<table>
<thead>
<tr>
<th>Title</th>
<th>GENOTYPE IDENTIFICATION OF HEPATITIS C VIRUS (HCV) ISOLATED FROM A SINGLE JAPANESE CARRIER IN NAGASAKI PREFECTURE AND GENOME ANALYSIS OF E1 AND E2/NS1 ENVELOPE GLYCOPROTEIN REGIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Zheng, Wei-Yun</td>
</tr>
<tr>
<td>Citation</td>
<td>日本熱帯医学会雑誌, vol.22(4), pp.169-177; 1994</td>
</tr>
<tr>
<td>Issue Date</td>
<td>1994-12-15</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10069/22414">http://hdl.handle.net/10069/22414</a></td>
</tr>
<tr>
<td>Right</td>
<td>Japanese Society of Tropical Medicine</td>
</tr>
</tbody>
</table>
GENOTYPE IDENTIFICATION OF HEPATITIS C VIRUS (HCV) ISOLATED FROM A SINGLE JAPANESE CARRIER IN NAGASAKI PREFECTURE AND GENOME ANALYSIS OF E1 AND E2/NS1 ENVELOPE GLYCOPROTEIN REGIONS

WEI-YUN ZHENG

Abstract: The nucleotide (nt) sequence of a Hepatitis C virus genome (HCV-N) which was derived from a single Japanese patient's serum in Nagasaki Prefecture has been determined by multiple clones covering 22 overlapping regions of the HCV genome. The sequenced region consisted of 9295 nt, including 248 nt of 5'-untranslated region (UTR), a single large open reading frame (ORF) encoding a polyprotein of 3010 amino acids (aa) and a 17 nt of 3'-UTR. Phylogenetic analysis indicated that HCV-N belongs to II/1b genotype of group 1. Two other Nagasaki HCV strains (HCV-N1 and HCV-N2) were also sequenced in the E1 and N-terminus of the E2/NS1 regions. Two hypervariable regions (HVR1 and HVR2) were found in the N-terminus of E2/NS1 region among 3 Nagasaki strains and 7 other HCV strains with published sequences. Two well-conserved aa sequences were also identified among 10 HCV strains in the E1 and N-terminus of the E2/NS1 regions. The results will be useful for future understanding on the pathogenesis, virological diagnosis and development of vaccine for HCV.

INTRODUCTION

Hepatitis C virus (HCV) is the major cause of nonA-nonB (NANB) hepatitis, and chronic infection with HCV has been linked to the development of liver cirrhosis and hepatocellular carcinoma (Plagemann, 1991). Many entire and partial HCV sequence informations published so far have shown that HCV genome consisted of about 10 Kb single-stranded positive-sense RNA with 5'UTR, a single long ORF followed by 3'UTR. The viral genome organization resembles that of the flaviviruses and pestiviruses (Kato et al., 1990; Plagemann, 1991).

Since Choo et al. (1989) cloned the genome of an RNA virus from the plasma of a chimpanzee inoculated with plasma from patient with NANB hepatitis and designated it as HCV, entire sequences have been reported for at least 15 HCV strains. There are: HCV-J (Kato et al., 1990), HCV-1 (Choo et al., 1991), HCV-H (Inchauspe et al., 1991), HCV-BK (Takamizawa et al., 1991), HC-J6 (Okamoto et al., 1991), HC-J1 (Okamoto et al., 1992a), HC-J8 (Okamoto et al., 1992b), HC-J4/83, HC-J4/91 (Okamoto et al., 1992c), HCV-JT (Tana-ka et al., 1992), HCV-T (Chen et al., 1992), HCV-JK1 (Honda et al., 1993), HC-C2 (Wang et al., 1993), HC-G9 (Okamoto et al., 1994) and NZL1 (Sakamoto et al., 1994). Depending on the HCV sequence similarity, HCV genome can be classified into 6 genotypes: I, II, III, IV, V and VI as reviewed by Sakamoto et al. (1994), or into 6 major genotypes: 1 (a, b, c), 2 (a, b, c), 3 (a, b), 4, 5 and 6 (Simmonds et al., 1993b).

Recently, the newest classification has been summarized by Sakamoto et al. (1994) who classified HCV genome into 3 major groups with entire published sequences. The entire HCV sequence of Japanese strains have been assigned either to group 1 (I/1a genotype: HC-J1, II/1b genotype: HCV-J, HCV-BK, HCV-J4/83, HC-J4/91, HCV-JT and HCV-JK1) or group 2 (III/2a genotype: HC-J6; IV/2b genotype: HCV-J8) according to Sakamoto et al. (1994). In order to know whether there is particular HCV in the local area of Nagasaki, Japan, HCV patient serum which came from Nagasaki Prefecture was obtained and the genome sequence of the HCV-N in this serum was determined by multiple overlapping clones.
In order to see the genetic variability of HCV in the Nagasaki area, HCV genome sequence in 2 more HCV patient sera in Nagasaki Prefecture (HCV-N1 and HCV-N2) were also analyzed similarly in the E1 and N-terminus of the E2/NS1 envelope glycoprotein regions. Two hypervariable regions have been found among these 3 Nagasaki HCV strains using the sequence diversity comparison with published sequence data of 7 HCV patient sera in Nagasaki.

**RESULTS**

**The genome characterization of HCV-N**

Multiple cDNA clones were isolated from altogether 22 overlapping regions which covered almost entire genome of HCV-N, except extreme 5' and 3' terminals (Fig. 1). For each of the overlapping region, 3 independent cDNA clones were isolated from the RT-PCR product and sequenced. The cleavage site of the polyprotein coded by the ORF of the HCV-N genome was assigned according to the publications by Okamoto et al. (1992b) and Honda et al. (1993). Total length of the sequenced region of HCV-N consisted of 9295 nt and 3010 aa, respectively (Fig. 2). The sequenced region can be divided into 248 nt in the 5'UTR, 9030 nt in a single ORF and 17 nt in the 3'UTR, respectively. The ORF was considered to be translated into C (191 aa), E1 (192 aa), E2/NS1 (346 aa), NS2 (277 aa), NS3 (609 aa), NS4 (398 aa) and NS5 (997 aa) proteins.

**Genotype classification and genome homology comparison among HCV-N and 7 other HCV strains with published sequences**
Entire sequence information have been published for the following 7 HCV strains: HCV-1 (Choo et al., 1991), HCV-BK (Takamizawa et al., 1991), HCV-J (Kato et al., 1990), HC-J6 (Okamoto et al., 1991), HC-J8 (Okamoto et al., 1992b), HCV-JT (Tanaka et al., 1992) and HCV-T (Chen et al., 1992). Homology comparison between HCV-N and these 7 HCV strains was shown in Table 1. The most conserved region is the 5' UTR which showed nt homology more than 91.5% among 8 HCV isolates. The sequence homology of the C protein region was also highly conserved (nt > 81.1%, aa > 88.5%). In the putative E1 and E2/NS1 regions, HCV-N showed high homology with HCV-BK, HCV-J, HCV-JT and HCV-T (nt > 84.5%, aa > 85.8%). Whereas HCV-N showed low homology with HC-J6 and HC-J8 (nt < 68.8%, aa < 72.3%). Nt sequence homology between HCV-N and HCV-1 in the E1 and E2/NS1 regions was 73.7%, 74.5% and aa sequence homology was 76.0%, 81.2%, respectively. From NS2 to NS5 regions, it was evident that HCV-N is closer to HCV-BK, HCV-J, HCV-JT and HCV-T and remote from HC-J6 and HC-J8, while HCV-1 seems to be located intermediate. The total sequence homology comparison among 8 HCV isolates gave us a conclusion that HCV-N belongs to HCV-BK/HCV-J/HCV-JT/HCV-T genotype. In contrast, HCV-N is remote from J6 and J8 genotypes.

Table 1  Homology comparison of nucleotide and deduced amino acid sequence among 8 HCV strains (HCV-N, HCV-1, HCV-BK, HCV-J, HCV-JT, HCV-T, HC-J6 and HC-J8). The homology is indicated by %. The nucleotide sequence homology is shown in upper and amino acid sequence homology is shown in lower with parenthesis, respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>77.9 (84.5)</td>
<td>92.0 (94.1)</td>
<td>91.5 (93.7)</td>
<td>62.4 (70.7)</td>
<td>61.7 (69.2)</td>
<td>91.7 (93.9)</td>
<td>90.0 (92.1)</td>
</tr>
<tr>
<td>5'UTR</td>
<td>98.8</td>
<td>99.2</td>
<td>98.4</td>
<td>94.0</td>
<td>91.5</td>
<td>99.2</td>
<td>99.2</td>
</tr>
<tr>
<td>C</td>
<td>90.7 (96.3)</td>
<td>96.0 (97.4)</td>
<td>95.8 (97.9)</td>
<td>83.0 (90.1)</td>
<td>81.1 (88.5)</td>
<td>95.1 (96.3)</td>
<td>93.9 (93.7)</td>
</tr>
<tr>
<td>E1</td>
<td>73.7 (76.0)</td>
<td>89.0 (89.6)</td>
<td>91.7 (93.2)</td>
<td>60.1 (52.9)</td>
<td>56.7 (51.8)</td>
<td>92.2 (92.2)</td>
<td>91.1 (93.8)</td>
</tr>
<tr>
<td>E2/NS1</td>
<td>74.5 (81.2)</td>
<td>85.7 (88.1)</td>
<td>84.5 (86.1)</td>
<td>68.8 (71.1)</td>
<td>66.8 (72.3)</td>
<td>84.7 (85.8)</td>
<td>85.3 (87.0)</td>
</tr>
<tr>
<td>NS2</td>
<td>71.0 (74.4)</td>
<td>92.1 (91.7)</td>
<td>92.7 (94.9)</td>
<td>59.8 (58.7)</td>
<td>58.2 (56.2)</td>
<td>89.0 (92.4)</td>
<td>87.8 (91.3)</td>
</tr>
<tr>
<td>NS3</td>
<td>78.1 (89.0)</td>
<td>90.7 (93.4)</td>
<td>90.4 (91.8)</td>
<td>69.9 (79.3)</td>
<td>69.8 (78.8)</td>
<td>91.1 (94.9)</td>
<td>89.4 (92.6)</td>
</tr>
<tr>
<td>NS4</td>
<td>78.9 (88.2)</td>
<td>93.1 (97.5)</td>
<td>92.7 (97.2)</td>
<td>66.8 (74.1)</td>
<td>67.0 (72.9)</td>
<td>93.5 (97.0)</td>
<td>89.9 (94.0)</td>
</tr>
<tr>
<td>NS5</td>
<td>78.8 (83.7)</td>
<td>93.5 (96.3)</td>
<td>92.7 (95.2)</td>
<td>59.5 (71.8)</td>
<td>59.2 (71.5)</td>
<td>93.1 (95.2)</td>
<td>91.8 (92.5)</td>
</tr>
</tbody>
</table>
Figure 2  Nucleotide and deduced amino acid sequence of HCV-N strain. The nucleotide sequence of HCV-N is shown in the upper line and deduced amino acid sequence of HCV-N is shown in the lower line, respectively.
Figure 2

respectively. The sequence homology of HCV-N also showed its closer relationship to HCV-1 than to HC-J6 and HC-J8 isolates (Table 1). HC-J8 possessed 23\% divergence with HC-J6 and has been clearly identified as a separate genotype from HC-J6 (Okamoto et al., 1992b). In order to confirm this conclusion, phylogenetic trees were constructed base on nucleotide

Figure 3 Phylogenetic trees are based on nucleotide divergence (%). Sequence of all 8 HCV strains were compared and analyzed by the nearest neighbor method.
divergence throughout entire sequence using nearest neighbor method (Williams and Lance 1977) (Fig. 3). This result indicated that 8 HCV strains can be classified into 2 groups: the group 1 contains genotype II/1b (HCV-N, HCV-BK, HCV-J, HCV-JT, HCV-T) and genotype I/1a (HCV-1); while group 2 contains genotype III/2a (HC-J6) and genotype IV/2b (HC-J8).

Sequence variation in the E1 and N-terminus of E2/NS1 envelope glycoprotein regions among 3 HCV Nagasaki strains and 7 other HCV strains with published sequences

Recently, many papers reported that hypervariable regions existed in the N-terminus of E2/NS1 envelope glycoprotein region of HCV genome (Hijikata et al., 1991; Weiner et al., 1991; Honda et al., 1993). In order to know the HCV sequence diversity of the envelope glycoprotein in local area of Nagasaki, 2 other HCV Nagasaki strains (HCV-N1 and HCV-N2) were also cloned and sequenced for the region of nt 623-1988, aa 126-580 (nt and aa base on the HCV-N number). Alignment comparison of deduced aa sequence among 3 Nagasaki strains and other 7 HCV strains with published sequences were shown in Fig. 4. Two hypervariable regions (HVR) were discovered in the N-terminus of E2/NS1 region. HVR 1 (aa 384-411) was located directly downstream at the beginning of the E2/NS1, whereas HVR 2 (aa 475-480) was observed 64 aa downstream from the HVR 1. In the HVR 1 consisting of 28 aa, 10 aa residues were well-conserved among 3 Nagasaki strains. Whereas only 3 aa were conserved in this

**Figure 4** Alignment of amino acid sequence among 10 HCV strains (HCV-N, HCV-N1, HCV-N2, HCV-1, HCV-BK, HCV-J, HCV-JT, HCV-T, HC-J6 and HC-J8) in the E1 and N-terminus of E2/NS1 regions (aa 126-580). Two hypervariable regions (HVR 1 and HVR 2) are shown by boxes. Potential N-glycosylation sites (N-X-S/T) are underlined.
Table 2  HCV genotype classification including HCV-N from this study and 15 entire HCV strains which have been reported by Sakamoto et al. (1994).

<table>
<thead>
<tr>
<th>Group</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>I/1a Ic</td>
<td>II/1b II/2a</td>
<td>V/3a</td>
</tr>
<tr>
<td>Geographic Area</td>
<td>U.S.A. Japan</td>
<td>Indonesian Japan China Taiwan</td>
<td>Japan Japan New Zealand</td>
</tr>
</tbody>
</table>

region when a total 10 HCV strains were compared. In the HVR 2 consisting of 6 aa, none of the aa were conserved among 3 Nagasaki strains. This hypervariability was also observed when a total 10 HCV strains were compared simultaneously. In the E1 and N-terminus of the E2/NS1 regions, 2 well conserved aa sequences were observed among 10 HCV strains: G-H-R-M-A-W-D-M-M (aa 315-323) and W-F-G-C-T-W-M-N (aa 549-556). Although 14 N-glycosylation sites (N-X-S/T) were identified in HCV-N, one of them (NNS at aa No. 475-477) was unique to this strain and not present in 2 other Nagasaki strains and 7 other HCV strains with published sequences. Instead, HCV-N1 and HCV-N2 possessed another N-glycosylation site (NFS or NSS at aa No. 250-252). This glycosylation site is found also in HCV-BK, HCV-J, HCV-JT, and HCV-T (Fig. 4). Altogether, only 9 N-glycosylation sites were conserved in the E1 and N-terminal of the E2/NS1 regions among 10 HCV strains shown in Fig. 4.

**DISCUSSION**

Relationship between genotype and geographic area of HCV strains including HCV-N was shown by their sequence comparison using the classification by Sakamoto et al. (1994) (Table 2). The sequence homology indicated that HCV-N strain did not possess its local genotypic character, and similar to HCV-BK, HCV-J and HCV-JT strain which were isolated from other areas of Japan. All these strains apparently belong to the same genotype II/1b of group 1. This genotype also included HC-J4/83; HC-J4/91; HCV-JK1 of Japanese HCV strains, HC-C2 of Chinese HCV strain and HCV-T of Taiwan strain. The genotype 1c of group 1 includes HC-G9 which was derived from Indonesian strain. While, the genotype I/1a of group 1 includes HCV-1 and HCV-H which were derived from American strains. Although another HC-J1 was Japanese strain, it was supposed to have originated from the US, because HC-J1 was derived from a Japanese haemophiliac who developed hepatitis C after receiving US-made factor VIII (Okamoto et al., 1992a). In group 2, HC-J6 and HC-J8 belong to the III/2a and the IV/2b genotype, respectively. Both of them were Japanese strains. In group 3, NZLI that came from New Zealand belongs to V/3a genotype (Table 2). Simmonds et al. (1993b) have classified 6 major genotypes of HCV from 76 HCV isolates, which were almost worldwide collection, using phylogenetic analysis of the NS5 region. The 76 HCV isolates contained entire sequence of HCV-1, HCV-H, HCV-J, HCV-BK, HCV-T, HCV-JT, HC-J6 and HC-J8 strains and other partial sequence of HCV strains. From this classification, only genotype 4, 5 and 6 showed highly restricted geographical distributions, being apparently confined to Egypt, South Africa and Hong Kong respectively (Simmonds et al., 1993b).

Two hypervariable regions have already been observed in the N-terminal of the E2/NS1 region. The number of conserved aa in the HVR 1 among 3 Nagasaki strains were higher than among a total 10 strains including 3 Nagasaki strains and 7 other HCV strains with published sequences. The numbers of conserved aa (5 aa/28 aa) among HCV strains of genogroup 1 (HCV-N/HCV-1/HCV-BK/HCV-J/HCV-JT/HCV-T) were higher than those (3 aa/28 aa) among 10 HCV strains which included genogroup 2 (HC-J6 and HC-J8). This result may give us an idea that aa conservation in the HVR 1 of the same genotype or genogroup of HCV are
higher than among the different genotype or genogroup of HCV strains. Therefore, the 3 Nagasaki HCV strains may have been originated from the same ancestor.

In contrast, none of the aa was conserved among 3 Nagasaki strains nor among a total 10 HCV strains in the HVR 2 which showed higher variability than the HVR 1. The reason why mutations occur so frequently in only limited regions such as HVR is not known (Tanaka et al., 1992). The higher degree of divergence in HCV E2/NS1 region might reflect the immune selection and suggest that this region cannot probably be an ideal target for future vaccine development.

Regarding the N-glycosylation sites in the E1 and N-terminus of the E2/NS1 region, only 9 sites were conserved out of the 14 sites seen in HCV-N. Some of the glycosylation sites, therefore, would not be essential for the survival, transmission and maintenance of HCV in nature. On the other hand, 2 well-conserved aa sequences were found in the E1 and N-terminus of the E2/NS1 regions among the 10 HCV strains compared. These conserved sequences may be better targets of vaccine development if they were related with protective immunity.

The nt sequence homology of HCV-N in the 5'UTR shows high conservation comparing with other 7 HCV isolates (nt > 91.5%) and was used as an ideal target for PCR amplification to detect HCV RNA (Okamoto et al., 1990). But recently, some papers reported that several HCV strains have sequence variation in the 5'UTR (Lee et al., 1992; Bukh et al., 1992; Simmonds et al., 1993a). Therefore additional informations would be required to select optimal primer for PCR diagnosis on HCV. The functional motifs of the putative encoded proteins of HCV (Plagemann et al., 1991; Tanaka et al., 1992) have also been found in HCV-N. There are consensus sequences of RNA helicase in the NS3 region: Gly-Ser-Gly-Lys-Ser-Thr (aa 1233-1238) and Gln-Arg-Gly-Arg-Arg-Thr (aa 1486-1493), while NS5 region possessed consensus sequence of RNA-dependent RNA polymerase: Gly-Asp-Asp (aa 2736-2738). The sequence heterogeneity in putative structural proteins and nostructural proteins of HCV provided significant evidence for genotype classification. The genetic informations of HCV-N genome obtained in this study will be useful in future understanding on the pathogenesis, diagnosis and development of vaccine for HCV.

ACKNOWLEDGEMENTS

The author would like to express thanks to Prof. A. Igarashi for valuable advice and encouragement, careful reading and comments of the manuscript.

Special thanks are due to Dr. K. Morita, Dr. S. Kurihara, Dr. K.Z. Thant and Dr. Y. Tsuda for useful suggestion and kindly assistance.

This study was supported by Health Science Research Grants (Non A-NonB Hepatitis Research Grants) from the Ministry of Health and Welfare of Japan.

REFERENCES