencies of aa substitutions in related proteins which is a function of the physical (size, charge, and hydrophathy) and chemical similarities between aas. Using this algorithm, similarity scores range up to 1500/aa for identity (maximum score = (1500 L), where L = sequence length). Other similarity scoring matrices were used including an identity matrix, PAM250 and the BLOSUM62 matrices. All the matrices produced results supporting similar trends of change for HVR1 evolution (not shown). The three-dimensional model structure of the N-terminal portion of E2 was made using the SWISS-MODEL server (Schwede et al., 2003) for HCV genotype 2a (AAF59944) relative to the tertiary structure of the variable domain of humanized antibody 4D5 (PDB:1FVD). The model image was generated with both VMD v1.8.2 and Raster3D v2.6.

**Statistical analysis**

Population prevalence of differences in mutant populations was determined by Chi-squared analysis and differences in Ig similarity scores were determined using the Student’s t-test.

**Expression of recombinant HCV E2 proteins**

Full length (aa1–322) or amino terminal regions of E2 genes (aa1–113 or 123) from HCV 1a and 3a variants were inserted into the baculovirus expression vector using the BaculoGold kit (BD Biosciences, Mississauga, Canada) as described previously (Hu et al., 1996) or the pET17a vector for expression in E. coli as indicated by the manufacturer (Novagen, San Diego). A methionine and six histidines were added to the N-terminus of each construct. Forms of E2 protein isolated from Patient A and B were expressed in Spodoptera frugiperda (Sf9) insect cells and in E. coli BL21 cells induced with IPTG. Expressed protein sequences are in a Supplementary note.

**Western blotting**

Western blotting was carried out using equal amounts of crude protein lysates as previously described (Hu et al., 1996). Recombinant protein expression levels were assessed using anti-histidine (His) antibody staining (1:5000 dilution of 6× His monoclonal antibody, BD Biosciences, US) and secondary antibody, alkaline phosphatase conjugated anti-mouse polyclonal immunoglobulins (1:15000 dilution, Sigma, US) or alternatively were stained with Coomassie blue. Recombinant E2 proteins were also reacted with commercial preparations of alkaline phosphatase conjugated goat antibodies raised against the Fab fragment of human IgG, (binding was also obtained with goat antibodies against whole human IgG, not shown) (catalogue no. A8542 and A3312, respectively, Sigma-Aldrich Canada Ltd.). Antibody was visualized by development with NBT and BCIP.

**Acknowledgments**

Signed consent was obtained from human subjects to whom we are sincerely grateful. This work was funded by the Canadian Blood Services and the Natural Sciences and Engineering Research Council of Canada to EGB 41771–2001. Helpful criticism was provided by J. Webb.

**Appendix A. Supplementary data**

References


Harpaz, Y., Chothia, C., 1994. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. J. Mol. Biol. 238, 528–539.


charges of hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment. J. Virol. 75, 5703–5710.