Prevalence of hepatitis C virus sequence variants in south-east Asia

W. K. Greene, M. K. Cheong, V. Ng and K. W. Yap*

Scitech Genetics Ltd, 12 Science Park Drive #04-04, The Mendel, Singapore 0511

The nature and distribution of hepatitis C virus (HCV) genotypic variants present in south-east Asia have not been extensively investigated. We analysed HCV RNA obtained from 67 clinical serum samples from Singapore, Thailand, Indonesia, the Philippines and South Korea. All samples were amplified by semi-nested RT-PCR and the nucleotide sequence determined for four regions within the E1, E2/NS1, NS4 and NS5 genes. Each isolate had a unique nucleotide and deduced amino acid sequence, consistent with the genetic heterogeneity of this virus. There was remarkably little amino acid sequence variation between isolates of the same genotype, apart from variable domains within putative envelope glycoproteins that are likely to be under immune pressure. All isolates could be classified according to the currently recognized genotypes of HCV, with the exception of one Singapore isolate that defined a new group 3 subtype. The 1b genotype, which predominates in Japan, was the most widely distributed genotype and accounted for 58% of all isolates sequenced. Regional variations in HCV genotype distribution were observed, with type 3a being found almost exclusively in Thailand. By contrast, the 1a genotype, which predominates in the USA was the most prevalent genotype in the Philippines. Genotype 1a was found less commonly among the Thai isolates, presumably having been introduced from the West in stored blood products or by sporadic transmission. The significant prevalence of HCV types 2 and 3 restates the need for variant genotypes to be included in immunodiagnostic and vaccine development strategies. This study reveals that the 1b genotype of HCV, previously found to be the major variant present in east Asia, also predominates in the south-east Asian region, and may be the major HCV type found worldwide.

Hepatitis C virus (HCV), a distant relative of the pesti/flaviviruses, is a hepatotropic virus first identified by molecular cloning (Choo et al., 1989, 1991). Like human immunodeficiency virus type 1, HCV has a highly plastic RNA genome capable of extensive sequence variation, both among isolates from different geographical areas and within a single individual (Okamoto et al., 1991, 1992a; Martell et al., 1992). Consequently, the analysis and classification of genetic variants is important for understanding the molecular basis of viral virulence and drug resistance, as well as routes of HCV transmission. Knowledge of antigenic variants will in addition aid in the selection of target antigens/epitopes for sensitive immunodiagnosis and ultimately vaccine development.

An enormous amount of sequence data of different HCV isolates has recently been reported. The complete genomic sequences of 13 HCV isolates have been published, including two from the USA (Choo et al., 1991; Inchauspe et al., 1991), nine from Japan (Kato et al., 1990; Takamizawa et al., 1991; Okamoto et al., 1991, 1992a, b, c; Honda et al., 1992; Tanaka et al., 1992; Hayashi et al., 1993) and one each from Taiwan (Chen et al., 1992) and China (Wang et al., 1993). In addition, partial sequences have been made available from a wide range of geographical locations throughout the world (Enomoto et al., 1990; Tsukiyama-Kohara et al., 1991; Weiner et al., 1991; Kremsdorf et al., 1991; Cha et al., 1992; Stuyver et al., 1993a; Simmonds et al., 1993a; Bukh et al., 1993). Collectively, these studies have revealed that the HCV family is a heterologous group, clustering into at least 14 distinct genotypes, some of which are widely distributed around the world while others are confined to particular geographical locations. However, little information on the nature of HCV isolates from south-east Asia has so far been forthcoming, with the exception of partial sequences of the NS5 gene of isolates from Thailand (Mori et al., 1992).

To determine the nature and distribution of HCV genotypic variants in south-east Asia, semi-nested RT–PCR was performed using RNA isolated (by the acid guanidium thiocyanate–phenol–chloroform method)
Table 1. HCV genotype distribution

<table>
<thead>
<tr>
<th>Country</th>
<th>1a</th>
<th>1b</th>
<th>2a</th>
<th>3a</th>
<th>3b</th>
<th>3c</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore</td>
<td>–</td>
<td>10</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Thailand</td>
<td>3</td>
<td>8</td>
<td>10</td>
<td>1</td>
<td>–</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td>–</td>
<td>11</td>
<td>2</td>
<td>–</td>
<td>–</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Philippines</td>
<td>6</td>
<td>4</td>
<td>–</td>
<td>1</td>
<td>–</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>South Korea</td>
<td>–</td>
<td>6</td>
<td>4</td>
<td>–</td>
<td>–</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>39</td>
<td>6</td>
<td>11</td>
<td>1</td>
<td>67</td>
<td></td>
</tr>
</tbody>
</table>

(13%) (58%) (9%) (16%) (1-5%) (1-5%) 

from 0.1 ml of serum from each of 67 HCV-infected individuals. Sera were confirmed RNA positive by PCR using Tbr polymerase (Finnzymes) with primers specific for the 5' untranslated region (Bukh et al., 1992). Eleven sera were from Singapore, 22 from Thailand, 13 from Indonesia, 11 from the Philippines and 10 from South Korea. Four sets of primers were designed in order to amplify genomic regions within the HCV E1, E2/NS1, NS4 and NS5 genes. These four primer sets were carefully chosen to match highly conserved sequences and included degeneracy, in order to enable amplification of variant HCV genomes. The primer sequences were: for the E1 region, 5' CGGSGTKAAYTWYGCAACAGG 3' from nt 480-500 (external sense primer E1Fo), 5'YGGTTGCTCYTTYTCTTCCT 3' from nt 510-530 (internal sense primer E1Fi), 5' CCAGTCATCATCATRTCCCA 3' from nt 978-958 (antisense primer E1R); for the E2/NS1 region, 5' GGAATGATGATGAACTGGTC 3' from 960-980 (sense primer E2F), 5' DGGBGSGTWRTGCCARCARTA 3' from nt 1473-1453 (internal antisense primer E2Ri), 5' GAARCARTAYACYGGRCRCA 3' from nt 1527-1507 (external antisense primer E2Ro); for the NS4 region, 5' VTGTTGYYGYAGGGCYMAROC 3' from nt 4776-4796 (external sense primer 4Fo), 5' VTTGGGAYSWVTGTTGGAAGTG 3' from nt 4809-4829 (internal sense primer 4Fi), 5' SCCRCTGATGAAR- TCCACAT 3' from nt 5313-5293 (antisense primer 4R); for the NS5 region, 5' TGAACGCGMTGY- TTYGACTC 3' from nt 7917-7939 (sense primer 5F), 5' RGGRTCWCGKGTSAGRTAGTA 3' from nt 8424-8404 (internal antisense primer 5Ri), 5' KCCTAGCCAKGTRACWGG 3' from nt 8490-8470 (external antisense primer 5Ro). All sequences are numbered according to the HCV-1 isolate (Choo et al., 1991). Amplified HCV cDNA was purified using Prep-A-Gene (Bio-Rad) and directly sequenced using Hot Tub and [α-32P]dATP (Amersham).

At least one PCR-amplified region for each of the 67 HCV isolates obtained was sequenced. Each isolate analysed had a unique nucleotide and deduced amino acid sequence. The results (summarized in Table 1) showed that with the exception of one Singapore isolate, all isolates could be classified within currently recognized genotypes. Classification was based on the characteristic segregation of amino acid residues between HCV types that is maintained throughout the genome. The nomenclature for designating HCV genotypes is that proposed by Simmonds et al. (1993a) that takes into account the two clearly distinct levels of sequence variation which differentiate HCV into several types and subtypes.

Of the known genotypes, only the widely distributed types 1 (a and b), 2a and 3 (a and b) were identified. Genotypes 2b, 4, 5 or 6 were not present in any sample. Although the sample sizes of virus isolates drawn from each population were relatively small, several salient points could be inferred from these data. Firstly, the results revealed that genotype 1b predominated in southeast Asia, where it accounted for 58% of all isolates. It was also widely distributed, being the only genotype found in all five countries investigated. In Singapore, Indonesia and South Korea, the main HCV genotype found was 1b, where it accounted for 10/11, 11/13 and 6/10 isolates in each country, respectively. Such a result is consistent with the high prevalence of type 1 b found in east Asia (Nakao et al., 1991; Wang et al., 1993). Genotype 1b also accounted for a significant proportion of the HCV isolates from Thailand (8/22) and the Philippines (4/11). These findings, along with the relatively high prevalence of type 1b reported in Europe and the Americas (Bukh et al., 1993; Simmonds et al., 1993a; Stuyver et al., 1993b), indicate that type 1b may be the major HCV genotype worldwide.

Other genotypes appeared to be more restricted in their geographical distribution. Similar to the reported prevalence in Japan (Nakao et al., 1991), HCV type 2 was found as a less common variant in South Korea (4/10) and Indonesia (2/13). The absence of type 2 from other southeast Asian countries implies that the incidence of this type is not consistent throughout east Asia. In agreement with this observation, Wang et al. (1993) found a much higher incidence of type 2 HCV in northern China than in the south, where type 1b was
highly predominant. A high proportion of HCV type 3 (a and b) was found in Thailand (11/22 isolates). Type 3 sequences were also found as a minor variant in Singapore (1/11) and the Philippines (1/11). The type 3 Singaporean isolate (S48) shared sequence identity with both of the previously reported 3a and 3b subtypes, but was sufficiently different from both to warrant tentative classification as a new subtype (3c). Type 3 HCV has also been reported in Europe, the USA and South America (Simmonds et al., 1993a; Stuyver et al., 1993b), but appears to be rare elsewhere in east Asia.

The 1a genotype also appeared to be restricted in its distribution, presumably reflecting its low prevalence in Asia. However, in contrast to other isolates from the various countries investigated, type 1a predominated in the Philippines (6/11 isolates). This genotype was also present in Thailand as a minor variant (3/22), as reported in Japan (Nakao et al., 1991). The low incidence of HCV type 1a in Japan has been associated with haemophilia patients (Hijikata et al., 1990). Likewise, in Italy genotype 1a was predominant among HCV-infected haemophiliacs but not in those patients with community-acquired hepatitis C (Pistello et al., 1994). Many countries have relied on commercial plasma products largely derived from paid donors from the USA. This indicates that type 1a HCV, common in many Western countries, may have spread to other countries via imported blood products. The relatively high incidence of type 1a in the Philippines also suggests possible spread by inadvertent modes of transmission.

A comparison of the deduced amino acid sequences of the four regions was made with other HCV sequences found worldwide. Two structural (E1, E2/NS1) and two non-structural (NS4, NS5) regions were analysed, which together constitute 22% of the HCV polyprotein sequence. Despite their geographically diverse origins, remarkably little variation in amino acid sequence was found between isolates of the same subtype and amino acid substitutions were often conservative. This was particularly the case in the NS4 and NS5 regions where amino acid sequence identity within a subtype always exceeded 90% (Table 2). Significantly more variation was found in the E1 region representing the N-terminal half of this molecule, particularly within residues 216–257. Notably, one isolate (T145) had a four amino acid deletion within E1 (residues 298–301). The most striking intratypic variation was found in E2/NS1 which contained two hypervariable regions (HVRs), HVR1 at the N terminus (residues 384–410) and HVR2 (residues 475–480). HVR1 exhibited only 26–89% amino acid identity, such that no two isolates, even of the same subtype, were identical within this region. Amino acid sequence variation within the variable, and particularly the hypervariable domains of the envelope regions, is most likely the result of selection imposed by the immune response (Weiner et al., 1992). That the envelope glycoproteins of HCV appear to be subject to a positive selection for changes is indicated by the high rate of nonsynonymous mutations found in this region (Ogata et al., 1991; Kurosaki et al., 1993).

A striking segregation of amino acid sequences was evident on comparison of the different HCV genotypes. The extent of identity between HCV types was dependent on the region used for comparison (Table 2). Invariant residues were more common in the less variable NS4 and NS5 regions, comprising 53/138 and 72/140 residues respectively, as compared with the E1 (39/128) or E2/NS1 regions (43/135). The NS5 region, although displaying considerable variability, was the most conserved of the four regions analysed with 72–82% amino acid identity between types. E1 was the least conserved (51–76% amino acid identity between types). Genotype-specific residues, which were highly conserved within the same HCV subtype, were found to cluster and intersperse throughout all four regions sequenced. Most notable was a relatively conserved type-specific motif within E1 (residues 248–258) that alone could predict the division of all isolates into their respective genotypes. The significance of these regions is not yet known but they would be expected to account for any biological differences between genotypes. The degree of sequence variation implies that the various HCV types diverged quite some time ago, and is similar to that found for serotypes of various flaviviruses (Chambers et al., 1990).

The sequence differences between HCV types might be significant for HCV screening given the present high degree of dependence on serological methods. Current anti-HCV assays commonly employ type 1a core, NS3, NS4 and NS5 antigens as recombinant proteins or peptides for the detection of anti-HCV antibodies. The high variability displayed by NS4 and NS5 in particular would be expected to result in a relatively low degree of serological cross-reactivity between HCV types. Indeed, recent evidence has revealed the existence of type-specific antibody responses. Several studies have shown that blood donors infected with HCV types 2, 3 or 4 react

### Table 2. Identity of sequenced regions between HCV isolates

<table>
<thead>
<tr>
<th>Region</th>
<th>Amino acid sequence identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Within a subtype</td>
</tr>
<tr>
<td>E1</td>
<td>88–99</td>
</tr>
<tr>
<td>E2/NS1</td>
<td>74–95</td>
</tr>
<tr>
<td>NS4</td>
<td>93–99</td>
</tr>
<tr>
<td>NS5</td>
<td>91–100</td>
</tr>
</tbody>
</table>

The sequence identity between isolates of the same subtype and amino acid substitutions were often conservative. This was particularly the case in the NS4 and NS5 regions where amino acid sequence identity within a subtype always exceeded 90% (Table 2). Significantly more variation was found in the E1 region representing the N-terminal half of this molecule, particularly within residues 216–257. Notably, one isolate (T145) had a four amino acid deletion within E1 (residues 298–301). The most striking intratypic variation was found in E2/NS1 which contained two hypervariable regions (HVRs), HVR1 at the N terminus (residues 384–410) and HVR2 (residues 475–480). HVR1 exhibited only 26–89% amino acid identity, such that no two isolates, even of the same subtype, were identical within this region. Amino acid sequence variation within the variable, and particularly the hypervariable domains of the envelope regions, is most likely the result of selection imposed by the immune response (Weiner et al., 1992). That the envelope glycoproteins of HCV appear to be subject to a positive selection for changes is indicated by the high rate of nonsynonymous mutations found in this region (Ogata et al., 1991; Kurosaki et al., 1993).

A striking segregation of amino acid sequences was evident on comparison of the different HCV genotypes. The extent of identity between HCV types was dependent on the region used for comparison (Table 2). Invariant residues were more common in the less variable NS4 and NS5 regions, comprising 53/138 and 72/140 residues respectively, as compared with the E1 (39/128) or E2/NS1 regions (43/135). The NS5 region, although displaying considerable variability, was the most conserved of the four regions analysed with 72–82% amino acid identity between types. E1 was the least conserved (51–76% amino acid identity between types). Genotype-specific residues, which were highly conserved within the same HCV subtype, were found to cluster and intersperse throughout all four regions sequenced. Most notable was a relatively conserved type-specific motif within E1 (residues 248–258) that alone could predict the division of all isolates into their respective genotypes. The significance of these regions is not yet known but they would be expected to account for any biological differences between genotypes. The degree of sequence variation implies that the various HCV types diverged quite some time ago, and is similar to that found for serotypes of various flaviviruses (Chambers et al., 1990).

The sequence differences between HCV types might be significant for HCV screening given the present high degree of dependence on serological methods. Current anti-HCV assays commonly employ type 1a core, NS3, NS4 and NS5 antigens as recombinant proteins or peptides for the detection of anti-HCV antibodies. The high variability displayed by NS4 and NS5 in particular would be expected to result in a relatively low degree of serological cross-reactivity between HCV types. Indeed, recent evidence has revealed the existence of type-specific antibody responses. Several studies have shown that blood donors infected with HCV types 2, 3 or 4 react
relatively infrequently to NS4-derived commercial antigens (McOmish et al., 1993; Stuyver et al., 1993b). In a more detailed study, type-specific antibody was directly demonstrated and found to form a substantial proportion of total anti-NS4 reactivity within each serum sample (Simmonds et al., 1993b). Together, these results demonstrate that the extent of amino acid variation between HCV types may be sufficient to alter antigenicity and thus detection by immunoassay. The development of assays equally sensitive for the various HCV genotypes is therefore required and would have the added advantage of enabling serotyping. Impetus for an inexpensive and rapid means of determining HCV genotype has come from studies which indicate that HCV genotype may be related to clinical or biological characteristics, including severity of liver damage and drug sensitivity. Genotype 1b, in particular, has been associated with more advanced disease in infected patients (Pozzato et al., 1991; Takada et al., 1992; Dusheiko et al., 1994) and is less responsive to therapy with interferon-α (Pozzato et al., 1991; Kanai et al., 1992; Yoshioka et al., 1992).

In summary, the results of this study have confirmed and extended observations that genotype 1b, the major HCV type found worldwide, also predominates in east Asia. In addition, the significant prevalence of types 2 and 3 and the discovery of yet another HCV subtype (3c) further emphasizes the need to take variant genotypes into account, both for reliable serological diagnosis as well as for the future development of an effective vaccine.

We gratefully acknowledge Associate Professor R. Guan (National University of Singapore), Professor C. Wasi (WHO Collaborating Centre on AIDS, Mahidol University), Professor A. Sulaiman (University of Indonesia), Dr J. Lao (Cebu Doctors/Chong Hua Hospitals) and Korea Green Cross Corporation for the collection of sera. We also thank Academic Associate Professor R. Guan (National University of Singapore) for proofreading, Ms X. H. Yu and Ms G. L. Goh for excellent technical assistance and Ms A. Ahmad for preparation of the manuscript. This work was supported in part by a grant from the Singapore National Science and Technology Board (NSTB/16/4/4).

References


Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Lizuka, H.,


(Received 14 July 1994; Accepted 21 September 1994)