Long-term evolution of the 5’UTR and a region of NS4 containing a CTL epitope of hepatitis C virus in two haemophilic patients

H. L. Devereux,1 D. Brown,2 G. M. Dusheiko,2 V. C. Emery3 and C. A. Lee1

The Haemophilia Centre and Haemostasis Unit, Haematology Department1, Academic Department of Medicine2 and Department of Virology3, Royal Free Hospital and School of Medicine, Pond Street, London NW3 2QG, UK

Haemophilic patients exposed to unsterilized clotting factor concentrates prior to 1985 have become infected with hepatitis C virus (HCV). We have studied the sequence evolution of the 5’UTR and a region of NS4 over 12 years in one human immunodeficiency virus (HIV) positive haemophilic patient and 14 years for one HIV negative haemophilic patient. One sample each year from the date of HCV infection to 1994 was analysed for genotype, virus load and nucleotide sequence of the two genetic loci. Both patients were infected with HCV genotype 1 throughout the study period. The virus load profiles were similar except that the profile for the HIV infected patient was displaced 4 years earlier relative to the other patient. Mean divergence of the quasispecies at both the 5’UTR and NS4 loci was higher in the HIV coinfected patient. Phylogenetic analysis indicated that evolution of the 5’UTR was host independent, whereas the NS4 region containing a CD8 restricted CTL epitope evolved in a host specific fashion.

Introduction

Hepatitis C virus (HCV) was first identified in 1989 (Choo et al., 1989) and subsequently extensive investigations have revealed the clinical importance of HCV infection (Esteban et al., 1991; Colombo et al., 1991; Di Bisceglie et al., 1991). The virus exists as a quasispecies in the human host and up to 11 major genotypes have been characterized to date (Simmonds et al., 1993; Tokita et al., 1995, 1996). Epidemiological evidence suggests that genotypes 1, 2 and 3 are predominant in the indigenous United Kingdom and United States populations (McOmish et al., 1994).

Haemophilic patients have a particularly high sero-prevalence of HCV due to the use of unsterilized multi-donor pool clotting factor concentrates prior to 1985 (Makris et al., 1990). In contrast to the extensive investigations carried out on human immunodeficiency virus (HIV) (Ait-Khaled et al., 1995; Delassus et al., 1991; Wolfs et al., 1991; Holmes et al., 1992) there is a paucity of longitudinal HCV sequence data on frequently sampled patients over long periods of time (> 10 years). In order to redress this issue, we have assessed the HCV sequence evolution in one HIV positive haemophilic patient and in one HIV negative haemophilic patient over 12 and 14 years, respectively. HCV quantification and sequence analysis of the highly conserved 5’UTR and a CD8 restricted CTL epitope in NS4 were carried out on one sample each year from the date of infection to 1994. Both patients had mild haemophilia and, therefore, have received very little unsterilized clotting factor concentrate. At our centre, comprehensive clinical records for each patient have been kept since registration and, from 1979, serum samples were taken for storage at −40 °C. These protocols have facilitated the study of longitudinal evolution of the HCV quasispecies and the influence of HIV coinfection over prolonged periods of time. We reasoned that the results from this study would provide insight into the pathogenesis of HCV and allow the relative effects of declining T-helper cell functions on HCV replication and diversity to be assessed.

Methods

Patient details. Patient 1 was anti-HIV positive. During 1982 he received two batches of commercial factor VIII concentrate (1·66 l of one batch and 20 ml of the other batch) manufactured by Immuno in the United States. It is highly likely that he became infected with HCV after administration of this concentrate as his aspartate aminotransferase (AST)
The DNA from five clones of each sample was then purified using the TA Cloning Kit (Invitrogen) into the modified plasmid pCR II.

The PCR products were cloned using the Sequenase version 2.0 sequencing kit (United States Biochemical).

The Oligonucleotides used to amplify the NS4 region are listed in Table 1.

### Table 1. Oligonucleotides used to amplify the NS4 region

<table>
<thead>
<tr>
<th>First round primers (5′–3′)</th>
<th>Strand</th>
<th>Nucleotide position in genome</th>
<th>Second round primers (5′–3′)</th>
<th>Strand</th>
<th>Nucleotide position in HCV genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGTGGCGACCCCTTGAGGCCCTCTGGGCGAAGCAGCATGT*</td>
<td>+</td>
<td>5258–5297</td>
<td>GGAATTCTACGCGAGGTACA</td>
<td>+</td>
<td>5298–5319</td>
</tr>
<tr>
<td>TTAACCGGGCTCTGGTGAGGGACACATCCTGACTGGA</td>
<td>−</td>
<td>5604–5643</td>
<td>CAGAATGTCACAAAGCCTT</td>
<td>−</td>
<td>5537–5557</td>
</tr>
<tr>
<td>GCCCTCTGGGGAAGCATATGT</td>
<td>+</td>
<td>5276–5297</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGTAGGACCAGCTCTCCTGGGACACATCCTGACTGGA</td>
<td>−</td>
<td>5604–5643</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACATTCTGGGAAAAACACATGT</td>
<td>+</td>
<td>5255–5276</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAGGGGACCTCAGCGGCTCATGGA</td>
<td>−</td>
<td>5604–5625</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAAATTGGGCAAACACATG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAGGGGACCCTGACACTGAGA</td>
<td>−</td>
<td>560b–5625</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* First round primers which amplified the NS4 region from the majority of samples. The other first round primers were used in various combinations to amplify the remaining samples. The nucleotide positions are quoted on the basis of the HCV sequence published by Choo et al. (1991).

#### Factor VIII concentrate analysis

Sequence analysis of the concentrates used the same methods as described above for serum, except that ultracentrifugation was required prior to the extraction of RNA. Commercial concentrate (4 ml) and NHS concentrate (9 ml) was centrifuged at 20,000 r.p.m. for 3 h at 4 °C in an SW41 Beckman rotor. The supernatant was removed and the pellet was resuspended in 100 μl of guanidinium thiocyanate and the extraction continued as described above.

#### Computer analysis of the HCV sequences

The DNA sequences were analysed using the Sequence Analysis Software Package (version 7) by Genetic Computers Inc. from the University of Wisconsin’s Department of Genetics (GGC) (Devereux et al., 1984). Multiple sequence alignments were produced using the CLUSTAL software and alignments produced as PHYLP-lib-compatible files (Felsenstein, 1989). The following alignment parameters were used: pairwise parameters with a gap penalty of 3, K-tuple 1, a window size of 5 and multiple sequence parameters with a gap penalty of 10 and weighted transitions.

Maximum parsimony trees were generated using the DNAPARS program, available within the PHYLP-lib 3.5 package. A randomized input order of sequences was used along with ordinary parsimony methods. In order to test the robustness of the trees, bootstraps data for each multiple alignment were produced in SEQBOOT using 100 replicates. The consensus tree was generated using CONSENSE and unrooted trees were displayed using DRAWTREE.

Genetic distances were computed via the DISTANCES program within the GCG suite of programs using the Kimura 2-parameter algorithm to yield an estimate of the number of substitutions per 100 bases. The mean genetic distance was calculated for the entire quasispecies present during follow-up and at individual time-points for each patient. Comparison of means was performed using t-tests and differences are quoted together with their 95% confidence intervals (CI).

### Results

Patients 1 and 2 have both been infected with HCV for similar lengths of time (12 and 14 years, respectively) and were matched for age and factor VIII usage. However, the clinical course of HCV infection and disease in these two haemophilic patients was markedly different. Whilst the HIV coinfected...
patient suffered continuous hepatic abnormalities, the HIV negative patient remained relatively healthy following seroconversion to HCV.

The longitudinal variations in HCV RNA load, AST levels and CD4 cell counts in the two patients investigated are shown in Figs 1 and 2. During the period of follow-up both patients have exhibited intermittently elevated AST levels with the highest AST level at seroconversion. However, the spiking AST pattern which characterized patient 1, and frequently involved substantial increases, was not apparent in patient 2. This may reflect the increased frequency of AST measurement in patient 1, although it is interesting to note that both patients possessed comparable AST levels at 12 years post-infection. There was no clear relationship between HCV RNA loads and fluctuations in AST levels over time for either patient. At seroconversion patient 1 had a virus load of $3.3 \times 10^6$
Fig. 3. Unrooted maximum parsimony phylogenetic tree for 125 NS4 nucleotide sequences derived for patient 1, patient 2 and concentrate. Sequences for each patient are shown by the year from which they were derived. Bootstrap values are quoted at the major nodes. Ca and Cb, commercial concentrate (1 and 2, respectively).

genomes/ml which reduced to \(4.4 \times 10^4\) at year two and subsequently increased over the following 9 years to a maximum load of \(2.2 \times 10^7\) genomes/ml at year 11, remaining above \(10^7\) genomes/ml at year 12. The virus load profile observed in patient 2 paralleled the profile of patient 1 except that the curve was temporally displaced by 4 years. Thus, a virus load of \(10^7\) genomes/ml was reached by patient 1 at 9-25 years post-seroconversion whereas patient 2 reached the same virus load at 13-25 years. However, there was no statistically significant difference in the cumulative virus load between either patient.

RFLP analysis showed that both patients were infected with HCV genotype 1 throughout the period of follow-up. Only the concentrate received by patient 1 was available for analysis. Both batches of commercial concentrate contained HCV genotype 1 whilst only one of the four NHS concentrates contained HCV RNA, which was genotype 3. The concentration of HCV in one of the commercial concentrates and in the NHS concentrate was less than \(10^4\) genomes/ml (HCV RNA PCR positive but undetectable using the Amplicor quantification kit) and in the second commercial concentrate was \(1 \times 10^4\) genomes/ml.

In order to investigate the evolution of HCV strains in vivo we constructed unrooted maximum parsimony trees of the 5’UTR and NS4 nucleotide sequences for patients 1 and 2, respectively. The 5’UTR sequences formed a relatively homogeneous collection of variants, irrespective of source, with most branches associated with a low bootstrap value (data not shown). In contrast, the NS4 sequences segregated into three clades, one for each patient and one for the concentrate, all of which were associated with high bootstrap values (Fig. 3). Although there was no identical genetic sequence between concentrate and the earliest samples for patient 1 at either the 5’UTR or NS4 loci, a maximum parsimony tree for NS4 revealed the presence of variants with identical amino acid sequences in the concentrate and the earliest sample obtained from patient 1 following HCV seroconversion in 1982 (data not shown).

To determine whether the clustering shown in Fig. 3 was due specifically to amino acid alterations within the CD8\(^+\) CTL epitope, the frequency of amino acid substitution in this epitope and the flanking regions was determined. Although amino acid variation occurred in the CTL epitope at various stages over the course of infection in both patients, no individual CTL variant persisted in the population subsequent to its appearance (Fig. 4). In addition, at the nucleotide level, there was no evidence for increased sequence variation in the epitope region relative to the flanking sequences.

The temporal fluctuations in the mean divergence of the NS4 region post-infection are shown in Fig. 5 and aptly illustrate the oscillatory nature of divergence throughout the course of infection of HCV in the human host. In patient 2 a gradual homogenization of the quasispecies occurred during years 4 to 9 and was followed by rapid fluctuations whilst patient 1 exhibited a more uniform oscillatory pattern throughout infection. In patient 1, the divergence of the
Fig. 4. Amino acid substitutions in the 44 amino acid region of NS4 amplified by RT–PCR, including a 10 amino acid CD8 restricted CTL epitope (underlined). N/A, not applicable as the sample was PCR negative; ND, not determined as the sample was not available.
Discussion

There has been a lack of data examining the longitudinal sequence evolution of HCV in the human host over long periods of time (in excess of 10 years) with frequent sampling. The present study has analysed two patients with similar factor VIII usage and duration of HCV infection to show that the mean divergence of the HCV quasispecies oscillates during the course of infection, the evolution of NS4 occurs in a patient specific manner, the effect of HIV co-infection appears to result in a temporal shift of the virus load–time curve and the generation of a more divergent quasispecies.

Multiple infection with different HCV genotypes has been documented in haemophiliac patients (Devereux et al., 1995; Jarvis et al., 1994) but whether these reflect simultaneous coinfection or sequential reinfections is not known. Patient 1 received two batches of concentrate infected with genotype 1 and one infected with genotype 3. However, he only became infected due to treatment with one of the batches infected with genotype 1. He received 1·66 l of this batch corresponding to a total exposure to approximately 16·6 million HCV genomes. Quantification of the remaining batches revealed that he was exposed to less than 2000 genomes from the other genotype 1 concentrate and approximately 288,000 genomes from the genotype 3 concentrate. This suggests that the concentrate with the highest virus load initiated infection whereas the inocula with lower virus loads were unable to initiate a superinfection. However, we cannot exclude the possibility that there was a protective effect of genotype 1 infection against genotype 3 superinfection.

The results of the phylogenetic analysis of the 5′UTR were consistent with data showing a low level of divergence within genotypes (Simmonds et al., 1993). Thus, the genotype 1 5′UTR nucleotide sequences from both the patients and the concentrate intermingled within the phylogenetic tree whilst the genotype 3 5′UTR sequences formed a unique clade. In contrast, phylogenetic analysis of the nucleotide sequences present in NS4 showed a patient specific clustering with the concentrates also forming a separate clade. This patient specific clustering could not be attributed to the evolution of CTL escape mutants since the CD8 epitope amino acid sequences were fairly well conserved over time in both patients. Since HCV strains can differ by up to 30% in overall amino acid identity (Simmonds et al., 1993), it suggests that this epitope may be relatively resistant to immune pressure. Weiner et al. (1995) have documented the emergence of a single amino acid CTL escape mutant in an NS3 epitope by 28 weeks post-infection. Such a rapid appearance of an escape mutant was not observed in the NS4 sequences of the two patients analysed here and the likely infecting strain (at the amino acid level) in patient 1 has persisted throughout the entire period of infection.

There was no significant difference in the cumulative HCV load present in either patient although in the HIV infected patient it increased faster than in the HIV negative individual. Notwithstanding this observation, the overall divergence of the HCV quasispecies in patient 1 was significantly greater at both the 5′UTR and NS4 loci when compared with patient 2. As expected, both loci were relatively divergent within the concentrates presumably due to the large number of donors that make up each batch. The oscillatory pattern of divergence in both patients over time presumably reflects host–parasite dynamics resulting in the selection of strains that are most fit. Whether the fitness resides in the NS4 region or another region of the genome in linkage disequilibrium requires further study. However, it has been noted that there is a high degree of conservation in the 3′ half of NS4 in sequential analysis of HCV infected chimpanzees (Okamoto et al., 1992).

In contrast to the frequent sampling over > 10 years performed in the present study, the majority of previous studies in patients have assessed HCV sequence evolution over 1 to 3 years (Martell et al., 1994; Kurosaki et al., 1993, 1994; Enomoto et al., 1994; Nakazawa et al., 1994; Higashi et al., 1993; Kumar et al., 1993; Weiner et al., 1992) or have only studied two time-points separated by a greater number of years (Ogata et al., 1991; Hohne et al., 1994). Four studies have assessed HCV sequence changes in chimpanzees by sampling
at time 0 and one 9 years later without intervening samples (Abe et al., 1992) or by analysis of between three and five samples over 9 to 10 years (van Doorn et al., 1994, 1995; Okamoto et al., 1992). These studies have predominantly involved analyses of the hypervariable region in E2/NS1 (HVR1) and have provided evidence for the immune selection of HVR1 variants. Although this study has been performed on only two patients, it has given an insight into the long-term evolution of the hepatitis C virus in vivo. Somewhat similar to HIV (Ho et al., 1995; Wei et al., 1995) HCV appears to be a dynamic host-dependent infection. Hence, it will be important to assess many different patients using sequence analysis of different genomic regions over long periods of time with frequent sampling before correlations between the pathogenic potential of different genotypes, influence of host factors and their pathogenic potential and the effects of HIV co-infection on HCV replication can be elucidated.

We wish to thank Dr. Sebastian Bonhoeffer at the Department of Zoology, University of Oxford for helpful discussions.

References


Received 27 June 1996; Accepted 15 November 1996