

Acta Medica Okayama

Volume 45, Issue 5

1991

Article 9

OCTOBER 1991

Sequence variations in the envelope protein of the hepatitis C virus: comparison with partial cDNA sequence of a new variant virus obtained by the polymerase chain reaction.

Hajime Hada*

Norio Koide[†]

Hiroyuki Takabatake[‡]

Tadashi Hanafusa**

Takao Tsuji^{††}

*Okayama University,

[†]Okayama University,

[‡]Okayama University,

**Okayama University,

^{††}Okayama University,

Sequence variations in the envelope protein of the hepatitis C virus: comparison with partial cDNA sequence of a new variant virus obtained by the polymerase chain reaction.*

Hajime Hada, Norio Koide, Hiroyuki Takabatake, Tadashi Hanafusa, and Takao Tsuji

Abstract

It has been reported that the envelope region located at the 3' portion of the structural protein coding region is one of the most variable regions at both nucleotide and amino acid sequence levels in the hepatitis C virus (HCV) genome. We cloned HCV cDNA fragments of an envelope protein coding region (HCVNK), which were derived from serum of a Japanese patient with hepatocellular carcinoma and were amplified by polymerase chain reaction. After determining the nucleotide sequence, deduced amino acid sequence of the envelope protein region was compared with those of six HCV strains already published (HCJ1, HCVUS, HCJ4, HCVJH, HCVJ and HCVBK). Homology analysis among the strains revealed that the seven strains were classified into two subtypes; a US subtype (HCJ1 and HCVUS) and a Japanese subtype (HCJ4, HCVJH, HCVJ, HCVBK and HCVNK), since percentage homologies between two subtypes (70.3-77.3%) were significantly lower than those within each subtype (83.9-93.5%). Detailed analysis of the amino acid sequences also indicates that the region at aa246-aa258, tentatively named intersubtype variable region-1, may distinguish the US subtype from the Japanese subtype.

KEYWORDS: hepatitis C virus, envelope, DNA sequencing, homology, intersubtype variable region

*PMID: 1661558 [PubMed - indexed for MEDLINE]

Sequence Variations in the Envelope Protein of the Hepatitis C Virus: Comparison with Partial cDNA Sequence of a New Variant Virus Obtained by the Polymerase Chain Reaction

Hajime Hada*, Norio Koide, Hiroyuki Takabatake, Tadashi Hanafusa^a and Takao Tsuji

First Department of Internal Medicine, Okayama University Medical School, ^aRadioisotope Research Center, Okayama University, Okayama 700, Japan

It has been reported that the envelope region located at the 3' portion of the structural protein coding region is one of the most variable regions at both nucleotide and amino acid sequence levels in the hepatitis C virus (HCV) genome. We cloned HCV cDNA fragments of an envelope protein coding region (HCVNK), which were derived from serum of a Japanese patient with hepatocellular carcinoma and were amplified by polymerase chain reaction. After determining the nucleotide sequence, deduced amino acid sequence of the envelope protein region was compared with those of six HCV strains already published (HCJ1, HCVUS, HCJ4, HCVJH, HCVJ and HCVBK). Homology analysis among the strains revealed that the seven strains were classified into two subtypes; a US subtype (HCJ1 and HCVUS) and a Japanese subtype (HCJ4, HCVJH, HCVJ, HCVBK and HCVNK), since percentage homologies between two subtypes (70.3-77.3 %) were significantly lower than those within each subtype (83.9-93.5 %). Detailed analysis of the amino acid sequences also indicates that the region at aa246-aa258, tentatively named intersubtype variable region-1, may distinguish the US subtype from the Japanese subtype.

Key words : hepatitis C virus, envelope, DNA sequencing, homology, intersubtype variable region

In 1989 Choo *et al.* cloned a blood-borne non-A, non-B hepatitis virus genome (1), and also succeeded in detecting the antibody to non-structural protein of the virus (C100-3) expressed in yeast (2). This virus, now called hepatitis C virus (HCV), has a single positive strand RNA that is about 10,000 nucleotides long as genome. By means of analysis of the polyprotein deduced

from the nucleotide sequence HCV is thought to have a structure homologous with the flaviviruses as schematically shown in Fig. 1 (1). However, since they published only about 75 % of the 3' terminal portion of the genome (3), analyses of the HCV genome by polymerase chain reaction (PCR) had been limited within the region encoding non-structural protein until recently (4, 5).

In 1990, Okamoto *et al.* disclosed a sequence of 1863 nucleotides of the remaining 5' terminal portion of the HCV genome by screening cDNA libraries made by using published sequences as

* To whom correspondence should be addressed.

nt: nucleotide number

aa: amino acid number

primers (6). Takeuchi *et al.* also showed the nucleotide sequence of the 5' structural protein coding region (7, 8). Recently Kato *et al.* and Takamizawa *et al.* determined the whole nucleotide sequence of HCV, which are 9413 and 9416 nucleotides long, respectively (9, 10). According to those reports, HCV has only one open reading frame of 3010 amino acids, which is preceded by 5' non-coding region of 329-332 nucleotides and followed by 3' non-coding region of 54 nucleotides (9, 10).

Okamoto *et al.* obtained four HCV clones and suggested that there are at least 2 major HCV strains represented by HCJ1 and HCJ4 (6). HCVJH by Takeuchi *et al.*, HCVJ by Kato *et al.*, and HCVBK by Takamizawa *et al.* were derived from sera taken from Japanese patients, and showed evident differences from the HCJ1 or HCV prototype clone reported by Houghton *et al.* (HCVUS)(3, 11) at nucleotide sequence level (7-10). These findings are likely to lead to a speculation that most HCV strains isolated in Japan might belong to the Japanese subtype (HCJ4, HCVJH, HCVJ and HCVBK) as opposed to the US subtype (HCJ1 and HCVUS) (9).

It has also been reported that nucleotide or amino acid sequence homologies among strains vary from location to location, and that the envelope protein coding region is one of the most variable regions within the genome of HCV (7-9). Therefore, comparing nucleotide or amino acid sequences of the envelope coding region among HCV strains is very useful for determining whether HCV strains can be classified into subtypes.

In this paper, we report on the nucleotide sequence for HCV cDNA fragment of the envelope protein coding region amplified by PCR and cloned (HCVNK), and compare the nucleotide and amino acid sequences with those of other HCV strains published in the literature (6-11). Thereby, a region, which might be useful for distinguishing the US subtype from the Japanese subtype, appeared to be identified.

Materials and Methods

Isolation of RNA from serum. Serum was obtained from a 67-year-old female Japanese patient with hepatocellular carcinoma (N. K.). Two milliliters of serum, positive for antibody to C100-3, were diluted to 10 ml with 50 mM Tris-Cl, pH 8.0, 1 mM EDTA, 100 mM NaCl, and ultracentrifuged in an RP55T rotor (Hitachi Co., Ltd., Tokyo, Japan) at 40,000 rpm for 4 h at 4°C. Collected virus particles were dissolved in 2 ml of guanidinium thiocyanate solution containing 25 mM sodium citrate, pH 7.0, 0.5 % sodium sarcosinate, 0.2 M β -mercaptoethanol and overlaid onto 2.5 ml cesium chloride cushion containing 0.5 M EDTA, pH 7.5. We pelleted RNA through ultracentrifugation in an RPS 50-2 rotor (Hitachi Co., Ltd., Tokyo, Japan) at 35,000 rpm for 21 h at 20°C and redissolved in 20 μ l of distilled water after precipitation with 80 % ethanol.

PCR. Five microliters of RNA solution were subjected to reverse transcription to cDNA, amplification by PCR, cloning into plasmid vector and sequencing according to the procedures in the manual (12). In brief, reverse transcription was done at 42°C for 1 h using commercially available random primer (Takara shuzo Co., Ltd., Kyoto, Japan). Based on the nucleotide sequence of HCJ1 (6), two sets of primers for PCR below were synthesized with DNA Synthesizer Model 380B (Applied Biosystems, Ltd., Foster City, CA, USA): 5'-GTCCGGGTTCTGGAAGACGGCGTGAAC-TAT-3' (nt 787-nt 816), 5'-GCCCCGCCAGGACT-CCCCAGTGAGC-3' (nt 1398-nt 1375) for 5' half of the envelope protein coding region E612, and 5'-CAGCTGCTCCGGATCCCACAAGCC-3' (nt 1330-nt 1353), 5'-GGGGTGTTGGAGTGAAGCAATATAC-3' (nt 1863-nt 1840) for 3' half of the envelope coding region E534. We used relatively conserved regions between HCJ1 and HCJ4 to make these primers. The positions of those primers and the expected PCR products were schematically illustrated in Fig. 1. With these sets of primers and 2 units of Taq polymerase (Takara Shuzo Co., Ltd., Kyoto, Japan), cDNA was amplified by repeating 30 cycles of incubation at 94°C for 1 min, at 42°C for 1.5 min, and at 72°C for 1.5 min in Programmable Temperature Control System PC-700 (Astec, Ltd., Fukuoka, Japan). After precipitation with ethanol, these PCR products were again subjected to PCR by repeating another 30 cycles with the same condition above.

Cloning and sequencing. Final products were visualized on 3 % NuSieve™ GTG-1 % Seakem™ ME agarose (FMC Corporation, Rockland, ME, USA)

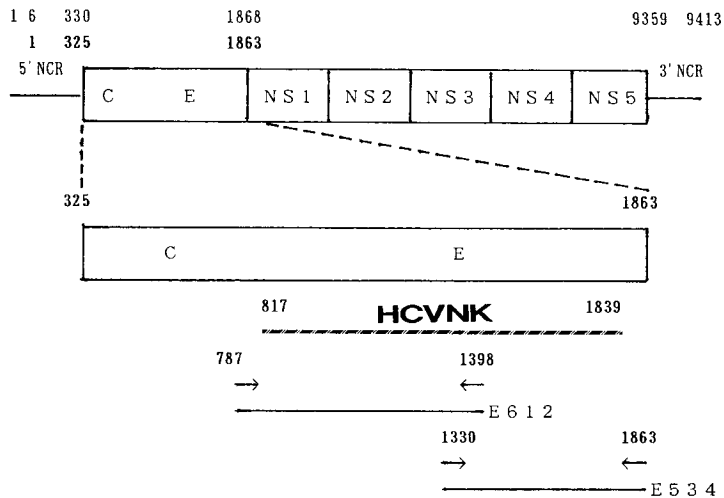


Fig. 1 Schema showing the positions of primers used for PCR and cloned PCR products. The top schema shows the whole structure of the genome of hepatitis C virus proposed by analogy to flaviviruses (1), of which the enlarged structural protein region is shown in the bottom schema. C: core protein. E: envelope protein. NS1, 2, 3, 4 and 5: non-structural proteins 1, 2, 3, 4 and 5. NCR: non-coding region. Open boxes indicate an open reading frame. Bold bar: HCVNK. Bars of E612 and E534: PCR products. Arrows indicate the positions of primers used for PCR. Bold numbers show the nucleotide number corresponding to those reported by Okamoto *et al.* (6), and to those reported by Kato *et al.* (9), which are indicated at the top of this figure.

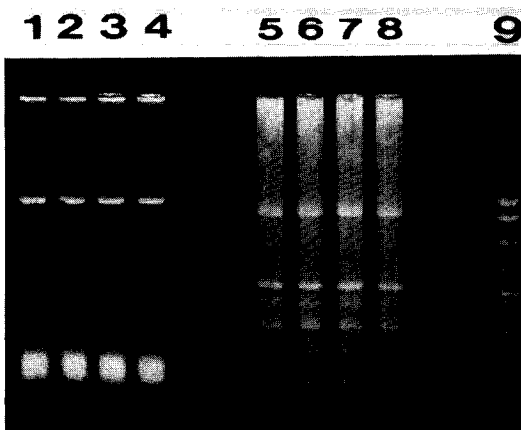


Fig. 2 cDNA fragments of envelope protein coding region of hepatitis C virus amplified by PCR. PCR products were electrophoresed on 3 % NuSieveTM GTG-1 % SeakemTM ME agarose composite gel at 100 V for 1 h and stained with 0.5 µg/ml ethidium bromide. Lanes 1-4: E612. Lanes 5-8: E534. Lane 9: pBR 322 DNA digested with *Msp* I as a size marker. Upper 4 bands on Lane 9 are 622, 527, 404, and 309 nucleotides long, respectively, from the top.

composite gel by staining with ethidium bromide (Fig. 2). Bands having the expected size were excised and electroeluted in dialysis tubing at 240 mA for 1 h under monitoring with UV lamp. After extraction with phenol/chloroform and precipitation with ethanol, cDNA fragments were cloned into the *Sma* I site of pUC19 plasmid vector by blunt end ligation. Out of the many recombinants screened by checking the size of the insert, the nucleotide sequences of one clone each from E612 and E534 were determined by dideoxynucleotide chain termination using M13/pUC sequencing primerTM and M13/pUC reverse sequencing primerTM (Boehringer Mannheim GmbH, Ltd., Mannheim, Germany). All these experiments were performed under the permit by the Committee of Okayama University on Recombinant DNA Technology and Safety (No. 9041).

Positions of nucleotides and amino acids. In this paper we adopted the nucleotide sequence numbers reported by Okamoto *et al.* (6). According to this numbering, nt1 equals nt-325 of HCVJH, nt6 of HCVJ, and nt9 of HCVBK. Because the same ATG at nt325 is used as the starting codon, amino acid sequence numbers in all HCV strains are identical.

Computer analysis and statistical analysis. Computer analysis was performed with GENETYX program (Software development, Inc., Tokyo, Japan). Statistical analysis was performed by the Student's *t*-test.

Results

Fig. 3 shows the whole nucleotide sequence of our HCVNK clone, composed of E612 and E534, and the aligning sequences of HCJ1 and

HCJ4 reported by Okamoto *et al.* (6). The nucleotide sequences of primers used for PCR, which were located at nt 787-nt 816 and nt 1840-nt 1863, were excluded from this investigation. Even though we used nucleotide sequences of HCJ1 as primers for PCR, our HCVNK clone showed greater homology to HCJ4 (89.1 %) than to HCJ1 (70.1 %). Twenty-one overlapping nucleotides (nt 1354-nt 1374)(Fig. 1) were completely identical between E612 and E534.

The ratios of nucleotide changes at the third

```

      817
HCJ1  GCAACAGGGAACCTTCCTGGTGTCTTTCTCTATCTTCCTCTGGCCCTGCTCTCTTGC
      *
HCVNK GCAACAGGGAACCTGCCCGGTGTCTCTCTATCTTCCTCTGGCTCTGCTGTCTCTGT
      *
HCJ4  GCAACAGGGAACCTGCCCGGTGTCTTTCTCTATCTTCCTCTGGCTTGTCTGTCTCTGT
      877
HCJ1  CTGACTGTGCCCGCTTCAGCCTACCAAGTGGCCAACCTCCACAGGGCTTTATCATGTCCAC
      *
HCVNK TTGACCACCCAGCCTCCGCCTATGAAGTGGCCAACGTGTCCGGGTATACCATGTCCAG
      *
HCJ4  TTGACCATCCAGCTTCCGCTTATGAAGTGGCCAACGTGTCCGGGTATACCATGTCCAG
      937
HCJ1  AATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGCAGATGCCATCCTGCATACTCCG
      *
HCVNK AACGACTGCTCCAACCTCAAGTATTGTGTATGAGGCAGCGACGTGATTATGCATACTCCC
      *
HCJ4  AACGACTGCTCCAACCTCAAGCATTTGTGTATGAGGCAGCGACATGATCATGCATACTCCC
      997
HCJ1  GGGTGTGTCCCTTGGCTTCGCGAGGGCAACGTCTCGAGGTGTTGGGTGGCGATGACCCCC
      *
HCVNK GGGTGTGTCCCTTGTCTTCGGGAGAACATTCCTCCCGTTGCTGGGTAGCGCTCACTCCC
      *
HCJ4  GGGTGTGTCCCTTGGCTTCGGGAGGACAACAGCTCCCGTTGCTGGGTAGCGCTCACTCCC
      1057
HCJ1  ACGGTAGCCACCAGGACGGCAAACTCCCCGCGACGAGCTTCGACGTCACATCGATCTG
      *
HCVNK ACGCTCGCGCCAGGAATACCAGCATCCCTACTACGACAATAAGACGCCACGTCGACTTG
      *
HCJ4  ACGCTCGCGCCAGGAATGCCAGCGTCCCACTACGACAATAACGACGCCACGTCGACTTG
      1117
HCJ1  CTTGTGCGGAGCGCCACCCTCTGTTCGGCCCTCTACGTGGGGGATCTGTGCGGGTCCGTC
      *
HCVNK CTCGTTGGGCGAGCTGCTTTCTGCGCCGAATGTACGGGTGGGACCTCTGCGGGTCTGCT
      *
HCJ4  CTCGTTGGGCGGCTGCTTTCTGCTCCGCTATGTACGTGGGGGATCTCTGCGGATCTGTT
      1177
HCJ1  TTGCTTATTGGTCAACTGTTTACCTTCTCTCCAGGCGCCACTGGACAACGCAAGGCTGC
      *
HCVNK TTCTTCGTCTCCAGTTGTTACCTTTTCGCTAGTGGGATGAGACAGTACAGTACTGC
      *
HCJ4  TTCTTCGTCTCCAGCTGTTACCTTCTCGCTAGCCGGCATGAGACAGTGCAGGACTGC
      1237
HCJ1  AATTGTCTATCTACCCGCGCATATAACGGGTATCGCATGGCATGGGATATGATGATG
      *
HCVNK AAATGCTCAATTATCCCGCCACTTAACCGGCATCGTATGGCATGGGATATGATGATG
      *
HCJ4  AACTGCTCAATCTATCCCGCCATTATCAGGTCACGCGATGGCTGGGATATGATGATG
      1297
HCJ1  AACTGGTCCCTACGCGCGGCTTGGTAATGGTCAGTGCTCCGGATCCCAAGGCCATC
      *
HCVNK AACTGGTACCCGAGCAGCTCTAGTGGTGTCCGAGTTACTCCGGATCCCAAGGCCGTC
      *
HCJ4  AACTGGTACCTACAACAGCCCTAGTGGTGTCCGAGTTGCTCCGGATCCCAAGGCCATC

```

Fig. 3-1

Fig. 3 Nucleotide sequence of the envelope protein coding region of HCVNK and a comparison with those of HCJ1 and HCJ4 (6). Numbers at the upper left corners of the sequences show nucleotide positions corresponding to those reported by Okamoto *et al.* (6). Asterisks show homologous nucleotides in HCJ1 and HCJ4 with HCVNK.

codon to all nucleotide changes in HCVNK were 58.0 % (151/260) and 54.8 % (57/104) for HCJ1 and HCJ4, respectively. If nucleotide changes were random, the ratio of the third codon changes to all nucleotide changes should be around one third. Therefore, nucleotide changes in HCV are likely to occur at the third codon more frequently as observed in other strains (6–10).

Amino acid sequences deduced from nucleotide sequences of 7 HCV strains, including our HCVNK, are shown in Fig. 4. Table 1 shows the calculated nucleotide and amino acid sequence homologies among HCV strains.

Homologies between HCJ1 and HCVUS, both of which are thought to belong to the US subtype, were 93.0 % and 93.5 % at nucleotide and amino acid sequence levels, respectively. Average homologies among HCJ4, HCVJH, HCVJ, HCVBK and HCVNK, all of which are thought to belong to the Japanese subtype, were 87.9 % and 87.2 % at nucleotide and amino acid sequence levels, respectively. On the other hand, HCJ1 and HCVUS showed only 71.7 % and 75.0 % homologies on the average at nucleotide and amino acid levels, respectively, with each strain of the Japanese subtype. These homologies were significantly lower than those observed within

HCJ1	1357	TTGGATATGATCGCTGGTGCTCACTGGGGAGTCCTGGCGGGCATAGCGTATTCTCCATG
HCVNK		*****
HCJ4		GTGGATATGGTGGCGGGGCCCACTGGGGAATCCTGGCGGGCTTGCTACTATTCCATG

HCJ1	1417	GTGGGAAGTGGGCGAAGGTCCTGGTAGTGCTGTTGTGCTGTTGGCGGGCTGACGCGGAA
HCVNK		*****
HCJ4		GTGGGAAGTGGGCTAAGGTTCTGATTGTGATGCTACTTTTGGCGGGCTGACGAGGCA

HCJ1	1477	GTAGGGAAGTGGGCTAAGGTCCTGATTGTGGCGCTACTCTTCGCGGGCTTGACGCGGAG
HCJ1		ACCATCGTCTCGGGGGACAAGCCGCGGCCATGTCTGGACTTGTTAGTCTCTTCACA
HCVNK		*****
HCJ4		ACCAACATGACAGGGGGAGTCCGGCCCGCACCACTTACAAGCTTACGACCCCTCTCTCA

HCJ1	1537	ACCTACAGCTCGGGGGGGCGGCCAGCCACACCACTCCACGCTGCGCTCCTCTCTCA
HCJ1		CCAGGCGCTAAGCAGAACATCCAAGTATCAACACCAACGGCAGTTGGCACATCAATAGC
HCVNK		*****
HCJ4		TATGGGGCGTCTCAGAAATCCAGCTCATAATACCAATGGCAGCTGGCATCAACAGG

HCJ1	1597	CCTGGGGCGTCTCAGAGAATCCAAGCTTGTGAATACCAACGGCAGCTGGCACATCAACAGG
HCJ1		ACGGCCTTGAAGTGAATGAAAGCCTTAACACCGGCTGGTAGCAGGGCTTATCTATCAA
HCVNK		*****
HCJ4		ACTGCCCTAAAGTGAATGACTCCCTCAACACCGGGTTCTTGCCGGCTGTTCTACACA

HCJ1	1657	ACTGCCCTAAAGTGAATGACTCCCTCCACACTGGGTTCTTGCCGGCTGTTCTACACA
HCJ1		CACAAATCAACTCTTCGGGCTGTCCCGAGAGGTTGGCCAGCTGCCAGCGCTTACGAT
HCVNK		*****
HCJ4		CACAGGTTCAACAGTCCGGATGTCCAGAGCGCTTGCCAGCTGCCGCTCAATTGACAAG

HCJ1	1717	CACAGGTTCAACTCGTCCGGGTGCCGGAGCGCATGGCCAGCTGCCCGCCATTGACTGG
HCJ1		TTTGACCAAGGGTGGGGCCCTATCAGTCATGCCAACGGAAGCGGCCCGACCAACGCCCC
HCVNK		*****
HCJ4		TTTCGATCAGGGGTGGGGTCCCATCACCTATACTGAGCCTGGCAACAAGGACAGAGGCT

HCJ1	1777	TTTCGCCAGGGATGGGGCCCATCACCTATACTGAGCCTGACAGCCCGGATCAGAGGCT
HCJ1		TATTGTTGGCACTACCCCCAAAACCTTGGGTATCGTGCCCGCAAGAGCGTATGTGGC
HCVNK		*****
HCJ4		TATTGTTGGCACTACGCGCCCGACAGTGTGGTATCGTACCCGCGTGGAGGTGTGTGGT

HCJ1	1837	TATTGCTGGCATTACGCGCTCGACCGTGTGGTATCGTACCCGCGTGGAGGTGTGTGGT
HCJ1		CCG
HCVNK		CC
HCJ4		CCC
		CC
		CCA

Fig. 3-2

Table 1 Nucleotide and amino acid sequence homologies in envelope protein region among hepatitis C virus strains

		HCVUS	HCVJH ^a	HCVJ	HCVBK	HCVNK
HCVUS	nt (%) ^b	93.0	71.8	72.4	70.3	70.5
	aa (%) ^c	93.5	75.7	76.0	73.6	74.5
HCVJH ^a	nt (%) ^b		73.1	73.3	72.4	71.7
	aa (%) ^c		76.5	77.3	75.1	76.0
HCVJ	nt (%) ^b			91.3	88.1	88.2
	aa (%) ^c			89.8	88.6	88.2
HCVBK	nt (%) ^b				88.0	89.4
	aa (%) ^c				88.7	88.8
HCVNK	nt (%) ^b					86.4
	aa (%) ^c					87.1

a: Only 831 nucleotides and 277 amino acids in the corresponding region were published on HCVJH.

b: Homology at nucleotide sequence level.

c: Homology at deduced amino acid sequence level.

165	HCVUS	ATGNLPGGCSFISFLALLSCLTVPASAYQVANSTGLYHVTNDPCNSSIVYEAHDAILHTP
	HCVJH	R.....A.....M.....
	HCVJ	I.....E.....VS.....I.....S.....A.....M.....
	HCVBK	I.....E.....R.....VS.....I.....S.....A.....M.....
	HCVNK	T.....E.....H.....VS.....I.....S.....A.....L.....M.....
	HCVNK	T.....E.....VS.....V.....S.....A.....V.....M.....
225	HCVUS	GCVPVREGNVSRCVAMTPTVATRDGKLPATQLRRHIDLLVGSATLCSALYVGDLGCGSV
	HCVJH	A.....L.....A.....NASV.....T.....V.....A.....AF.....M.....
	HCVJ	D.....S.....L.....A.....NASV.....T.....V.....T.....AF.....M.....
	HCVBK	N.....S.....L.....A.....NASV.....T.....V.....T.....AF.....M.....
	HCVNK	S.....F.....L.....A.....NTSI.....T.....V.....A.....A.....M.....
	HCVNK	S.....S.....L.....A.....NVTI.....T.....V.....A.....AF.....M.....
	HCVNK	N.....S.....L.....A.....NTSI.....T.....V.....A.....AF.....M.....
285	HCVUS	FLIGQLFTFSPRRHWTTCGCNCISIYPGHITCHRMADWMMNWSPTAALVMAQLLRIPQAI
	HCVJH	V.....S.....E.....V.....D.....LS.....T.....VS.....V.....
	HCVJ	S.....S.....E.....V.....D.....VS.....T.....VS.....V.....
	HCVBK	VS.....YE.....V.....D.....VS.....T.....VS.....V.....
	HCVNK	VS.....V.....L.....D.....VS.....T.....VS.....V.....
	HCVNK	VS.....SG.....E.....V.....Y.....K.....L.....VS.....V.....
345	HCVUS	LDMIAGAHWGLAGIAYFSMVGWNAKVLVLLLFAGVDAETIVSGQAARAMSGLVSLFT
	HCVJH	V.....L.....Y.....I.....A.....G.....YT.....A.....SHTT.....T.....A.....S
	HCVJ	V.....V.....L.....Y.....I.....M.....GH.....R.....T.....VQGHVT.....T.....R
	HCVBK	V.....V.....L.....Y.....I.....M.....GH.....H.....T.....RV.....SSTQS.....WLS
	HCVNK	V.....V.....L.....Y.....I.....M.....GD.....H.....T.....AQ.....KTTNR.....M.....A
	HCVNK	V.....V.....L.....I.....L.....Y.....I.....M.....EA.....NMT.....AP.....TTYK.....T.....S
405	HCVUS	PGAKNQIQLINTNGSWHINSTALNCNESLNTGWLGLIYQHKFNSSGCPERLASCRRLTD
	HCVJH	V.....L.....D.....D.....F.....H.....P.....P.....
	HCVJ	S.....R.....V.....R.....D.....H.....F.....A.....F.....T.....R.....M.....PIDW
	HCVBK	S.....K.....V.....R.....D.....Q.....F.....A.....
	HCVNK	Q.....PS.....K.....V.....R.....D.....Q.....FI.....A.....F.....AHT.....A.....M.....PIDE
	HCVNK	S.....PS.....K.....R.....D.....Q.....F.....A.....F.....T.....S.....M.....Q.....TIDK
	HCVNK	Y.....S.....K.....R.....D.....D.....F.....A.....F.....T.....R.....T.....SIDK
465	HCVUS	FDQGGWPISTHANGSGFDQRPYCWHPKPGCIVPAKSVCGP
	HCVJH	Y.....T.....Y.....EPDS.....A.....R.....SQ.....
	HCVJ	A.....T.....DMPES.....A.....R.....SQ.....
	HCVBK	T.....Y.....ES.....RS.....PQ.....T.....SE.....
	HCVNK	T.....Y.....EPGN.....A.....RQ.....SE.....

Fig. 4 Deduced amino acid sequences of the envelope protein region of HCVNK and 6 other HCV strains in the literature (6-11). Amino acids were expressed as one character. Numbers at the upper left corners of the sequences show amino acid positions from the starting codon at nt325. Asterisks indicate the same amino acid as that of HCVJ1 (6). ISVR-1: intersubtype variable region-1. HVR-1 and -2: hypervariable regions-1 and -2 (13). In HCVJH only 277 amino acids of the corresponding region have been published (7, 8).

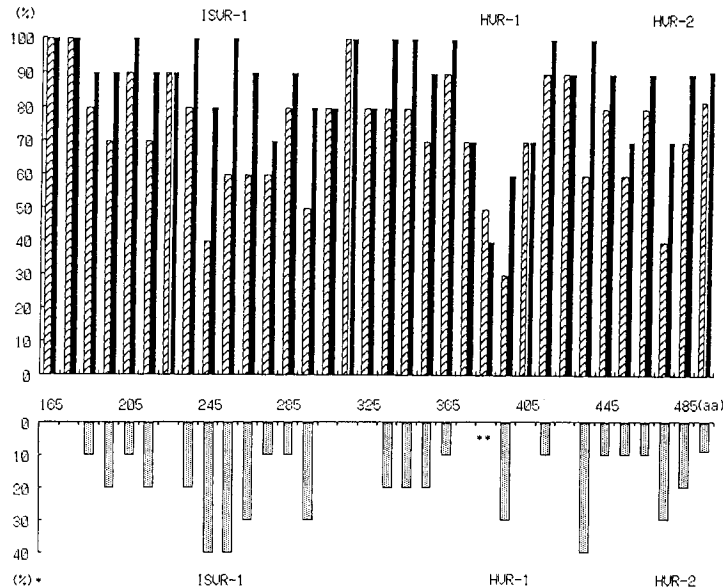


Fig. 5 Distribution of amino acid variation in the envelope protein region of hepatitis C virus. Except for aa495-aa505, every 10 consecutive amino acids in the envelope protein region of HCVNC were compared with those of HCVNC and HCVNC (6). Percentages of amino acid sequence homologies were shown as hatched bars (to HCVNC) and black bars (to HCVNC) in the upper half of this figure. Subtraction of percent homologies to HCVNC from those to HCVNC ((%)*) were shown as meshed bars in the lower half of this figure. * *: -10 % was shown as 0. aa: amino acid position from the starting codon at nt 325. ISVR-1, HVR-1 and -2: see Legend for Fig. 4.

each subtype ($p < 0.001$).

To clarify the distribution of amino acid replacements in HCV envelope protein, we compared the amino acid sequence of HCVNC at every 10 consecutive amino acids with those of HCVNC (as a representative of the US subtype) and HCVNC (as a representative of the Japanese subtype)(Fig. 5). Three regions with the lowest homology to HCVNC were located at aa245-aa254 (40 %), aa395-aa404 (30 %) and aa475-aa484 (40 %). Hypervariable region-1 (aa391-aa400) and -2 (aa474-aa480)(HVR-1 and -2) reported by Hijikata *et al.* (13) were involved in the latter 2 regions (Figs. 4 and 5). The largest difference between homologies to HCVNC and to HCVNC were also observed in the region at aa245-aa254 (40 %)(Fig. 5). Detail analysis of the amino acid sequences around this region of all 7 HCV strains revealed the following: in the region at aa246-aa

258, i) 13 consecutive amino acids were completely identical between HCVNC and HCVNC, ii) 9 of 13 amino acids were identical among the Japanese subtype and 2 out of the remaining 4 amino acids were in common among 4 of 5 strains of the Japanese subtype, iii) 8 of 13 amino acids were different between the two subtypes (Fig. 4). We tentatively named this region intersubtype variable region-1 (ISVR-1)(Figs. 4 and 5).

Discussion

It is speculated that a long polypeptide of 3010 amino acids coded by an HCV genome might be processed by cellular protease or by its own protease encoded within the HCV genome (9). Since N-terminal 120 amino acids of the polypeptide are rich in basic amino acids or proline,

this portion is thought to encode viral core protein (8, 9). Compared with amino acid sequences between HCVJH and HCVUC (US subtype), N-terminal 190 amino acids showed 97.4 % homology, whereas the following 190 amino acids showed only 75.3 % homology (8). N-glycosylation sites characteristic for envelope protein were clustered within aa196-aa645 (9). From these findings, the C-terminal portion of the structural protein region following aa190 is supposed to encode viral envelope protein (7-9). Therefore, the region encompassing aa165-aa505 investigated in this paper seems to be mostly included in the envelope protein region. Because Saiki *et al.* calculated the ratio of misincorporation during amplification by PCR as only 0.25 % (14), this ratio may not affect the results in this paper.

The fact that the homologies of HCJ1 and HCVUS to HCV strains isolated in Japan (6-10) were significantly lower than those between HCJ1 and HCVUS and those among the strains isolated in Japan at both nucleotide and amino acid sequence levels suggests the possibility that HCV could be classified into at least 2 major subtypes, a US subtype (HCJ1 and HCVUS) and a Japanese subtype (HCJ4, HCVJH, HCVJ, HCVBK and HCVNK). Although from the analysis of the non-structural protein coding region where the nucleotide sequence is more conservative, Ulrich *et al.* reported a variation of nucleotide sequence of less than 10 % among 4 HCV clones isolated from patients in the USA (15). In 3 out of 4 Japanese hemophiliac patients who had received blood products from the USA, HCV clones from their sera were reported to be more homologous to the US subtype than to the Japanese subtype at the nucleotide level (16). These two reports also support the classification of HCV into a US subtype and a Japanese subtype.

In this paper, by analyzing in detail the distribution of amino acid variation among strains, an intersubtype variable region (ISVR-1), in which the amino acid sequence was conserved within

each subtype and was clearly different between the two subtypes, was revealed. We might be able to predict the subtype of newly isolated strains of HCV by analyzing the amino acid sequence of this region. Although there seems to be a similar region to ISVR-1 at aa432-aa441, the difference of amino acid sequence between the US and the Japanese subtypes in this region is not as clear (4/10) as that in ISVR-1 (8/13) (Figs. 4 and 5). HVR-1 and -2 (13) are different from ISVR-1 in that remarkable amino acid variations were observed even among the Japanese subtypes in these 2 regions.

However, partly because nucleotide changes were realized more commonly at the third codon of translation frame, amino acids tend to be replaced by similar amino acids throughout the HCV genome. If we regard this type of change as homologous, amino acid homologies among all HCV strains, including HCJ1 and HCVUS, are as high as 95 % even in the envelope protein region. Therefore, we do not think that the Japanese subtype is a different virus from the US subtype.

Now that antibodies to non-structural protein (C100-3)(2) and core proteins (CP-9, CP-10, and GOR47-1)(17, 18) have been found. In the research field of HCV, researchers are now turning their attention to finding antibodies with virus neutralizing capabilities and developing vaccines. Because HCV-RNA can be detected in sera positive for antibodies above, those antibodies can not be neutralizing antibodies (17, 18). Although the putative envelope protein has not yet been purified, the envelope protein is considered to be one of the most promising candidates. And, in the development of vaccines, the high variability of this virus must be taken into consideration.

Acknowledgment. We thank Professor Shuui Seki, Associate Professor Ken Tsutsui and Dr. Kousuke Akiyama of the Department of Molecular Biology, Institute of Cellular and Molecular Biology, Okayama University Medical School for helpful discussions. And we are also grateful to Associate Professor Yayoi Koide of the Faculty of Education, Okayama University,

and Dr. Kozo Fujio of the First Department of Internal Medicine, Okayama University Medical School, for their kind suggestions on statistical analysis and computer surveys, respectively.

This study was supported in part by a research grant from the study group for Non-A, non-B hepatitis, Ministry of Health and Welfare, Japan.

References

1. Choo Q-L, Kuo G, Weiner AJ, Overby LR, Bradley DW and Houghton M: Isolation of cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* (1989) **244**, 359-362.
2. Kuo G, Choo Q-L, Alter HJ, Gitnick GL, Redeker AG, Purcell RH, Miyamura T, Dienstag JL, Alter MJ, Stevens CE, Tegtmeier GE, Bonino F, Colombo M, Lee W-S, Kuo C, Berger K, Shuster JR, Overby LR, Bradley DW and Houghton M: An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* (1989) **244**, 362-364.
3. Houghton M, Choo Q-L and Kuo G: NANBV diagnostics and vaccines. *Eur Patent Appl* (1988) 88-310, 922, 5, Publ 0, 318, 216.
4. Weiner AJ, Kuo G, Bradley DW, Bonino F, Saracco G, Lee C, Rosenblatt J, Choo Q-L and Houghton M: Detection of hepatitis C viral sequences in non-A, non-B hepatitis. *Lancet* (1990) **335**, 1-3.
5. Kubo Y, Takeuchi K, Boonmar S, Katayama T, Choo Q-L, Kuo G, Weiner AJ, Bradley DW, Houghton M, Saito I and Miyamura T: A cDNA fragment of hepatitis C virus isolated from an implicated donor of post-transfusion non-A, non-B hepatitis in Japan. *Nucl Acids Res* (1989) **17**, 10367-10372.
6. Okamoto H, Okada S, Sugiyama Y, Yotsumoto S, Tanaka T, Yoshizawa H, Tsuda F, Miyakawa Y and Mayumi M: The 5'-terminal sequence of the hepatitis C virus genome. *Jpn J Exp Med* (1990) **60**, 167-177.
7. Takeuchi K, Boonmar S, Kubo Y, Katayama T, Harada H, Ohbayashi A, Choo Q-L, Kuo G, Houghton M, Saito I and Miyamura T: Hepatitis C viral cDNA clones isolated from a healthy carrier donor implicated in post-transfusion non-A, non-B hepatitis. *Gene* (1990) **91**, 287-291.
8. Takeuchi K, Kubo Y, Boonmar S, Watanabe Y, Katayama T, Choo Q-L, Kuo G, Houghton M, Saito I and Miyamura T: The putative nucleocapsid and envelope protein genes of hepatitis C virus determined by comparison of the nucleotide sequences of two isolates derived from an experimentally infected chimpanzees and healthy human carries. *J Gen Virol* (1990) **71**, 3027-3033.
9. Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T and Shimotohno K: Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc Natl Acad Sci USA* (1990) **87**, 9524-9528.
10. Takamizawa A, Mori C, Fuke I, Manabe S, Murakami S, Fujita J, Onishi E, Andoh T, Yoshida I and Okayama H: Structure and organization of the hepatitis C virus genome isolated from human carriers. *J Virol* (1991) **65**, 1105-1113.
11. Houghton M: NANBV diagnostics and Vaccines. *Eur Patent Appl* (1990) 90-302, 866.0, Publ 0, 388, 232, A1.
12. Sambrook J, Fritsch EF and Maniatis T: *Molecular Cloning, A laboratory manual*, Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989) pp 7.6-14.21.
13. Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Ohkoshi S and Shimotohno K: Hypervariable regions in the putative glycoprotein of hepatitis C virus. *Biochem Biophys Res Commun* (1991) **175**, 220-228.
14. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA and Arnheim N: Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* (1985) **230**, 1350-1354.
15. Ulrich PP, Romeo JM, Lane PK, Kelly I, Daniel LJ and Vyas GN: Detection and semiquantitation, and genetic variation in hepatitis C virus sequences amplified from the plasma of blood donors with elevated alanine aminotransferase. *J Clin Invest* (1990) **86**, 1609-1614.
16. Hijikata M, Kato N, Mori S, Ootsuyama Y, Nakagawa M, Sugimura T, Ohkoshi S, Kojima H, Meguro T, Taki M, Takayama K, Yamada K and Shimotohno K: Frequent detection of hepatitis C virus US strain in Japanese hemophiliacs (Rapid communication). *Jpn J Cancer Res* (1990) **81**, 1195-1197.
17. Okamoto H, Munekata E, Tsuda F, Takahashi K, Yotsumoto S, Tanaka T, Tachibana K, Akahane Y, Sugai Y, Miyakawa Y and Mayumi M: Enzyme linked immunosorbent assay for antibodies against the capsid protein of hepatitis C virus with a synthetic oligopeptide. *Jpn J Exp Med* (1990) **60**, 223-233.
18. Mishihiro S, Hoshi Y, Takeda K, Yoshikawa A, Gotanda T, Takahashi K, Akahane Y, Yoshizawa H, Okamoto H, Tsuda F, Peterson DA and Muchmore E: Non-A, non-B hepatitis specific antibodies directed at host-derived epitope: implication for an autoimmune process. *Lancet* (1990) **336**, 1400-1403.

Received May 23, 1991; accepted July 1, 1991.