J Mol Evol (2002) 54:465–473
DOI: 10.1007/s00239-001-0037-6

JOURNAL OF MOLECULAR EVOLUTION

Sequence Variation in the Gene Encoding the Nonstructural 3 Protein of Hepatitis C Virus: Evidence for Immune Selection

Huiru Wang,[1] Tonghua Bian,[1] Stephen J. Merrill,[2] David D. Eckels[1]

[1] Blood Research Institute, The Blood Center, P.O. Box 2178, Milwaukee, WI 53201-2178, USA
[2] Department of MSCS, Marquette University, Milwaukee, WI 53233, USA

Received: 2 November 2000 / Accepted: 1 October 2001

Abstract. To determine whether the persistent nature of hepatitis C infection is related to the emergence of antigenic variants driven by immune selection, we examined the sequence heterogeneity in a portion of the hepatitis C virus (HCV) nonstructural 3 (NS3) gene of a patient infected over the course of more than 2 years. By PCR amplification, cloning, and sequencing, we observed several variable and conserved regions in the NS3 segment of the HCV genome. All variable regions had higher ratios of nonsynonymous/synonymous mutations and encompassed immunodominant epitopes, and their locations were not essential to maintain the known function of HCV RNA helicase. In contrast, the regions that are critical for HCV RNA helicase activity were found to be conserved with lower heterogeneity or lower ratios of nonsynonymous/synonymous mutations, and none except one of these regions was encoded within immunodominant epitopes. Our results are consistent with immune selection of viral variants at the epitope and molecular levels that may enable HCV to evade host defenses over time. Plotting the relatedness of sequence variants revealed a star topology suggesting that a wild-type HCV sequence is maintained, unlike HIV.

Key words: HCV-NS3 — Chronic hepatitis C — Sequencing — Escape variant — Positive selection

Introduction

Hepatitis C virus (HCV) is a positive-strand RNA virus that appears to be related in structure to human *Flavivirus* and animal *Pestivirus* (Chambers et al. 1990; Choo et al. 1991). The 9.4-kb HCV genome directs the synthesis of a single polyprotein, which is proteolytically cleaved into at least seven functional proteins, consisting of a core region (C), two envelope protein encoding regions (E1 and E2/NS1), and four regions encoding for the nonstructural (NS) proteins NS2, NS3, NS4, and NS5. The amino-terminal portion of the polyprotein forms the structural proteins C, E1, and E2, which show substantial variation among different viral strains, whereas the carboxyl-terminal region constitutes a variety of viral enzymes, including NS2, NS3, NS4, and NS5, the most conserved of which is NS3 (Hijikata et al. 1991; Houghton et al. 1991; Takamizawa et al. 1991; Clarke 1997).

The NS3 protein is believed to be a multifunctional protein with two known catalytic activities, consisting of a serine protease in the first 180 amino acids, which is involved in polyprotein processing, followed by an NTPase-dependent RNA helicase in the C-terminal region (Grakoui et al. 1993a; Manabe et al. 1994; Jin and

The nucleotide sequence data reported in this paper can be found in the DDBJ, EMBL, and GenBank nucleotide sequence databases under accession numbers AF035122 through AF035150.

Informed consent was obtained from patient and human experimentation guidelines of the U.S. Department of Health and Human Services, and those of our institution were followed in the conduct of clinical research.

Correspondence to: David D. Eckels, Ph.D.; *email:* ddeckels@bcsew.edu

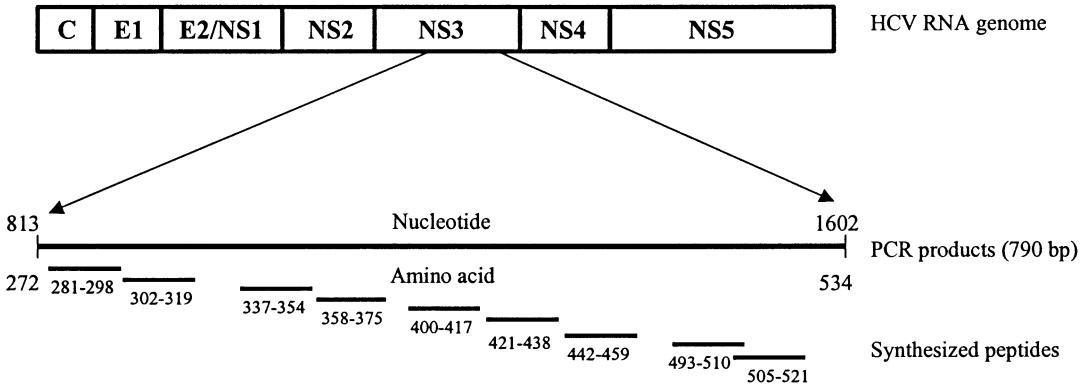


Fig. 1. Strategy for sequence analysis of heterogeneity of the HCV NS3 gene. cDNA templates synthesized from RNA of isolates P.B3019.1 and P.B3019.3 were amplified by PCR. The amplified fragments encompassed several immunodominant T-cell epitopes derived from HCV NS3. The PCR products were cloned and sequenced.

Peterson 1995; Morgenstern et al. 1997; Hijikata et al. 1993). NS3 is therefore critical to HCV infection and replication.

RNA viruses evolve at very rapid rates, a characteristic that is recognized as the basis of their ubiquity and adaptability (Holland et al. 1992). It is believed that an RNA virus such as HCV is not homogeneous in an infected patient but consists of variants that may differ from the population average at one or more positions, and the whole population of phylogenetically related viral variants observed within a single individual forms quasispecies (Ina et al. 1994). Current hypotheses relate the persistent nature of hepatitis C infection to the existence of virus quasispecies and the emergence of antigenic variants driven by immune selection (Alter 1995; Franco et al. 1995; Weiner et al. 1995; Kaneko et al. 1997; Tsai et al. 1998; Frasca et al. 1999; Wang and Eckels 1999). The quasispecies distribution is attributed to the observation that RNA polymerases lack a proofreading activity. However, the same argument should apply to all RNA viruses, yet only a minority of RNA viruses establishes persistent infections. Hence, the persistence of HCV infections must result from something other than the error-prone replication of the RNA-dependent RNA polymerase of HCV (Ina et al. 1994).

We and others have identified immunodominant T-cell epitopes in HCV NS3 that may be under immune selection pressure and therefore subject to a high incidence of viral mutation (Diepolder et al. 1997; Tabatabai et al. 1999; Eckels et al. 1999b; Wang and Eckels 1999). We also observed, by focusing on individual T-cell epitopes, that antigenic variants may have important consequences for T-cell recognition. In this study, we examined sequence heterogeneity in a portion of NS3 that encodes a series of such immunodominant epitopes that are functionally distinguished by the release of different cytokines. Our observations suggest that the high heterogeneity in the HCV genome may be the result of immune selection.

Materials and Methods

Samples

Peripheral blood samples were collected from a patient (P.B3019) with chronic HCV at about 12 (P.B3019.1) and 28 (P.B3019.3) months after infection. The presence of HCV-specific antibodies (Ab) and HCV RNA was determined as described previously (Eckels et al. 1999a). Blood was collected in acid citrate dextrose anticoagulant, centrifuged at 400*g* for 15 min, and divided into plasma and buffy coat fractions. After isolation of peripheral blood mononuclear cells (PBMC) over Lymphocyte Separation Medium (Organon/Teknika), plasma and PBMC were stored at −70°C or in liquid nitrogen, respectively.

RNA Extraction and Synthesis of Complementary DNA (cDNA)

HCV genomic RNA was isolated from 1 ml of P.B3019.1 and P.B3019.3 sera using RNA_{Zol} B (TEL-TEST, Inc.). cDNA was synthesized using reverse transcriptase from Moloney's murine leukemia virus (Gibco BRL, Gaithersburg, MD), and random hexadeoxynucleotide primers (Pharmacia Biotech Inc.).

Amplification of HCV cDNA by PCR

Specific primers were designed to amplify the HCV NS3 region based on the previously reported HCV sequence of a 1a genotype (accession No: I08294). Primers NS3-3A (5′-CGGACCTTTACCTGGTCACG-3′) and NS3-2M (5′-CGCCCTCCCAAAATCCAAGATGG-3′) were used for the first PCR amplification, in which HCV cDNA was subjected to 35 cycles of amplification in a buffer with 10 m*M* Tris (pH 8.3), 50 m*M* KCl, 2.5 m*M* $MgCl_2$, 0.1 mg/mL gelatin, 0.02% NP40, and 5% DMSO. Each cycle consisted of 30 s at 94°C, 30 s at 55°C, and 60 s at 72°C, with an initial denaturation (94°C, 5 min) and a final extension step (72°C, 7 min). Two microliters of the products from the first PCR was used for the second PCR amplification under the same conditions using primers NS3-3C (5′-CAAGTTCCTTGCCGACGCCGG-3′) and NS3-2M. The amplified fragments were 790 base pairs (bp) long and encompassed nucleotides 813 to 1602 of the NS3 region (Fig. 1).

Cloning and Sequencing

The PCR products were cloned using a T-vector cloning kit (Invitrogen, San Diego, CA). Ligation and transformation were performed

essentially according to the manufacturer's instructions. Recombinant clones were screened and selected by specific PCR amplification of the inserted fragment. Plasmid DNA was prepared from the selected clones with a Wizard Plus Miniprep DNA Purification Kit (Promega, Madison, WI) according to the standard protocol. Fifteen independent clones for each sample of B3019.1 and B3019.3 were sequenced in both directions by dye-termination cycle sequencing using an automated DNA sequencer (373A; Applied Biosystems). The sequencing results were analyzed using the ABI Gene Works software (IntelliGenetics Inc.). Plasmid DNA from the Hutchinson strain (1a) of HCV (Grakoui et al. 1993b) was diluted to 10^{-14} g/mL, then amplified, cloned, and sequenced as a control for polymerase errors. This concentration of plasmid DNA produced PCR amplification bands with intensities comparable to those obtained from viral cDNA under essentially identical conditions. The Taq error rate under such conditions was calculated as:

$$(\text{number of sporadic changes}) \div [(\text{number of clones}) \times (\text{sequence length}) \times (\text{PCR cycles})]$$

according to Smith et al. (1997).

Peptide Synthesis

Peptides were synthesized according to the sequence of NS3 amino acids 272–534 by Chiron Mimotopes, Inc., as 18 oligomers, or locally using Fmoc chemistry, and purified by HPLC. Peptide powder was dissolved in a drop of DMSO and adjusted to approximately 1 mg/mL with RPMI 1640 tissue culture medium. Peptide was used at 5 μ*M* to stimulate PBMC in cytokine assays.

Cytokine Analysis

To measure secreted cytokines, PBMC from patient P.B3019 at 18 months after infection were resuspended at a concentration of 1×10^6/mL in RPMI 1640 tissue culture medium containing 25 m*M* HEPES, 2.0 m*M* L-glutamine, 1.0 m*M* Na-pyruvate, 10 U/mL Na-heparin, 100 U/mL penicillin, 100 μg/mL streptomycin, 5.0 μg/mL gentamycin sulfate, and 10% pooled human plasma (PHP). To 96-well round-bottom plates, 100-μL aliquots of PBMC were added to synthetic peptides at 5 μ*M* in 100 μL of 10% PHP tissue culture medium. Cultures were incubated at 37°C in 5% humidified CO_2 for 72 h, and supernatants were characterized for secreted IFN-γ, IL-2, and IL-10 using commercial antigen-capture human IFN-γ, IL-2, and IL-10 ELISA sets from Pharmingen Inc. Experimental values were compared to a standard curve derived using recombinant cytokines. Negative controls consisted of background levels derived from cultures of PBMC in the presence of tissue culture medium alone.

Sequence Analysis

Sequence data were configured for analysis using maximal likelihood as implemented in DNAML version 3.573c as part of the PHYLIP package [http://evolution.genetics.washington.edu/phylip.html (Felsenstein 1981)]. The analysis was run using default parameters and the unmodified topology was reproduced here with associated accession numbers using the Phylodendron program available at http://iubio.bio.indiana.edu/treeapp/.

Results

Variable and Conservative Regions in HCV NS3

To determine whether mutations were accumulating in HCV sequences, cloning and sequencing analysis of PCR products from epitope regions of NS3 were performed using cDNA templates synthesized from RNA isolates of P.B3019.1 and P.B3019.3. Extensive mutation was observed throughout the sequences from the 30 clones. All but one clone had at least one nucleotide substitution. Only one clone from P.B3019.3 had a sequence identical to that of a full-length clone of Hutchinson strain 1a (Grakoui et al. 1993b).

Table 1. Mutations in HCV NS3 clones from isolates P.B3019.1 and P.B3019.3[a]

Sample	P.B3019.1	P.B3019.3	Total
Clone number	15	15	30
Clones with mt	15	14	29
NM	51	32	83
SYN mt	17	8	25
NSY mt	34	24	58

[a] mt, mutations; NM, nucleotide mutation; SYN, synonymous; NSY, nonsynonymous.

There were 83 nucleotide substitutions in the sequences of the 29 clones with mutations, 58 of which induced amino acid alterations (Table 1). One nucleotide substitution was identical between two isolates from P.B3019.1, two changes were found to be the same among four isolates from P.B3019.3, and three were identical among three isolates from P.B3019.1 and P.B3019.3. Therefore, the remaining 77 nucleotides differed between the early isolates from P.B3019.1 and the latter isolates from P.B3019.3. The mutation rate was calculated as 2.5×10^{-3} base substitutions per site per year ($77 \div 790$ bp $\div$ 1.3 years $\div$ 30 clones or molecules), although this figure may be somewhat misleading, as it does not take into account a more extensive history of sequence variation.

Notably, most observed nonsynonymous (NSY) mutations (52/58; 90%) were not randomly scattered but unevenly clustered in four regions, called variable regions here, consisting of NS3 amino acids 331–376, 387–414, 425–454, and 486–515. On the other hand, there were few or no NSY mutations in four other regions, called conservative regions here, consisting of NS3 amino acids 281–330, 377–386, 415–424, and 455–485 (Fig. 2 and Table 2).

To distinguish actual viral mutations from possible sporadic nucleotide substitutions caused by Taq misincorporation errors, we calculated the error rate using 0.01 pg of a known plasmid template from which amplified bands were comparable to those obtained from viral cDNA. As reported previously (Wang and Eckels 1999), we calculated a polymerase error rate of 2.16×10^{-5}, which was well within the usual Taq error range of 0.2–2 $\times 10^{-4}$ reported by others (Lundberg et al. 1991; Barnes 1992). Thus we discount the possibility of underestimating the number of artifacts due to saturation reached after only a few PCR cycles. The expected numbers of sporadic amino acid substitutions were calculated and com-

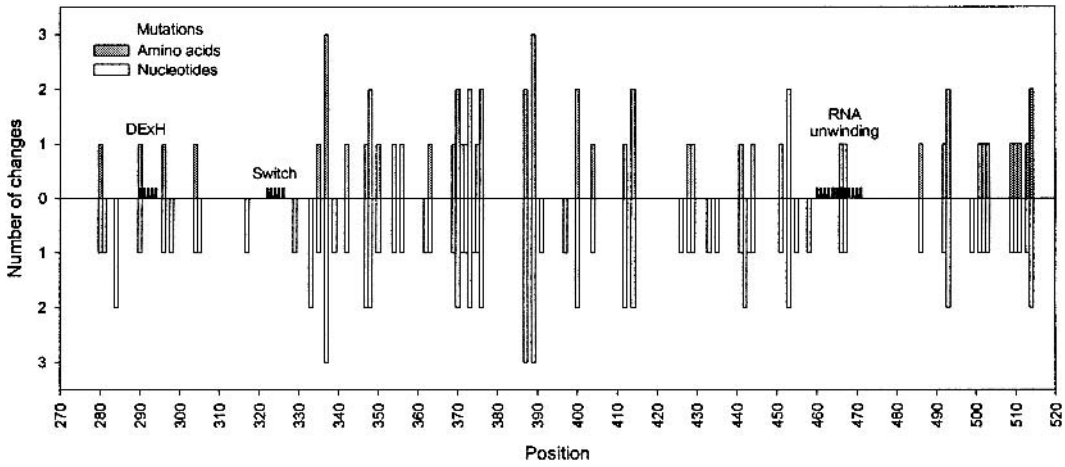


Fig. 2. The nucleotide mutations and amino acid alterations in a partial NS3 sequence were clustered in different regions. Fifteen independent clones (30 in total) from each serum sample of P.B3019.1 and P.B3019.3 were sequenced in both directions. The sequencing results were analyzed using the ABI Gene Works software (IntelliGenetics Inc.).

Table 2. Expected sporadics and observed nonsynonymous mutations in different regions of HCV NS3

Region (amino acids)	Length of sequence (bp)	PCR cycles	Clone number	Expected sporadic mutations[a]	Observed nonsynonymous mutations	p[b]
NS3 281–330	150	70	30	4.5	3.0	>0.05
NS3 331–376	138	70	30	4.2	22	<0.001
NS3 377–386	30	70	30	1.0	0	>0.05
NS3 387–414	84	70	30	2.5	11	<0.001
NS3 415–424	30	70	30	1.0	0	>0.05
NS3 425–454	90	70	30	2.7	7	<0.001
NS3 455–485	93	70	30	2.8	2	>0.05
NS3 486–515	90	70	30	2.7	12	<0.001

[a] The expected number of sporadic nonsynonymous mutations is calculated as Taq error rate (2.16×10^{-5}) × length of compared region × number of PCR cycles × number of clones sequenced × proportion of sporadics expected to produce amino acid substitutions (2/3)/number of samples (1) according to Smith et al. (1997).
[b] The agreement between hypothesis and observation was tested by χ^2 goodness of fit.

pared with the actually observed NSY mutations in both conservative regions and variable regions as shown in Table 2 and Fig. 2. We observed that in all such variable regions, the observed NSY mutation rate was significantly higher than expected. In contrast, the observed NSY mutation frequerncy in so-called conservative regions was lower than that expected from sporadic amino acid substitutions. Taken together, our observations support the hypothesis that the vast majority of NSY mutations in the regions of HCV NS3 represent natural changes to the HCV RNA genome rather than artificial PCR errors.

Evidence for Positive and Negative Selection

A powerful method to distinguish positive Darwinian selection from neutral selection is to compare the rates of NSY and synonymous (SYN) nucleotide substitution per site. In the case of positive selection favoring diversity at the amino acid level, the rate of NSY substitution is found to exceed that of SYN substitution and heterogeneity will increase more quickly, whereas at sites subject to negative selection, the NSY/SYN ratio is <1 and the heterogeneity will be lower (Hughes and Nei 1988, 1989; Kimura 1977; Smith et al. 1997).

To determine if the observed mutations were consistent with an immune selection model, we compared the ratios of NSY/SYN mutations in each region, according to the method reported by Nei and Gojobori (1986). Among the total 83 nucleotide substitutions observed in 263 codons, 26 occurred in the first nucleotide position of the codons, and 32 and 25 were at the second and third nucleotides, respectively. We calculated theoretical SYN mutations (mt) as

$$0.05 \times 26 + 0.72 \times 25 = 19.3$$

and the total SYN mutational rate (MR_{SYN}) as

$$19.3 \div 263 = 0.07$$

Table 3. Synonymous (SYN) and nonsynonymous (NSY) mutations (mt) in different regions of HCV NS3 among 30 clones

Region (codons)	Codons	SYN mt			NSY mt			NSY/SYN
		Exp[a]	Obs[b]	p[c]	Exp	Obs	p	
NS3 281–330	50	3.5	7	>0.05	10.85	3	<0.05	0.43
NS3 331–376	46	3.2	6	>0.05	9.98	22	<0.001	3.67
NS3 377–386	10	0.7	0	>0.05	2.17	0	>0.05	0
NS3 387–414	28	2	4	>0.05	6.08	11	<0.05	2.75
NS3 415–424	10	0.7	0	>0.05	2.17	0	>0.05	0
NS3 425–454	30	2.1	5	<0.05	6.51	7	>0.05	1.4
NS3 455–485	31	2.2	2	>0.05	6.73	2	>0.05	1
NS3 486–515	30	2.1	1	>0.05	6.51	12	<0.05	12

[a] The expected number of SYN and NSY mt is calculated as rate of SYN or NSY mt × length of the compared region.
[b] Observed mt.
[c] The agreement between hypothesis and observation was tested by the χ^2 goodness-of-fit test.

SYN mt per codon. NSY mt were calculated as

$$0.95 \times 26 + 1 \times 32 + 0.28 \times 25 = 63.7$$

and the total NSY mutational rate (MR_{NSY}) as

$$63.7 \div 263 = 0.24$$

NSY mt per codon. The theoretical NSY number was very close to our observed number of 58, while the theoretical SYN number was lower than our observed 25. The expected numbers of SYN mt and NSY mt and the NSY/SYN ratio were calculated and compared with the observed numbers in both conservative and variable regions. As shown in Table 3, all the NSY/SYN ratios within the variable regions ranged from 1.4 to 12.0, consistent with the hypothesis that such mutations resulted from positive selection. In contrast, the NSY/SYN ratio within region 281–330 was as low as 0.43, consistent with negative selection. No mutations were observed in regions 377–386 and 415–424, and the two mutations in region 455–485 could be considered sporadic because the expected number of sporadic mutations is calculated as 2.7.

A simple pairwise analysis of SYN mt and NSY mt does not adequately present all of the information present in the data. Therefore, a phylogeny was created using the DNAML program as implemented in the PHYLIP package and converted to a graphic format using the Phylodendron program at the University of Washington and University of Indiana web sites, respectively. A star-like topology was obtained and sequences did not appear to cluster by time point. Interestingly, a similar analysis of a much more limited segment, consisting of the region encoding NS3 amino acids 358–375, produced virtually identical results (Harris 2000).

We further analyzed secreted cytokines using synthetic 18-amino acid peptides corresponding to epitopes encoded within residues 280–521 of NS3 by ELISA to determine whether particular regions of NS3 might be recognized by peripheral blood T cells from donor P.B3019. As shown in Figs. 3 and 4 (Kim et al. 1996; Yao et al. 1997), each variable region encompassed at least one T-cell epitope that could stimulate different cytokine secretions. In contrast, only one epitope, $NS3_{281–298}$, was found within the conservative regions (Figs. 3 and 4).

We also evaluated whether the residues in conservative and variable regions are essential to the functional integrity of HCV helicase. As shown in Fig. 4, none of the variable regions are located either near or in the known functionally important parts of the HCV helicase, whereas two of four conservative regions, NS3 281–330 and NS3 454–485, encompassed the functionally important DExH box and Switch region as well as the RNA unwinding region of the HCV helicase (Fig. 4). Furthermore, NS3 415–424 is located close to the RNA unwinding site of the helicase based on crystallographic evidence (Yao et al. 1997) and thus may also be involved in the preservation of functional protein folding.

Our findings collectively support the possibility that the high substitution rates in the observed variable regions of HCV NS3 have been accelerated by host immune defenses against the virus. These data also suggest that negative selection might occur in a region, NS3 281–330, including residues maintaining functional activities of the HCV helicase.

Immune Selective Pressures Earlier and Later in Infection

More nucleotide substitutions were observed in the earliest isolate (P.B3019.1; 34 amino acid changes) compared to the latter (P.B3019.3; 24 amino acid changes), although the localizations of these changes appeared to be similar, suggesting higher immune pressure in early infection.

Over time, comparable levels of mutation were observed in the regions encoding IL-2 epitopes such as $NS3_{358–375}$ (Fig. 5). In contrast, for IL-10 inducing epitopes such as $NS3_{400–417}$ and $NS3_{421–438}$, the majority

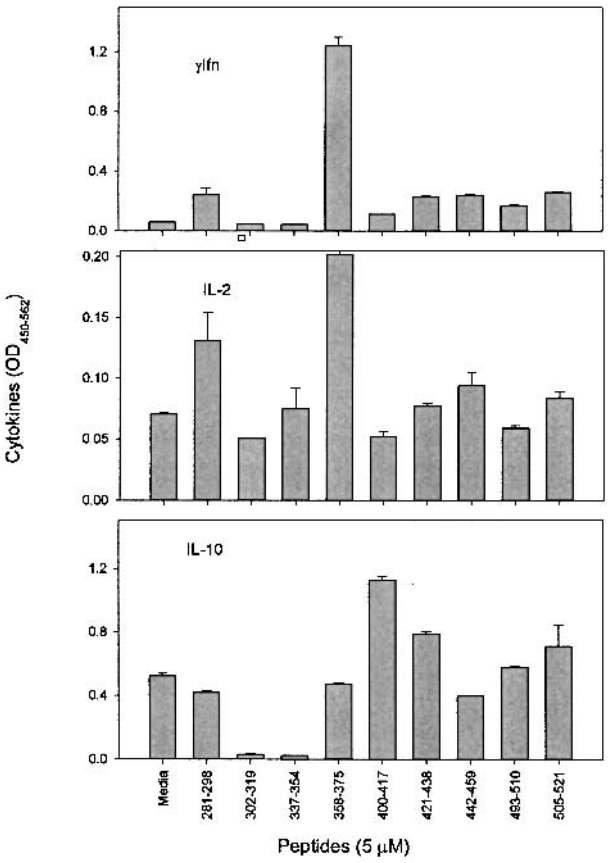


Fig. 3. The results of cytokine analysis. One hundred-microliter aliquots of PBMC (at a concentration of 1×10^6/mL in RPMI 1640 tissue culture medium with 10% pooled human plasma) from patient P.B3019 at about 18 months after infection were cultured for 72 h with synthetic 18-amino acid peptides at 5 μ*M* in 100 μL of the medium mentioned above. The supernatants were characterized for secreted IFN-γ, IL-2, and IL-10 using commercial ELISA sets from Pharmingen Inc. Negative controls consisted of background levels derived from cultures of PBMC in the presence of tissue culture medium alone.

of the nucleotide substitutions was found earlier in infection (Fig. 5). Furthermore, the NSY/SYN ratio for region 331–376, which includes the immunodominant $NS3_{358–375}$, was higher (3.67) than those for regions 387–414 (2.75) and 425–454 (1.4), which covered epitopes $NS3_{440–417}$ and $NS3_{421–438}$. This observation suggests that immune selective intensity may vary with respect to time and the type of epitope found at a given position. Thus, for IL-2 inducing epitopes, immune selective pressure is stronger and persists longer than that for IL-10 producing epitopes, for which selective pressure may weaken in the later stages of infection as chronicity is established.

Discussion

We have observed extensive sequence variation in the HCV NS3 region during a short infective course in a chronic HCV patient. Most nucleotide mutations were found to be clustered within specific regions of NS3, and the NSY/SYN ratios computed for these variable regions were consistent with models of positive Darwinian selection (Hughes and Nei 1988, 1989). Furthermore, we show that all of the four NS3 variable regions occurred in areas recognized by host T cells. The contribution of these positions to the HCV helicase structure seems not to be essential, because none is located in the known functionally important regions of the helicase. Therefore, our observations support the hypothesis that mutations in variable regions of HCV NS3 may be driven by a positive selective force, possibly host immune responses.

In contrast, we observed low heterogeneity in four conservative regions, 281–330, 377–386, 415–425, and 456–485, which are located in what are likely to be functionally important regions of the HCV helicase. No evidence of T-cell recognition was observed within the latter three regions. When we compared these regions to corresponding sequences of different HCV types in GenBank, we found that all these conservative regions included certain fixed amino acids which were the same or similar within a given strain but differed from strain to strain. Most of these sites are functionally important.

The NS3 281–330 region, however, is interesting because it covers an IL-2 epitope, $NS3_{281–298}$, as well as the DExH box and Switch region of the NS3 helicase. Our observations suggest that such an IL-2 epitope should be under positive immune selection and thus susceptible to more NSY mutations than other sorts of epitopes. However, the need to maintain functional structures may result in negative selection (Hughes and Nei 1988, 1989), which is consistent with our results showing a NSY/SYN ratio of 0.43 (Table 3). Notably, the number of SYN mutations in this region was among the highest of all the regions we compared, and the evidence taken together suggests the possibility that most variants with NSY mutations affecting the function of the RNA helicase would not survive, whereas only variants with amino acid changes that do not seriously impair RNA helicase activity would. Stimulation of IL-2 and the low NSY mutation rate make it reasonable to consider the $NS3_{281–298}$ segment as a candidate peptide vaccine epitope, with the caveat that effective immunity against one epitope presented by a limited number of allelic MHC molecules is not likely in isolation to immunize efficiently against virus.

Although peptide $NS3_{337–354}$ induced no IFN-γ, IL-2, or IL-10, to the extent that mutation is driven by immune selection, frequent mutations in the $NS3_{337–354}$ segment suggest that it may represent an epitope capable of inducing other cytokines, binding to MHC molecules, or serving as a recognition target for other immune cells, CTL, for example, which can be addressed only by further experimentation.

According to the number of nucleotide and amino acid changes during the early and later stages of infec-

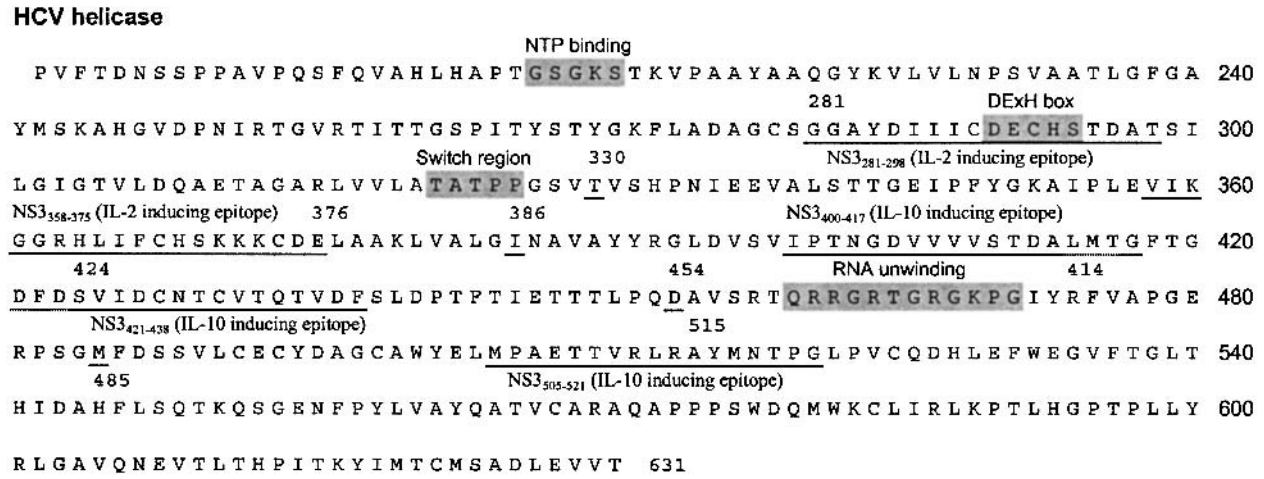


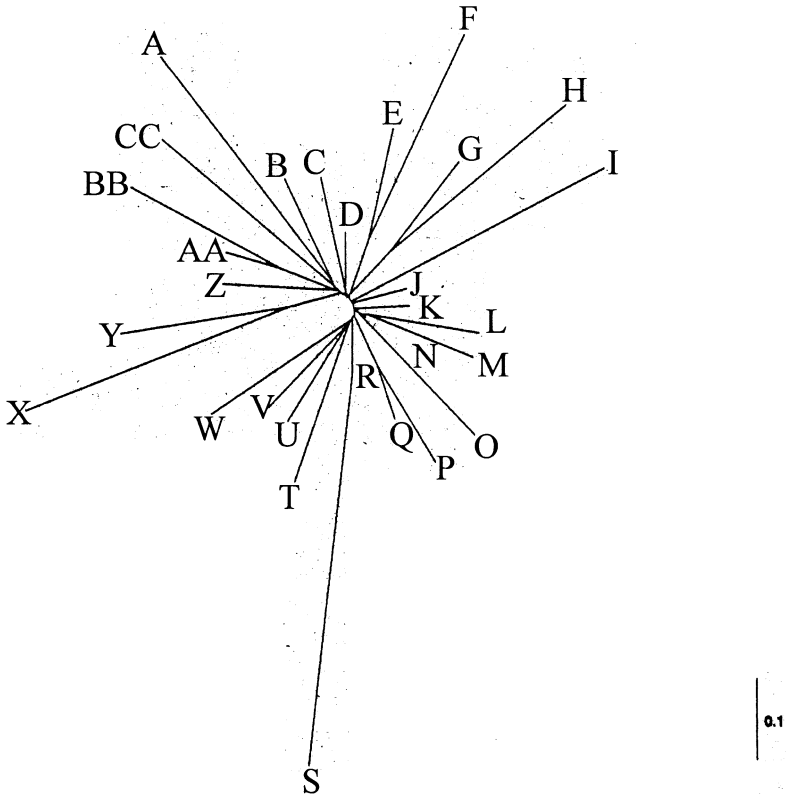
Fig. 4. The structure of HCV helicase and the different regions observed in the helicase, as well as the location of immunodominant peptides. At least one immune epitope occurred within each variable region, and no variable region was located either close to or within known functionally important sites of the HCV helicase. Two conservative regions, 281–330 and 455–485, encompass the DExH box and Switch region, as well as the RNA unwinding region.

	358-375								400-417					421-438			505-521				335-356							
AA No.	336	339	340	341	342	343	345	346	348	400	401	404	411	421	422	424	505	511	513	514	335	337	342	347	348	350	354	355
	R	H	S	K	K	K	D	E	V	I	G	D	L	C	N	D	T	V	L	R	N	E	S	I	P	Y	I	L
Clone No.																												
1-1	-	-	-	-	-	-	-	-	-	-	-	-	H	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-
1-2	-	-	-	-	-	-	-	-	-	-	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1-3	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1-4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1-5	-	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-
1-6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S	-	-	-	-	-	-	-	-	-	-	-	-	-
1-7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S	-	-	-	-	-	-	-
1-8	-	-	-	E	-	-	-	-	A	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1-9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-
1-12	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1-13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-	-	-	-	-
1-14	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	-	P
1-17	-	-	P	-	-	-	-	-	A	V	-	-		-	-	-	-	-	-	-	-	-	-	T	-	-	-	-
1-21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-
3-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3-3	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3-4	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3-5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3-6	-	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3-7	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-
3-8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-
3-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3-14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	-	-
3-15	-	-	-	-	R	R	-	-	-	-	-	-	-	Y	-	-	-	-	-	-	-	V	P	-	-	-	-	-
3-17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-
3-18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S	-	-	-	-
3-19	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Q	-	-	-	-	-	-	-	-	-
3-22	-	-	-	-	-	R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Fig. 5. Amino acid alterations in the regions encoding for known immunodominant epitopes of NS3, as well as a region with frequent mutations. Peptide $NS3_{358–375}$ could stimulate strong IL-2 and IFN-γ secretion, while peptide $NS3_{400–417}$, $NS3_{421–438}$, and $NS3_{505–521}$ could stimulate IL-10 secretion. Although region 335–356 induced none of the three tested cytokines, the frequent mutations in this region suggested the possibility that they may represent other epitopes such as CD8 T-cell epitopes to the extent that mutation is driven by immune selection.

tion, we can divide the variable regions into two types. First, we see IL-2 inducing regions such as $NS3_{358–375}$, in which the NSY/SYN ratios are much higher than 1, and similar levels of mutation have been maintained over the course of infection, perhaps indicative of high levels of immune selection. Second, we see IL-10 inducing regions as $NS3_{400–417}$, which also have a NSY/SYN rate higher than 1, but lower than that of an IL-2 epitope, along with fewer mutations as infection progresses (Fig. 5), perhaps indicative of decreasing immune selection pressure as chronicity is established. Consistent with immune selection, three $NS3_{358–375}$ (Th1 epitope) and one $NS3_{505–521}$ (Th2 epitope) variants failed to stimulate T-cell proliferation, and two other variants weakly stimulated T-cell responses (Wang and Eckels 1999).

Phylogenetic analysis revealed sequence diversification radiating from the wild type and producing a star rather than a tree topology (Fig. 6). This rather unexpected result differs from those obtained in similar studies of HIV evolution in the presence of host immune responses (Zanotto et al. 1999; Crandall et al. 1999), both of which report tree-like architectures. In each of these studies, a simple pairwise analysis of NSY and SYN mutation was inadequate and a phylogenetic analysis was required to demonstrate positive selection effectively. That such a simple analysis of our data revealed



	Clone	Accession Number
A	12.1	AF035122
B	28.18	AF035148
C	12.4	AF035125
D	28.5	AF035143
E	12.12	AF035127
F	12.17	AF035130
G	18.3	AF035133
H	18.14	AF035138
I	12.14	AF035129
J	28.14	AF035145
K	12.21	AF035131
L	18.22	AF035141
M	28.10	AF035144
N	18.1	AF035132
O	28.2	AF035142

	Clone	Accession Number
P	12.5	AF035126
Q	18.10	AF035136
R	28.22	AF035150
S	28.15	AF035146
T	12.10	AF035124
U	12.13	AF035128
V	18.6	AF035134
W	18.7	AF035135
X	18.13	AF035137
Y	18.15	AF035139
Z	28.17	AF035147
AA	18.19	AF035140
BB	28.19	AF035149
CC	12.2	AF035123

Fig. 6. Phylogenetic tree of the 29 HCV sequences, GenBank accession numbers AF035122 through AF035150, as computed by DNAML 3.57c (http://evolution.genetics.washington.edu/phylip.html). The tree was drawn with Phylodendron (http:iubio.bio.indiana.edu/treeapp/). The star-like topology of the tree is in contrast to that expected in HIV (Plikat et al. 1997).

highly significant differences from expected variation in the HCV NS3 gene may provide strong evidence that positive selection plays a role in HCV diversification.

The difference in architecture, however, is also intriguing. One of the most directly comparable analyses examined sequence variations in the 206-amino acid *nef* gene of HIV over 30 months of infection in one patient (Plikat et al. 1997). A distinct tree phylogeny was obtained, unlike our results with HCV NS3. These different modes of viral evolution imply dramatically different mechanisms of host–viral interaction. Rather puzzling is the fact that our data seem to indicate that viral sequence diversification, though positively selected, seems not to deviate far from wild type, thus the star topology.

We have recently found that naturally occurring, single-amino acid mutations in the region encompassing amino acids 358–375 produce alterations in the type of helper T-cell response obtained. The vast majority of changes represents loss of function mutations, in that the ability to induce Type 1 cytokines is lost or attenuated, while simultaneously the ability to induce Type 2 cytokines is gained (manuscript submitted for publication). Type 2 cytokines such as IL-4 and IL-10 counterregulate antiviral Type 1 cytokines and may be conducive to viral persistence. Thus relatively few changes in immunodominant epitopes of NS3 may lead in the host to a functional tolerance of the wild-type HCV, in any given sequence of which only a very limited number of mutations is permitted.

According to our knowledge, this work is the first example of both positive and negative selection established at the molecular and epitope levels for HCV. The

information on heterogeneity in a single carrier is very important to help elucidate immunopathogenic mechanisms influenced by viral genomic changes and the information on variants is essential for consideration of effective vaccine constructs for HCV.

Acknowledgments. We thank Donor P.B3019 for heroic donation of blood. This work was supported by National Institutes of Health grants RO1-AI47347, RO1-DK57732, KO1-DK02970, PO1-HL44612 and The Blood Center Research Foundation. We also thank, especially, the two anonymous reviewers of our manuscript; their lucid criticisms and suggestions are sincerely appreciated.

References

Alter HJ (1995) To C or not to C: These are the questions [Review]. Blood 85:1681–1695

Barnes WM (1992) The fidelity of *Taq* polymerase catalyzing PCR is improved by an N-terminal deletion. Gene 112:29–35

Chambers TJ, Hahn CS, Galler R, Rice CM (1990) Flavivirus genome organization, expression, and replication. Annu Rev Microbiol 44: 649–688

Choo QL, Richman KH, Han JH, et al. (1991) Genetic organization and diversity of the hepatitis C virus. Proc Natl Acad Sci USA 88: 2451–2455

Clarke B (1997) Molecular virology of hepatitis C virus. J Gen Virol 78:2397–2410

Crandall KA, Kelsey CR, Imamichi H, Lane HC, Salzman NP (1999) Parallel evolution of drug resistance in HIV: Failure of nonsynonymous/synonymous substitution rate ratio to detect selection. Mol Biol Evol 16:372–382

Diepolder HM, Gerlach JT, Zachoval R, et al. (1997) Immunodominant $CD4^+$ T-cell epitope within nonstructural protein 3 in acute hepatitis C virus infection. J Virol 71:6011–6019

Eckels DD, Tabatabai N, Bian T-H, et al. (1999a) *In vitro* human Th-cell responses to a recombinant hepatitis C virus antigen: Failure in IL-2 production despite proliferation. Hum Immunol 60:187–199

Eckels DD, Zhou H, Bian T-H, Wang H (1999b) Identification of antigenic escape variants in an immunodominant epitope of hepatitis C virus. Int Immunol 11:577–583

Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17:368–376

Franco A, Ferrari C, Sette A, Chisari FV (1995) Viral mutations, TCR antagonism and escape from the immune response. Curr Opin Immunol 7:524–531

Frasca L, Porto PD, Tuosto L, et al. (1999) Hypervariable region 1 variants act as TCR antagonists for hepatitis C virus-specific $CD4^+$ T cells. J Immunol 163:650–658

Grakoui A, McCourt DW, Wychowski C, Feinstone SM, Rice CM (1993a) Characterization of the hepatitis C virus-encoded serine proteinase: Determination of proteinase-dependent polyprotein cleavage sites. J Virol 67:2832–2843

Grakoui A, Wychowski C, Lin C, Feinstone SM, Rice CM (1993b) Expression and identification of hepatitis C virus polyprotein cleavage products. J Virol 67:1385–1395

Harris FR (2000) A phylogenetic map of HCV-1A and its various mutations in an individual over time. Technical Report 465, Department of MSCS, Marquette University, Milwaukee, WI

Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Shimotohno K (1991) Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. Proc Natl Acad Sci USA 88:5547–5551

Hijikata M, Mizushima H, Akagi T, et al. (1993) Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. J Virol 67:4665–4675

Holland JJ, de la Torre JC, Steinhauer DA (1992) RNA virus populations as quasispecies. Curr Topics Microbiol Immunol 176:1–20

Houghton M, Weiner A, Han J, Kuo G, Choo QL (1991) Molecular biology of the hepatitis C viruses: Implications for diagnosis, development and control of viral disease. Hepatology 14:381–388

Hughes A, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Hughes A, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: Evidence from overdominant selection. Proc Natl Acad Sci USA 86:958–962

Ina Y, Mizokami M, Ohba K, Gojobori T (1994) Reduction of synonymous substitutions in the core protein gene of hepatitis C virus. J Mol Evol 38:50–56

Jin L, Peterson DL (1995) Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA helicase. Arch Biochem Biophys 323:47–53

Kaneko T, Moriyama T, Udata T, et al. (1997) Impaired induction of cytotoxic T lymphocytes by antagonism of a weak agonist borne by a variant hepatitis C virus epitope. Eur J Immunol 27:1782–1787

Kim JL, Morgenstern KA, Lin C, et al. (1996) Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. Cell 87:343–355

Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:275–276

Lundberg KS, Shoemaker DD, Adams MWW, Short JM, Sorge JA, Mathur EJ (1991) High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus.* Gene 108:1–6

Manabe S, Fuke I, Tanishita O, et al. (1994) Production of nonstructural proteins of hepatitis C virus requires a putative viral protease encoded by NS3. Virology 198:636–644

Morgenstern KA, Landro JA, Hsiao K, et al. (1997) Polynucleotide modulation of the protease, nucleoside triphosphatase, and helicase activities of a hepatitis C virus NS3-NS4A complex isolated from transfected COS cells. J Virol 71:3767–3775

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Plikat U, Nieselt-Struwe K, Meyerhans A (1997) Genetic drift can dominate short-term human immunodeficiency virus type 1 *nef* quasispecies evolution in vivo. J Virol 71:4233–4240

Smith DB, McAllister J, Casino C, Simmonds P (1997) Virus "quasispecies": Making a mountain out of a molehill? J Gen Virol 78: 1511–1519

Tabatabai N, Bian T-H, Rice CM, Yoshizawa K, Gill J, Eckels DD (1999) Functionally distinct T-cell epitopes within the hepatitis C virus non-structural 3 protein. Hum Immunol 60:105–115

Takamizawa A, Mori C, Fuke I, et al. (1991) Structure and organization of the hepatitis C virus genome isolated from human carriers. J Virol 65:1105–1113

Tsai SL, Chen YM, Chen MH, et al. (1998) Hepatitis C virus variants circumventing cytotoxic T lymphocyte activity as a mechanism of chronicity. Gastroenterology 115:954–966

Wang H, Eckels DD (1999) Mutations in immunodominant T cell epitopes derived from the non-structural 3 protein of hepatitis C virus have the potential for generating escape variants that may have important consequences for T cell recognition. J Immunol 162:4177–4183

Weiner A, Erickson AL, Kansopon J, et al. (1995) Persistent hepatitis C virus infection in a chimpanzee is associated with emergence of a cytotoxic T lymphocyte escape variant. Proc Natl Acad Sci USA 92:2755–2759

Yao NH, Hesson T, Cable M, et al. (1997) Structure of the hepatitis C virus RNA helicase domain. Nature Struct Biol 4:463–467

Zanotto PM de A, Kallas EG, deSouza RF, Holmes EC (1999) Geneological evidence for positive selection in the *nef* gene of HIV-1. Genetics 153:1077–1089