Simplified Hepatitis C Virus Genotyping by Heteroduplex Mobility Analysis

PETER A. WHITE, 1* XINYAN ZHAI, 1, 2 IAN CARTER, 1 YUE ZHAO, 1, 2 AND WILLIAM D. RAWLINSON 1

Virology Division, Department of Microbiology, SEALS, Prince of Wales Hospital, Randwick, Sydney, New South Wales 2031, Australia, 1 and Beijing Biochemical and Immune Reagents Centre (BBIRC), Yang Fang Cun, An Ding Men, Beijing 100012, China 2

Received 16 July 1999/Returned for modification 16 September 1999/Accepted 22 October 1999

Heteroduplex mobility analysis (HMA) was used to genotype hepatitis C viruses (HCV) with PCR fragments derived from the 5′ untranslated region (5′-UTR) or the NS5b region. HCV 5′-UTR fragments were amplified from 296 serum samples by use of a combined reverse transcription-PCR assay, and the genotypes of isolates were determined by sequencing. HCV genotype distributions in Australia were 39% for genotype 1a, 15% for 1b, 3% for 1a/b, <1% for 2a/c, 5% for 2b, 34% for 3a, <1% for 3b, and 1% for 4, and 1% of patients were infected with more than one genotype. Pairwise HMA of subtypes 1a, 1b, 2a/c, 2b, 3a, 3b, 4a, and 6a demonstrated that five distinct heteroduplex patterns were formed between the eight subtypes. A reference panel that contained a representative of each pattern (1a, 2b, 3a, 4a, and 6a) was used for genotyping. The pattern of heteroduplexes formed when a test isolate was mixed with the five reference isolates was correlated with the genotype, as determined by sequencing. Genotypes determined by HMA correlated exactly with sequencing results within the groups 1, 2, 3a, 3b/4, and 6. HMA was also used to simplify the identification of mixed infection with two HCV genotypes. In further studies, with amplicons from the NS5b region, HMA classified isolates into their respective subtypes, and the heteroduplex mobility ratio correlated closely with nucleotide sequence variation at the isolate, subtype, and genotype levels. HMA provides an adaptable, inexpensive, and rapid method of genotyping HCV that requires fewer resources than DNA sequencing.

Hepatitis C virus (HCV) is the major cause of non-A, non-B hepatitis (5, 16). It is a highly variable positive-stranded RNA flavivirus whose genetic diversity is exhibited on three levels. The first level differs by >30% in nucleotide sequence over the whole genome, currently enabling classification into 11 genotypes. The second level divides genotypes into more than 80 subtypes whose nucleotide sequences differ by >20%. The third level defines isolates within a subtype which differ by up to 10%. Genetic variation is not uniformly distributed across the genome, with the region encoding envelope glycoproteins being the most variable, whereas the 5′ untranslated region (5′-UTR) and the 3′ untranslated region are highly conserved (15, 26).

Genotype, viral load, and liver histology are important parameters used in selecting antiviral therapy with the greatest chance of success. Genotype 1 isolates, in particular, 1b, are known to respond poorly to interferon treatment (23), whereas other HCV genotypes, such as 3, respond more favorably. The highly conserved 5′-UTR is almost exclusively used for routine reverse transcription (RT)-PCR detection of HCV; this method is currently the most sensitive and reliable for establishing ongoing infection. The 5′-UTR also exhibits specific polymorphisms between types and subtypes, allowing classification into six genotypes (28). Genotyping assays which have used the 5′-UTR PCR product include direct sequencing (12, 21), restriction fragment length polymorphism analysis (20), and the use of genotype-specific probes (29). Other genotyping methods include type-specific PCR with primers for NS5b (3) or core regions (13) and the use of genotype-specific probes for the core region (32). The widely accepted reference standard for genotyping HCV is nucleotide sequencing of NS5b amplicons. However, in all assays, the initial RT-PCR detection step or the downstream processing of the PCR product remains too complicated, costly, or time-consuming for routine genotyping. Many methods have been published, although no simple, inexpensive, and accurate method has yet been developed (27).

Heteroduplex mobility analysis (HMA) relies on the formation of mismatches when two divergent DNA molecules are mixed, denatured, and allowed to reanneal. This process results in the formation of homoduplexes and heteroduplexes that migrate at different speeds by polyacrylamide gel electrophoresis (PAGE). The mismatches reduce the mobility of the heteroduplexes, which are retarded approximately in proportion to the divergence between the two sequences. Unpaired nucleotides produce larger shifts than mismatched nucleotides (8). Genotyping by HMA involves mixing a PCR product of unknown genotype separately with a panel of reference products of each genotype and separating the resultant heteroduplexes by PAGE. Ideally, the sequences of the subtypes in the panel should adhere as closely as possible to the consensus sequence of each subtype. Genotype determination relies on the identification of heterologous genotypes in lanes that contain heteroduplexes with reduced mobility. HMA applications have included assessment of quasispecies in human immunodeficiency virus (8) and HCV (24, 30, 33) and screening of influenza B virus variants (35). In the present study, we genotyped HCV by HMA with PCR products derived from two regions of the genome, the highly conserved 5′-UTR and the NS5b region, and compared the results directly with sequencing results.

MATERIALS AND METHODS

Serum samples. Serum was obtained from 296 patients around Australia who were chronically infected with HCV. Serum HCV RNA was detected by RT-PCR of the 5′-UTR by use of either the AMPLICOR HCV diagnostic kit according to the manufacturer’s instructions (Roche Diagnostic Systems, Branchburg, N.J.) (256 samples) or an in-house assay (40 samples).
RNA extraction. RNA used for the in-house RT-PCR was extracted from 100 μl of serum by a modification of the method of Chomczynski and Sacchi (4). Serum (100 μl) was added to 300 μl of solution (4 M guanidinium thiocyanate, 25 mM sodium citrate [pH 7.0], 0.1 M 2-mercaptoethanol), 300 μl of citrate-buffered saturated phenol (pH 4.8), and 150 μl of chloroform-isomyl alcohol (24:1 [vol/vol]). Samples were vortexed briefly and centrifuged at 13,000 × g for 10 min. The aqueous supernatant was precipitated with 3 M sodium acetate (pH 4.8), 10 μl of tRNA (1:10 [vol/vol]), and an equal volume of isopropanol and centrifuged for 15 min at 13,000 × g. The pellet was washed with 70% ethanol, air dried, and suspended in 50 μl of water.

RT-PCR. PCR and RT were performed in a one-step reaction, like previous methods used to detect HCV (14, 34). Positive and negative controls consisted of known positive and negative sera and water and were included in all reactions from the extraction step. Strict experimental procedures were observed to avoid false-positives (17). The conserved 5′-UTR was amplified with first-round oligonucleotide primer pairs described previously (34): KY78, sense (5′-GCGTCTACACTCATCAGGAG-3′); hep22, antisense (5′-GCRCACAACCTCGCTGCTG-3′). Oligonucleotide primer pairs designed to amplify the NS5b region were as follows: hep31b, sense outer (5′-CATAGC-3′); hep33b, sense inner (5′-AYACCCGMTGYTTTGACTC-3′); hep34b, antisense (5′-CCGARTACCTGGTCTCTCDTATGAYACC-3′); hep35b, sense inner (5′-CATAGC-3′); hep36b, antisense (5′-UTR inner PCR fragment used for genotype determination by HMA and sequencing. This fragment corresponds to nucleotides 98 to 272 (5). (A) Sequence alignment of the six genotypes and 12 main subtypes. DNA sequences are compared by use of a Prism DyeDeoxy terminator cycle sequencing kit (Perkin-Elmer). Consensus sequences for subtypes 1a, 1b, 2a/c (sequences of subtypes 2a and 2c are identical in this region), 2b, 3a, 3b, 4a, 4c/d (sequences of subtypes 4c and 4d are identical in this region), and 6a were determined from database searches and previously published data (28). HCV subtypes were established by comparison with the consensus sequences of the main subtypes (Fig. 1A).

Database searches were conducted with BLAST (1). Pairwise alignments of DNA sequences were carried out with the GCG program GAP (9), and multiple alignments were carried out with Clustal W (31). Evolutionary distances between sequences were determined with the DNADIST program (Kimura two-parameter method) of the PHYLIP package (version 3.57) (10). The computed distances were used for the construction of phylogenetic trees by the neighbor-joining method (25) of the program NEIGHBOR. Trees were plotted with the program TREEVIEW (version 1.5) (22).

Selection of 5′-UTR reference PCR products. In order to investigate the mobility of heteroduplexes formed between different subtypes, 5′-UTR amplimers from subtypes 1a, 1b, 2a/c, 2b, 3a, 3b, 4a, and 6a (of known DNA sequence) were chosen for pairwise HMA. These reference isolates were selected from the 296 samples sequenced in this study and 12 other sequences from a separate cohort of rarer subtypes. All reference subtypes were identical to 175-bp consensus sequences, except for subtypes 1a and 3b, which differed from their
TABLE 1. Distribution of HCV subtypes in Australia, as assessed by sequencing and HMA of 5′-UTR PCR products

<table>
<thead>
<tr>
<th>HCV subtype</th>
<th>No.</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a/b</td>
<td>8</td>
<td>2.7</td>
</tr>
<tr>
<td>1a</td>
<td>116</td>
<td>39.2</td>
</tr>
<tr>
<td>1b</td>
<td>44</td>
<td>14.9</td>
</tr>
<tr>
<td>2a/c</td>
<td>2</td>
<td>0.7</td>
</tr>
<tr>
<td>2b</td>
<td>14</td>
<td>4.7</td>
</tr>
<tr>
<td>3a</td>
<td>101</td>
<td>34.1</td>
</tr>
<tr>
<td>3b</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>4a</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>4c/d</td>
<td>2</td>
<td>0.7</td>
</tr>
<tr>
<td>6a</td>
<td>4</td>
<td>1.4</td>
</tr>
<tr>
<td>Mixed infection</td>
<td>3</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Sequence analysis of the 5′-UTR. Sequence alignments and phylogenetic analysis of consensus sequences of the 175-bp 5′-UTR fragment used in this study are shown in Fig. 1A for the 12 main subtypes. This region of the HCV genome is highly conserved. The average distance between genotypes is 0.064 substitution per nucleotide, and 85% of the nucleotides demonstrated no variation between the major subtypes (Fig. 1A). The sequences of subtypes 2a and 2c and of subtypes 4c and 4d are identical over this 175-bp region, and other subtypes are differentiated by 1 to 6 bp. These include subtypes 1a and 1b, which differ by only 1 bp at nucleotide position 243, and subtypes 4a and 4c/d, which differ by 2 bp at nucleotide positions 264 and 243 (according to the numbering system of Choo et al. [5]). Subtypes 2a and 2b are differentiated by 4 bp, and subtype 3b diverges by 6 bp from subtypes 3a and 4a. Genotype 6 contains a 2-bp insertion at position 199 and a 1-bp insertion at position 207 compared to all other genotypes (Fig. 1A). Phylogenetic analysis of the region reveals three divergent groups, subtypes 2b, 3a, and 6a. Subtypes 2b and 3a are the most divergent (86.3% identity), and genotypes 1, 4, and 5 and subtype 3b are closely related, clustering together in the phylogenetic analysis (Fig. 1B).

Sequencing of the 5′-UTR of HCV to determine genotype. Serum from 296 HCV RNA-positive Australian patients were collected for genotyping. HCV 5′-UTR fragments were amplified from purified RNA, and genotypes were determined by sequencing PCR products. Table 1 shows the distribution of genotypes and subtypes in Australia. Sequencing and HMA identified 56.8% of isolates as genotype 1, of which 39.2% were subtype 1a and 14.9% were subtype 1b. Subtype could not be assigned to eight genotype 1 isolates (2.7%) due to the presence of both G and A nucleotides of significant amplitude at the key polymorphic position (nucleotide 243 [5]), possibly because of coinfection with subtypes 1a and 1b. The second most common subtype found in Australia is subtype 3a, representing 34.1% of isolates. Other genotypes are comparatively rare in Australia, with genotypes 2, 4, and 6 representing only 5.4, 1.0, and 1.4% of isolates, respectively. Three serum samples (1%) contained two different genotypes (Table 1).

HMA of the 5′-UTR. (i) Pairwise HMA. Pairwise HMA with the 175-bp internal PCR product was carried out between each of the eight main subtypes. The pattern of heteroduplexes formed with the subtypes was recorded (Fig. 2). Heteroduplexes formed between closely related isolates could not be separated by PAGE from the homoduplexes, including those formed between 1a and 1b, 2a/c and 2b, and 3b and 4a (Fig. 2). These pairs of subtypes (1a-1b, 2a/c-2b, and 3b-4a) formed very similar or identical heteroduplex patterns and could not easily be distinguished from each other by HMA. Subtypes 3a and 6a formed heteroduplexes with all heterologous subtypes; these heteroduplexes could be separated from homoduplexes by PAGE (Fig. 2 and Fig. 3D and F). Larger shifts produced by unpaired nucleotides occurred with the subtype 6a product, which contained three additional nucleotides (Fig. 2 and Fig. 3F). Heteroduplexes formed between genotype 1 isolates and either subtype 3b or subtype 4a could not be separated from homoduplexes, consistent with the close relationship between these 5′-UTR sequences (Fig. 1). In most cases, sequence differences correlated closely with the HMA, where the most divergent pairs of subtypes formed heteroduplexes with the greatest reduction in mobility (Fig. 2). The exception was HMA between genotype 2 products and either subtype 3b or subtype 4a products. For example, no visible heteroduplexes were identified between subtype 2b and 4a sequences (Fig. 2, lane 25) although they exhibit 10.5% sequence diversity (14 mismatches). In contrast, a 5.3% sequence divergence between subtype 3a and 3b amplicons (seven mismatches) produced visible heteroduplexes (Fig. 2, lane 28).

(ii) Genotyping by HMA. Pairwise HMA demonstrated that five distinct heteroduplex patterns were formed between the

![FIG. 2. Ethidium bromide-stained 8% polyacrylamide gel showing pairwise HMA of subtypes 1a, 1b, 2a/c, 2b, 3a, 3b, 4a, and 6a. Heteroduplexes formed between HCV subtypes can be seen as bands with reduced mobilities in the gel. Subtypes used in the pairwise analysis are shown along the top. Lanes containing reference subtypes are indicated by large bold typeface and show the homoduplex bands; subsequent lanes show HMA of the reference subtype and other subtypes (small normal typeface). Lane M, molecular size markers. All possible combinations of the subtypes are shown.](http://jcm.asm.org/10.1128/JCM.38.4.476-485.2000)
eight subtypes tested; pattern 1—genotype 1, pattern 2—genotype 2, pattern 3—subtype 3a, pattern 4—subtypes 3b and 4a, and pattern 5—subtype 6a (Fig. 2 and 3). The reference panel used for genotyping contained a representative of each pattern (1a, 2b, 3a, 4a, and 6a). The pattern of heteroduplexes formed by mixing a test isolate with the five reference isolates was correlated with the genotype. For example, a genotype 3a isolate was easily identified by the presence of heteroduplexes in all lanes except the lane containing the 3a reference (Fig. 3D). Figure 3 shows examples of HMA of each genotype and the five heteroduplex patterns identified. Despite minor sequence variation, the ability of HMA to correctly identify genotypes was not affected. Figure 3A shows HMA of a subtype 1a isolate that exhibited 4 nucleotide differences from the 1a reference but was assigned the correct genotype. Figure 3B and C show HMA of genotype 2 isolates; larger shifts are indicative of subtype 2b rather than subtype 2a/c (Fig. 2 and Fig. 3B and C). HMA of several 3a subtypes revealed a minor shift in the 3a reference lane; however, the subtype could still be determined, as this lane revealed the heteroduplex with the greatest mobility. Figure 3E shows an example of a subtype 4c/d isolate that demonstrates a heteroduplex pattern identical to that of a subtype 4a isolate (Fig. 2). The sequences of many of the 296 5’-UTR products were identical. The ability of HMA to genotype HCV was evaluated by testing multiple identical sequences and all unique sequences. Genotyping by HMA correlated exactly with sequencing results within groups 1, 2, 3a, 3b/4, and 6 (Table 1).

The reference panel was reduced to include only subtypes 1a and 3a. HMA with this panel was still able to identify unknown isolates, generating one of four patterns (Fig. 4). These patterns correlated with the following genotypes and subtypes: pattern A—genotype 1, pattern B—subtype 2b or genotype 6, pattern C—subtype 3a, and pattern D—subtypes 2a/c and 3b and genotype 4 (Fig. 4). Pattern A is distinguished from pattern D by a much larger heteroduplex shift with the 3a reference (Fig. 4). With this refined reference panel, the most common genotypes, 1 and 3a (representing 90.9% of isolates), could be distinguished quickly and easily (Fig. 4).

(iii) Mixed infection. Sequence data of 6 of the 296 5’-UTR PCR products were ambiguous, with electropherograms demonstrating overlapping peaks. In three cases, this result was due to an insufficient or impure template, and the PCR and sequencing were repeated. Three samples reproducibly had overlapping peaks, and careful inspection of the electropherograms suggested that the samples were most likely coinfected with two different genotypes. Subsequent HMA of these three samples revealed multiple heteroduplexes in several reference lanes. The PCR products were then subjected to denaturation and annealing (that is, HMA with no reference), and subsequent PAGE identified heteroduplexes in all three cases, confirming mixed infection.

HMA of the NS5b region. Genotype determination by HMA of the 5’-UTR was able to distinguish genotypes but was unable to differentiate between subtypes in the majority of cases. Therefore, subtypes were determined by HMA of the more variable NS5b region with 12 isolates (S1 to S12) representing 1a, 2b, 2c, 3a, 3b, and 4a. All 12 of the 366-bp NS5b PCR products were sequenced on both strands. The genotypes of the isolates were established by phylogenetic analysis with consensus sequences of eight subtypes (222 nucleotides, 8316 to 8537 [5]). The phylogenetic tree clearly shows the three levels of genetic diversity among HCV genotypes, subtypes, and isolates (Fig. 5). Initial experiments to establish optimal NS5b HMA conditions demonstrated that an annealing temperature of 60°C resulted in weak heteroduplex bands formed between genotypes when PCR products diverged by more than 40%. Therefore, the annealing temperature was decreased to 50°C, producing clearer bands after HMA; this temperature was used in all subsequent NS5b HMA experiments.

As has previously been reported, there is a relationship between the gel shift distances (HMR) of the homoduplex and heteroduplex bands which is proportional to the nucleotide
The average HMRs for isolates (0.95), subtypes (0.69), and genotypes (0.28) were calculated following HMA of genotypes, differences between the two DNA molecules (8, 24). The average HMR was 0.95 ± 0.06, corresponding to an average genetic distance between sequences of 7.5% (n = 8) (Fig. 6). Larger differences between isolates were reflected by a greater shift in mobility, as shown in Fig. 6, lane 3, for two 4a subtypes that demonstrated 11.2% nucleotide divergence. As shown in Fig. 6, lanes 4 and 5, HMA between different subtypes within a genotype resulted in heteroduplexes with significantly lower mobility; the average HMR was 0.69 ± 0.03, corresponding to 28.0% (n = 4) sequence divergence. Fig. 6, lanes 6 and 7, show heteroduplexes formed between genotypes where sequences diverged, on average, by 45.2% (n = 54), resulting in a larger reduction in mobility; the average HMR was 0.55 ± 0.04 (Fig. 6). The proportional reduction of shift is not linear, because larger genetic distances are not reciprocated by a proportional reduction in mobility (8). HMA of the NS5b region was found to be a more accurate indicator of subtype than 5'-UTR HMA. Moreover, the HMR could be used to predict the genotype and subtype of the isolate, and the accuracy of this genotyping method was confirmed by sequencing.

### DISCUSSION

HCV genotype 1 is the most prevalent worldwide, representing 77% of isolates in China, 82% in Spain, 90% in Brazil (13), and 85% in the United States (21). In contrast, we found a lower proportion of HCV genotype 1 isolates (56.8%) and a higher proportion of subtype 3a isolates (34.1%) in Australia consistent with other Australian studies (12, 19). This result is of importance in terms of a successful outcome of chemotherapy, as subtype 3a is more responsive to current antiviral agents (23).

The highly conserved nature of the 5'-UTR means that identification of subtype by HMA or sequencing is not possible with all HCV isolates. Examination of other HCV sequences in the database showed that subtypes 7a, 7b, 8a, 9a, and 11a differ from the consensus sequence of subtype 1b by less than 3 bp. These sequences would be grouped by HMA as genotype 1 and by sequencing as subtype 1b. In this study, we found no genotype 5 isolates, which are prevalent in South Africa (7). Genotype 5 is most closely related to genotype 1, demonstrating 5- and 4-bp sequence differences from subtypes 1a and 1b, respectively, over the 5'-UTR PCR product used in this study. This result suggests that HMA of a genotype 5 isolate would classify it as a genotype 1 isolate. Subtype 6b differs from subtype 6a by only 1 bp in this 175-bp region, and subtype 10a is closest to subtype 3b, differing by 4 bp from the consensus sequence. Amplification and subsequent analysis of the more variable NS5b region would be required for accurate subtype determination of these less common subtypes.

To date, the majority of reports describing the applications of HMA for the study of HCV infection have focused on quasispecies evolution (11, 24, 30, 33). Only one report, by Calvo and coworkers, describes HMA to determine the HCV genotype for 15 patients by use of PCR products derived from the core/E1 region (2). In the present study, we applied HMA to determine genotype by using PCR products derived from the 5'-UTR and NS5b region.

Amplicons from the 5'-UTR were chosen for HMA genotyping, as this region is almost universally used for routine detection of HCV in commercial assays. Unlike sequencing, HMA was unable to differentiate between subtypes 1a and 1b, subtypes 2a/c and 2b, and subtypes 3b, 4a, and 4c/d. Clinically, the most important differentiation not detectable by HMA was between subtypes 1a and 1b, as subtype 1b isolates are not as responsive to interferon therapy as subtype 1a isolates (23). Although sequencing can differentiate between these subtypes...
due to a 1-bp difference, 3 to 5% of isolates are mistyped (27). Our HMA results matched exactly the results obtained by sequencing of 5′-UTR PCR products. We found HMA genotyping to be extremely simple and rapid, and it did not involve the use of expensive reagents or equipment. The assay was also found to be robust and able to cope with minor sequence variations, unlike restriction enzyme analysis, where single point mutations can adversely affect the results. Furthermore, HMA was easily adapted; the reference panel could be altered to identify specific subtypes or to increase the number of samples that could be analyzed on a gel. We found that mixed infection was more easily identified by HMA than by sequencing. By HMA, three samples were found to be coinfected, and two of these were identified by sequencing. Lau and coworkers suggested that direct sequencing of PCR products would identify coinfection only if one variant represents more than 10 to 20% of the HCV genomes amplified (18).

Preliminary HMA studies with 12 isolates demonstrated that PCR products derived from the NS5b region formed heteroduplexes that migrated relative to the genetic diversity of the isolates. This genetic diversity was exhibited on three levels in HCV, corresponding to genotype, subtype, and isolate. These three levels correlated with three distinct levels of mobility of the heteroduplexes in HMA. HMA genotyping of HCV isolates allows for rapid and simple genotyping in less than 3 h post-PCR and should prove useful in screening large numbers of samples during antiviral therapy.

ACKNOWLEDGMENTS

We thank Evelyn Crewe, Alison Doughty, Geoff McCaughan, George Marinus, and Jenean Spencer for serum samples used in this study; Michele Taylor for organization of the serum database; and Gillian Scott and Yi-Mo Deng for critical reading of the manuscript. We also thank AusAID for funding of X.Z. and Y.Z.

REFERENCES