Analysis of the Evolutionary Forces in an Immunodominant CD8 Epitope in Hepatitis C Virus at a Population Level

Christoph Neumann-Haefelin,1 David N. Frick,2 Jing Jing Wang,2 Oliver G. Pybus,4 Shadi Salloum,3 Gagandeep S. Narula,2 Anna Eckart,3 Andrea Biezynski,3 Thomas Eiermann,2 Paul Kleer,5 Sergei Viazov,3 Michael Roggendorf,3 Robert Thimme,1 Markus Reiser,1 and Jörg Timm3*

Department of Medicine II, University of Freiburg, Hugstetter Str. 55, 79106 Freiburg, Germany;1 Department of Biochemistry and Molecular Biology, New York Medical College, Valhalla, New York;2 Department of Virology, University of Essen, Hufelandstrasse 55, 45122 Essen, Germany;3 Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom;4 Department of Transfusion Medicine, University of Hamburg, Martinistr. 52, 20246 Hamburg, Germany;5 Nuffield Department of Medicine, South Parks Road, Oxford OX1 3SY, United Kingdom; and Paracelsus Hospital Marl, Department of Medicine, Lipper Weg 11, 45770 Marl, Germany

Received 6 August 2007/Accepted 12 January 2008

Failure of the adaptive immune response to control infection with the hepatitis C virus (HCV) can result from mutational escape in targeted T-cell epitopes. Recent studies suggest that T-cell immune pressure is an important factor in the evolution of the nonstructural proteins in HCV. The aim of this study was to characterize the forces that contribute to viral evolution in an HLA-A*01-restricted epitope in HCV NS3. This epitope represents a potentially attractive target for vaccination strategies since it is conserved across all genotypes. In our cohort of subjects with chronic HCV infection (genotype 1b or 3a), it is a frequently recognized CD8 epitope in HLA-A*01-positive subjects. Viral sequence data reveal that an escape variant is the dominant residue in both genotypes. The predominant Y1444F substitution seemingly impairs binding to the HLA-A*01 molecule, which may have an important impact on the ability to prime a functional CD8 response upon infection. Interestingly, a case of evolution toward the prototype sequence was observed during chronic infection, possibly because the helicase activity of the protein containing the Y1444F substitution is reduced compared to the prototype sequence. Comparison of HCV sequences from Asia and Europe suggests that the frequency of the HLA-A*01 allele in a population may influence the frequency of the escape variant in circulating strains. These data suggest a complex interaction of multiple forces shaping the evolution of HCV in which immune pressure both within the individual and also at the population level in addition to functional constraints are important contributing factors.

A major characteristic of hepatitis C virus (HCV) is the tendency to set up persistent infection. An important role of the cellular immune response for viral control has been highlighted (reviewed in reference 6). Failure of HCV-specific T cells to result from mutational escape in targeted T-cell epitopes (10). Several studies have demonstrated selection of escape mutations during acute HCV infection (18,39,41) and suggested T-cell immune pressure as an important driving force for the evolution of the nonstructural proteins in HCV (8). Moreover, population-based studies revealed that at least in some CD8 epitopes selection of escape mutations is reproducible in subjects sharing the same restricting HLA alleles (16,34,35,42). However, functional constraints on the encoded proteins clearly limit the degree of variation of the viral genome (33,36). Accordingly, reversion of CD8 escape mutations back to the prototype sequence upon transmission to a host who does not express the restricting allele has been demonstrated in human immunodeficiency virus (HIV) and HCV (1,14,27,35,41). In contrast, some escape mutations can also be stable upon transmission, either if the associated fitness costs are low or if subsequent compensatory evolution occurs. For HIV it has been described that these stable mutations potentially persist in circulating viruses in a population to an extent where the escape variant appears as the predominant consensus sequence (3,26).

For the design of effective immunotherapeutic strategies it is important to identify attractive immune targets in the pathogen. HCV epitopes that are conserved across genotypes are particularly valuable because they potentially provide protection against different genotypes. The HLA-A*01-restricted epitope NS3 1436-1444 (ATDALMTGY) in the helicase portion against different genotypes. The HLA-A*01-restricted epitope NS3 1436-1444 (ATDALMTGY) in the helicase portion against different genotypes. The HLA-A*01-restricted epitope NS3 1436-1444 (ATDALMTGY) in the helicase portion against different genotypes. The HLA-A*01-restricted epitope NS3 1436-1444 (ATDALMTGY) in the helicase portion against different genotypes. The HLA-A*01-restricted epitope NS3 1436-1444 (ATDALMTGY) in the helicase portion against different genotypes. The HLA-A*01-restricted epitope NS3 1436-1444 (ATDALMTGY) in the helicase portion against different genotypes. The HLA-A*01-restricted epitope NS3 1436-1444 (ATDALMTGY) in the helicase portion against different genotypes.

1 Corresponding author. Mailing address: University of Essen, Department of Virology, Hufelandstrasse 55, 45122 Essen, Germany. Phone: 49 201 723 2306. Fax: 49 201 723 5929. E-mail: joerg.timm@uni-due.de.
2 Published ahead of print on 23 January 2008.
chronic HCV genotype 1b or 3a infection. We show that this epitope is a target of the CD8 T-cell response in HLA-A*01-positive subjects infected with HCV. Viral sequence data reveal that the potential escape variant is the dominant residue in both genotypes studied in Western populations, which are rich in HLA-A*01. This indicates that reproducible immune pressure on this epitope may drive continuous selection and potentially accumulation of the escape variant in populations with a high frequency of HLA-A*01. We demonstrate that the selected Y144F substitution impairs binding to the HLA-A*01 molecule, which may have an important impact on the ability to prime a functional CD8 response against this epitope upon infection.

MATERIALS AND METHODS

Subjects. Subjects with chronic HCV infection were recruited from the hepatology outpatient clinics of the University of Bochum and the University of Freiburg. Subjects were selected from the serum bank of the German Hepatitis Network (Kompetenznetz Hepatitis). Subjects were included if they presented with a positive HCV RNA test in serum and no clinical evidence for acute infection. Only subjects infected with genotypes 1b and 3a were chosen. The study was approved by the local Institutional Review Board, and all subjects gave written informed consent.

HLA typing. HLA-A and -B alleles were serotyped (85 subjects) or genotyped (12 subjects). The analysis was performed by the blood bank service of the University Hospital Freiburg or the Department of Transplantation Medicine at the University of Hamburg.

Polynuclear expansion of HCV-specific T cells. A total of 4 × 10^5 peripheral blood mononuclear cells (PBMC) were resuspended in 1 ml of complete medium (RPMI 1640 containing 10% fetal calf serum, 1% streptomycin-penicillin, and 1.5% HEPES buffer [1 mol/liter]) and stimulated with peptide (10 µg/ml) and anti-CD28 (0.5 µg/ml; BD Pharmingen). On day 3 and 10, 1 ml of complete medium and recombinant interleukin-2 (IL-2; 20 U/ml; Hoffmann-La Roche) were added. On day 7, the cultures were restimulated with the corresponding peptide (10 µg/ml) and 10^6 irradiated autologous feeder cells. On day 14, the cells were tested for gamma interferon (IFN-γ) secretion after 5 h of stimulation with the corresponding synthetic peptides.

Intracellular cytokine staining (ICS). Procedures were performed as described previously (40). Briefly, cells were stimulated with peptides (10 µg/ml) or peptide loaded HLA-A*01 allogeneic Epstein-Barr virus (EBV) immortalized B-cell lines (B-CLL) in the presence of 50 U of recombinant IL-2/ml and 1 µl of brefeldin A (BD Pharmingen)/ml. For comparison of peptides, increasing concentrations of the prototype and variant peptides were used. After 5 h of incubation (37°C, 5% CO2), cells were blocked with immunoglobulin G1 antibodies and permeabilized with Cytoperm (BD Pharmingen). The cells were stained with antibodies to IFN-γ (BD Pharmingen) and fixed in 100 µl of CellFIX (BD Pharmingen) per well before fluorescence-activated cell sorting analysis.

Amplification, sequencing, and cloning of HCV RNA. Viral RNA was extracted from plasma samples by using a Quiagen (Hilden, Germany) vRNA purification kit and population sequenced on an ABI 3730 XL auto-gel (Qiagen gel extraction kit) or directly purified utilizing a QIAquick PCR purification kit (Qiagen) and population sequenced on an ABI 3730 XL automated sequencer. Some PCR products were also cloned (TOPO TA; Invitrogen) and sequenced. Sequences were aligned and edited by using CodonCode Aligner (Codemac, MA) and Se-Al (http://evolve.zoo.ox.ac.uk). Sequence data are available at GenBank under accession numbers EU078744 to EU078840.

Phylogenetic analysis. All available reference sequences >300 bp in length which spanned the sequenced region were identified and downloaded from the HCV sequence database (21). Sequences with known countries of origin were retained. All sample sequences and reference strains were aligned by hand using Se-Al. Phylogenies were estimated by using the maximum-likelihood approach implemented in PAUP*4 (38). The HKY model (19) with gamma-distributed among-site rate heterogeneity was used. A heuristic search for the maximum-likelihood tree was performed by using the tree-branch-reconnection and nearest-neighbor-interchange algorithms. The resulting estimated trees were midpoint rooted, and branches were colored according to the Fy polymorphism in the epitope using FigTree 1.0 (http://evolve.zoo.ox.ac.uk). A neighbor-joining bootstrap analysis was performed with distances calculated by using maximum likelihood under the “HKY+gamma” model, as described above. Bootstrap scores of >0.60 are shown in boxes on the appropriate branches.

Expression of the HCV genotype 3a helicase, mutagenesis, and purification. HCV cDNA derived from patient 3a-Ch (accession no. EU287946) was amplified in PCRs using the primers Nhe-3a (5’-GGC GCC GCT AGC CTA CAA TTC ATA CCA GGA GTA AGG CAC C-3’) and Bam-3a (5’-GGC GCC GCA TCC CCA AGC ACC AAC CAG GTG GTG GTG GTC-3’). The PCR product was digested with BamHI and NheI and ligated into a similarly treated pET24 vector (Novagen). The resulting p24Hel-3a V144F was subject to site-directed mutagenesis using a Quikchange kit (Stratagen) and the primers F1444Y(+) (5’-GGC CCT CAT GAC TGG ATA CCA CCG AGA AGC ATG TCG AAG TCT CCG GTG CAT CCA GTG AGG GGC-3’). The sequence of the resulting plasmid (p24Hel-3a proto) was confirmed, and both p24Hel-3a proto and p24Hel-3a V144F were transformed into BL21 (DE3) cells. The transformed cells were grown to express the 53-kDa protein containing NS3 amino acids 116 to 631 (the complete HCV helicase) flanked by an N-terminal MAS sequence and a C-terminal STWLYL GDPNSSSVKDLAELHSHHHHHH sequence (His tag).

Proteins were purified after suspension in 5 volumes of buffer A (20 mM Tris [pH 8.0], 0.5 M NaCl, 5 mM imidazole), sonicated, and centrifuged. The lysate was applied to a 5 ml of nickel-NTA column (Novagen) and eluted with a linear imidazole gradient. Fractions containing the corresponding proteins were pooled, dialyzed against 50 mM Tris-HCl (pH 8), and subjected to gel filtration chromatography, followed immediately by ion-exchange chromatography using DEAE-Phenylarose (Pharmacia). Protein concentrations were determined from calculated extinction coefficients at A280 (Table 2). Protein concentration and purity were determined by analyzing Western blots with anti-NS3 antibodies (Scantibodies, San Diego, CA) and stained with antibodies to CD8. After protein purification, the Nterminal His-tagged protein by amplying cDNA from the plasmid pBDL2492pI+ (29) using the PCR primers N3si (+) (5’-GGC GCC GCG GGT GGC GCT CCT GTC ATG ACC GAC C-3’) and N3si (-) (5’-GGC GCC GAA TTC CCT GAC GGT ACC TCC AG-3’). The PCR product was purified, digested with Nhel and EcoRl (sites underlined), and ligated into a similarly treated pET28 vector (Novagen). The NS3 protein was expressed and purified as described previously for truncated NS3 except that the final DEAE column was substituted with a Fractogel EMD COO- column (EMD Biosciences).

The prototype Y1444 sequence was generated in a conl background by subjecting plasmids expressing either Hel-1b (conl) or NS3 (conl) to site-directed mutagenesis using the primers F1444-Conl (+) (5’-GAC CTA CTA ATG AGC GAC TAT ACC GCC GAT TTC TCC AGC-3’) and F1444-Conl (-) (5’-GTC GAA TTC CCT GAC GGT ACC TAC AGC-3’). The resulting proteins Hel-1b (F1444Y) and NS3 (F1444Y) were purified as described above.

Helicase assay. HCV helicase-catalyzed DNA unwinding was monitored by using a novel fluorescent “self-trapping” unwinding assay. This assay monitors the displacement of a dual-labeled DNA oligonucleotide from a longer immobilized oligonucleotide. Upon displacement, both strands form hairpin structures, making the reaction essentially irreversible. Hairpin formation leads to a fluorescent increase because a Cy5 fluorophore approaches an Iowa Black quencher moiety (Integrated DNA Technologies). To generate the substrate, 25 µM concentrations of the labeled short strand (Cy5-GCT CCCCAA TCG ATG AAC GGC GAG C-IBQ) and long strand (3’-TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT CCGA GGG GTG AGC TAC TTG CCC CTG5-3’) were...
heated to 95°C and then cooled to room temperature. The annealed substrate was diluted to 5 mM in reaction buffer (25 mM MOPS [pH 6.5], 2 mM MgCl$_2$, 0.1% Tween 20), and fluorescence was monitored continually at 37°C. The helicase protein was then added (final concentration 5 to 100 nM) and equilibrated with the substrate for 2 min before initiation of the unwinding reaction by adding ATP (4 mM). Initial unwinding rates were measured by using linear regression over the initial phase of the reaction (0.5 to 2 min).

ATPase assays. The initial rates of HCV helicase-catalyzed ATP hydrolysis were measured by using a modified colorimetric “malachite green” assay. Reactions were performed with catalytic amounts of HCV helicase (5 to 50 nM) at 37°C in 25 mM morpholinepropanesulfonic acid (pH 7.5). After 10 to 20 min, 25 l of a reaction was added to 200 l of a freshly prepared color reagent (3 parts 34% sulfuric acid, 1 part 4% ammonium molybdate in 4.0 N HCl, 0.01 part 10% Tween 20). Within 1 min, solutions were quenched with 25 l of a 34% sodium citrate and incubated for 20 min at room temperature, and the resulting absorbance was read at 625 nm. Phosphate concentrations were determined from a standard curve generated with the substrate for 2 min before initiation of the unwinding reaction by adding ATP (4 mM). Initial unwinding rates were measured by using linear regression over the initial phase of the reaction (0.5 to 2 min).

**RESULTS**

CD8 epitope ATDALMTGY$_{1436-1444}$ is frequently targeted in subjects infected with HCV genotype 1b or 3a. First, we determined the frequency of recognition of the HLA-A*01-restricted epitope ATDALMTGY$_{1436-1444}$ during chronic HCV infection. PBMC from HLA-A*01-positive subjects infected with genotype 1b or genotype 3a were stimulated with the synthetic peptide to expand antigen-specific T cells. In 5 of 10 subjects (50%) infected with genotype 1b and in 4 of 10 subjects (40%) infected with genotype 3a specific T cells targeting ATDALMTGY$_{1436-1444}$ were detectable (Table 1). In one subject (1b-Ch5) the number of specific T cells targeting this single epitope could be expanded up to 20% of all CD8 T cells. Of note, this response was detectable in only 3 of 10 subjects infected with genotype 1b before antigen-specific expansion (data not shown), suggesting that some low-level responses may be missed if only ex vivo techniques are used. In summary, 9 of 20 HLA-A*01-positive subjects with HCV genotype 1b or 3a infection had a detectable CD8 response against the ATDALMTGY$_{1436-1444}$ epitope, indicating that it is a frequent target even during chronic infection.

**Impact of Y1444F variant on T-cell targeting.** Selection of the Y1444F substitution has been described in a subject during acute HCV genotype 1a infection (8). The Y1444F substitution in the HLA-A*01 epitope is located in the anchor position for MHC class I binding (9), suggesting that the mutation from tyrosine to phenylalanine would affect peptide binding to the HLA-A*01 molecule and, consequently, efficient targeting by specific T cells. Interestingly, in the previous escape study no differential impact on T-cell recognition between the prototype and the variant was observed when tested ex vivo or after antigen-specific expansion of T cells (8). To further evaluate this finding, we determined the impact of the Y1444F substitution on T-cell recognition in all nine subjects targeting this epitope. In line with the previous observation there was substantial cross-recognition between the prototype and the variant in most cases (Fig. 1). Only T cells expanded from subject 1b-Ch6 did not target the variant sequence. In three cases (1b-Ch2, 1b-Ch5, and 1b-Ch8) the secretion of IFN-γ was diminished compared to the prototype. However, in all remaining subjects no difference between the prototype and the Y1444F variant was observed. Interestingly, when PBMC were cultivated in the presence of the variant peptide no or only a minor fraction of specific T cells were detectable against both the prototype sequence and the Y1444F variant. As a possible mechanism, we suggest that the Y1444F variant does not stimulate expansion of antigen-specific T cells against this epitope in vitro.

One possible explanation for this observation is an accelerated off-rate of the variant peptide from the major histocompatibility complex (MHC) molecule that is not unmasked in a

---

**TABLE 1. HLA-A1-positive subjects included for the analysis of ATDALMTGY$_{1436-1444}$specific T cells**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Subject</th>
<th>Age (yr)</th>
<th>Sex</th>
<th>ALT (U/liter)</th>
<th>Viral load (IU/ml)</th>
<th>HLA alleles</th>
<th>% IFN-γ/CD8$^+$ ATDALMTGY$_{1436-1444}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>1b-Ch1</td>
<td>72</td>
<td>F</td>
<td>91</td>
<td>1,500,000</td>
<td>A1, A23, B35, B49</td>
<td>Neg$^d$</td>
</tr>
<tr>
<td>1b</td>
<td>1b-Ch2</td>
<td>32</td>
<td>M</td>
<td>104</td>
<td>375,000</td>
<td>A1, A11, B7, B8</td>
<td>0.41</td>
</tr>
<tr>
<td>1b</td>
<td>1b-Ch3</td>
<td>38</td>
<td>M</td>
<td>68</td>
<td>1,140,000</td>
<td>A1, A11, B52, B61</td>
<td>6.62</td>
</tr>
<tr>
<td>1b</td>
<td>1b-Ch4</td>
<td>19</td>
<td>M</td>
<td>88</td>
<td>30,000</td>
<td>A1, A31, B60</td>
<td>Neg</td>
</tr>
<tr>
<td>1b</td>
<td>1b-Ch5</td>
<td>54</td>
<td>M</td>
<td>76</td>
<td>22,000</td>
<td>A1, B8</td>
<td>20.53</td>
</tr>
<tr>
<td>1b</td>
<td>1b-Ch6</td>
<td>41</td>
<td>M</td>
<td>55</td>
<td>&gt;8,000,000</td>
<td>A1, B8, B17</td>
<td>0.59</td>
</tr>
<tr>
<td>1b</td>
<td>1b-Ch7</td>
<td>64</td>
<td>F</td>
<td>120</td>
<td>566,000</td>
<td>A1, A29, B8, B44</td>
<td>Neg</td>
</tr>
<tr>
<td>1b</td>
<td>1b-Ch8</td>
<td>40</td>
<td>M</td>
<td>238</td>
<td>340,000</td>
<td>A1, A3, B35, B63</td>
<td>8.58</td>
</tr>
<tr>
<td>1b</td>
<td>1b-Ch9</td>
<td>22</td>
<td>M</td>
<td>55</td>
<td>580,000</td>
<td>A1, A3, B8, B35</td>
<td>Neg</td>
</tr>
<tr>
<td>1b</td>
<td>1b-Ch10</td>
<td>30</td>
<td>M</td>
<td>197</td>
<td>2,800,000</td>
<td>A1, A2, B4, B35</td>
<td>Neg</td>
</tr>
<tr>
<td>3a</td>
<td>3a-Ch1</td>
<td>47</td>
<td>M</td>
<td>141</td>
<td>1,000,000</td>
<td>A1, A3, B7, B35</td>
<td>0.33</td>
</tr>
<tr>
<td>3a</td>
<td>3a-Ch2</td>
<td>18</td>
<td>M</td>
<td>450</td>
<td>2,000,000</td>
<td>A1, A3, B44</td>
<td>Neg</td>
</tr>
<tr>
<td>3a</td>
<td>3a-Ch3</td>
<td>37</td>
<td>F</td>
<td>58</td>
<td>61,000</td>
<td>A1, A28, B13, B51</td>
<td>Neg</td>
</tr>
<tr>
<td>3a</td>
<td>3a-Ch4</td>
<td>32</td>
<td>F</td>
<td>142</td>
<td>13,000</td>
<td>A1, A2, B7, B61</td>
<td>Neg</td>
</tr>
<tr>
<td>3a</td>
<td>3a-Ch5</td>
<td>41</td>
<td>F</td>
<td>139</td>
<td>188,000</td>
<td>A1, A26, B8, 35</td>
<td>1.23</td>
</tr>
<tr>
<td>3a</td>
<td>3a-Ch6</td>
<td>41</td>
<td>M</td>
<td>170</td>
<td>5,800</td>
<td>A1, A24, B7, B8</td>
<td>Neg</td>
</tr>
<tr>
<td>3a</td>
<td>3a-Ch7</td>
<td>41</td>
<td>M</td>
<td>61</td>
<td>220,000</td>
<td>A1, A3, B8, B44</td>
<td>6.37</td>
</tr>
<tr>
<td>3a</td>
<td>3a-Ch8</td>
<td>34</td>
<td>F</td>
<td>78</td>
<td>33,000</td>
<td>A1, A2, B7, B13</td>
<td>Neg</td>
</tr>
<tr>
<td>3a</td>
<td>3a-Ch9</td>
<td>48</td>
<td>M</td>
<td>24</td>
<td>3,080,000</td>
<td>A1, A32$^e$</td>
<td>Neg</td>
</tr>
<tr>
<td>3a</td>
<td>3a-Ch10</td>
<td>21</td>
<td>M</td>
<td>184</td>
<td>1,100,000</td>
<td>A1, A2, B8, B44</td>
<td>4.27</td>
</tr>
</tbody>
</table>

$^a$ F, female; M, male.

$^b$ After 14 days of in vitro stimulation.

$^c$ B alleles were not determined.

$^d$ Neg, negative.
standard ICS assay, where peptide is present continuously. To address this, we loaded the variant or prototype peptide on allogeneic HLA-A*01-positive B cells, washed the cells extensively, and used them in an ICS assay as targets for specific T cells that have been expanded in the presence of the prototype (Fig. 2). B cells pulsed with the prototype peptide were efficiently targeted. In contrast, B cells pulsed with the variant peptide were not targeted after the washing step. In this case the stability of the MHC-peptide complex may not have been sufficient, and the variant peptide was removed during the washing procedure. In summary, the Y1444F substitution represents a true escape mutation in the HLA-A*01 restricted epitope that may act through a decrease in binding to the MHC class I molecule.

Putative escape variant Y1444F is predominant in two cohorts infected with HCV genotype 1b or 3a. To determine the
frequency of the Y1444F escape mutation in circulating viruses in our cohort, the region of NS3 containing this epitope was sequenced from 49 subjects infected with HCV genotype 1b and 48 subjects infected with genotype 3a (Fig. 3). In both cohorts the most common residue at position 1444 is the Y1444F substitution, and this putative escape variant represents the consensus sequence. The prototype sequence ATDALMTGY is only present in 9 of 49 (18.4%) subjects infected with HCV genotype 1b and 12 of 48 (25.0%) subjects infected with genotype 3a. All HLA-A*01-positive subjects harbor a virus with the escape sequence (ATDALMTGF), including the subjects that were previously tested for the presence of the corresponding CD8 response (Table 1). Interestingly, the majority of HLA-A*01-negative subjects are also infected with the escape variant: 22 of 31 (71.0%) of genotype 1b patients and 19 of 31 (61.3%) of genotype 3a patients. This indicates that the escape variant is relatively stable also in the absence of HLA-A*01-associated immune pressure and persists in these populations. Of note, we tested PBMC from 10 HLA-A*01-negative subjects before and after antigen-specific stimulation with the prototype sequence (stimulated with ATDALMTGY) in an ICS assay for IFN-γ. Dot plots from two subjects (1b-Ch5 and 3a-Ch7) are shown.

FIG. 2. Accelerated off-rate of the Y1444F variant. HLA-A*01-positive allogeneic EBV immortalized B-cell lines were incubated without peptide (left panels) or in the presence of the prototype peptide (middle panels) or the Y1444F variant (right panel). After six washes with medium the B cells were used as targets for polyclonally expanded T cells (stimulated with ATDALMTGY) in an ICS assay for IFN-γ. The frequency of Y1444F escape mutation was identified in position 1444 in the quasispecies population (data not shown), suggesting that this residue had achieved fixation within the virus populations infecting these eight individuals.

Next we investigated whether longitudinal analysis of HCV sequences reveals ongoing evolution in the ATDALMTGY Y1436-1444 epitope during chronic infection. Follow-up samples were therefore collected and sequenced. Most subjects received treatment after their first presentation and were therefore excluded from longitudinal analysis. Follow-up samples from four untreated HLA-A*01-negative subjects between 32 and 78 weeks after the first time point were available. No sequence evolution was observed in two subjects (Fig. 4). A third subject infected with genotype 1b showed one variant (G1443D) in the epitope in 1 of 11 clones; however, position 1444 was fixed and did not evolve between these two time points. A fourth subject, 3a-Ch37, harbored a virus with the Y1444F substitution at the first time point. Interestingly, sequence analysis at follow-up 40 weeks later revealed a mixed base in the bulk sequence coding for both the prototype sequence and the Y1444F polymorphism, suggesting that either neutral evolution or selectively driven reversion of the Y1444F site was under way. Analysis of the clonal sequences confirmed the presence of a mixed quasispecies population with 7 of 12 clones harboring the prototype sequence in the second time point (Fig. 4). Even though the prototype sequence might have been present in a very low frequency at the previous time point the shift in the clonal frequency indicates that evolution toward the prototype sequence can occur in genotype 3a.

Y1444F substitution is associated with decreased NS3 helicase function. The Y1444F substitution is located in the helicase domain of the NS3 protein. Since the amino acid substitution is near the presumed ATP and RNA binding sites of the protein, such genetic variation could impair its functional performance. To test this hypothesis, we cloned NS3/4A from patient 3a-Ch7 infected with HCV genotype 3a and subjected it to site-directed mutagenesis to generate both variants. The helicase domain of NS3 containing either the prototype (Hel-3a proto) or the Y1444F substitution (Hel-3a Y1444F) was then expressed as a recombinant protein and tested in a DNA unwinding assay, which monitors the displacement of a dual-labeled DNA oligonucleotide from a longer complementary oligonucleotide. Upon displacement both strands form hairpin structures, making the reaction essentially irreversible. Hairpin formation is monitored by a fluorescence decrease of a Cy5 fluorophore (Fig. 5A). The unwinding rates were compared to rates obtained using similar HCV helicase proteins from other genotypes (22). An alignment of the proteins used, showing the region surrounding the ATDALMTGY epitope, is shown in Fig. 5B. Notably, helicase containing the escape mutant (Hel-3a Y1444F) unwound the DNA substrate about five times slower than the prototype sequence (Hel-3a proto), suggesting that the Y1444F variant results in less active helicase (Fig. 5C). This difference in helicase activity was reproducible at different concentrations. At each concentration Hel-3a proto unwind DNA faster than Hel-3a Y1444F (Fig. 5D). To exclude that this observed difference is due to variations in the amount of active protein in either preparation the HCV helicase catalyzed ATP hydrolysis was determined for both helicases under optimal conditions. The maximum rate of ATP hydrolysis divided by the total protein concentration revealed...
that Hel-3a proto was slightly more active than Hel-3a Y1444F ($k_{\text{cat}} = 37 \pm 13$ versus $25 \pm 15$). However, this difference was not statistically significant. When ATP hydrolysis was measured at various substrate concentrations, nearly identical results were obtained for Hel-3a proto and Hel-3a Y1444F ($K_m = 188 \pm 22$ versus $193 \pm 26$). Thus, both the turnover rate ($k_{\text{cat}}$) and the catalytic efficiency ($k_{\text{cat}}/K_m$) of HCV helicase-catalyzed ATP hydrolysis appeared to be unchanged by the escape mutation.

To examine the impact of the escape mutation in the genotype 1b background, the plasmid encoding Hel-1b (con1), which contains the escape mutation Y1444F, was subjected to site-directed mutagenesis to generate the prototype sequence. The resulting protein (Hel-1b [F1444Y]) was then compared to the other helicases studied here (Fig. 5C and D). The con1 helicase with the prototype sequence (Hel-1b [F1444Y]) was slightly more active than the helicase with the escape mutation (Hel-1b [con1]). However, the difference in genotype 1b was minor compared to genotype 3a. Hel-3a (proto) has a specific activity (i.e., nM DNA unwound/min/nM protein) five times higher than that of the escape mutant Hel-3a (Y1444F), but Hel-1b (F1444Y) is only ca. 20% (1.2-fold) more active than the same protein with the escape mutation (Hel-1b [con1]). Similar differences between the proteins were observed when
they were assayed for their ability to unwind RNA (data not shown).

Since it has been established that the protease domain of NS3 plays a role in helicase action (13), we also examined the impact of the escape mutation on the helicase activity of the full-length NS3 protein. To this end, we subjected a plasmid encoding a full-length NS3 protein from the genotype 1b (con1) strain to mutagenesis. The resulting substitution (F1444Y) again reverted the escape mutation in the con1 background. The results with this protein were similar to those with the truncated NS3 protein (Fig. 5C). The full-length protein is about 2.5 times more active than truncated NS3 in this assay, and again the prototype sequence (NS3 [F1444Y]) is about 1.2 times more active than the escape mutant (NS3 [con1]). Again, a similar difference was observed when RNA was used as a substrate (not shown).

Phylogenetic analysis reveals numerous unambiguous changes between Tyr and Phe at position 1444. Recent reports suggest that the ability to detect HLA-associated sequence polymorphisms in HIV and HCV may be reduced by so-called "founder effects" in the phylogeny (5, 42). Sequences with a particular polymorphism may be closely related and form a distinct phylogenetic cluster. In this case, it is argued that it is less likely that the escape variant evolved by immune pressure independently within each individual; instead, it is considered more likely that a limited number of evolutionary events occurred in the founders of the lineage or cluster in question.

In order to address this in our data set a maximum-likelihood phylogeny was calculated for all genotype 1b and 3a sequences, upon which was mapped all unambiguous changes between Tyr and Phe in position 1444. A total of 16 unambiguous changes between Tyr and Phe at position 1444 were identified (nodes labeled with a black dot), including four cases with sequences from our cohort, again suggesting multiple mutational events at this site in genotype 1b. One of these sequence changes that is supported with a high bootstrap value (98%; indicated by an arrow in the figure) includes a prototype...
sequence from an HLA-A*01-negative subject (1b-Ch42) and a sequence with the Y1444F substitution from an HLA-A*01-positive subject (1b-Ch06). However, the tree also contains clusters that contain solely either prototype sequences or escape variant sequences. The majority of sequences from our cohort (30 of 49) are included in the middle section of the tree (boxed in dark gray) divided into three clades. Interestingly, these three clades predominantly contain sequences with the Y1444F substitution (42 of 46) but also include the sequence change that is supported by a high bootstrap value (arrow). Four more sequences from our cohort are part of another striking cluster (boxed in light gray) that predominantly contains the prototype. The remaining 15 sequences are distributed throughout the tree without a striking pattern. In summary, despite clade-specific differences in sequence variation at site 1444, which suggest that some local founder effect do exist, the observation of multiple unambiguous sequence changes indicate that this site is continuously evolving.

Region-specific differences in the frequency of the prototype and Y1444F variant correlate with HLA-A*01 frequency. One cluster in the tree of HCV genotype 1b is unique since it contains predominantly prototype sequences (boxed in light gray). Interestingly, most sequences included are from isolates of Asian origin. Twelve of the twenty sequences are from Japan, China, or Taiwan. The eight remaining sequences are from our cohort (four sequences), France, Australia, or the United States. In contrast, Asian sequences are absent from the three clades that predominantly contain Y1444F variant sequence.

**FIG. 5.** Impact of the Y1444F substitution on NS3 helicase activity. (A) The helicase portion of NS3 from various isolates was expressed as a recombinant protein, and DNA unwinding was monitored by using a fluorescent “self-trapping” unwinding assay. This assay monitors the displacement of a dual-labeled DNA oligonucleotide, which then forms hairpin structures. Hairpin formation leads to a fluorescence decrease. (B) An alignment of sequences spanning the ATDALMTGY1436-1444 epitope from the various proteins tested is shown. The substitution in position 1444 is the only difference between the two proteins derived from 3a-Ch7 (genotype 3a) and con1 (genotype 1b). Of note, con1 has phenylalanine in position 1444 and was mutated to contain the prototype epitope sequence (ATDALMTGY). (C) The unwinding rates were compared between the various helicase proteins (100 nM), and the helicase activity is shown in nM/min. White columns are results of the truncated helicas, and gray columns results of the full-length NS3 protein. (D) The unwinding rates of Hel-1b (con1), Hel-1b (F1444Y), Hel-3a (proto), and Hel-3a (Y1444F) were compared at different concentrations.
sequences from our cohort (boxed in dark gray). This raised the possibility that there are region-specific differences in the frequency of the prototype and the Y1444F substitution. A total of 149 sequences are included in the tree: 49 sequences are from our cohort recruited in Germany (Table 2), an additional 35 previously published sequences are from Europe (France, Germany, Russia, Ireland, and Switzerland), 35 sequences are of Asian origin (Japan, China, Taiwan, and Korea), 17 sequences are from the United States, 12 sequences are from Australia, and 1 more is from Turkey. The frequency of the Y1444F substitution is regionally different, with the highest frequency in published sequences from Europe (88.6%) and the lowest frequency in isolates from Asian countries (42.9%). These region-specific differences in the distribution of the prototype and the Y1444F variant correlate with the frequency of the HLA-A*01-allele in different populations.
FIG. 7. Maximum-likelihood phylogeny for HCV genotype 1b sequences. A maximum-likelihood phylogeny was calculated for the genotype 1b sequences from our cohort and all available reference sequences from the HCV sequence database >300 bp in length that span the sequenced region. The accession numbers and countries of origin of the reference sequences are indicated in the branch labels. The branches of the tree are color-coded: blue branches represent the prototype sequence, and red branches represent the Y1444F substitution. Sixteen different internal nodes in the tree split into a blue and a red branch (labeled with a black dot), indicating unambiguous and independent changes between Tyr and Phe. A case of an unambiguous change that includes two sequences from our cohort is indicated with an arrow. The tree also contains unique clusters with predominantly prototype sequences (boxed in light gray) or predominantly the Y1444F variant (boxed in dark gray). Bootstrap scores of >60% are shown in boxes on the appropriate branches.
HLA-A*01 is a common allele in Europe, with phenotype frequencies between 20 and 35% (Fig. 8). In contrast, HLA-A*01 is rare in Asians, with reported frequencies of <6%. Interestingly, only in isolates of Asian origin in the consensus residue in position 1444 is the prototype residue tyrosine (Table 2), suggesting that isolates from this region are potentially less well adapted to HLA-A*01-mediated selection pressure.

**DISCUSSION**

The propensity of HCV to evolve with a high mutation rate, enabling selection of escape variants in targeted regions, makes the design of effective immune therapies difficult. Here we describe the impact of MHC class I-associated selection pressure on the frequency of an escape mutation in an HLA-A*01-restricted CD8 epitope in NS3. Continuous HLA-A*01-mediated selection pressures in two cohorts infected with HCV genotype 1b or 3a replace the prototype epitope sequence (ATDALMTGY) with an escape variant (ATDALMTGF) that seems to impair MHC class I binding.

Recent studies in HIV-1 suggest that escape variants can become predominant in circulating strains in a population (3, 26, 31). In an analysis of HLA class I-associated sequence polymorphisms in viral genomes, these escape mutations represent “negative associations” (16, 31, 42), which means that a particular HLA allele is associated with presence of the consensus sequence. As a driving mechanism for these negative associations, reproducible selection in the presence of CD8 immune pressure, stability of the mutation upon transmission, and the lack of reversion in the absence of immune pressure have been proposed (26). Under these conditions continuous accumulation of the escape variant ultimately replaces the “wild-type” sequence as the most frequent strain in a population, and the escape residue becomes the predominant consensus residue. It has been highlighted that this mechanism may potentially lead to complete deletion of epitopes from the population after sufficient time has elapsed and all circulating “wild type” sequences are replaced (26). Here we similarly demonstrate for HCV that an escape mutation in an HLA-A*01-restricted epitope in HCV NS3 is predominant in circulating isolates. Mutational escape during acute infection has been described for this epitope in a subject infected with genotype 1a (8). Moreover, significant associations between the Y1444F substitution and expression of the HLA-A*01 allele in larger cohorts with chronic infection have been previously reported (16, 42). However, different frequencies of this escape mutation have been observed. The predominant residue in a cohort infected with genotype 1a in North America was the prototype, whereas in a mixed cohort from Switzerland and Western Australia infected with genotype 1a or 1b the escape mutation Y1444F was predominant, indicating regional differences and potentially differences between genotypes or subtypes.

Even though the Y1444F substitution in the HLA-A*01 epitope is located in the anchor position for MHC class I binding, Cox et al. (8) observed in a previous escape study no differential impact on stimulation of IFN-γ secretion both ex vivo and using expanded specific T cells from a subject with acute HCV infection. The hydrophilic tyrosine is the common residue for the anchor in position 9 of the HLA-A*01 motif (9). Substitution with the similar but hydrophobic phenylalanine at this crucial site likely preserves the peptide’s structure but impairs binding to the MHC molecule. This is supported by another study suggesting that the tyrosine-to-phenylalanine substitution in this anchor position reduces HLA-A*01 binding (20) and is in line with a previously reported >200-fold-reduced binding affinity of the Y1444F variant to the HLA-A*01 molecule compared to the prototype (8). We observed substantial cross-recognition between the two variants when added in an ICS assay, where the peptide is present throughout the whole incubation period. However, when loaded on B cells that are washed after the incubation period, the presentation of the variant peptide is significantly impaired. This suggests that MHC class I binding is likely affected, resulting in a higher drop-off rate of the variant peptide. A similar subtle effect of an escape mutation that only unmasked when the off-rate from pulsed B cells was determined has been previously described for HIV (17, 26).

Accumulation of escape mutations in circulating strains may have important implications for the ability to prime a functional antiviral CD8 immune response upon acute infection. Even though the plasticity of the T-cell response allows priming of CD8 T cells against various sequences including escape variants (2, 7), mutations of the anchor residue for MHC class

**TABLE 2. Regional differences in the frequency of the Y1444F substitution**

<table>
<thead>
<tr>
<th>Region</th>
<th>No. of sequences</th>
<th>Y/F</th>
<th>% F</th>
<th>Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>49</td>
<td>9/40</td>
<td>81.6</td>
<td>F</td>
</tr>
<tr>
<td>Europe</td>
<td>35</td>
<td>4/31</td>
<td>88.6</td>
<td>F</td>
</tr>
<tr>
<td>Asia</td>
<td>35</td>
<td>20/15</td>
<td>42.9</td>
<td>Y</td>
</tr>
<tr>
<td>United States</td>
<td>17</td>
<td>5/12</td>
<td>70.6</td>
<td>F</td>
</tr>
<tr>
<td>Australia</td>
<td>12</td>
<td>5/7</td>
<td>58.3</td>
<td>F</td>
</tr>
<tr>
<td>Remaining regions</td>
<td>1</td>
<td>0/1</td>
<td>50.0</td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 8. Frequency of the HLA-A*01 allele in populations in Europe and Asia.** The frequencies of the HLA-A*01 allele in different populations were retrieved from a public database (http://www.allelefrequencies.net). The observed phenotype frequencies in different cohorts are shown.
I binding as observed here make this scenario unlikely. Lack of efficient processing and presentation of the antigen completely prevents effective priming. In line with this we were reproducibly not able to expand specific T cells targeting the ATDALMTGY_{1436-1444} epitope upon stimulation with the escape variant. Transmission of a virus harboring an antigen-processing mutation or an escape mutation in the anchor residue resulted in the lack of priming of a normally immunodominant CD8 response during acute HIV-1 infection (1, 3). The ATDALMTGY_{1436-1444} epitope analyzed here represents an immunodominant CD8 response during acute and resolved HCV infection (23, 24). However, it is unclear how lack of such an immunodominant CD8 response may affect the functionality of the overall T-cell response and the ability to control viral replication. In animal models, conflicting results on the consequences of the lack of immunodominant CD8 epitopes on disease control have been reported. Infection of mice with HSV-1 or influenza virus lacking one immunodominant CD8 epitope resulted in a shifted epitope hierarchy, but enhancement of otherwise-subdominant CD8 responses compensated for the missing dominant response (4, 37). In contrast, mice infected with a lymphocytic choriomeningitis virus escape variant in one or two immunodominant CD8 epitopes presented with prolonged viremia and disease (32, 45). In HCV, expression of the HLA A*01-B*08-Cw07-DRB1*03011-DQB1*0201 haplotype was found in one study to be associated with viral persistence (30). That study was based on an analysis of a cohort infected by contaminated anti-D immune globulin from a single source in Ireland. Interestingly, the infecting virus harbored two escape mutations: one was the Y1444F mutation, and the other was located in an immunodominant HLA-B*08-restricted epitope (HSKKKCEDEL_{1395-1403}) in NS3 (35). It is tempting to speculate that adaptation in the ATDALMTGY_{1436-1444} epitope and that the B*08-epitope contributed to the failure of this HLA-haplotype to control viral replication.

We detected the CD8 response against the ATDALMTGY_{1436-1444} epitope in ca. 50% of all subjects. However, only 25% of circulating isolates contain the prototype sequence. Assuming that both variants can be transmitted upon infection, this high frequency of the CD8 response may be unexpected. As a possible explanation, it was hypothesized for a similar observation in the simian immunodeficiency virus-infected rhesus macaque model that viral variants may have reverted back to the prototype sequence and primed the response but were rapidly eliminated by specific T cells and therefore not detectable in serum (15). Transient reversion to the prototype as the basis for priming of a CD8 response after infection with an escape variant has been documented (1, 11), even though this may result in delayed priming of the response (1). In our study we were not able to find the prototype epitope sequence even in minor frequencies in HLA-A*01-positive subjects. However, we observed one case of evolution back to the prototype residue in the absence of the HLA-A*01 allele, indicating that such mutations occur at least in some isolates. It therefore seems possible that a low frequency of the prototype sequence was also present in some HLA-A*01-positive subjects but was missed in our quasispecies analysis. This may have driven the corresponding CD8 response at a very low detection level, and the immune response at the same time may have prevented full mutation back to the prototype.

We observed one case of evolution toward the prototype residue in a subject with chronic HCV genotype 3a infection. Similarly, reversion in this epitope was observed in 4 of 11 subjects in the Irish cohort infected by a single source of HCV genotype 1b (32). In contrast, in a similar single source outbreak with HCV genotype 1b in East Germany (46), only 1 of 19 sequences from HLA-A*01-negative subjects has reverted 20 years after infection (data not shown). This raises the possibility that the Y1444F substitution is at least in some but not all isolates associated with fitness costs to the virus. To address the impact of the Y1444F substitution on the function of the encoded HCV helicase, the protein from a genotype 3a isolate and from genotype 1b (con1) was expressed, and the helicase activity was determined in vitro. The helicase activity of the Y1444F variant was reproducibly diminished compared to the prototype in the genotype 3a isolate. Similarly, in the con1 context the helicase containing the prototype epitope sequence (ATDALMTGY) was more active compared to the variant; however, the effect was in this case only minor. The recently developed HCV culture systems that produce infectious viral particles are based on HCV genotype 2a (28, 44, 47). Therefore, direct evaluation of the impact of this escape mutation observed in HCV genotype 3a and 1b on viral replication capacity is at the moment difficult. It is unclear how this difference in helicase function translates into differences in overall replication capacity. In a transient-replication model utilizing a subgenomic replicon based on con1, both the prototype and the variant were replication competent on a similar level, even though subtle differences in the replication level were difficult to dissect (data not shown). If optimal performance of the helicase supports viral replication, it seems possible that the prototype sequence is advantageous overall, even though clear evidence of a fitness benefit is lacking from the in vitro assays. Potentially, even subtle differences in replication capacity result in a selection advantage of a variant in vivo. However, the observation in our cohort that the majority of HLA-A*01-negative subjects are infected with the escape variant indicates that the Y1444F substitution is relatively stable in many isolates and that evolution toward the prototype in the absence of immune pressure is not universal. Of note, there was no significant difference in viral load between subjects with the prototype and the Y1444F variant (data not shown).

The frequency of escape mutations in circulating strains in a population is not only determined by the associated fitness costs and consequently the rate of reversion. CD8 immune pressure mediated by frequent HLA class I alleles may contribute proportionally more to viral evolution at the population level compared to rare alleles. Hence, circulating strains in a population may be more adapted to immune pressure mediated by frequent HLA alleles and less adapted to immune pressure by rare alleles. In HIV it has been suggested that the expression of rare HLA alleles to which the virus is not yet adapted is even advantageous for containing viremia (43). HLA-A*01 is a frequent allele in a Caucasian population, with reported phenotype frequencies between 20 and 35%. Reproducible immune pressure mediated by this common allele may support a high prevalence of the Y1444F escape variant in circulating strains. Inter-
estingly, the frequency of the Y1444F escape variant in isolates from different regions correlates well with reported HLA-A*01 frequencies. Sequences from Asia, where HLA-A*01 is rare, predominantly harbor the prototype sequence, indicating that circulating viruses in Asia are less frequently exposed to immune pressure mediated by this allele. It will therefore be interesting to investigate whether HLA-A*01 is beneficial in such populations, i.e., in which this allele is rare and the circulating strains are predominantly prototype. However, the phylogenetic analysis of genotype 1b sequences suggests that in addition to numerous unambiguous sequence changes there are region-specific clusters that predominantly contain either prototype or the Y1444F variant. The interpretation of this observation is difficult. A recent phylogenetic reevaluation of HLA-associated sequence polymorphisms in HIV suggested that these clusters are unlikely to be selected by immune pressure; this phenomenon was termed “founder effect” (5). Many of the reported viral sequence differences in HIV corresponded to different subtypes, resulting in a detection bias toward associations between subtype-specific sequence polymorphisms and the predominant HLA allele of the population in which this subtype is common. However, in the present study, in which subtypes were analyzed separately, multiple sequence changes in position 1444 in HCV genotype 1b indicate that this site is continuously evolving. Even though we cannot completely rule out other mechanisms, we hypothesize that the Asian cluster predominantly contains prototype sequences because it was less exposed to HLA-A*01-associated selection pressure. In turn, sequences in clusters predominantly containing Y1444F sequences may have additional sequence changes that compensate for fitness costs, which is the subject of additional ongoing studies. This may prevent frequent reversion, giving the phylogeny a less dynamic picture. Additional studies are clearly needed to address the impact of regional differences in HLA frequencies on the global evolution of HCV.

In summary, we demonstrate here that several factors influence the evolution in an HLA-A*01-restricted epitope in NS3 in the individual but also at the population level. Importantly, this epitope is a frequent target. Specific CD8+ T cells are detectable in approximately half of all HLA-A*01-positive patients, even during chronic HCV infection, and reproducible immune pressure results in the selection of a CD8 escape variant that seemingly impairs HLA binding in all HLA-A*01-positive patients. However, mutational escape in this epitope is potentially associated with viral fitness costs at least in some isolates, and evolution toward the prototype residue in the absence of the restricting HLA allele occurs, albeit seemingly at a low rate. Interestingly, in this interplay of positive and negative selection pressures the frequency of the HLA-A*01-allele in a population seems to be an additional important factor that influences the overall frequency of the escape mutation in circulating isolates. In HLA-A*01-rich populations the majority of HCV isolates is adapted to immune pressure in this epitope, and further compensatory evolution may have tipped the balance for the predominant consensus residue from prototype to the escape variant. Adaptation of circulating viruses to HLA class I-associated selection pressure in a population may further complicate the design of T-cell-based immune therapies.

ACKNOWLEDGMENTS

C.-N.H. and J.T. are supported by the German Hepatitis Network (Hep-Net). J.T. is supported by the DFG (TI 323/3-1). D.N.F. is supported by the National Institutes of Health (AI052395). P.K. is supported by the Wellcome Trust and the James Martin School of the 21st Century, Oxford, England. We thank the German Hepatitis Network (Kompetenznetz Hepatit) for the contribution of samples.

REFERENCES


