Comparative Nucleotide and Amino Acid Sequences Among Seven Hepatitis C Virus cDNA Isolates in Japan and U.S.A.

Zheng, Wei-Yun; Kurihara, Shintaro; Morita, Kouichi; Tanaka, Mariko; Igarashi, Akira

熱帯医学 Tropical medicine 34(3). p105-112, 1992
Comparative Nucleotide and Amino Acid Sequences Among Seven Hepatitis C Virus cDNA Isolates in Japan and U.S.A.

Wei-Yun Zheng, Shintaro Kurihara, Kouichi Morita, Mariko Tanaka and Akira Igarashi

Department of Virology, Institute of Tropical Medicine, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852, Japan

Abstract: The nucleotide and deduced amino acid sequence homology of hepatitis C virus (HCV) among seven HCV cDNA isolates in Japan and U.S.A. have been compared and discussed. It was found that Okamoto's nucleotide sequence (J1) has high homology (>94.1%) with American (CHIRON) in both nucleotide and amino acid sequence. Kato's nucleotide sequence (JGC) was close to Takamizawa's nucleotide sequence (JBK), with similarity of >87.0% for both nucleotide and amino acid. CHIRON was closer to JGC and JBK than to Okamoto's nucleotide sequence (J6). J6 was different from each of the other six isolates. The high degree of sequence conservation in the 5' non-coding region among six isolates (>92.8%) except Takeuchi's nucleotide sequence (JYK) has been observed. The low level of nucleotide and amino acid sequence variation in the putative core protein and the variability of putative envelope glycoprotein among seven isolates have been observed.

Key words: Hepatitis C virus, Nucleotide and amino acid sequence homology

INTRODUCTION

Hepatitis C virus (HCV) is supposed to be an enveloped virus with about 10 Kb positive-sense, single-stranded RNA genome. Its genome organization is predicted to resemble that of the flavivirus and pestivirus, therefore classified as a genus in the family Flaviviridae (Wengler, 1991). A 5' non-coding segment of 341 nucleotides precedes a continuous open reading frame (ORF) of about 9030 nucleotides which is followed by a 54 nucleotides long 3' non-coding region. Genomic RNA is probably translated into a single polyprotein of about 3010 amino acids which is processed into functional proteins (Plagemann, 1991).

Recently, isolations of several HCV cDNA clones from human patient's sera and HCV infected chimpanzee have been reported in the U.S.A. and Japan as follows: CHIRON group (Choo et al., 1991) entire sequence of CHIRON 9,379 Kb; Kato's group (1990) entire sequence of JGC 9,413 Kb; Takeuchi's group (1990) partial sequence of JYK 1,323 Kb; Okamoto's group partial sequence of J1 and J4 1,863 Kb (Okamoto et al., 1990a) and entire sequence of J6
9,589 Kb (Okamoto et al., 1991); and Takamizawa’s group (1991) entire sequence of JBK 9,416 Kb, respectively. It is valuable and feasible to compare the homology of HCV cDNA sequences among these seven isolates, although some of them have been compared partly before (Kato et al., 1990; Okamoto et al., 1991; Takamizawa et al., 1991).

In this paper, significance of the genome variabilities among seven isolates of HCV will be discussed in order to identify and analyze the functions of HCV genome, and to understand immunology and epidemiology of HCV.

**MATERIALS AND METHODS**

**Information of HCV Nucleotide Sequence:** Nucleotide sequence data of seven HCV cDNA isolates were obtained from following publications: CHIRON (Choo et al., 1991), JGC (Kato et al., 1990), JBK (Takamizawa et al., 1991), J6 (Okamoto et al., 1991), J1 and J4 (Okamoto et al., 1990a), and JYK (Takeuchi et al., 1990).

**Comparison of Nucleotide and Amino Acid Sequences Among Seven HCV cDNA Isolates:** According to the method reported by Okamoto et al. (1991), the HCV nucleotide sequence was divided into following genes: 5’non-coding region, core (C) protein, envelope (E) protein, NS1-NS5 protein regions and 3’non-coding region.

The homology comparison of HCV nucleotide and deduced amino acid sequence were performed by using DNASIS Version 7.00 computer programing software (HITACHI, 1991).

**RESULTS**

Figure 1a shows the 5’ non-coding region nucleotide sequence homology of six HCV cDNA isolates. The homology among all compared isolates was very high (>92.8%) except JYK. The homology search indicated that JYK (1,323 Kb) nucleotide sequence start from putative core protein region. The putative C protein showed high conservation among all compared HCV cDNA isolates (nucleotide >80.9%, amino acid >90.6%; Fig. 1b). From putative E to NS1-NS5 protein region, J6 showed low degree homology in nucleotide and amino acid sequences compared with other isolates (nucleotide <69.5%, amino acid <73.6%) except NS3 region, which showed slightly higher homology (nucleotide 70.0%, amino acid 80.3%; Figs. 1, 2 and 3). The J6 showed very low percentage of nucleotide sequence homology especially in 3’ non-coding region (<42.9%), while the conservation between the JGC and JBK was very high (92.3%; Fig. 3C).

The nucleotide and amino acid sequence conservation of E protein was lower than C protein region among all isolates especially in J6 which was the lowest. In E protein region, the J4 is closely related to JYK, JGC and JBK (nucleotide >90.9%, amino acid >91.4%), whereas only 74.5% (nucleotide average count) and 77.3% (amino acid average count) homologies were conserved to J1 and CHIRON. The J1 is closely related to CHIRON (nucleotide 94.1%, amino acid 97.0%, whereas only 73.4% (nucleotide average count) and 76.8% (amino acid average count) homologies were conserved to JYK, JGC and JBK. The J6...
Fig. 1. Nucleotide sequence homology of 5' non-coding region (a), and nucleotide and amino acid sequence homology of C protein (b) and E protein (c) regions. Figures in the left lower half show nucleotide sequence homology and those in the right upper half show amino acid sequence homology, respectively, represented in homology percent.
NS1 protein region (a)

<table>
<thead>
<tr>
<th></th>
<th>J6</th>
<th>CHIRON</th>
<th>JGC</th>
<th>JBK</th>
</tr>
</thead>
<tbody>
<tr>
<td>J6</td>
<td>69.4</td>
<td>69.4</td>
<td>70.3</td>
<td>72.6</td>
</tr>
<tr>
<td>CHIRON</td>
<td>66.5</td>
<td>77.9</td>
<td>80.0</td>
<td></td>
</tr>
<tr>
<td>JGC</td>
<td>69.5</td>
<td>73.5</td>
<td></td>
<td>88.5</td>
</tr>
<tr>
<td>JBK</td>
<td>69.4</td>
<td>72.0</td>
<td>87.0</td>
<td></td>
</tr>
</tbody>
</table>

NS2 protein region (b)

<table>
<thead>
<tr>
<th></th>
<th>J6</th>
<th>CHIRON</th>
<th>JGC</th>
<th>JBK</th>
</tr>
</thead>
<tbody>
<tr>
<td>J6</td>
<td>57.8</td>
<td>57.8</td>
<td>60.6</td>
<td>59.6</td>
</tr>
<tr>
<td>CHIRON</td>
<td>56.7</td>
<td>77.3</td>
<td>76.2</td>
<td></td>
</tr>
<tr>
<td>JGC</td>
<td>59.6</td>
<td>71.0</td>
<td></td>
<td>92.4</td>
</tr>
<tr>
<td>JBK</td>
<td>59.4</td>
<td>71.1</td>
<td>91.4</td>
<td></td>
</tr>
</tbody>
</table>

NS3 protein region (c)

<table>
<thead>
<tr>
<th></th>
<th>J6</th>
<th>CHIRON</th>
<th>JGC</th>
<th>JBK</th>
</tr>
</thead>
<tbody>
<tr>
<td>J6</td>
<td>80.3</td>
<td>80.3</td>
<td>80.3</td>
<td>80.3</td>
</tr>
<tr>
<td>CHIRON</td>
<td>69.9</td>
<td>91.5</td>
<td></td>
<td>91.8</td>
</tr>
<tr>
<td>JGC</td>
<td>70.1</td>
<td>79.8</td>
<td></td>
<td>95.1</td>
</tr>
<tr>
<td>JBK</td>
<td>70.0</td>
<td>78.6</td>
<td>90.6</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Nucleotide and amino acid sequence homology of NS1 protein (a), NS2 protein (b) and NS3 protein (c) regions. Figures in the left lower half show nucleotide sequence homology and those in the right upper half show amino acid sequence homology, respectively, represented in homology percent.
Fig. 3. Nucleotide and amino acid sequence homology of NS4 protein (a) and NS5 protein (b) regions, and nucleotide sequence homology of 3' non-coding region (c). Figures in the left lower half show nucleotide sequence homology and those in the right upper half show amino acid sequence homology, respectively, represent by the homology percent.
has a highest degree of variation among seven isolates (nucleotide <60.8%, amino acid <53.3%; Fig. 1C). Homologies from NS1 to 3' non-coding region also showed moderate variable and hypervariable regions among each four isolates (Figs. 2 and 3).

**DISCUSSION**

According to the homology comparison, J1 showed that its nucleotide and amino acid sequence has high homology (more than 94.1%) with CHIRON. Therefore, J1 should belong to the CHIRON subgroup HCV. (Kato et al., 1990). J6 was different from each of the other three isolates (CHIRON, JGC and JBK) in entire sequence (Okamoto et al., 1991). JGC was close to JBK, similar in >87.0% for both nucleotide and amino acid sequence in entire sequence. J4 and JYK were also close to JBK and JGC (nucleotide >85.1%, amino acid >90.9%; Figs. 1, 2 and 3). Our homology comparison resembles with those by Okamoto et al. (1991) and Takamizawa et al. (1991). On the basis of sequence similarity therefore, there would be at least three subtypes of HCV genome. Recently, Okamoto et al. (1992) indicated that the HCV genome sequences can be classified into four genotypes: CHIRON and J1 as genotype 1; JGC, JBK, JYK and J4 as genotype 2; J6 as genotype 3; J8 as genotype 4 (this sequence was not included in our sequence comparison). Our homology comparison gave the result compatible with this classification in general.

Due to the high degree of nucleotide and amino acid sequence conservation in the 5' non-coding region among compared six isolates (>92.8%) except JYK, primers deduced from this region could be useful for the specific detection of HCV RNA by reverse transcription polymerase chain reaction (RT-PCR). HCV antibody detection system using core protein antigen have already been developed and available as a commercial assay kit (Abbot, U.S.A.; Ortho Diagnostic Systems, Tokyo, Japan). According to the low level of nucleotide and amino acid sequence variation in the core protein region, the products have been used to detect HCV antibody by the ELISA (Okamoto et al., 1990b, Chiba et al., 1991) and by western blotting method for early diagnosis of HCV infection. Using core protein antigen for HCV antibody detection is more sensitive than C-100 antigen which was derived from NS3 and NS4 protein region (Harada et al., 1991). The significance of the apparent nucleotide and amino acid sequence variabilities in the E protein region is unclear. It could be a consequence of immune selection and needs to be considered in the future development of HCV vaccines (Plagemann, 1991).

Because its genome organization and predicted virion structure closely resemble those of the flaviviruses and pestiviruses, HCV has been proposed to be classified in the Family Flaviviridae as a new third genus (Plagemann, 1991; Wengler, 1991). The nucleotide and amino acid sequence variation of the E protein region indicated the possibility of multiple subtypes assay. Recently, researcher in France and Thailand also published partial sequences of HCV (Kremsdorf et al., 1991; Mori et al., 1992). Each HCV subtype would be expected to have distinct epidemiological distribution, and this phenomenon would influence the choice of materials best suited for diagnosis and immunoprophylaxis in the geographical area at issue.
The authors are now determining the sequence of HCV cDNA which has been derived from the HCV infected patient in Nagasaki Prefecture in order to clarify the local strain of HCV in Japan.

ACKNOWLEDGMENTS

This study was supported by a Grant-in Aid for Non A-non B Hepatitis Study from Ministry of Health and Welfare of Japan. The first author was supported by Monbusho Scholarship of Japan.

REFERENCES


