Genomic characterization and mutation rate of hepatitis C virus isolated from a patient who contracted hepatitis during an epidemic of non-A, non-B hepatitis in Japan

Kenji Abe,1 Genevieve Inchauspe2 and Kiyoshi Fujisawa3

1Department of Pathology, National Institute of Health, 4-7-1, Gakuen Musashimurayama-shi, Tokyo 208, Japan, 2Laboratory of Virology and Parasitology, The Lindsley F. Kimball Research Institute, The New York Blood Center, New York, New York 10021, U.S.A. and 3Department of Internal Medicine, Jikei University School of Medicine, Tokyo 105, Japan

To investigate the genomic characterization of hepatitis C virus (HCV) isolated from a patient who contracted hepatitis during an epidemic of non-A, non-B (NANB) hepatitis in Shimizu city, Japan, we have cloned the nucleotide sequence of the viral genome (HCV-KF) spanning the structural domain. When compared to other previously reported HCV isolates, HCV-KF showed an overall identity at the amino acid level of 90.0 to 92.1% with Japanese isolates and 80.9 to 82.1% with American-like isolates. The HCV-KF genome displays an insertion of three nucleotides in-frame (corresponding to one amino acid) found at the junction between the E1 and E2/NS1 region. The mutation rate of the HCV-KF genome was assessed by comparing the nucleotide and deduced amino acid sequences of the viral RNA obtained from the serum of the original patient with viral sequences derived from the serum of a chimpanzee inoculated with the same serum 9 years previously. The substitution rate of the viral genome was estimated at 0.9 × 10⁻³ nucleotides per site per year for the HCV structural region. The highest mutation rate was found in the hypervariable region within the E2/NS1 domain. It is suggested that the outbreak in Shimizu city was caused by a strain of HCV closely related to the Japanese-like subgroup of isolates.

Hepatitis C virus (HCV) is the major cause of post-transfusion hepatitis throughout the world. The causative virus has recently been identified by cDNA cloning techniques using plasma from an experimentally infected chimpanzee (Choo et al., 1989). The virus contains a positive-strand RNA genome of about 9400 nucleotides which encodes a large polyprotein of at least 3000 amino acids (Kato et al., 1990). Analysis of sequences of various HCV cDNA clones obtained in the United States and in Japan reveals some similarities at the amino acid level and the structural level (analysis of the hydropathicity profile) with the non-structural proteins of flaviviruses, pestiviruses and some plant viruses (Miller & Purcell, 1990).

The main route of transmission of non-A, non-B (NANB) hepatitis virus is well documented as being parenteral (Kuo et al., 1989; Esteban et al., 1989). Sexual transmission and intrafamilial transmission have also been suggested (Hess et al., 1989; Kamitsukasa et al., 1989; Kiyosawa et al., 1991). However the route of transmission in more than 50% of the patients with community-acquired acute NANB hepatitis is unknown. In fact, HCV-specific antibody has been detected at high frequency among patients with NANB chronic liver disease with and without a history of transfusion (Kuo et al., 1989). These facts suggest that transmission of HCV other than via blood transfusion is also common.

A mass outbreak of epidemic hepatitis occurred in 1980 and 1981 in Shimizu city, Shizuoka Prefecture, Japan. Epidemiological and clinical studies have been reported by Yamauchi et al. (1983). We succeeded in transmitting the agent responsible for the epidemic of NANB hepatitis to chimpanzees, from patients who contracted hepatitis during the outbreak (Abe et al., 1986). A recent study suggested that this outbreak was caused by HCV, as serological markers for HCV were detected by anti-C100 ELISAs in the patient's serum (Fujisawa et al., 1990). However the route of infection is still unknown. To investigate in more detail the molecular nature of the infectious agent responsible for this epidemic of NANB hepatitis, we have cloned and sequenced most of the structural genes of the viral RNA
isolated from the serum of a patient who contracted hepatitis during the outbreak (isolate HCV-KF). We have also examined the mutation rate of the HCV-KF genome over a period of 9 years following transmission to the chimpanzee.

Two distinct sources of HCV-infected materials were used: human serum (denoted HCV-KF) taken during the acute phase of infection from a patient infected during the Shimizu outbreak, and a recipient chimpanzee's (no. J-59) serum obtained at 476 weeks post-infection (denoted HCV-476) of the chimpanzee with HCV-KF. Chimpanzee J-59 developed chronic persistent hepatitis (denoted HCV-476) of the chimpanzee with HCV-KF. The human patient had no history of blood transfusions, intravenous drug use or homosexual activity.

Methods for the extraction of RNA from serum, cDNA synthesis and amplification of HCV cDNA by nested polymerase chain reaction (PCR) were described previously (Abe et al., 1992). Primers for the PCR amplification were derived from the sequences of HC-J1 and HC-J4 (Okamoto et al., 1990), and HC-BK (Takamizawa et al., 1991). To determine the nucleotide and amino acid sequences of HCV from serum, PCR fragments were isolated and inserted into plasmid (Stratagene) as described previously (Ausubel et al., 1988), and sequenced using the dideoxynucleotide chain termination method in a DNA sequence analyser (Dupont). To avoid the possibility of misincorporation by Thermus aquaticus polymerase, we isolated at least three independent clones for each sequence amplified and the final consensus sequence was derived after comparison of the sequences derived from the three clones.

The derived consensus nucleotide sequences of the cDNAs for the isolates HCV-KF and HCV-476 are shown in Fig. 1. The percentage identities at the nucleotide and amino acid levels between the reported structural regions for HCV-KF (nucleotides 1 to 1557) and heterogeneous HCV isolates are shown in Table 1. Taken together these data indicate that the overall similarity between HCV-KF and Japanese isolates varies between 90-0 and 93-0% at both the nucleotide and amino acid levels. On the other hand, the overall similarity between HCV-KF and American isolates was approximately 78-1 to 78-9% at the nucleotide level and 80-9 to 82-1% at the amino acid level. Both isolates have an insertion of three nucleotides in-frame corresponding to one amino acid found at the junction of the E1 and E2/NS1 regions (nucleotides 1150 to 1152) when compared with the other reported HCV isolates.

When the HCV-KF and HCV-476 isolates were compared, a divergence of 12 nucleotides out of the 1560 nucleotides (0-8%) reported in this study was found and Table 1. Percentage nucleotide (nt) and amino acid (aa) identities of the structural regions of HCV-KF and heterologous HCV isolates

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Overall (1-1557)</th>
<th>C (1-573)</th>
<th>E1 (574-1149)</th>
<th>E2/NS1 (1150-1557)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nt</td>
<td>aa</td>
<td>nt</td>
<td>aa</td>
</tr>
<tr>
<td>Japanese</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HC-J4</td>
<td>92.6</td>
<td>90.6</td>
<td>96.7</td>
<td>96.9</td>
</tr>
<tr>
<td>HCV-J</td>
<td>90.5</td>
<td>90.8</td>
<td>96.0</td>
<td>96.3</td>
</tr>
<tr>
<td>HCV-BK</td>
<td>90.8</td>
<td>90.0</td>
<td>96.3</td>
<td>96.9</td>
</tr>
<tr>
<td>HCV-JH</td>
<td>93.0</td>
<td>92.1</td>
<td>95.8</td>
<td>96.3</td>
</tr>
<tr>
<td>American</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HC-J1</td>
<td>78.1</td>
<td>80.9</td>
<td>90.4</td>
<td>95.3</td>
</tr>
<tr>
<td>HCV-1</td>
<td>78.9</td>
<td>81.7</td>
<td>90.6</td>
<td>95.3</td>
</tr>
<tr>
<td>HCV-H</td>
<td>78.5</td>
<td>82.1</td>
<td>90.6</td>
<td>96.3</td>
</tr>
</tbody>
</table>

† Nucleotide positions deduced from HCV-1 (Houghton et al., 1991).
‡ The putative map position of the entire E2/NS1 region according to Houghton et al. (1991) is 1150 to 2250. Only a partial sequence is reported in our study.
chimpanzee at 50 weeks (denoted HCV-50) and 102 weeks (denoted HCV-102), respectively, after inoculation. As shown in Fig. 2, a point mutation was found in HCV-102 at one of the 78 nucleotides mentioned above, position, but this mutation was not seen in HCV-50.

Potential N-glycosylation (N-X-T/S) sites were found to be entirely conserved between the HCV-KF and HCV-476 isolates. In addition, there was perfect conservation of cysteine residues in the structural domains.

In this study, we cloned and examined the nucleotide sequences encompassing the C, E1 and partial E2/NS1 regions of the HCV genome obtained from a serum specimen of a human patient who contracted hepatitis during an epidemic of NANB hepatitis (HCV-KF). We also report the cloning and sequence analysis of the same genomic domains 9 years after inoculation of the original infectious material into a chimpanzee (HCV-476). The data indicated that both isolates are more closely related to Japanese-type HCV than to American-type isolates. As reported in previous studies, the most conserved domain was found in the capsid regions (95% identity) that have been shown to contain conserved epitopes useful for the diagnosis of HCV infection during the acute phase of the disease (Okamoto et al., 1990a; Nasoff et al., 1991).

The rate of evolution of the HCV-KF genome was determined by comparing the genomic sequences of the original human isolate and the chimpanzee isolate. The
changes in the structural domains of both isolates were exclusively nucleotide point mutations. The rate of mutation of the HCV-KF genome was estimated to be around $0.9 \times 10^{-3}$ nucleotides per site per year for the structural domain. This amounts to half of the mutation rate for strain H of HCV (Ogata et al., 1991) in the same domain, but the difference is probably not significant. It remains to be determined whether these differences are virus-dependent or host-dependent. It also remains to be studied whether these differences are reflected in the severity of liver disease.

The mutation rate of the HCV genome appears to be rather high. It is known that the mutation rate of RNA viruses ranges from $10^{-1}$ to $10^{-4}$ base substitutions per site per year (Holland et al., 1982; Imazeki et al., 1990). Thus the amino acid sequence of the viral protein is subject to considerable change within a relatively short evolutionary period. This high frequency of change may be similar to that found for human immunodeficiency virus (Saag et al., 1988). The high levels of diversity in the nucleotide sequence may be due to low fidelity of the viral RNA polymerase (Hahn et al., 1986).

The distribution of nucleotide substitutions in the HCV-KF and HCV-476 isolates was not uniform throughout the structural domain. Instead, most of the changes were clustered in the V region. In this region, we have observed not only a high proportion (50%) of G→A and A→G transitions but also a significant number of nucleotide substitutions occurring in the first codon position. The biological function of region V is still unknown, and it remains to be seen whether it encodes protective epitopes. The rapid rate of genetic change supports the view that HCV populations within infected patients may be extremely heterogeneous and that any single isolate from a patient may or may not represent the mutant most prevalent at that time. Future studies of viral pathogenesis will need to take these considerations into account.

We are grateful to Drs Toshio Shikata, Takeshi Kurata and Alfred M. Prince for continuous encouragement during this study. We are also very grateful to Drs Masayoshi Yamauchi, Ke Liu and Don-Hun Lee for their kind help, and the Green Cross Corporation for generous assistance with the chimpanzee experiments.

### References


(Received 13 April 1992; Accepted 17 June 1992)