In Silico Biology 11 (2011/2012) 163–173
DOI 10.3233/ISB-2012-0451
IOS Press

Differences in variability of hypervariable region 1 of hepatitis C virus (HCV) between acute and chronic stages of HCV infection

I.V. Astrakhantseva*,†, D.S. Campo†, A. Araujo, C.-G. Teo, Y. Khudyakov and S. Kamili
Laboratory Branch, Division of Viral Hepatitis, Center for Disease Control and Prevention, Atlanta, GA, USA

Received 16 January 2012
Revised 19 June 2012
Accepted 6 July 2012

Abstract. Distinguishing between acute and chronic HCV infections is clinically important given that early treatment of infected patients leads to high rates of sustained virological response. Analysis of 2179 clonal sequences derived from hypervariable region 1 (HVR1) of the HCV genome in samples obtained from patients with acute (n = 49) and chronic (n = 102) HCV infection showed that intra-host HVR1 diversity was 1.8 times higher in patients with chronic than acute infection. Significant differences in frequencies of 5 amino acids (positions 5, 7, 12, 16 and 18) and the average genetic distances among intra-host HVR1 variants were found using analysis of molecular variance. Differences were also observed in the polarity, volume and hydrophobicity of 10 amino acids (at positions 1, 4, 5, 12, 14, 15, 16, 21, 22 and 29). Based on these properties, a classification model could be constructed, which permitted HVR1 variants from acute and chronic cases to be discriminated with an accuracy of 88%. Progression from acute to chronic stage of HCV infection is accompanied by characteristic changes in amino acid composition of HVR1. Identifying these changes may permit diagnosis of recent HCV infection.

Keywords: HCV infection, acute, chronic, quasispecies, hypervariable region 1

1. Introduction

Hepatitis C virus (HCV) is a single-stranded RNA virus belonging to the *Flaviviridae* family [14]. It infects 2%–3% of the world's population and is a major cause of liver disease worldwide [2]. HCV infection progresses to chronicity in 70%–85% of infected adults [1].

Acute hepatitis occurs 2–12 weeks after exposure and may last up to 12 weeks [35]. It can be severe and prolonged but is seldom fulminant [34]. Viremia persisting for >6 months indicates establishment of chronic infection. The rate of spontaneous clearance after acute HCV infection is not known but is estimated to be 20%–25% [15]. Humoral immune responses seem to have little effect on viral clearance; no specific antibody responses have been found to predict the outcome of infection. Cellular immune responses have an important role in determining the outcome of acute HCV infection; studies on patients and experimentally infected chimpanzees suggest that clearance of viremia is associated with vigorous CD4+ and CD8+ cellular responses [16,18,32,51].

Discrimination between acute and chronic HCV infection is clinically important as identification of patients with early infection followed by their treatment with interferon and ribavirin can lead to eradication of the virus in >95% of cases, compared to 40%–80% of cases if the treatment starts during the chronic phase

*Corresponding author: I.V. Astrakhantsev, Laboratory Branch, Division of Viral Hepatitis, Center for Disease Control and Prevention, 1600 Clifton Rd, MS A-33, Atlanta, GA 30300, USA. Tel.: +1 404 639 1150; Fax: +1 404 639 1563; E-mail: irenastr@rambler.ru.
†Authors contributed equally.

of infection [42]. Examination for seroconversion of anti-HCV IgG antibodies or appearance in blood of HCV RNA in serial specimens is the only means for detecting acute infections. We previously reported two approaches that potentially allow the serological diagnosis of acute HCV infection. One approach is based on the measurement of avidity of IgG anti-HCV, and the other, a high-throughput microsphere immunoassay which simultaneously detects IgG anti-HCV responses to multiple structural and nonstructural HCV proteins [4,25].

It has been reported that clearance of acute HCV infection is associated with reduction of quasispecies diversity in the hypervariable region 1 (HVR1) of HCV [20]. The high sequence variability of HVR1 correlates with the emergence of antibody-escape mutants, and appears to play an important role in maintenance of viral persistence during chronic infection [19,29,36, 49,53,55]. Taking these observations into consideration, we investigated the HVR1 quasispecies in patients with acute and chronic infections to identify changes in the HVR1 structure that may be associated with progression of HCV infection to chronicity.

The results indicate the existence of patterns of sequence changes in HVR1 quasispecies that are specific to acute and to chronic stages of infection, indicating that chronic infection has a measurable effect on the amino acid (aa) composition of HVR1 sequences and their physicochemical properties.

2. Results

2.1. Genetic structure

The aa diversity of chronic samples was significantly higher than the diversity of acute samples ($p = 0.008$; CI 95% for acute infection: 0.02–0.07, for chronic infection: 0.07–0.11). Figure 1 shows the aa diversity for each polymorphic position, 23 of which had significantly different mean values of diversity between the two groups. The overall dN/dS ratio of the HVR1 segment was not significantly different between the two groups ($p = 0.065$). However, we found sequences from more chronic patients ($n = 24$) showing individual sites with positive selection than sequences from acute patients ($n = 2$). Brown et al. [7] found 6 sites of HVR1 being positively selected in four genotypes of HCV (sites 1, 8, 12, 14, 21 and 22). We have found that there are number of additional sites (11, 13, 15–19) (10 sites in total) to be positively selected in 15 of chronic patients, and two sites (14 and 21) to be positively selected in two acute patients. Lara et al. also observed that HVR1 contains 5 of these sites (8, 11, 14, 17, and 18) whose changes correlate with outcome of INF therapy [30].

We estimated the population differentiation between acute and chronic stages by means of AMOVA. A small but significant portion of the variance in genetic distances was found to be due to the differences between

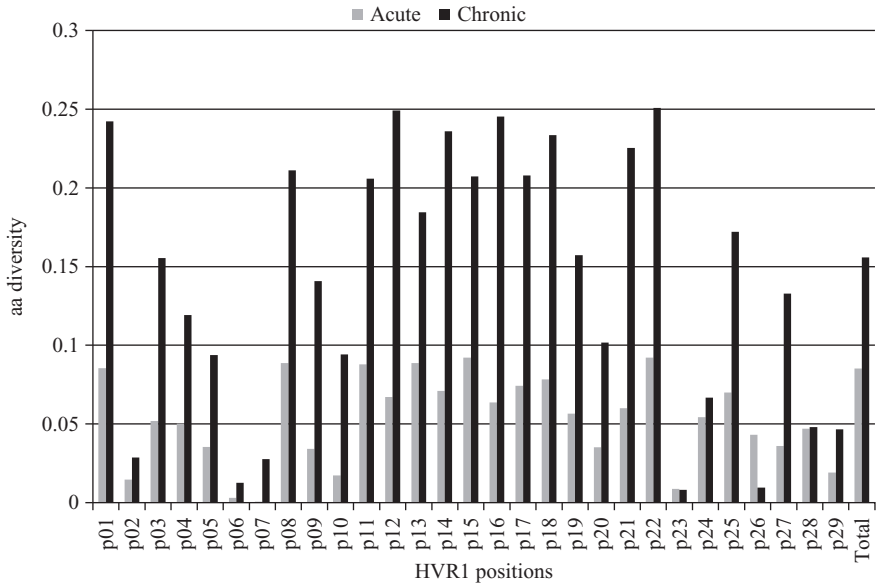

Fig. 1. Amino acid diversity of HVR1. Average diversity in samples of each group is shown, grey for acute samples and black for chronic samples.

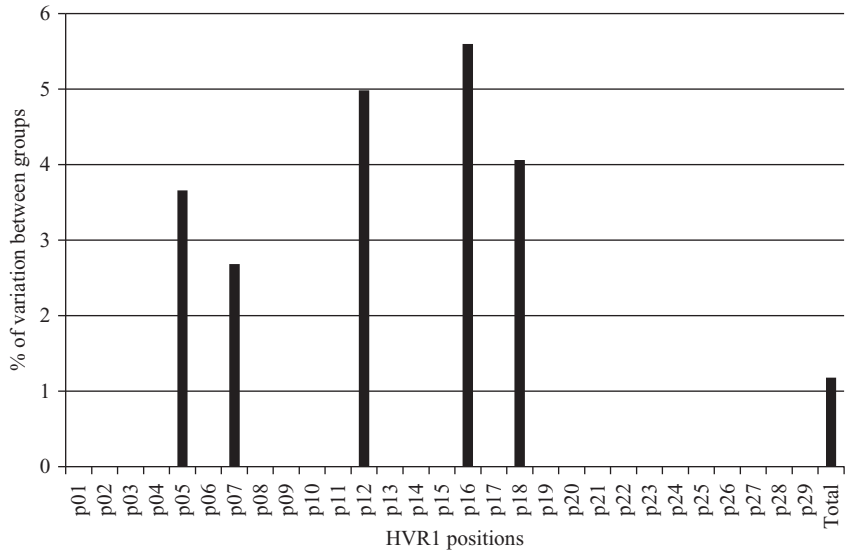

Fig. 2. Differences in HVR1 variance by AMOVA. Y axis relate to percentage of variance due to differences between acute and chronic groups. Only significantly different positions ($p < 0.05$) are shown.

the two groups (1.18% variations, $p = 0.0003$). To investigate further this difference, we carried out an AMOVA for each position (Fig. 2) and found 5 aa positions, which significantly contributed to differentiation of the two groups: positions 5 (4.71% variation, $p = 0.02$), 7 (1.76% variations, $p = 0.04$), 12 (4.49% variations, $p = 0.002$), 16 (4.2% variations, $p = 0.02$) and 18 (4.54% variations, $p = 0.007$). When this analysis was repeated using only genotype 1 samples (which constituted 88.6% of all samples) a similar result was produced (0.82%, $p = 0.01$).

2.2. Physicochemical properties

Given that there are small but significant differences between sequences from acute and chronic samples, we explored how the aa differences affected the physicochemical property of each position and of the whole segment (Fig. 3). There were significant differences observed between the two groups when all properties were considered (MRPP test; $p = 0.0109$). When only one given property is considered for all positions, the following properties showed significant differences between the two groups: number of basic residues (MRPP test; $p = 0.0090$), isoelectric point (MRPP test; $p = 0.0044$), volume (MRPP test; $p = 0.0087$), net charge (MRPP test; $p = 0.0117$) and accessible surface area (MRPP test; $p = 0.0187$). (All the p values were generated by the MRPP test.) Significant differences in polarity, volume and hydrophobicity were found for aa at 10 positions: 1, 4, 5, 12, 14–16, 21, 22 and 29 (Fig. 3). Three of these positions were identical to the ones which also showed significantly different aa frequencies in AMOVA (positions 5, 12 and 16). The strongest differences were found in surface accessibility, polarity and volume of aas located at the HVR1 center. The hydrophobicity ($p = 0.0015$), polarity ($p = 0.0423$), isoelectric point ($p = 0.0066$) and the number of basic residues ($p = 0.0023$) in the entire HVR1 showed significantly higher mean values in the chronic than in the acute samples.

2.3. Classification of consensus sequences

We explored the possibility of using the sequence information to create a classification model for distinguishing acute from chronic samples. Figure 4 shows the best model constructed using linear projection of the data based on 18 variables. This model uses the following properties: hydrophobicity (for positions 5 and 12, mean), flexibility (positions 12 and 16), polarity (position 29), alpha-helix (positions 4 and 12), isoelectric point (positions 16 and 19), beta sheet (positions 12 and 29), volume (positions 11, 15 and 29), protein composition (positions 16 and 18) and the number of basic residues of the whole segment. The separation between the two groups is clear, yielding a leave-one-out cross-validation accuracy of 88% (the area under the receiver

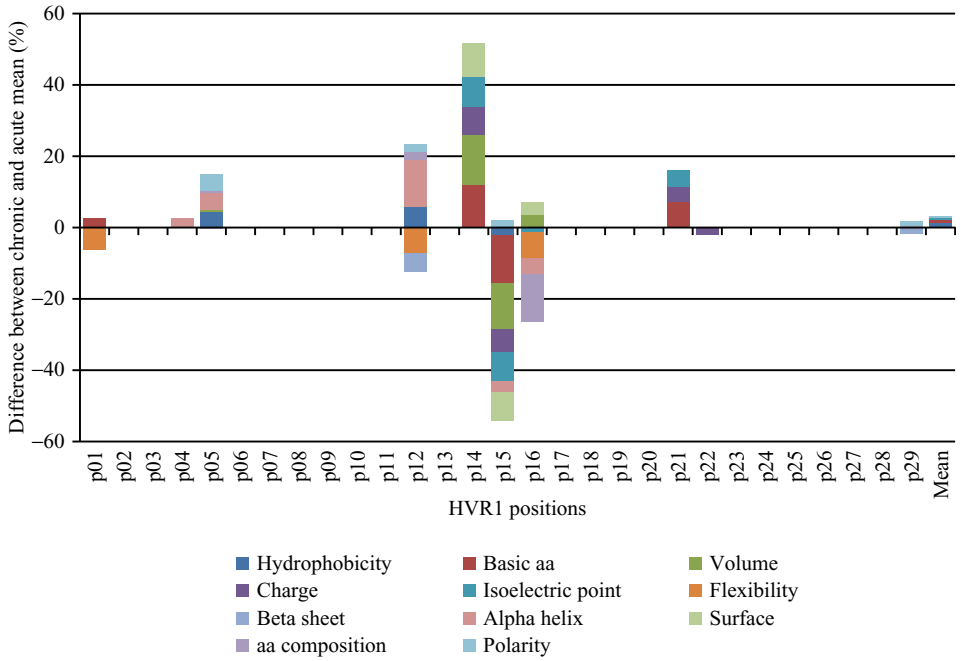

Fig. 3. Physicochemical properties of HVR1. Y axis relates to difference between mean of each property in chronic samples minus mean of the same property in acute samples. Only significantly different variables ($p < 0.05$) are shown.

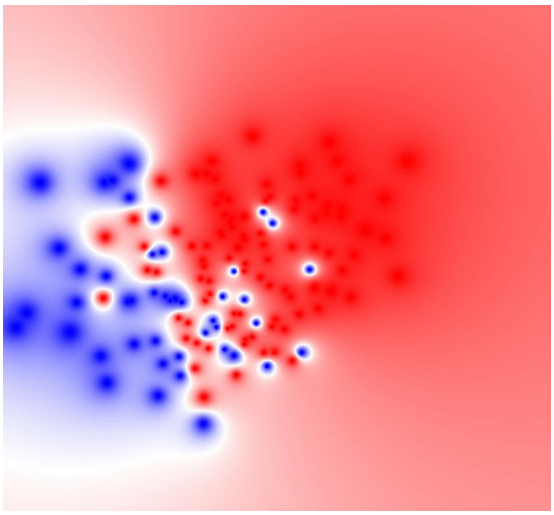

Fig. 4. RadViz projection of HVR1 sequences based on 18 variables. Blue corresponds to acute samples and red to chronic samples. Color density is proportional to probability that a sequence projected belongs to that class.

operating characteristic curve being 0.94). We couldn't find any pattern shared by the misclassified samples in terms of their source (previously published or produced in our lab), nor in its estimated time of infection.

3. Discussion

It has been previously reported that HVR1 diversity is lower in people with acute HCV infection compared to those with chronic infection, and this difference in is the most important molecular variable differentiating the two stages of HCV infection [20,46,54]. Although it has been shown that negative selection increases over time during late HCV intra-host evolution [44], we did not find significant differences in the overall dN/dS ratio between the acute and chronic samples studied here. However, many chronic patients showed evidence of natural selection at individual sites. The type of selection was not strictly connected to a particular site, as the same site can be under different types of selection in different patients (Table 1). The evidence of the HVR1 positive selection was found in a few acutely infected patients. This finding suggests the lack of selection or weak selection pressures acting on this HCV genomic region during early stages of infection, possibly owing to the absence or low avidity of anti-HVR1 activity [16,25]. However, the HCV HVR1 heterogeneity may be affected by many selection pressures during chronic infection, which is in accord with the observation of the increasing number of positively selected sites at the early stages of chronicity [52].

Table 1
Sample characteristics

ID	Source	Div	dN/dS	Group	HCV Genotype	Method	Clones n	Positively selected sites	Negatively selected sites
CA_2	Complex Antibody	0.0026	0.0911	Acute	1a	EPLD	35		
CA_5	Complex Antibody	0.0029	0.1987	Acute	1b	EPLD	40		
F_101	Farci et al, 2000	0.0584	1.4343	Acute	1a	Cloning	10		
F_11	Farci et al, 2000	0.0172	0.0000	Acute	1b	Cloning	10		15
F_111	Farci et al, 2000	0.0347	2.2190	Acute	1a	Cloning	10		
F_121	Farci et al, 2000	0.0623	0.8112	Acute	1b	Cloning	10		16
F_21	Farci et al, 2000	0.0218	1.5365	Acute	1a	Cloning	8		
F_31	Farci et al, 2000	0.6667	0.0000	Acute	2a	Cloning	9		
F_41	Farci et al, 2000	0.0513	1.0861	Acute	1a	Cloning	10		19
F_51	Farci et al, 2000	0.0350	2.0716	Acute	1a	Cloning	8		
F_61	Farci et al, 2000	0.0259	0.0461	Acute	1b	Cloning	11		5;20
F_71	Farci et al, 2000	0.0471	2.2367	Acute	1b	Cloning	10		
F_81	Farci et al, 2000	0.0096	3.0000	Acute	1b	Cloning	7		
F_91	Farci et al, 2000	0.0427	1.4720	Acute	1a	Cloning	10		
H_PD10017	Herring et al, 2005	0.0791	0.8094	Acute	1a	Cloning	14		
H_PD10022	Herring et al, 2005	0.1190	1.3230	Acute	1a	Cloning	10		
H_PD10051	Herring et al, 2005	0.0105	1.4776	Acute	1b	Cloning	11		
H_PD10081	Herring et al, 2005	0.1548	0.8686	Acute	1a	Cloning	10		
H_PD10082	Herring et al, 2005	0.0575	0.0000	Acute	1a	Cloning	12		
H_PD2003	Herring et al, 2005	0.0059	3.0000	Acute	1a	Cloning	11		
H_PD2004	Herring et al, 2005	0.0115	0.9361	Acute	1a	Cloning	8		
H_PD2013	Herring et al, 2005	0.0148	3.0000	Acute	1b	Cloning	12		
H_PD2019	Herring et al, 2005	0.0029	0.3851	Acute	1a	Cloning	16		
H_PD2021	Herring et al, 2005	0.0524	0.6129	Acute	1a	Cloning	9		
H_PD2023	Herring et al, 2005	0.0655	0.7895	Acute	1a	Cloning	5		
H_PD2024	Herring et al, 2005	0.0128	0.3906	Acute	1a	Cloning	9		
H_TR10	Herring et al, 2005	0.0099	3.0000	Acute	1a	Cloning	7		
H_TR17	Herring et al, 2005	0.0219	0.4667	Acute	1b	Cloning	7		
H_TR18	Herring et al, 2005	0.0799	2.1967	Acute	1b	Cloning	13		5
H_TR2	Herring et al, 2005	0.0100	0.1374	Acute	1b	Cloning	11		
H_TR7	Herring et al, 2005	0.0214	0.1838	Acute	1b	Cloning	10		
H_TR8	Herring et al, 2005	0.2222	0.0000	Acute	1b	Cloning	9		
N_15v11	Lui et al, 2010	0.0045	3.0000	Acute	1a	Cloning	4		
N_161v11	Lui et al, 2010	0.0085	0.5139	Acute	1a	Cloning	5		
N_29vSD	Lui et al, 2010	0.0026	0.3808	Acute	1a	Cloning	3		
N_30vSD	Lui et al, 2010	0.0021	3.0000	Acute	1a	Cloning	3		
N_56d03	Lui et al, 2010	0.0010	0.0000	Acute	1a	Cloning	2		
N_65v05	Lui et al, 2010	0.0025	3.0000	Acute	1a	Cloning	3		
VH_H_77	Von hahn et al, 2007	0.0431	2.3783	Acute	1a	Cloning	8		
Z_6212	Zeptometrix	0.0225	0.9573	Acute	1	EPLD	20		14;25
Z_6213	Zeptometrix	0.0812	0.7770	Acute	1a	EPLD	34		16;19;29
Z_6214	Zeptometrix	0.0294	3.0000	Acute	1	EPLD	43	21	26
Z_6215	Zeptometrix	0.0586	1.2611	Acute	1a	EPLD	27	14	
Z_6228	Zeptometrix	0.0039	0.8859	Acute	1a	EPLD	22		
Z_9041	Zeptometrix	0.0499	0.0000	Acute	1a	EPLD	27		
Z_9046	Zeptometrix	0.0000	0.0000	Acute	1a	EPLD	29		
Z_9047	Zeptometrix	0.0917	1.6694	Acute	1a	EPLD	44		
Z_9054	Zeptometrix	0.0690	0.0000	Acute	3a	EPLD	29		
Z_9058	Zeptometrix	0.0210	0.4097	Acute	1a	EPLD	25		
B_11	BBI	0.0259	0.7397	Chronic	1a	EPLD	39		
B_12	BBI	0.0613	0.4675	Chronic	1a	EPLD	3		
B_14	BBI	0.0665	0.9780	Chronic	1a	EPLD	22		
B_19	BBI	0.1796	1.0504	Chronic	1a	EPLD	27		
B_20	BBI	0.0370	0.3131	Chronic	1b	EPLD	23		
B_21	BBI	0.0070	0.4309	Chronic	1a	EPLD	9		
H_SP14	Herring et al, 2005	0.0777	0.9171	Chronic	1a	Cloning	10		
H_SP15	Herring et al, 2005	0.0431	0.8757	Chronic	1b	Cloning	8		

(*Continued*)

Table 1. Sample characteristics (Continued)

ID	Source	Div	dN/dS	Group	HCV Genotype	Method	Clones n	Positively selected sites	Negatively selected sites
H_SP2	Herring et al, 2005	0.1241	1.8962	Chronic	1a	Cloning	10		
H_SP4	Herring et al, 2005	0.0920	1.3677	Chronic	1a	Cloning	10		
H_SP5	Herring et al, 2005	0.0051	0.5101	Chronic	1a	Cloning	9		
H_SP9	Herring et al, 2005	0.1379	2.5136	Chronic	1a	Cloning	6		
N_512743	Alter et al, 1999	0.2026	0.7503	Chronic	1b	Cloning	22	21	23
N_512754	Alter et al, 1999	0.0180	2.2270	Chronic	1b	Cloning	13		
N_512834	Alter et al, 1999	0.1279	3.0000	Chronic	1b	Cloning	19		
N_512859	Alter et al, 1999	0.1279	2.0608	Chronic	2b	Cloning	19	22	
N_512863	Alter et al, 1999	0.0115	3.0000	Chronic	1a	Cloning	4		
N_513014	Alter et al, 1999	0.0180	3.0000	Chronic	1a	Cloning	13		
N_513125	Alter et al, 1999	0.0178	0.9687	Chronic	3a	Cloning	12		
N_513174	Alter et al, 1999	0.0192	1.9390	Chronic	1a	Cloning	10		
N_513492	Alter et al, 1999	0.2581	2.0061	Chronic	1a	Cloning	29	13	
N_513835	Alter et al, 1999	0.0583	2.0065	Chronic	1b	Cloning	17		
N_514366	Alter et al, 1999	0.0176	0.7309	Chronic	1a	Cloning	11		26
N_514374	Alter et al, 1999	0.0115	0.0000	Chronic	1b	Cloning	4		
N_514418	Alter et al, 1999	0.2733	1.2103	Chronic	1a	Cloning	9		
N_514640	Alter et al, 1999	0.0176	1.6855	Chronic	1b	Cloning	11		
N_514880	Alter et al, 1999	0.0173	1.3493	Chronic	1a	Cloning	15		19;20
N_515000	Alter et al, 1999	0.2381	1.6761	Chronic	1a	Cloning	8		
N_515010	Alter et al, 1999	0.2733	0.9559	Chronic	1a	Cloning	9		
N_515163	Alter et al, 1999	0.0192	3.0000	Chronic	1b	Cloning	10		
N_515260	Alter et al, 1999	0.2581	1.0411	Chronic	1b	Cloning	29	11	5;7
N_515467	Alter et al, 1999	0.2481	1.4605	Chronic	1b	Cloning	28		20
N_515603	Alter et al, 1999	0.2733	2.8912	Chronic	4a	Cloning	9	18	
N_515675	Alter et al, 1999	0.0184	0.6855	Chronic	1b	Cloning	14		20
N_515727	Alter et al, 1999	0.0178	1.4520	Chronic	1b	Cloning	12		
N_516476	Alter et al, 1999	0.0176	0.9422	Chronic	1a	Cloning	11		
N_516655	Alter et al, 1999	0.2381	0.9078	Chronic	1b	Cloning	8		24
N_516733	Alter et al, 1999	0.0184	0.3429	Chronic	1b	Cloning	14		20;26;28
N_516833	Alter et al, 1999	0.0176	3.0000	Chronic	1a	Cloning	11		27
N_517557	Alter et al, 1999	0.2186	2.3907	Chronic	2b	Cloning	23	12;17;19;22	6
N_518765	Alter et al, 1999	0.0178	1.2952	Chronic	1a	Cloning	23		
N_519281	Alter et al, 1999	0.1831	1.6514	Chronic	1a	Cloning	21	17;22	
N_519739	Alter et al, 1999	0.2186	0.7421	Chronic	1a	Cloning	13		
N_520124	Alter et al, 1999	0.0180	3.0000	Chronic	2b	Cloning	4		
N_520837	Alter et al, 1999	0.0115	2.4211	Chronic	2b	Cloning	11	22	
N_522063	Alter et al, 1999	0.0176	0.6844	Chronic	1a	Cloning	12		23
N_522754	Alter et al, 1999	0.0178	0.6602	Chronic	1a	Cloning	14		16;23
N_522760	Alter et al, 1999	0.0184	0.9272	Chronic	1a	Cloning	25		13
N_523281	Alter et al, 1999	0.2370	0.6355	Chronic	1a	Cloning	23		
N_523401	Alter et al, 1999	0.2186	0.9846	Chronic	1a	Cloning	22		
N_523730	Alter et al, 1999	0.2026	1.1265	Chronic	1a	Cloning	33		24
N_523777	Alter et al, 1999	0.2846	3.0000	Chronic	1a	Cloning	12		18
N_523866	Alter et al, 1999	0.0178	1.4639	Chronic	1a	Cloning	19	28	10
N_524253	Alter et al, 1999	0.1279	2.3781	Chronic	3a	Cloning	15		
N_524359	Alter et al, 1999	0.0173	2.0787	Chronic	1a	Cloning	12	11	
N_524547	Alter et al, 1999	0.0178	1.4821	Chronic	1a	Cloning	16	16	
N_524972	Alter et al, 1999	0.0177	0.1577	Chronic	1a	Cloning	8		13
N_525357	Alter et al, 1999	0.2381	0.7329	Chronic	1a	Cloning	17		
N_525442	Alter et al, 1999	0.0583	1.6124	Chronic	1b	Cloning	9		11;20
N_525738	Alter et al, 1999	0.2733	1.1545	Chronic	2b	Cloning	12		5;16;25
N_526075	Alter et al, 1999	0.0178	0.8120	Chronic	1a	Cloning	12		
N_526134	Alter et al, 1999	0.0178	1.8283	Chronic	3a	Cloning	9	1	
N_526135	Alter et al, 1999	0.2733	1.3078	Chronic	3a	Cloning	8		
N_526330	Alter et al, 1999	0.2381	2.6611	Chronic	1a	Cloning	12		
N_526936	Alter et al, 1999	0.0178	0.2053	Chronic	1a	Cloning	13		1;22
N_527026	Alter et al, 1999	0.0180	0.3222	Chronic	1b	Cloning	17		
N_527166	Alter et al, 1999	0.0583	1.2629	Chronic	1a	Cloning	28	14	25

Table 1. Sample characteristics (Continued)

ID	Source	Div	dN/dS	Group	HCV Genotype	Method	Clones n	Positively selected sites	Negatively selected sites
N_527183	Alter et al, 1999	0.2481	1.7260	Chronic	2b	Cloning	20	22	26
N_527517	Alter et al, 1999	0.1587	1.8093	Chronic	1b	Cloning	12		18;20
N_527518	Alter et al, 1999	0.0178	0.5738	Chronic	1b	Cloning	22		4;16;19;27
N_527962	Alter et al, 1999	0.2026	1.3832	Chronic	1b	Cloning	23	8;21;22;25	2;4;14;20
N_528039	Alter et al, 1999	0.2186	0.6266	Chronic	2	Cloning	10		
N_528044	Alter et al, 1999	0.0192	0.1568	Chronic	1a	Cloning	7		6
N_528140	Alter et al, 1999	0.1834	0.1183	Chronic	1b	Cloning	16		21;25
N_528772	Alter et al, 1999	0.0177	0.9095	Chronic	3a	Cloning	18		13;26
N_528912	Alter et al, 1999	0.0864	3.0000	Chronic	2b	Cloning	18	13	20
N_530382	Alter et al, 1999	0.0864	1.3519	Chronic	1a	Cloning	12		6
N_530446	Alter et al, 1999	0.0178	3.0000	Chronic	1a	Cloning	19		
N_532325	Alter et al, 1999	0.1279	3.0000	Chronic	2a	Cloning	4		
N_532352	Alter et al, 1999	0.0115	0.9146	Chronic	1a	Cloning	18	8;17	1;11;19;28
N_532354	Alter et al, 1999	0.0864	0.4757	Chronic	1a	Cloning	14		1;3;22
N_533676	Alter et al, 1999	0.0184	0.1438	Chronic	1b	Cloning	13		3;10;15;18;23;27
N_533678	Alter et al, 1999	0.0180	0.2704	Chronic	1b	Cloning	8		5;23
N_535154	Alter et al, 1999	0.0864	0.0627	Chronic	6a	Cloning	12		16
N_535318	Alter et al, 1999	0.0178	1.0962	Chronic	6x	Cloning	8	1	14;20;25
N_535363	Alter et al, 1999	0.2381	1.1902	Chronic	6a	Cloning	23	13	28
N_536048	Alter et al, 1999	0.2186	2.9551	Chronic	1b	Cloning	14		
N_537410	Alter et al, 1999	0.0184	0.1612	Chronic	1a	Cloning	17	17	20;23;9;16
N_537415	Alter et al, 1999	0.0583	0.4487	Chronic	1a	Cloning	10		15;22;23
N_538569	Alter et al, 1999	0.0192	0.4713	Chronic	1b	Cloning	16	22	19;23;24;27
N_539751	Alter et al, 1999	0.0177	0.5159	Chronic	1b	Cloning	24		6;20;23;28
N_539765	Alter et al, 1999	0.2325	0.4893	Chronic	1b	Cloning	10		6
N_540394	Alter et al, 1999	0.0192	0.4703	Chronic	1a	Cloning	12		
N_541805	Alter et al, 1999	0.0178	0.2831	Chronic	1a	Cloning	14		2;12
N_542587	Alter et al, 1999	0.0184	1.1558	Chronic	1a	Cloning	15	8	29
N_542594	Alter et al, 1999	0.0173	2.0982	Chronic	1a	Cloning	15	15;22	4;19
N_542732	Alter et al, 1999	0.0173	0.5496	Chronic	1b	Cloning	7		7;19;24
N_542751	Alter et al, 1999	0.1834	0.3024	Chronic	6e	Cloning	12		21
N_542971	Alter et al, 1999	0.0178	0.7402	Chronic	1a	Cloning	7		22
N_542975	Alter et al, 1999	0.1834	2.2710	Chronic	1a	Cloning	18	8;15	
N_544314	Alter et al, 1999	0.0864	0.4185	Chronic	1a	Cloning	5		
N_544735	Alter et al, 1999	0.0508	1.0404	Chronic	1a	Cloning	16		12;16

In the present study, we found that a small but significant portion of genetic differences among HVR1 variants was associated with distinction between acute and chronic stages. Furthermore, this difference was also significant for certain key HVR1 positions and were not related to the genotype distribution among HCV strains from each group of samples, as the pattern remained unchanged when analysis was repeated for genotype 1 strains alone.

There were significant changes in the aa frequencies at the HVR1 positions 5, 7 and 12, all of which belong to the antigenic epitope previously identified to be at positions 1–13 [43], suggesting that changes at these positions may be the result of immune selection. The observed differences in hydrophobicity and accessible surface area may also reflect variation in conformation or antigenic reactivity, or both, of the epitope. In support of this supposition, the HVR1 positions with the highest variation of aa physicochemical properties between acute and chronic infections were located in the central part of HVR1, which has been predicted to adopt a secondary structure [23,43]. Despite the significant sequence variability, HVR1 maintains a stable content of basic aas located at specific positions [10,43]. Changes in the basic properties of HVR1 have been linked to HCV infectivity *in vitro* [10]. The HVR1 sequences identified from the acute and chronic cases had statistically significant differences in basic aa composition, suggesting variation in requirements for HCV infectivity at different stages of infection.

There are strong structural constraints that contribute to conservation of the HVR1 physicochemical

properties and conformation [11,39,43]. Selection pressures seemed to restrict HVR1 sequence variability [12,36,48,50], indicating a greater importance of this region for virus replication other than being an antigenic decoy [36,50]. The functional significance of HVR1 was recently confirmed by the observation of attenuated infection of chimpanzee with an HCV strain lacking HVR1 [21]. Additionally, experiments in cell culture showed that HVR1 is important for virus replication and persistence [5]. Identification of such extensive epistatic connectivity of HVR1 sites across the entire HCV genome further validates functional importance of this region and its contribution to HCV adaptation [12], and also suggests that variation in HVR1 may reflect selection pressures acting somewhere else in the HCV genome [30].

In conclusion, the data obtained in this study show that the composition of HVR1 quasispecies is different between the acute and the chronic stages of infection. Such difference could be exploited to identify patients with recent HCV in order that they may be referred to early antiviral treatment.

4. Materials and methods

4.1. Samples

4.1.1. Samples from acutely and chronically infected patients

We analyzed 2179 HVR1 sequences from 151 samples. The "acute" group included 700 clonal sequences from 49 patients: sequences from 12 patients were generated in our laboratory and 37 were reported from previous studies [20,22,33,54]. The "chronic" group included 1479 clonal sequences from 102 patients: 96 whose sera were sequenced in our laboratory [3,13], and 6 from previous studies [22] (Table 1).

4.1.1.1. Details of the acute samples analyzed in our laboratory

Ten plasma samples were obtained from Zeptometrix (Buffalo, NY) and 2 samples from Complex Antibody (FL, USA). They originated from untreated patients with asymptomatic HCV infections. All samples were HCV-RNA-positive and anti-HCV-negative.

4.1.1.2. Details of the chronic samples analyzed in our laboratory

For 90 of the new chronic samples, details can be found in [3,13]. For the other six, anti-HCV-positive plasma samples were sourced from HCV-RNA-positive, first-time blood donors (SeroCare, MA). All anti-HCV-IgG results were confirmed by a recombinant immunoblot assay (Roche Diagnostics, Mannheim, Germany).

4.2. RNA extraction, reverse transcription and real-time nested PCR

Total nucleic acids from the specimens were extracted from serum by the use of the Roche MagNA Pure LC instrument and the MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche Diagnostics, Mannheim, Germany), and eluted with 50 μl of elution buffer according to the manufacturer's instructions. PCR quantification was determined by COBAS® AmpliPrep/ COBAS® TaqMan® HCV Test (Roche Diagnostics, Mannheim, Germany), and genotyping was done by VERSANT®HCV Genotype 2.0 Assay (LiPA) (Innogenetics NV, Gent, Belgium). RNA was precipitated and reverse-transcribed using both random and specific primers as previously described [45]. PCR conditions were optimized by manipulating only concentration of primers and thermal cycling conditions as previously described [45] with slight modifications. Fluorescence data were collected continuously during heating to monitor the dissociation of the product. The derivative melting curves were obtained with the LightCycler data analysis software.

4.3. End-Point Limiting-Dilution Real-Time PCR for clonal HVR1 sequences

We used the End-Point Limiting-Dilution Real-Time PCR (EPLD-PCR) protocol for sequencing multiple clones of HVR1, as previously described [40,45]. The number of clones amplified varied depending on the viral titer an average 31 ± 7 of replicates for acute samples and 20 ± 10 for chronic samples. The relatively fewer clones obtained from the chronic samples was related to the low viral titers in them. Preliminary sequence analysis was conducted using SeqMan and MegAlign programs from the Lasergene DNA & Protein analysis software (Version 8.0, DNASTAR Inc., Madison, WI). The HCV sequences generated were aligned by using ClustalW, and then were trimmed to encompass 87 base-pairs (29 aas) of the HVR1 region.

4.4. Statistical analysis

The extent of quasispecies heterogeneity in each sample was examined by unbiased estimates of diversity,

calculated according to Nei and Li [41] using the ARLEQUIN program [47]. We used analysis of molecular variance (AMOVA) to quantify the level of differentiation between sequences from acute and chronic patients. The genetic structure was analyzed with consideration of the aa differences between sequences in addition to differences in their frequencies, resulting in estimates of Φst (an Fst analogue). AMOVA was calculated using ARLEQUIN program [47], in which the significance levels of the genetic variance components are estimated by use of a permutation test (n = 10,000). In this permutation test, the group label of each population sample is randomly permutated in order to simulate the null hypothesis of no difference between the groups.

We estimated the mean number of non-synonymous (dN) and synonymous substitutions (dS) per site (ratio dN/dS) using the Fixed Effects Likelihood (FEL) analysis implemented in the program HyPhy 0.99 beta [28], which is available in a parallel computing fashion at the Datamonkey web interface [26]. The algorithm works in three phases [27,56]: first, the General Time Reversible nucleotide model was fitted to the data and tree using maximum likelihood to obtain branch lengths and substitution rates; second, a codon model was fitted to the data to obtain codon branch lengths for scaling dN and dS estimated subsequently from each site; and third, a site-by-site likelihood-ratio test was performed to assess whether dN is significantly different from dS.

Each HVR1 aa sequence was transformed into a string of values according to 10 physicochemical properties: hydrophobicity, volume, polarity, charge, isoelectric point, flexibility, propensity to form beta sheet, propensity to form alpha-helix, accessible surface, protein composition and the presence of basic aa. We tested the significance of the differences between the acute and chronic groups by means of the Multi-response Permutation Procedure (MRPP). MRPP is a non-parametric permutation test for testing the hypothesis of no difference between two or more groups of entities [37]. Permutation tests represent the ideal situations in which the exact probabilities associated with a test statistic may be derived, rather than approximated from common probability distributions, such as t, F and Chi square [9]. In the majority of studies, the population distribution is unknown; assuming a normal distribution is inappropriate for many biological datasets, which often are skewed, discontinuous, and multi-modal. The distance-functions that form the basis of the MRPP are used to detect differences in distributions that are sensitive to both dispersion (variation) and shifts in central tendency (median) [8]. We implemented MRPP [38] in MATLAB (Mathworks, Natick, MA) using 10,000 permutations. In this permutation test, the group label of each population sample (a set of HVR1 sequences) is randomly permutated in order to simulate the null hypothesis of no difference between the groups. MRPP test was calculated for each individual variable (univariate MRPP) and for multivariate sets (multivariate MRPP) taking into consideration: (i) all positions of a given physicochemical property; and (ii) all variables together.

Physicochemical mapping of the data was conducted using Radviz [24], a non-linear multi-dimensional visualization technique that can display data on several attributes in a 2-dimensional projection. We used VizRank [31] to search for the best Radviz projections, the combination of the physicochemical variables that most accurately separates HCV HVR1 sequences into two classes: acute and chronic. For each sample, the consensus sequence was transformed into a vector of the same physicochemical variables. The classification accuracy of the model was evaluated with leave-one-out cross-validation, which involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data and the classification accuracy is averaged across all runs. The class of each instance was assigned based on k-Nearest Neighbor (kNN) classifier, which consists of two steps: first, the closest k neighbors of each instance in the 2-D representation are defined; second the class of the instance is determined as the majority class of these k nearest neighbors. Suppose (s_1,s_2,…..s_n) are the n instances in a training dataset, the class of the query instance will be predicted to belong to a given class if most of its k nearest neighbors belong to it. We explored the results with different values of k, with k = 7 showing marginally better results. We used the VizRank implementation available in the open-source data mining suite ORANGE [17,31].

Acknowledgments

We thank Guo-liang Xia for alignments of the NHANES dataset, Sumathi Ramachandran and Yulin Lin for advice regarding the EPLD protocol and Zoya Dimitrova for analysis of physical chemical properties.

Author contributions

The research theme was defined by CGT and KS. The methods and experiments were design by AA and IA.

IA was carried out the laboratory experiment. IA and DSC were analyzed the data. The results were interpreted and manuscript was written by IA, DSC and YK. AA, CGT and KS were discussed analyses, interpretation, and presentation. All authors have contributed to, seen, review and approved the manuscript.

References

[1] Alberti A, Chemello L, Benvegnu L, Natural history of hepatitis C, J Hepatol, v. **31** (1999), p. 17–24.

[2] Alter M, Epidemiology of hepatitis C virus infection, World J Gastroenterol, **13** (2007), 2436–2441.

[3] Alter M, Kruszon-Moran D, Nainan O, McQuillan G, Gao F, Moyer L, et al., The prevalence of hepatitis C virus infection in the United States, 1988 through 1994, N Engl J Med, **341** (1999), 6.

[4] Araujo A, Astrakhantseva I, Fields H, Kamili S, Distinguishing acute from chronic hepatitis C virus (HCV) infection based on antibody reactivities to specific HCV structural and nonstructural proteins, J Clin Microbiol, **49** (2011), 54–57.

[5] Bankwitz D, Steinmann E, Bitzegeio J, Ciesek S, Friesland M, Herrmann E, et al., Hepatitis C virus hypervariable region 1 modulates receptor interactions, conceals the CD81 binding site, and protects conserved neutralizing epitopes, J Virol, **84** (2010), 5751–5763.

[6] Bowen D, Walker C, Adaptive immune responses in acute and chronic hepatitis C virus infection, Nature, **436** (2005), 946–952.

[7] Brown RJ, Tarr AW, McClure CP, Juttla VS, Tagiuri N, Irving WL, et al., Cross-genotype characterization of genetic diversity and molecular adaptation in hepatitis C virus envelope glycoprotein genes., J Gen Virol, **88** (2007), 458–470.

[8] Cade B, Richards J, User manual for BLOSSOM statistical software, Midcontinent Ecological Science Center US Geological Survey, (2001).

[9] Cai L, Multi-response permutation procedure as an alternative to the analysis of variance: An SPSS implementation, Depart Psychol University N Carolina, (2004).

[10] Callens N, Ciczora Y, Bartosch B, Vu-Dac N, Cosset FL, Pawlotsky JM, et al., Basic residues in hypervariable region 1 of hepatitis C virus envelope glycoprotein E2 contribute to virus entry, J Virol, **79** (2005), 15331–15341.

[11] Campo D, Dimitrova Z, Khudyakov Y, Physicochemical correlation between amino acid sites in short sequences under selective pressure, In: Mandoiu I, Sunderraman R, Zelikovsky A, editors, Bioinformatics research and applications, Atlanta, GA, Springer; 2008, pp. 146–158.

[12] Campo D, Dimitrova Z, Mitchell R, Lara J, Khudyakov Y, Coordinated evolution of the hepatitis C virus, PNAS, **105** (2008), 9685–9690.

[13] Campo DS, Dimitrova Z, Yokosawa J, Hoang D, Perez NO, Ramachandran S, et al., Hepatitis C virus antigenic convergence, Sci Rep, **2** (2012), 267–277.

[14] Choo Q, Kuo G, Weiner A, Overby L, Bradley D, Houghton M, Isolation of a cDNA clone derived from a bloodborne non-A, non-B viral hepatitis genome, Science, **244** (1989), 359–362.

[15] Chung RT, Acute hepatitis C virus infection, Clinical Infection Disease, **41 Suppl 1** (2005), 14–17.

[16] Coppola N, Pisapia R, Marrocco C, Martini S, Vatiero L, Messina V, et al., Anti-HCV IgG avidity index in acute hepatitis C, J Clin Virol, **40** (2007), 110–115.

[17] Demsar J, Leban G, Zupan B, FreeViz–an intelligent multivariate visualization approach to explorative analysis of biomedical data, J Biomed Inform, **40** (2007), 661–671.

[18] Eckels DD, Wang Y, Bian TH, Tabatabai N, Gill JC, Immunoboilogy of hepatitis C virus (HCV) infection: the role of CD4 T-cells in HCV infection., Immunol Rev, **174** (2000), 90–97.

[19] Erickson AL, Kimura Y, Igarashi S, Eichelberger J, Houghton M, Sidney J, et al., The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes, Immunity, **15** (2001), 883–895.

[20] Farci P, Shimoda A, Coiana A, Diaz G, Peddis G, Melpolder J, et al., The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies, Science, **288** (2000), 339–344.

[21] Forns X, Thimme R, Govindarajan S, Emerson S, Purcell R, Chisari F, et al., Hepatitis C virus lacking the hypervariable region 1 of the second envelope protein is infectious and causes acute resolving or persistent infection in chimpanzees, PNAS, **97** (2000), 13318–13323.

[22] Herring B, Tsui R, Peddada L, Wide range of quasispecies diversity during primary hepatitis C virus infection, J Virol, **79** (2005), 4340–4346.

[23] Hino K, Korenaga M, Orito E, Katoh Y, Yamaguchi Y, Ren F, et al., Constrained genomic and conformational variability of the hypervariable region 1 of hepatitis C virus in chronically infected patients, J Viral Hepat, **9** (2002), 194–201.

[24] Hoffman PE, Grinstein GG, Marx K, Grosse I, Stanley E, DNA visual and analytic data mining, Proceedings of IEEE Visualization ISBN, Phoenix, AZ, 1997, pp 437–441.

[25] Klimashevskaya S, Obriadina A, Ulanova T, Bochkova G, Burkov A, Araujo A, et al., Distinguishing acute from chronic and resolved hepatitis C virus (HCV) infections by measurement of anti-HCV immunoglobulin G avidity index, J Clin Microbiol, **45** (2007), 3400–3403.

[26] Kosakovsky S, Frost S, Datamonkey: rapid detection of selective pressure on individual sites of codon alignments, Bioinformatics, **21** (2005), 2531–2533.

[27] Kosakovsky S, Frost S, Not so different after all: A comparison of methods for detecting amino acid sites under selection, Mol Biol Evol, **22** (2005), 1208–1222.

[28] Kosakovsky S, Frost S, Muse S, HyPhy: hypothesis testing using phylogenies, Bioinformatics, **21** (2005), 676–679.

[29] Kurosaki M, Enomoto N, Marumo F, Sato C, Evolution and selection of hepatitis C virus variants in patients with chronic hepatitis C, Virology, **205** (1994), 161–169.

[30] Lara J, Xia G, Purdy M, Khudyakov Y, Coevolution of the hepatitis C virus polyprotein sites 1 in patients on combined pegylated interferon and ribavirin therapy, J Virol, **85** (2011), 3649–3663.

[31] Leban G, Bratko I, Petrovic U, Curk T, Zupan B, VizRank: finding informative data projections in functional genomics by machine learning, Bioinformatics, **21** (2005), 413–414.

[32] Lechner F, Wong DK, Dunbar PR, Analysis of successful immune responses in person infected with hepatitis C virus, J Exp Med, **191** (2000), 1499–1512.

[33] Liu L, Fisher BE, Dowd KA, Astemborski J, Cox AL, Ray SC, Acceleration of hepatitis C virus envelope evolution in humans is consistent with progressive humoral immune selection during the transition from acute to chronic infection, J Virol, **84** (2010), 5067–5077.

[34] Maheshawari A, Ray S, Thuluvath PJ, Acute hepatitis C, Lancet, **372** (2008), 321–332.
[35] Marcellin, Hepatitis C: the clinical spectrum of the disease, J Hepatol, **31** (1999), 9–17.
[36] McAllister J, Casino C, Davidson F, Power J, Lawlor E, Yap P, et al., Long-term evolution of the hypervariable region of hepatitis C virus in a common-source-infected cohort, J Virol, **72** (1998), 4893–4905.
[37] McCune B, Grace J, Analysis of ecological communities, Gleneden Beach, MjM Software Design; 2002.
[38] Mielke P, Berry K, Permutation methods: A distance function approach. New York, Springer-Verlag; 2001.
[39] Mondelli MU, Cerino A, Segagni L, Meola A, Cividini A, Silini E, et al., Hypervariable region 1 of hepatitis C virus: immunological decoy or biologically relevant domain? Antiviral Research, **52** (2001), 153–159.
[40] Nainan OV, Khristova ML, Byun K, Xia G, Taylor PE, Stevens CE, et al., Genetic variation of hepatitis B surface antigen coding region among infants with chronic hepatitis B virus infection, J Med Virol, **68** (2002), 319–327.
[41] Nei M, Li W, Mathematical model for studying genetic variation in terms of restriction endonucleases, PNAS, **76** (1979), 5269–5273.
[42] Pawlotsky J, Therapy of hepatitis C: from empiricism to eradication, Hepatology, **43** (2006), 207–220.
[43] Penin F, Combet C, Germanidis G, Frainais P, Deleage G, Pawlotsky J, 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** (2001), 5703–5710.
[44] Ramachandran S, Campo D, Dimitrova Z, Xia G, Purdy M, Khudyakov Y, Temporal variations in the hepatitis C virus intra-host population during chronic infection, J Virol, **85** (2011), 6369–6380.
[45] Ramachandran S, Xia G, Ganova-Raeva L, Nainan O, Khudyakov Y, End point limiting dilution real time PCR assay for evaluation of HCV quasispecies in serum, Journal of Virological Methods, **151** (2008), 217–224.
[46] Rispeter K, Lu M, Lechner S, Zibert A, Roggendorf M, Cloning and characterization of a complete open reading frame of the hepatitis C virus genome in only two cDNA fragments, J Gen Virol, **78** (1997), 2751–2759.
[47] Schneider S, Roessli D, Excoffier L, ARLEQUIN version 2000: Software for population genetic data analysis, Geneva, Genetics and Biometry Laboratory, University of Geneva; 2000, 120 p.
[48] Sheridan I, Pybus O, Holmes E, Klenerman P, High-resolution phylogenetic analysis of hepatitis C virus adaptation and its relationship to disease progression, J Virol, **78** (2004), 3447–3454.
[49] Shimizu YK, Hijikata M, Iwamoto A, Alter HJ, Purcell RH, Yoshikura H, Neutralizing antibodies against hepatitis C virus and the emergence of neutralization escape mutant viruses, J Virol, **68** (1994), 1494–1500.
[50] Smith D, Evolution of the hypervariable region of hepatitis C virus, J Viral Hepat, **6** (1999), 41–46.
[51] Thimme R, Oldach D, Chang KM, Determinants of viral clearance and persistence during acute hepatitis C virus infection, J Exp Med, **194** (2001), 1395–1406.
[52] Thomson E, Smith J, Klenerman P, The natural history of early hepatitis C virus evolution; lessons from a global outbreak in human immunodeficiency virus-1-infected individuals, Journal of General Virology, **92** (2011), 2227–2236.
[53] Van Doorn L, Capriles I, Maertens G, DeLeys R, Murray K, Kos T, et al., Sequence evolution of the hypervariable region in the putative envelope region E2/NS1 of hepatitis C virus is correlated with specific humoral immune responses, J Virol, **69** (1995), 773–778.
[54] von Hahn T, Yoon J, Alter H, Rice C, Rehermann B, Balfe P, et al., Hepatitis C virus continuously escapes from neutralizing antibody and T-cell responses during chronic infection in vivo, Gastroenterology, **132** (2007), 667–678.
[55] Weiner AJ, Geysen HM, Christopherson C, Hall JE, Mason TJ, Saracco G, et al., Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections, PNAS, **89** (1992), 3468–3472.
[56] Yang Z, Swanson W, Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes, Mol Biol Evol, **19** (2002), 49–57.