Use of NS-4 peptides to identify type-specific antibody to hepatitis C virus genotypes 1, 2, 3, 4, 5 and 6


1 Edinburgh and South East Scotland Blood Transfusion Service, Royal Infirmary of Edinburgh, Lauriston Place, Edinburgh EH3 9HB, UK, 2 Murex Diagnostics Ltd, Central Road, Temple Hill, Dartford DA1 5LR, Kent, UK, 3 Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK, 4 Unit of Hepatology, Aker University of Hospital, 0514 Oslo 5, Norway, 5 Cairo Liver Centre, 5 El-Girgawy Street, Dokki, Cairo, Egypt, 6 University of Witwatersrand, Johannesburg 2193, South Africa, 7 The Natal Institute of Immunology, Robin Road, 3610, South Africa, 8 Hong Kong Red Cross, Kowloon, Hong Kong, 9 Institute of Virology, Church Street, Glasgow G11 5JR, UK and 10 Riyadh Armed Forces Hospital, Riyadh 11159, Saudi Arabia

The 5' end of the NS-4 protein of different genotypes of hepatitis C virus (HCV) is highly variable in nucleotide and inferred amino acid sequence, with frequent predicted amino acid substitutions between all six of the major HCV genotypes described to date. This region has been shown to be antigenic by epitope mapping, and elicits antibody in HCV-infected individuals with a detectable type-specific component. We have used this sequence data to specify branched peptides for an indirect binding/competition assay to detect type-specific antibody to each major genotype. A total of 183 out of 210 samples (87%) from blood donors and patients with chronic hepatitis C infected with genotypes 1 to 6 showed detectable type-specific antibody to NS-4 peptides that in almost all cases (> 97%) corresponded to the genotype detected by a PCR typing method. These findings demonstrate the existence of major antigenic differences between genotypes of HCV, and indicate how infection with different variants of HCV may be detected by a serological test.

Introduction

Hepatitis C virus (HCV) is the major causative agent of post-transfusion non-A, non-B hepatitis. The viral genome is a single-stranded, positive-sense RNA of about 9400 nucleotides in length (Choo et al., 1991), encoding a single, continuous polyprotein that is cleaved into putative core, E1, E2/NS-1 (envelope), NS-2, NS-3, NS-4 and NS-5 (non-structural) proteins. Comparisons of published complete genome sequences have found that isolates fall into distinct virus types (Choo et al., 1991; Kato et al., 1990; Okamoto et al., 1991; Takamizawa et al., 1991; Okamoto et al., 1992b; Sakamoto et al., 1994). According to recent proposals (Simmonds et al., 1994a), based upon extensive sequence comparisons of HCV in the core (Bukh et al., 1994), E1 (Bukh et al., 1993) and NS-5 regions (Simmonds et al., 1993a), variants of HCV can be usefully classified into six major genotypes (designated 1 to 6), many of which can be further subdivided into a number of more closely related subtypes.

The variation in nucleotide sequences between HCV isolates is expressed in the translated amino acid sequences and may alter the antigenic properties of viral proteins. For example, sera from individuals infected with non-type 1 genotypes react infrequently with recombinant proteins derived from NS-4 (5-1-1 and c100-3) used in current screening and confirmatory assays for HCV (McOmish et al., 1993; Chan et al., 1991); there is also evidence for less frequent reactivity to the NS-3 protein (c33c) amongst those infected with types 2 to 6, leading to a statistically significant over-representation of 'indeterminate' results from these variants in confirmatory assays such as the 2nd generation recombinant immunoblot assay (RIBA-2) (McOmish et al., 1994).

At present PCR-based methods are the principal means by which variants of HCV may be identified (Okamoto et al., 1992c; Stuyver et al., 1993a; Chayama et al., 1993; McOmish et al., 1994). The major drawbacks of PCR-based methods, however, are the expense and complexity of the procedure which makes difficult the processing of large numbers of clinical samples on a routine basis. Serological methods for HCV typing, based upon the ability to distinguish antibody elicited by
infection by different genotypes of HCV are necessarily
indirect, and not always appropriate (as they rely upon a
humoral immune response in the infected individual),
but do have important practical advantages over PCR-
based methods in ease of use and cost. The problem with
developing such tests lies in finding protein sequences
which are both type-specific and antigenic.

Epitope mapping of the core and NS-4 proteins has
revealed the existence of linear antigenic determinants in
regions of variability between HCV genotypes
(Tsukiyama Kohara et al., 1993; Simmonds et al., 1993c).
ELISAs based on peptides corresponding to
these regions have been used to identify type-specific
antibody elicited by genotypes 1 and 2 (core or NS-4;
Tsukiyama Kohara et al., 1993; Tanaka et al., 1994;
Mondelli et al., 1994), and 1, 2 and 3 (NS-4; Simmonds
et al., 1993c). In the latter assay, cross-reactivity due to
epitopes shared between genotypes was minimized by co-
incubation of serum with peptides in solution hetero-
logous to those on the solid phase.

In this paper, we have obtained nucleotide sequences
of the antigenic region of NS-4 for the more recently
described genotypes of HCV (types 1c, 2c, 3b, 4a, 5a
and 6a) and have developed a competition ELISA to detect
type-specific antibody to each of the six genotypes. The
frequency with which type-specific antibody to NS-4
could be detected was measured by testing sera from
individuals infected with a range of different genotypes.

Methods

Samples. HCV sequences from the 5' non-coding region (5'NCR)
were amplified by PCR from plasma or serum samples from anti-HCV
positive blood donors from Egypt, South Africa and Hong Kong.
Genotypes were identified by sequence comparisons with previously
published 5'NCR sequences for identification of genotypes (see Fig. 1).
Sequences were obtained in the NS-4a region from those infected with
genotypes 4, 5 and 6, as well as from samples from individuals
previously shown to be infected with type 1c and 2c (Simmonds et al.,
1993).

Investigation of the frequency of type-specific antibody to NS-4
peptides was carried out using a combination of anti-HCV positive
blood donors (n = 39) and patients with HCV-induced liver disease
(n = 171) infected with genotypes 1 to 6 as identified by RFLP of the
5'NCR as previously described (McOmish et al., 1994; Davidson et al.,
1995).

PCR amplification of NS-4 sequences. RNA was reverse transcribed
using an antisense primer with the sequence AACTCGAGTCATGCAGCTGATCTGG
(5' base: 4941) and antisense primer TTTGGATCCATGCATGTCAGCTGATCTGG (5'
base: 5278). Both primers used in the second PCR step included
terminal BamHI restriction sites (in bold) to allow cloning via cohesive
ends. The NS-4 sequence of genotype 6 was found to amplify better
using a combination of primers that matched genotypes 1, 2 and 3
(Simmonds et al., 1993c) for the first and second rounds of PCR
followed by a third round employing primers 5351 and 594 to enable
cohesive-end cloning.

Cloning of amplified NS-4 sequences. All cloning steps followed
standard protocols (Ausubel et al., 1992) using Boehringer Mannheim
enzymes except where stated. The amplified NS-4 DNA bands
(50 kb) were excised from 1% agarose gels and the DNA was
recovered by low-speed centrifugation through glass wool (Heery et al.,
1990). Extracted DNA was cleaved with BamHI and ligated into
BamHI-digested pUC18 vector (Yanisch-Perron et al., 1985) at 16°C
overnight, and used to transform competent Escherichia coli cells (XL1-
Blue from Stratagene). Plasmid DNA from colonies containing inserts
was sequenced with forward and reverse pUC/M13 primers using the
United States Biochemical Sequenase kit. The manufacturer's protocol
for double-stranded sequencing with alkaline denaturation was followed
except that the reactions were carried out in 10% dimethyl
sulphoxide.

Sequence analysis of NS-4. Phylogenetic analysis of NS-4 sequences
was carried out with the programs DNADIST, NEIGHBOR and
DRAWTREE in the PHYLP package version 3.5c kindly provided by
Dr J. Felsenstein (Felsenstein, 1993). Unrooted trees were
constructed by the neighbour-joining algorithm as previously described
for analysis of NS-5 sequences (Simmonds et al., 1993a).

Synthesis of peptides for HCV serotyping assay. Synthetic peptides
were prepared on p-hydroxymethylphenoxymethylpolystyrene (HMP)
resin bearing a branched-lysine core to give eight copies of the same
sequence per molecule, using the multiple antigenic peptide (MAP)
system (Fmoc MAP 8-Branch; ABI Ltd; Tam, 1988). Synthesis was
performed using Fastmoc chemistry on an Applied Biosystems model
432A automated peptide synthesizer (ABI Ltd). Purity of the peptides
was determined by reverse-phase HPLC. Integration of peaks by
computer analysis (PE Nelson, Perkin Elmer) showed all peptides to be
greater than 80% pure. Peptides were not purified further prior to use
in the assay.

Serotyping of serum samples using the 1–6 ELISA. Each peptide was
reconstituted in distilled water at an approximate concentration of
1 mg/ml. A small volume of 30% acetic acid was added to the type 3
peptides to help them dissolve, while the remaining peptides dissolved
more easily after bubbling ammonia vapour through the solution.

Polypropylene microtitre wells (Immunoplate, GIBCO BRL) were
coated with 100 µl volumes of phosphate buffered saline (PBS)
containing all 21 peptides, each at a concentration of 50 ng/ml. The
plates were incubated at 4°C overnight, washed in PBS and blocked
with 125 µl volumes of blocking solution (PBS, 2% bovine serum
albumin) for a minimum of 1 h at room temperature. After blocking,
plates were air-dried and kept at 4°C prior to use. Plates prepared in
this way could be stored for several months without loss of reactivity.

Prior to the addition of the serum to the wells, blocking solutions of
competing peptides (in a 100-fold excess over those used for coating)
were added to the appropriate wells on the microtitre plate in volumes
of 10 µl per well. Plasma, diluted in 800 µl of sample diluent (Murex
Diagnostics) at a concentration of 1:20, was immediately added to each
of the eight wells used per sample (100 µl per well). This was followed
by incubation overnight at 4°C.

Plates were washed four times in PBS+0.05% Tween 20 (PBST),
and then incubated with 100 µl of conjugate (Murex; VK 47-
horseradish peroxidase-labelled mouse monoclonal antibody to human IgG) for 1 h at 37 °C. The plates were washed again, four times in PBST, before adding 100 μl substrate to each well (TMB, Murex) and incubating for 30 min at 37 °C. Sulphuric acid (8 μl, 50 μl) was added to each well before reading the plate against a negative control blank at 450 nm, with a reference wavelength of 690 nm.

Statistical analysis. Analysis of the median and variance of the absorbance values recorded for plasma samples was carried out using standard statistical software (SYSTAT). Box plots represent the 25th and 75th quartiles of a non-parametric distribution as the boundaries of the box, and the median as the line dividing the box (Tukey, 1977). The whiskers represent the range of values that lie within 1-5 times the interquartile range (inner fence). Values between 1.5 and 3 times the interquartile range are plotted as asterisks (*), while open circles (O) represent outlying values greater than three times the interquartile range (outer fence). For calculation of median values and ranges of serological data, a value of 24 was substituted for samples with absorbance values greater than 2.

Results

Phylogenetic analysis of the NS-4 region

HCV types 4, 5 and 6 were identified by sequence comparison of the 5'NCR with those of published variants (Fig. 1). In order to obtain cloned NS-4 nucleotide sequences from these genotypes, RNA was reverse transcribed and amplified using conserved primers from the NS-4 region, as described in Methods. Sequences were obtained from six type 4a samples (EG-1, -13, -21, -24, -25 and -33), four type 5a (T478, SC-6, SC-23 and SC-24) and two type 6a (HK-4 and T3950). NS-4 sequences were also amplified from previously identified samples of genotypes lc (4TY4), 2c (T983), 3b (B4, B9) identified by sequence comparisons in NS-5 (Simmonds et al., 1993a) and unpublished data).

The NS-4 region used in the phylogenetic analysis covers the nucleotide sequence positions 4941-5282 (numbered as in Choo et al., 1991) and is 341 bases in size. Previously published nucleotide sequences, and the corresponding regions of complete genomes (numbered 1–33; Fig. 2) were included in the comparison with the newly isolated NS-4 sequences (nos 34–50; Fig. 2). The unrooted phylogenetic tree displays three levels of branching which reflects different levels of variability between isolates. The major branches correspond to the six major genotypes of HCV based on analysis of NS-5 (Simmonds et al., 1993a) and E1 sequences (Bukh et al., 1993). The second level of branching occurs between variants of HCV previously described as subtypes (Simmonds et al., 1993a). As found in NS-5 and E1, NS-4a sequences show discontinuous, ‘hierarchical’, as opposed to continuous, variation between isolates. The grouping of the individual isolates was in all cases consistent with genotype classifications obtained using the 5’NCR (Fig. 1), and E1 and NS-5 (Simmonds et al., 1994b).

Analysis of the distribution of pairwise distances

The relationship between sequence diversity between major genotypes and subtypes was investigated by comparison of the distributions of evolutionary distances of separate pairwise comparisons between major genotypes, subtypes, and variants within a subtype (isolates) (Table 1a). The median nucleotide distance of comparisons between members of different major genotypes was 0.381, greater than those calculated for subtypes and isolates (0.168 and 0.037). Although the range of values of type, subtype and isolate overlapped, the 95th percentile ranges did not. The range of nucleotide distances corresponded closely to three discrete normal distributions, as previously described for NS-5 (Simmonds et al., 1993a), as shown by the similarity of mean and median values, and between the actual 95th percentile ranges with those calculated from the mean ±2 standard deviations (Table 1a).

The analysis was extended to the distribution of pairwise distances calculated between inferred amino acid sequences (expressed as the proportion of different residues). Such comparisons are more likely to reflect the degree of antigenic relatedness between different HCV variants than evolutionary distances between nucleotide sequences. A similar set of distinct approximately normally distributed values was observed for type, subtype and isolates comparisons (Table 1b), each showing numerically distinct median values and non-overlapping 95th percentile ranges. The amino acid sequence differences between major genotypes (38%; range 25–50%) are consistent with antigenic differences greater than those found between subtypes and isolates (median values 17% and 4%), although this analysis cannot compute the differing contributions to changes in antigenicity produced by biochemically similar or dissimilar amino acid substitutions (see below).

Description of NS-4 amino acid sequences for HCV types 4 to 6

The inferred amino acid sequence of 111 residues translated from the NS-4a sequences between positions 1653–1764 (Choo et al., 1991) includes the region corresponding to the immunoreactive clone 5-1-1 (1683–1735) (Fig. 3). In the region compared, only 47 sites are conserved between all variants analysed, while many of the variable sites contain substitutions of amino acids possessing side-chains with different biochemical properties, and which could potentially alter the structure and antigenic properties of the protein (highlighted in bold). Previous epitope mapping of this region for genotypes 1a, 1b, 2b, and 3a (Simmonds et al., 1993c) identified regions of greatest antigenicity at positions 1691–1708...
Fig. 1. Comparison of sequences in the 5'NCR of different genotypes of HCV from variants used for sequence determination in NS-4 (listed in the legend to Fig. 2); differences from the HCV prototype strain HCV-PT (Choo et al., 1991) are indicated. The genotype assigned to each sequence, following proposed nomenclature (Simmonds et al., 1994a) on the basis of the sequence in the 5'NCR is shown in the first column. Nucleotide sequences -232 to -225, -214 to -168 and -153 to 145 are omitted as they are invariant between the sequences shown. (.), Sequence identity to HCV-PT; (-), gap introduced into the sequence to preserve alignment; (N), sequence not determined.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>TGGAGTGCGGTGCCTGGACGACCTGGCAATTCGGGACGAGCCCTGGAGATTTGGGCGTGCCTGGCACTGCTAGCCAGGGTCGCG</td>
</tr>
<tr>
<td>1b</td>
<td>J</td>
</tr>
<tr>
<td>1c</td>
<td>HCVUK</td>
</tr>
<tr>
<td>1d</td>
<td>HC-G9</td>
</tr>
<tr>
<td>2a</td>
<td>T6</td>
</tr>
<tr>
<td>2b</td>
<td>J8</td>
</tr>
<tr>
<td>2c</td>
<td>T983</td>
</tr>
<tr>
<td>3a</td>
<td>T1787</td>
</tr>
<tr>
<td>3b</td>
<td>B4</td>
</tr>
<tr>
<td>4a</td>
<td>EG-1</td>
</tr>
<tr>
<td>4b</td>
<td>EG-13</td>
</tr>
<tr>
<td>4c</td>
<td>EG-21</td>
</tr>
<tr>
<td>4d</td>
<td>EG-24</td>
</tr>
<tr>
<td>4e</td>
<td>EG-25</td>
</tr>
<tr>
<td>4f</td>
<td>EG-33</td>
</tr>
<tr>
<td>5a</td>
<td>T478</td>
</tr>
<tr>
<td>5b</td>
<td>SC6</td>
</tr>
<tr>
<td>5c</td>
<td>SC23</td>
</tr>
<tr>
<td>5d</td>
<td>SC24</td>
</tr>
<tr>
<td>6a</td>
<td>HK-4</td>
</tr>
</tbody>
</table>

Genotypes 3a, 3b, 4a, 5a, and 6a show similar amino acid sequences in antigenic region 1, although the type 4a sequences contain a possible antigenically significant change to a glutamine residue at position 1691 in place of basic amino acids (Fig. 3). Type 3b has a glutamine in place of an otherwise universal glutamate at position...
Fig. 2. Phylogenetic analysis of a 342 base region of NS-4 (nucleotide positions 4941–5282) of sequences from HCV genotypes 1 to 6. 1–33, previously published sequences; 34–50, new NS-4 sequences obtained in this study corresponding to more recently described genotypes. Phylogenetic groups are labelled according to proposed nomenclature; sequence 13 (labelled 1d) from Indonesia is described as ‘1c’ in Okamoto et al. (1994) and ‘1d’ in Hotta et al. (1994). Type 4 sequences from Egyptian blood donors correspond to variants described as ‘4a’ in Simmonds et al. (1993a), but are different from ‘4a’ in Bukh et al. (1993). Sources of published NS-4A sequences (number, clone name, geographical origin, genotype, source reference) are as follows: (1) PT-1, USA, 1a (Choo et al., 1991); (2) HCV-H, USA, 1a (Inchauspe et al., 1991); (3–7) T16, T42, T77, T1801, T1825, Scotland, UK, 1a (Simmonds et al., 1993c); (8) HCV-J, Japan, 1b (Kato et al., 1990); (9) HCV-BK, Japan, 1b (Takamizawa et al., 1991); (10) HPCTJA, Japan, 1b (Tanaka et al., 1992); (11, 12) 2483, 2491, Japan, 1b (Okamoto et al., 1992a); (13) HCVJK1G, 1b (M. Honda and others, GenBank number X61596); (14) T, Taiwan, 1b (Chen et al., 1992); (15) HPCGENOM, China, 1b (X. Bi and others, GenBank number L02836); (16) HC-J9, Indonesia, 1d (Okamoto et al., 1994); (17) HC-J6, Japan, 2a (Okamoto et al., 1991); (18) K2A, Japan, 2a (Enomoto et al., 1990); (19) D11353, Japan, 2a (Sarashina et al., 1993); (20) T351, Scotland, UK, 2a (Simmonds et al., 1993c); (21) Clone A, Clone E, Japan, 2b (Tsukiyama et al., 1991); (22) HC-J8, Japan, 2b (Okamoto et al., 1992b); (24–26) T59 (E-b12), T810, T940, Scotland, UK, 2b (Simmonds et al., 1993c); (27, 28) BR36-20, HD10-1, Brazil, 3a (Stuyver et al., 1993b); (29–33) T26 (E-b3), T32 (E-b7), T38 (E-b2), T40 (E-b1), T787, Scotland, UK, 3a (Simmonds et al., 1993c). Sources of the new NS-4A sequences are as follows: (34) HCVUK, UK, 1b (Garson et al., 1990); (35) 4TY4, Lebanon, 1c (Simmonds et al., 1993a); (36) T893, Scotland, UK, 2c (Simmonds et al., 1993a); (37) B4, Bangladesh, 3b; (38) B9, Bangladesh, 3b; (39–44) EG-1, EG-13, EG-21, EG-24, EG-25, EG-33, 1699 and, along with 3a, a tyrosine residue instead of phenylalanine at 1705. Region 2 displays a high degree of antigenically significant type-specific variation between all the genotypes. The type 5a sequences contain a glutamate-alanine-arginine motif (EAR) at position 1720–1722, while type 6 has a specific motif alanine-glutamate-residue x-glutamine-glutamate (AE-QQ) at 1719–1723. Genotype 4a contains four amino acid changes of biochemical significance in this region, i.e. lysine at position 1713, leucine at 1721, histidine at 1720 and glutamine at 1724. A lesser degree of sequence variability is found amongst sub-types of HCV. The single examples of types 1c and 1d possess specific histidine residues at positions 1720 and 1723 respectively, while type 2c has threonine, histidine and arginine residues at positions 1715, 1722 and 1724. Sub-type 3b contains specific amino acid residues at positions 1720 (proline) and 1726 (glutamine).

**Development of an ELISA to detect genotype-specific antibody**

Peptides corresponding to the two previously identified antigenic regions of NS-4 (underlined in Fig. 3) for genotypes 1 to 6 were synthesized in branched form and used as antigens for detection of genotype-specific antibody. Consensus sequences for regions 1 and 2 were decided upon for each genotype, although it was clear from the outset that the region 1 peptides for genotypes 4 to 6 were all very similar to 3a (Fig. 4). In some cases, it was necessary to synthesize more than one peptide for a particular genotype to cover intra-typic sequence heterogeneity, but not to introduce combinations of amino acid polymorphisms that did not exist in native sequences. For example, different consensus sequence peptides were synthesized for each of the three subtypes of type 2 (Fig. 4).

The format of the current assay differs considerably from those of previously described serotyping assays that were based upon the coating of different antigens corresponding to different genotypes into different wells (Tsukiyama Kohara et al., 1993; Yunomura et al., 1994; Tanaka et al., 1994; Mondelli et al., 1994). In the current assay, all wells were coated with peptides from all six genotypes. Type-specific antibody was detected by competition between the solid phase antigens with a 100-fold molar excess of peptides added in solution with the plasma or serum sample. For example, detection of genotype-specific antibody to type 1 is achieved by
addition of peptides from genotypes 2 to 6 to the test serum or plasma samples (see Methods). This method guarantees that only those antibodies that can bind to the solid phase are those that do not cross-react with heterologous genotypes to HCV.

The frequency of antibody reactivity to type-specific antibody to this region of NS-4 was assayed using a total of 210 PCR-positive plasma samples from blood donors or patients with chronic hepatitis, and whose circulating genotype was determined by RFLP as recently described (Davidson et al., 1995). To investigate the frequency of non-specific reactivity to the NS-4 peptides, we also tested a total of 20 anti-HCV negative blood donor control sera. For each sample, we recorded the absorbance produced upon reaction with all of the peptides coated onto the solid phase (unblocked control), in competition with all peptides in solution (blocked control), and finally a series of six wells for detection of type-specific antibody to each of the genotypes individually (Fig. 5).

The relative proportion of type-specific and cross-reactive antibody to the epitopes presented by the NS-4 peptides was estimated by comparison of the median and interquartile ranges of the absorbance values upon reaction with homologous NS-4 peptides with the unblocked control (Fig. 5). The absorbance value from the blocked control was in each case subtracted from the values obtained from the other wells to control for any non-specific binding to the solid phase. For all genotypes, antibody reactivity was almost always confined to peptides of the same genotype as determined by PCR-RFLP typing, whereas binding was not observed amongst any of the anti-HCV negative controls (Fig. 5A). For the anti-HCV positive sera, a wide range of antibody reactivity was observed, both against the homologous type peptide and in the unblocked control well. For each genotype, the median values were greater in the unblocked wells than they were to the homologous peptide, indicating a mixture of antibody to type-specific and cross-reactive epitopes. Precise measurement of the relative proportions could not be carried out without titration of the samples, because there was not a linear relationship between absorbance and amount of antibody for absorbance values greater than 1 (unpublished observations), and because absorbance values recorded for many of the samples were greater than 2.0 (see Methods).

For genotypes 4, 5 and 6, levels of type-specific antibody were markedly lower than for types 1 to 3, while total anti-NS-4 antibody (unblocked control) was similar to that found in the other genotypes. These findings could be explained by the greater similarity of the first antigenic region of genotypes 4, 5 and 6; it is possible that the principal component of the type-specific reactivity is directed solely to the more divergent second region. Little difference was found in total reactivity to NS-4 or to homologous peptides between subtypes 1a and 1b, or between 2a and 2b, apart from the slightly weaker reactivity of type 1b antibody to the type 1 peptides (Fig. 5C). As described above, this could be the consequence of a greater similarity of type 1b NS-4 sequences to other genotypes than type 1a. For example, the alanine (A), serine (S) and glutamine (Q) substitutions at positions 1711, 1712 and 1722 (all in region 2) found in the 1b sequences are shared with type 2 (Fig. 3), and antibody reactivity with epitopes dependent upon these amino acid residues would be more likely to be absorbed by heterologous peptides.

### Table 1. Distribution of pairwise nucleotide (a) and amino acid (b) sequence distances between HCV types, subtypes and isolates

<table>
<thead>
<tr>
<th></th>
<th>Type</th>
<th>Subtype</th>
<th>Isolate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. of pairwise comparisons</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(a) Nucleotide distances</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.617</td>
<td>0.275</td>
<td>0.083</td>
</tr>
<tr>
<td>Range</td>
<td>Min</td>
<td>0.359</td>
<td>0.180</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>0.888</td>
<td>0.380</td>
</tr>
<tr>
<td>95% percentile</td>
<td>Lower</td>
<td>0.449</td>
<td>0.228</td>
</tr>
<tr>
<td></td>
<td>Upper</td>
<td>0.815</td>
<td>0.364</td>
</tr>
<tr>
<td>Mean</td>
<td>0.617</td>
<td>0.279</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>±0.113</td>
<td>±0.030</td>
<td>±0.031</td>
</tr>
<tr>
<td><strong>(b) Amino acid divergence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.381</td>
<td>0.168</td>
<td>0.037</td>
</tr>
<tr>
<td>Range</td>
<td>Min</td>
<td>0.230</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>0.550</td>
<td>0.240</td>
</tr>
<tr>
<td>95% percentile</td>
<td>Lower</td>
<td>0.245</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>Upper</td>
<td>0.502</td>
<td>0.221</td>
</tr>
<tr>
<td>Mean</td>
<td>0.374</td>
<td>0.164</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>±0.071</td>
<td>±0.029</td>
<td>±0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
</tbody>
</table>
| a  | PT-1 | LEVVTSTSVLVGVVLAAAYCLSTOCVTVVGRVLSKPAIIPDRFVLRIFDSEMCCSQHPLYIPOGMMRALQKOKALGILLQTSRQAEVIAAPAVQTNQW
|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| b  | HCV-J | A... | T.S... | L... | V... | Q... | AS... | Q... |       |       |       |       |       |       |       |       |       |
|    | HCVUK | T.S... | T.S... | R... | V... | Q... |       |       |       |       |       |       |       |       |       |       |
| c* | 4TY4 | A... | I... | N... | VV... | Q... | AA... | A... | H... | Q... |       | K... | A.V.| V... | S... |       |       |
| d* | HC-G9 | I... | V... |       |       |       | AA... | I... | L... |       |       |       | T... | H... |       |       |       |       |
| 2a | HC-J6 | M... | A... | V... | A... | S.I... | L.HINQAVVA.K... | RA... | ASRAAAL.B.QRI.ML.S.IQ... | Q... | K.QD.Q... | AS.P.V.Q... |       |       |       |       |
|    | D11353 | M... | R.A... | V... | A... | S.I... | L.HINQAVVA.K... | RA... | ASKATL.E.QRI.ML.S.IQ... | Q... | K.QD.Q... | S.F.V.E... |       |       |       |       |
| 2b | HC-J8 | I.M... | A... | V... | A... | I.S.I... | L.NDRVVA.K... | I... | RA... | ASKAAL.E.QRM.ML.S.IQ... | Q.T... | Q.D.Q... | I.S... |       |       |       |
|    | T59 | I.M... | A... | V... | A... | I.S.I... | L.NDRVVVT.K... | I... | RA... | ASKAAL.E.QRM.ML.S.IQ... | Q.T... | Q.D.Q... | V.S... | Q... |       |       |
| 2c | T983 | I.M... | A... | S.I... | G.H... | NQTAVAA... | RA... | ASRTAL.E.HRR.ML.S.IQ... | M.Q... | K.QG... | Q.V... |       |       |       |       |       |       |
| 3a | T40 | T... | L... | V... | HIE.G... | V... | K... | QQY... | AA... | A.QV.I.H... | E.V... | R.TQ..QA... | E.I.A... |       |       |       |       |
|    | BR36_20 | T... | L... | V... | HIE.G... | V... | K... | QQY... | AA... | A.QV.I.H... | E.V... | R.TQ..QA... | E.I.T... | A... | H... |       |       |
| 3b | B4 | T... | RA... | V... | HIE.G... | V... |      | QQY... | AA... | A.QAV.Q... | P.II... | R... | Q... | BAD.R.I... | P.SH... | V.K... | BQ... |
|    | B9 | T... | DA... | V... | HIE.G... | V... |      | QQY... | AA... | P.AQV.Q... | D.V... | R... | H.EAD.E.I... | A.SH... | V.A... |       |       |
| 4a | EG-1 | R... | V... | S... | Q... | V... |       |       |       |       |       |       |       |       |       |       |       |       |
|    | EG-13 | S... | V... | S... | Q... | V... |       |       |       |       |       |       |       |       |       |       |       |       |
|    | EG-21 | R... | V... | S... | Q... | V... |       |       |       |       |       |       |       |       |       |       |       |       |
|    | EG-24 | F... | V... | T.V... | S... | G... | V... |       |       |       |       |       |       |       |       |       |       |       |
|    | EG-25 | R... | V... | S... | V... | Q... |       |       |       |       |       |       |       |       |       |       |       |       |
|    | EG-33 | R... | V... | S... | T.L... | Q... |       |       |       |       |       |       |       |       |       |       |       |       |
| 5a | T478 | I... | V... |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|    | SC6 | I... | V... | T.V... | S... | G... | V... |       |       |       |       |       |       |       |       |       |       |       |
|    | SC23 | I... | V... |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|    | SC24 | I... | V... |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 6a | HK-4 | I... | T... | V... | C... | I.T... | VV... | I... | QQ... |       |       |       |       |       |       |       |       |       |
|    | T3950 | I... | I... | V... | C... | I.T... | VV... | I... | QQ... |       |       |       |       |       |       |       |       |       |

Fig. 3. Comparison of deduced amino acid sequences from the NS-4 region with the HCV prototype sequence (HCV-PT; Choo et al., 1991). Antigenic regions identified by previous epitope mapping are underlined (Simmonds et al., 1993c). Substitutions of biochemically distinct amino acid residues that could affect the antigenicity of the encoded protein are highlighted in bold. (.), Sequence identity to HCV-PT; (-), gap introduced into the sequence to preserve alignment; (?), sequence not determined.
The presence of type-specific antibody within an individual sample was scored by comparison of the antibody reactivity within each well with that of the blocked control. Samples with absorbance values two or more times that of the blocked control were considered positive, provided the absorbance was greater than 0.1. Using this scoring system, we compared the detection of type-specific antibody by ELISA with the circulating genotype detected by PCR (Table 2). We found an overall concordance of 97% (177 out of 182 samples reactive in the ELISA contained type-specific antibody to the genotype detected by PCR), with no significant difference between genotypes. Only single instances were found where samples contained antibody to a different genotype in the RFLP assay. No sample showed antibody to more than one genotype. The frequency of detection of type-specific antibody was 87%, although variation was observed between genotypes. High rates (92–100%) were found for genotypes 1a, 2a, 2b and 3a, while type 1b and 4 were lower at 83% and 81% respectively. Detection rates for types 5 and 6 were based upon relatively fewer samples (16 and 6), and require further investigation with more extensive panels of genotyped samples.

Discussion

Sequence variability in the NS-4 region

Variability in the NS-4a region of the HCV genome between positions 4941–5282 appeared comparable to relationships between genotypes and subtypes of HCV in other coding regions of the genome (Simmonds et al., 1993a; Bukh et al., 1993, 1994). In each case, three levels of sequence variability were found, with a clear distinction into major genotypes and subtypes that has been proposed as the basis for classification of HCV (Simmonds et al., 1994a). The region of NS-4a analysed in this report shows similar degrees of sequence divergence to the other non-structural proteins of HCV, NS-3, NS-4B, NSSA and -5B. It is more conserved than E1 with nucleotide evolutionary distances ranging from 0.46–0.80 between major genotypes, compared with 0.43–0.85 for E1 (positions 574–1149, Bukh et al., 1993), and much more than core (0.13–0.25, Chan et al., 1992, and our unpublished observations). Similar relationships were found between the inferred amino acid sequences of the different parts of the genome (data not shown).

Pairwise comparisons of amino acid sequences of major genotypes showed 23–55% variation, and did not significantly overlap with any of those between subtypes (11–24%) nor between individual variants within a subtype (0–12%). Furthermore, genotype identifications of samples on the basis of sequence data obtained elsewhere in the genome, such as NS-5 (Fig. 2, seq. nos 6, 7, 20, 24, 25, 29–33) or the 5′NCR (Fig. 1; Fig. 2, seq. nos 34–50) are in each case consistent with the groupings observed in the NS-4 tree (Fig. 2), providing further evidence that sequence variability in NS-4a is representative of the virus as a whole.

Frequent amino acid substitutions were found between genotypes in regions of NS-4a previously shown to be antigenic by epitope mapping (Simmonds et al., 1993c), apart from the unexpected similarity found within the first antigenic region for genotypes 4, 5 and 6 (Figs 4 and 5). It is possible that this accounts for the slightly reduced sensitivity of the ELISA for type-specific antibody to these genotypes, and suggests the use of peptides from other regions of NS-4 as additional antigens for the ELISA. For example, we have been able to detect type-specific antibody using peptides from part of NS-4 downstream from the 2nd antigenic region that is highly variable between genotypes (unpublished data), and are currently investigating the possibility of using this region in a modified and more sensitive serotyping ELISA.

Detection of genotype-specific antibody

The incubation of samples with blocking concentrations of heterologous peptides provides an effective and sensitive assay for genotype-specific antibody. The method largely eliminates the possibility of false detection of antibody through shared epitopes between different genotypes, a problem that is inevitable and limiting for assays based upon indirect ELISAs (Tanaka et al., 1994; Tsukiyama Kohara et al., 1993; Yunomura et al., 1994; Mondelli et al., 1994). In contrast, the detection of reactivity to more than one genotype in the assay described in this report is more likely to indicate antibody produced by multiple infection with different HCV variants. This is shown by the previous finding of
Fig. 5. Box plot of the net absorbance value, \( A \) (A blocked control) for negative controls (A, \( n = 20 \)), and samples representing each HCV genotype [1a-B (\( n = 24 \)), 1b-C (\( n = 9 \)), 2a-D (\( n = 13 \)), 2b-E (\( n = 17 \)), 3-F (\( n = 50 \)), 4-G (\( n = 48 \)), 5-H (\( n = 10 \)), 6-I (\( n = 6 \))] against type-specific NS-4 epitopes from each genotype (wells 1-8) and to the unblocked control (well 7). Median value and interquartile ranges are indicated by the internal line and the upper and lower sides of the box; the single line represents values lying within 1.5 times the interquartile range. Outliers with values greater than 1.5 or 3 times the interquartile range are indicated by (*) and (○) symbols respectively (see Methods).

a high frequency of antibody to genotypes 1 and 2, or 1 and 3 in haemophiliacs exposed to multiple contaminated batches of clotting factor (Jarvis et al., 1994), which contrasts with the reactivity to single genotypes found in this study for all 182 reactive samples from blood donors and hepatitis C patients.

Few of the amino acid differences between HCV subtypes, for example between 1a and 1b, would be expected to lead to significant antigenic variability. We have previously found that reactivity to peptides derived from type 1a and 1b sequences produced almost identical patterns of reactivity upon epitope-mapping the NS-4a region, (Simmonds et al., 1993c). Similarly, sera from type 2a- and 2b-infected blood donors showed no consistent differences in reactivity with overlapping peptides from type 2b. This antigenic similarity (particularly between 1a and 1b) has provided an obstacle for the development of an ELISA to detect infection with different subtypes for all but a small proportion of samples. It remains possible that other (more variable) regions of the genome might be more suitable for such an assay. Alternatively, the sequence variability that does exist between subtypes might lead to antigenic differences between native proteins absent from linear peptides. However, the apparent inability of an NS-4 based serotyping assay based upon recombinant polypeptides to separately identify antibody to subtypes (Tanaka et al., 1994) argues against the second possibility. Fur-
Table 2. Detection of type-specific antibody in plasma samples of HCV-infected patients and blood donors of known genotype

<table>
<thead>
<tr>
<th>Serotype</th>
<th>Sensitivity (%)</th>
<th>Concordance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype*</td>
<td>1 2 3 4 5 6 NTS NR Total</td>
<td>11 All</td>
</tr>
<tr>
<td>a</td>
<td>24 - - - - 1 1 2 25 96.0 100</td>
<td>4</td>
</tr>
<tr>
<td>b</td>
<td>9 - - - - 1 1 1 12 83.3 90.0</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>All</td>
</tr>
<tr>
<td>Neg.</td>
<td>- - - - - - 20 20 - -</td>
<td></td>
</tr>
</tbody>
</table>

* Genotype determined by RFLP of 5'NCR sequences as previously described (McOmish et al., 1994; Davidson et al., 1994).
† Serotype: detection of type-specific antibody to HCV genotypes 1–6 in these samples; no sample showed type-specific antibody to more than one genotype. Results concordant with genotyping by RFLP shown in bold. NR, non-reactive; NTS, non-type-specific antibody reactivity to NS-4 peptides.
‡ Number typeable by ELISA/number tested.
§ Number concordant with genotype determination by RFLP/number typeable by ELISA.
[] Totals for type 1 and 2 samples; results for subtypes 1a, 1b and 2a, 2b are shown in italics.

thermore, there is little evidence that antigens derived from other regions of the HCV genome might be more appropriate for the detection of antibody to subtypes. For example, the degree of amino acid sequence variability in the E1 gene is similar to that found in NS-4 (see above), while E2 contains antigenic regions that are so variable that they are not conserved within a subtype.

The frequent detection of specific antibody to the circulating genotype identified by PCR indicates a potential role for this serotyping assay as an alternative to conventional PCR-based typing assays. However, the assay is dependent upon a normal immune response to virus infection, and would not necessarily be able to detect NS-4a antibody in individuals who are severely immunosuppressed, or in samples collected around the time of seroconversion following primary infection. In individuals who are multiply-exposed, antibody to more than one genotype was frequently found, even though only one circulating genotype was detected by PCR (Jarvis et al., 1994), indicating that the ELISA may detect past as well as current infection. The appearance of new type-specific antibody in those haemophiliacs in whom the circulating genotype spontaneously changed was frequently delayed by several months, providing a further cause of discrepant results in this risk group (Jarvis et al., 1994).

Information on risk behaviour and previous exposure to HCV was not available for the blood donor and hepatitis C patients in this study, and it is possible that past infection with different genotypes accounted for the discrepant results in Table 2. It is also possible that the peptides used do not represent the full range of antigenic variation within each genotype; to take one example from Table 2, epitopes present on an unusual variant of type 1 might also be found in type 6 NS-4a sequences. This possibility could be explored by sequence analysis of the NS-4a region sequences of samples producing such discrepant results. It would also exclude the possibility that the differences between genotyping in the 5'NCR by PCR and serotyping by ELISA result from infection with recombinants of HCV, with sequences corresponding to different genotypes in different parts of the genome. However, sequence analysis of variants of HCV in more than one region of the genome has so far failed to produce any evidence for the existence of recombinant viruses in vivo (Figs 1 and 2) (Simmonds et al., 1994b; Lau et al., 1994), and must be considered an unlikely explanation for the discrepant results reported here.

Clinical utility of genotyping

Although there is no evidence for genotypes of HCV that are completely non-pathogenic or perhaps non-hepatotropic, there are many reports of differences between them in the rate of disease progression, and particularly in the probability of achieving a sustained response to anti-viral treatment (reviewed in Dusheiko, 1994). These clinical differences indicate the potential utility for identification of the infecting genotype in patients for
patient selection, and in calculating the most effective duration and dose of interferon treatment to achieve a long-term response. The serological typing method described here may be of major value in service laboratories that require a quick and simple assay for identification of HCV genotypes, without the necessity to purchase expensive items of equipment for PCR or to introduce unfamiliar working techniques into the laboratory. The use of different combinations of competing peptides in the ELISA allows the design of different formats of the assay to maximize its clinical utility and ease of use. For example, in Europe it may be more practical to screen patients for genotypes 1, 2 and 3 separately, and to detect other genotypes (4, 5 and 6) together by competition with peptides from genotypes 1, 2 and 3 only. It would also be possible to incorporate additional peptides corresponding to newly discovered major genotypes (e.g. Apichartpiyakul et al., 1994) into the ELISA for screening populations where these variants occur.

The authors would like to thank Dr E. A. C. Follrett, Brian Dow and staff at the SNBTS Microbiology Reference Laboratory, Ruchill Hospital, Glasgow, Dr R. Crookes, South African Blood Transfusion Service, and Dr C. K. Lai, Hong Kong Red Cross Blood Transfusion Service, Kowloon, Hong Kong, for providing samples for sequence analysis and serotyping.

References


(Received 19 December 1994; Accepted 27 February 1995)