Arch Virol (1994) 137: 25–34

Archives of Virology

Sequence variability in the env-coding region of hepatitis C virus isolated from patients infected during a single source outbreak

M. Höhne[1], **E. Schreier**[1], and **M. Roggendorf**[2]

[1] Robert Koch-Institute, Berlin, [2] Institute of Virology, University of Essen, Federal Republic of Germany

Accepted March 16, 1994

Summary. The variability of the hepatitis C virus genome was investigated in a group of German patients who developed chronic hepatitis C after parenteral administration of contaminated immunoglobulin to prevent Rh sensitization after pregnancy. The nucleotide and deduced amino acid sequence alterations of the E1 and the first hypervariable region of the E2 gene of the hepatitis C virus (HCV) genome from sera of two randomly selected patients were studied by comparison of HCV sequences obtained from the original inoculum (anti Rh immunoglobulin) and from patient sera collected in 1979 and 1989. All isolates were classified as subtype 1b but showed nucleotide insertions of up to 12 nucleotides at the cleavage site of E1/E2. Microheterogeneity of HCV genomes was found in the immunoglobulin supporting the quasispecies model of HCV distribution. Remarkable nucleotide exchanges over the 10 year period in the E1 region ($0.9 - 5.2 \times 10^{-3}$ base substitutions per genome site per year) and especially in the first hypervariable region of the E2 gene (about 1.5×10^{-2}) occurred. The HCV genome undergoes a selection of variants, though it is not known if this derives from mutation or selection of pre-existing rare variants.

Introduction

Since the identification of Hepatitis C virus (HCV) as the major causative agent of parenterally transmitted non-A, non-B hepatitis [7] several isolates of HCV have been sequenced completely [8, 11, 12, 31] or in part (for reviews see [22, 23, 28]). HCV is a 9.4-kb positive-strand RNA virus encoding a large polyprotein of about 3010–3033 amino acids [7, 12]. The genome organization and the putative amino acid sequence of the polyprotein revealed that HCV constitutes a new genus of the *Flaviviridae* family [12, 23]. Considerable genetic heterogeneity has been reported among isolates from different individuals and within the same individual [21, 23, 34]. Based on the comparison of published full length and partial sequences several attempts have been undertaken to

classify the different known HCV variants [4, 5, 17, 19, 23]. Recently, a system discriminating types and subtypes was proposed which will allow the classification of new groups of HCV genotypes together with their related subtypes [28, 30]. Type 1a, 1b, 2a, 2b were found to coexist in various geographic regions all over the world [4] but show different prevalences. A recently described type 4 sequence seems to have a broad distribution in Africa but to be absent or rare in Europe and the Far East [3, 27].

The degree of sequence variability differs throughout the genome, showing some relatively conserved regions (5′ untranslated region and core region) and some highly variable regions (E1, E2). Especially, the amino terminus of the putative E2 protein is known to contain a 27–31 amino acid hypervariable region (HVR1) which has characteristics typical of a rapidly evolving protein domain like the V3 loop of the HIV-1 gp120 protein [14, 34]. The HCV genome was also shown to evolve rapidly in vivo with different rates of variability for different gene regions [21, 24].

In 1978 an outbreak of hepatitis C occurred in women in former Eastern Germany following administration of contaminated immunoglobulin to prevent Rh-sensitization after pregnancy. The antibody response to HCV and the long-term persistence of these antibodies have been reported [9]. About 60% of the patients developed chronic hepatitis. We investigated the evolution of a part of the E1 and E2 genes in two randomly selected patients of this group developing chronic disease by comparing sequences obtained from the original inoculum and from patient sera drawn in 1979 and 1989. The degree of sequence variation by mutation and/or selection over the period analysed and the microheterogeneity of HCV in the infection source will be discussed.

Materials and methods

Patients and serology

Patients 1 and 2 are German women who developed chronic hepatitis C after administration of anti-Rh-immunoglobulin. Sera were obtained 6 months (1979) and 10 years (1989) after infection. All samples were anti-HCV positive (HCV Abbott EIA 2nd generation test). There was no known re-exposure to any blood or blood products during the ten-year interval.

RNA preparation

HCV RNA was extracted from pelleted HCV particles [13] or from 200 µl serum by the acid guanidinium thiocyanate-phenol-chloroform method according to Chomczynski and Sacchi [6] and finally resuspended in 10 µl of diethylpyrocarbonate treated water.

Reverse transcription-polymerase chain reaction (RT-PCR)

Five µl of the RNA preparation was reverse transcribed with the antisense primer 9 (Table 1) using 100 U of Moloney murine leukaemia virus reverse transcriptase according to the manufacturer's instruction (GIBCO, BRL Life Technologies Inc., Gaithersburg, MD) at 37 °C for 1 h and then heated at 95 °C. One fourth of the cDNA synthesis was first amplified by

Table 1. Synthetic oligonucleotides used for cDNA synthesis, first and second PCR

Primer	Sequence (5′–3′)	Position
9, antisense	TTGAACTTGTGGTGATAGAA	1653–1672
10, sense	TGCTCTTTCTCTATCTTCCT	843–862
12, antisense	GCCGGCAAAKAGYAGCA	1452–1470
30, sense	CGCTTACGAAGTGCGCAA	897–916
36, sense	TCCATGGTGGGGAACTGGGC	1416–1435
46, antisense	TTGCAGTTCAGGGCAGTCCTGTTGATGTGCCA	1587–1619

Nucleotide positions according to [12], S: G + C, K: G + T; Y: C + T

40 cycles of PCR (1 min 92 °C, 1 min 37 °C, 2 min 72 °C) using primers 9 and 10. In general, a second round of amplification under the same conditions was done by using one to three microliters of first PCR products and three different pairs of inner region specific primers. For reamplification of the E1/E2 region the primer pair 10/46 was used and for the separate amplification of the E1 and 5′ E2 region primer pairs 12/30 and 9/36 were used, respectively (Fig. 1). The PCR products were analyzed by agarose gel electrophoresis and ethidium bromide staining. The PCR fragments obtained from the immunoglobulin were cloned using the TA cloning system (Invitrogen, San Diego, U.S.A.) into the plasmid vector pCRII according to the manufacture's instruction, and sequenced by the dideoxynucleotide chain termination reaction using sequenase version 2.0 (USB, Cleveland, U.S.A) of alkaline denatured recombinant plasmid DNA. The nucleotide sequences of PCR fragments obtained from sera of patients were determined by direct sequencing of single-stranded PCR products after asymmetric amplification and purification by absorption on glasspowder (USB, Cleveland, U.S.A.).

Results

HCV RNA present in the anti-Rh-immunoglobulin (G) and in serum samples from 1979 and 1989 from two patients (P1, P2) were amplified by RT-nested PCR as shown in Fig. 1. The 664 bp sequences corresponding to the nucleotide position 923 to 1586 (numbering according to [12] and the putative amino acid sequences (position 199–419) are shown in Figs. 2 and 3, respectively. The nucleotide sequence data obtained from the immunoglobulin have been

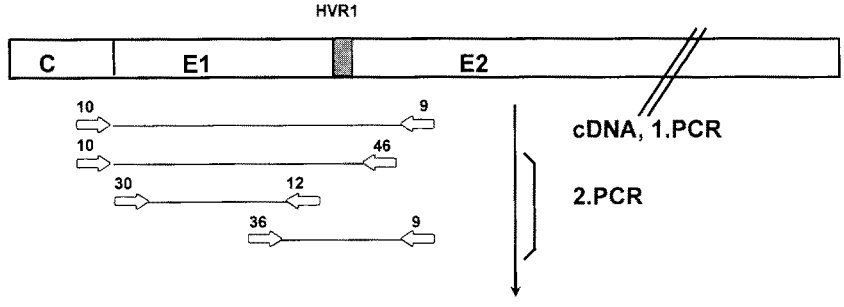

Fig. 1. Scheme of primer positions used for cDNA synthesis and amplification

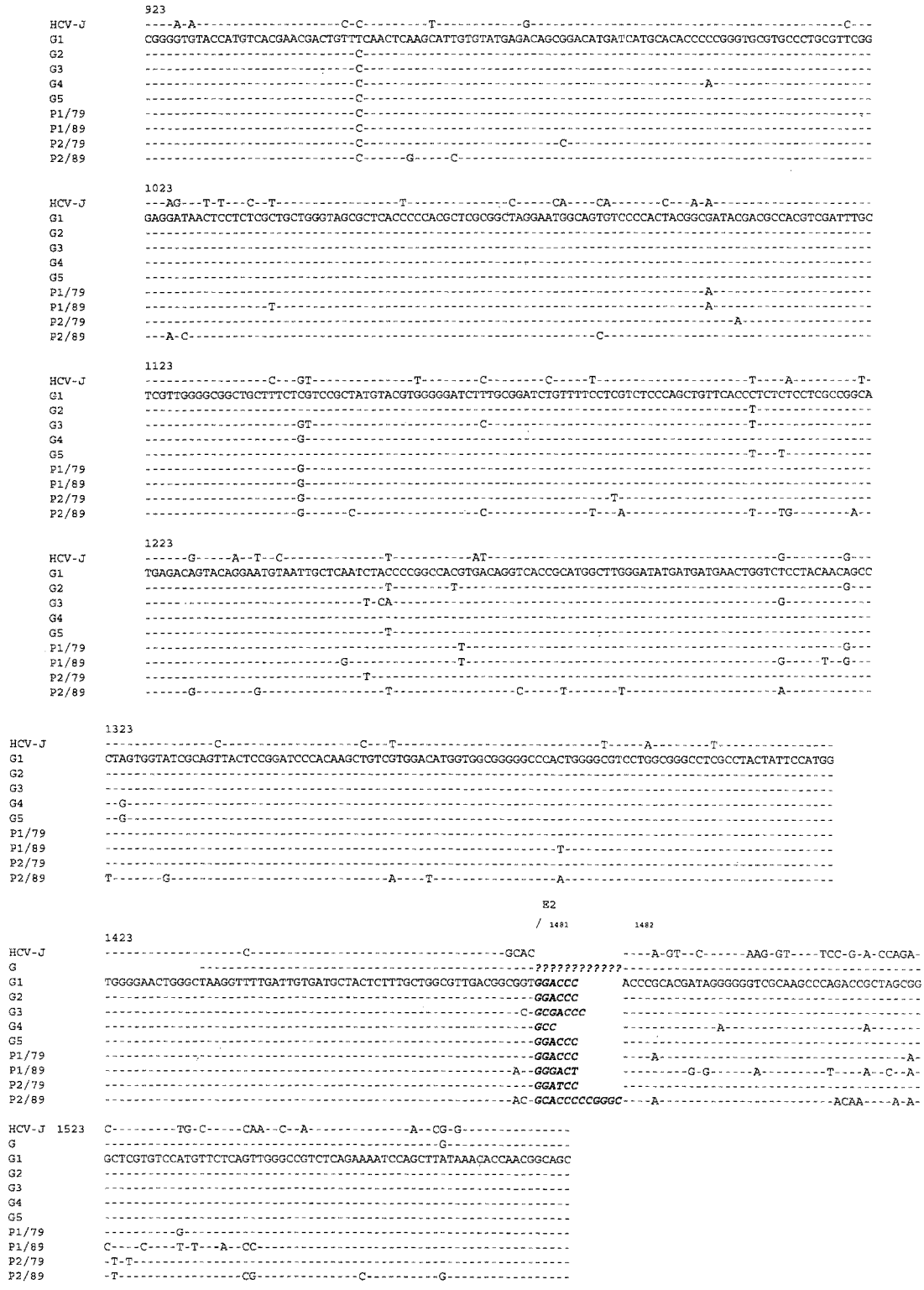

Fig. 2. Comparison of nucleotide sequences of the E1 and the 5′-terminal E2 gene coding region to subtype 1b sequence of HCV-J [12]; *G1-G5* sequences derived from 5 independent clones from the immunoglobulin; *G*, *P1*, *P2* sequences from direct sequencing of PCR products. Dashes represent nucleotides identical with G1

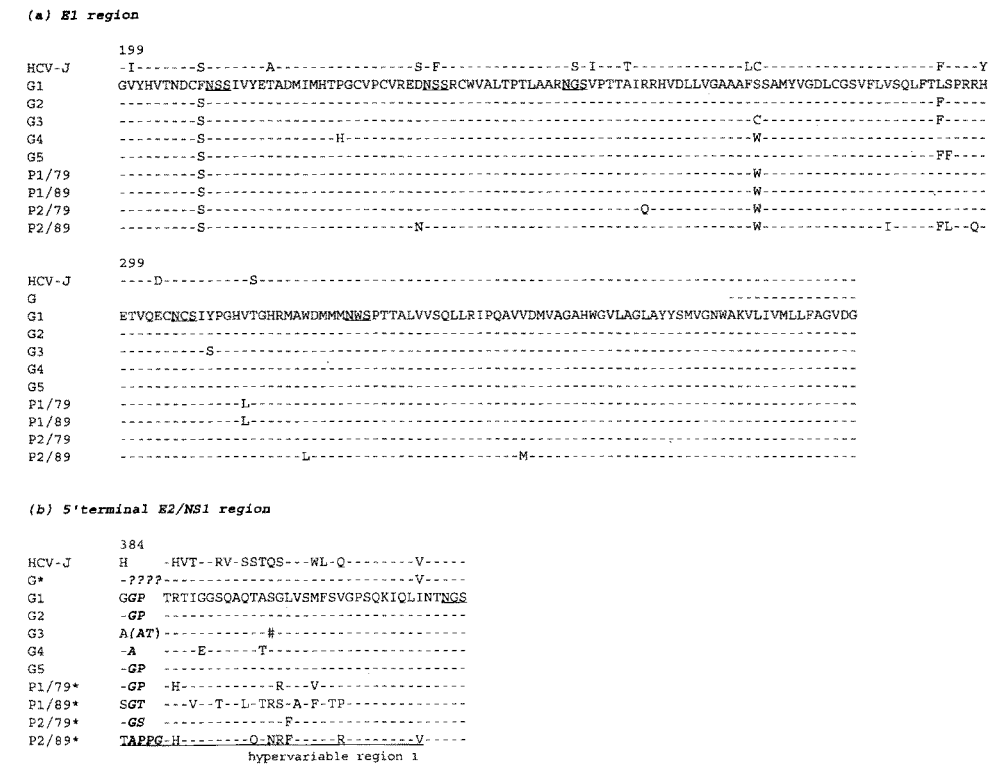

Fig. 3. Comparison of deduced amino acid sequences of the E1 (**a**) and N-terminal E2 region (**b**); description as in Fig. 2. Amino acid numberings are according to [12]. Amino acids written in italics indicate insertions and in brackets indicate that there are 7 nucleotides inserted. # indicates translation termination after frame shift. Underlined amino acid indicate potential N-linked glycosylation sites

deposited in the EMBL data library under the accession numbers X 67299 for the E1 gene region and X 72945 for the 5′-terminus of E2 gene region. The nucleotide sequence analysed from the anti-Rh-immunoglobulin (G1) showed on overall homology of 88.4% to the subtype 1b sequences e.g. HCV-J [12] but only about 75% to subtype 1a (e.g. HCV-1 [8]) and less than 63% to subtypes 2a (e.g. HC-J6 [22]) and 2b (e.g. HC-J8 [23]). Considering the regions separately 92.8% of nucleotides in the E1 region (nt 923–1478) but only 67% in the 5′ region of E2 (nt 1479–1586) were identical to those of HCV-J.

By direct sequencing of amplification products from this outbreak all showed nucleotide insertions of up to 12 nucleotides at the 5′ end of the E2 gene region between nt position 1481 and 1482. A six-nucleotide insertion was observed in sera of patient 1 (P1) from 1979 and 1989 and in sera of patient 2 (P2) from 1979. In P2 from 1989 a 12-nucleotide insertion was detected. In the immunoglobulin a mixture of insertions of up to 12 nucleotides were found by direct sequencing after amplification (Fig. 2: G*). Comparing the overall nucleotide sequences (E1/HVR1 of E2) from patient 1 and patient 2 over the 10-year-period a considerable exchange can be observed. A nucleotide exchange

Table 2. Nucleotide substitution rates (nt substitutions per genome site per year) and number of nucleotide and amino acid exchanges after 10 years for the E1 and E2 HVR1 region calculated from sequences obtained from patient 1 and patient 2 in 1979 and 1989

	E1 region			E2 HVR1		
	subst. rate	nt exchange (10 years)	aa exchange (10 years)	subst. rate	nt exchange (in 10 years)	aa exchange (in 10 years)
Patient 1	0.9×10^{-3}	5	0	1.6×10^{-2}	15	12
Patient 2	5.2×10^{-3}	29	8	1.5×10^{-2}	14	7

Nucleotide insertions at the cleavage site of E1/E2 are omitted. E1 region: nt 923–1478; aa 199–378, HVR1: nt 1479–1571; aa 384–414

was observed at 20 of 664 nucleotide positions for P1 (3%) and 43 of 664 nucleotide positions for P2 (6.5%) during the period observed. The base and the deduced amino acid exchanges were not uniformly distributed throughout the sequence analysed. In the E1 region 5 nucleotides (no aa) and 29 (8 aa) of 556 nucleotides were different for P1 and P2, respectively. In the HVR1 region 15 nt (12 aa) and 14 nt (7 aa) of 93 nucleotides were exchanged in P1 and P2, respectively. We estimated nucleotide exchange rates for patient 1 and 2 (omitting the insertions at E1/E2 cleavage site) for the HVR1 of the E2 region of 1.6×10^{-2} and 1.5×10^{-2} base substitutions per genome site per year, respectively (Table 2). The rates estimated for the E1 region are about 18 and 3 times lower than those observed those for the HVR1. 73% and 86% of codon exchanges in the HVR1 region of patient 1 and patient 2, respectively resulted in nonconservative amino acid exchanges. All potential N-glycosylation sites including the subtype 1b specific N-glycosylation site at position 234–236 and all cysteine residues are conserved.

To determine the microheterogeneity of HCV genomes in the source of infection in more detail PCR products of the HVR1 were cloned and 5 independent clones (G1–5) were sequenced. They differed in the E1 region at 4–9 nucleotide positions resulting in 2 to 4 amino acid changes. In the HVR1 region 3 clones (G1, G2, G5) showed identical nucleotide sequences, and G3 and G4 showed one and 2 exchanges resulting in one and two amino acid changes, respectively. At the cleavage site of E1/E2 insertions of 6 nucleotides were found in 3 clones (G1, G2, G5) and 3 nucleotides were inserted in G4. In clone G3 an insertion of 7 nucleotides was found.

Discussion

In this paper we analysed the sequence variation of HCV in patients infected from the same source of contaminated immunoglobulin. Therefore, the E1 gene and the 5′-terminus of the E2 gene of HCV from the contaminated immunoglobulin and from sera of chronically infected patients over a period of 10 years were analysed by RT-PCR and sequencing.

All isolates were classified as type 1b [28, 30], the predominant genotype in Europe [25, 26]. The E1 and the E2 gene regions analysed exhibited remarkable differences of nucleotide and predicted amino acid sequences over the 10 years. Especially, the 5'-terminus of the E2 gene known as the hypervariable region 1 showed a considerable dynamics concerning selection of different genome variants in the two patients from the source of infection.

As shown by sequencing of five independent clones derived from the immunoglobulin and by direct sequencing the source of infection contained a mixture of HCV genomes with different nucleotide insertions at the cleavage site of E1/E2 (no abundant species of insertion) but relatively few differences in the HVR1. In contrast, as seen by direct sequencing of the HVR1 each isolate from patient 1 and patient 2 seemed to contain one major insertion type. A lack of sequence diversity due to a selective process has been described for HCV transmission from mother to infant [36]. Recently, different level of heterogeneity of HCV genomes in different isolates have been reported [10]. This is in agreement with our results from sequencing of the HVR1 of a great number of isolates from HCV infected patients (not published). Thus, by demonstrating simultaneous presence of multiple variant genomes and their rapid selection our findings support the quasispecies model of HCV distribution [10, 18]. The observation of a 7 nucleotide insertion resulting in a frame shift which would produce a termination codon at nt position 1517 may support the assumption that co-circulation of defective genomes is a general feature of HCV infection [10, 18]. Nevertheless, the possibility of a PCR or cloning artefact has to considered. An insertion of nucleotides at the junction between E1/E2 has been recently reported [1, 15] but the role of nucleotide insertions at this site remains unclear.

The E1 and HVR1 sequences of samples from 1979 were more similar to the infective source than the 1989 samples (7 nucleotide differences in 1979, compared to 24 and 40 differences in 1989 for P1 and P2, respectively). Sixty to 75% of the nucleotide substitutions in the HVR1 region occurred in the first or second codon position resulting in substitution of amino acids. Only some of the alterations observed after the 10-year-period were already found in sequences of the five independent immunoglobulin derived clones. The HCV genome seems to undergo a strong selection of variants resulting in the occurrence of different abundant sequences in both patients, though it is not known if this derives from mutation or selection of pre-existing rare variants.

Although the biological function of the HVR1 of E2 gene is still unknown, it has been speculated to be subjected to strong immune selective pressure resulting in escape from the host immune surveillance. Recently, the existence of sequence specific immunological epitopes in this region was shown [16, 17, 32, 35] inducing the production of specific antibodies directed to this variants. Since the alterations of amino acids in this region have been shown to occur sequentially in the chronic phase of hepatitis [14], it was suggested that HVR1 antibodies are involved in the immune selection of mutants. To examine the HRV1 as a possible neutralising epitope and to investigate the variability of

the HCV genome due to spontaneous mutations additional studies using recently reported in vitro replication systems [2, 20, 29] will be helpful.

References

1. Abe K, Inchauspe G, Fujisawa K (1992) 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. J Gen Virol 73: 2725–2729
2. Bertolini L, Iacovacci S, Ponzetto A, Gorini G, Battaglia M, Carloni G (1993) The human bone-marrow-derived B-cell line CE, susceptible to hepatitis C virus infection. Rev Virol 144: 281–285
3. Bukh J, Purcell RH, Miller RH (1992) Sequence analysis of the 5' noncoding region of hepatitis C virus. Proc Natl Acad Sci USA 89: 4942–4946
4. Cha TA, Beall E, Irvine B, Kolby J, Chien D, Kuo G, Urdea MS (1992) At least five related, but distinct, hepatitis C viral genotypes exist. Proc Natl Acad Sci USA 89: 7144–7148
5. Chan SW, McOmish F, Holmes EC, Dow B, Peutherer JF, Follett E, Yap PL, Simmonds P (1992) Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants. J Gen Virol 73: 1131–1141
6. Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162: 156–159
7. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244: 359–361
8. Choo QL, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina-Selby A, Barr PJ, Weiner AJ, Bradley DW, Kuo G, Houghton M (1991) Genetic organization and diversity of the hepatitis C virus. Proc Natl Acad Sci USA 88: 2451–2455
9. Dittmann S, Roggendorf M, Dürkop J, Wiese M, Lorbeer B, Deinhard F (1991) Long-term persistence of hepatitis C virus antibodies in a single source outbreak. J Hepatol 13: 323–327
10. Higashi Y, Kakumi S, Yoshioka K, Wakita T, Mizokami M, Ohba K, Ito Y, Ishikawa T, Takayanagi M, Nagai Y (1993) Dynamics of genome change in the E2/NS1 region of hepatitis C virus in vivo. Virology 197: 659–668
11. Inchauspe G, Zebedee S, Lee DH, Sugitani M, Nasoff M, Prince AM (1991) Genomic structure of the human prototype strain of hepatitis C: Comparison with American and Japanese isolates. Proc Natl Acad Sci USA 88: 10292–10296
12. Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, Shimotohno K (1990) Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc Natl Acad Sci USA 87: 9524–9528
13. Kato N, Yokosuka O, Omata M, Hosoda K, Ohto M (1990) Detection of hepatitis C virus ribonucleic acid in the serum by amplification with polymerase chain reaction. J Clin Invest 86: 1764–1767
14. Kato N, Ootsayama J, Ohkoshi S, Nakazowa T, Sekiya H, Hijikata M, Shimotohno K (1992) Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus. Biochem Biophys Res Commun 189: 119–127
15. Kato N, Ootsuyama Y, Tanaka T, Nakagawa M, Nakazawa T, Muraiso K, Ohkoshi S, Hijikata M, Shimotohno K (1992) Marked sequence diversity in the putative envelope proteins. Virus Res 22: 107–123
16. Kato N, Sekiya H, Ootsuyama Y, Nakazowa T, Hijikata M, Ohkoshi S, Shimotohno

K (1993) Humoral immune response to hypervariable region 1 of the putative envelope glycoprotein (gp 70) of hepatitis C virus. J Virol 67: 3923–3930
17. Lesniewski RR, Boardway KM, Casey JM, Desai SM, Devare SG, Leung TK, Mushahwar IK (1993) Hypervariable 5′ terminus of hepatitis C virus E2/NS1 encodes antigenetically distinct variants. J Med Virol 40: 150–156
18. Martell M, Esteban J, Quer J, Genesca J, Weiner A, Esteban R, Guardia J, Gomez J (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: Quasispecies nature of HCV genome distribution. J Virol 66: 3225–3229
19. Mori S, Kato N, Yagyu A (1992) A new type of hepatitis C virus in patients in Thailand. Biochem Biophys Res Commun 183: 334–342
20. Nissen E, Höhne M, Schreier E (1994) In vitro replication of hepatitis C virus in a human lymphoid cell line (H9). J Hepatol 20: 437
21. Ogata N, Alter HJ, Miller RH, Purcell RH (1991) Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. Proc Natl Acad Sci USA 88: 3392–3396
22. Okamoto H, Okada S, Sugiyama Y, Kurai K, Iizuka H, Machida A, Miyakawa Y, Mayume M (1991) Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. J Gen Virol 72: 2697–2707
23. Okamoto H, Kurai K, Okada SI, Yamamoto K, Lizuka H, Tanaka T, Fukuda S, Tsuda F, Mishiro S (1992) Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. Virology 188: 331–341
24. Okamoto H, Kojima M, Okada S, Yoshizawa H, Iizuka H, Tanaka T, Muchmore EE, Peterson D, Ito Y, Mishiro S (1992) Genetic drift of hepatitis C virus during an 8.2 year infection in a chimpanzee: variability and stability. Virology 190: 894–899
25. Roggendorf M, Lu M, Fuchs K, Ernst G, Höhne M, Schreier E (1993) Variability of the envelope regions of hepatitis C in European islolates and its significance for diagnostic tools. In: Kaaden OR, Eichhorn W, Czerny CP (eds) Unconventional agents and unclassified viruses. Recent advances in biology and epidemiology. Springer, Wien New York, pp 27–39 (Arch Virol [Suppl] 7)
26. Schreier E, Höhne M, Fuchs K, Wiese M, Deinhardt F, Roggendorf M (1992) Variability and subtyping of hepatitis C virus genome. 3rd International Symposium on Positive Strand RNA Viruses, Clearwater, U.S.A., pp 1–33
27. Simmonds P, McOmish F, Yap PL, Chan SW, Lin CK, Dusheiko G, Saeed AA, Holmes EC (1993) Sequence variability in the 5′non-coding region of hepatitis C virus: identification of a new virus type and restrictions on sequence diversity. J Gen Virol 74: 661–668
28. Simmonds P, Holmes EC, Cha TA, Cham SW, McOmish F, Irvine B, Beall E, Yap PL, Kolberg J, Urdea MS (1993) Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. J Gen Virol 74: 2391–2399
29. Shimizu YK, Iwamoto A, Hijikata M, Purcell RH,Yoshikura H (1992) Evidence for in vitro replication of hepatitis C virus genome in a human T-cell line. Proc Natl Acad Sci U.S.A. 89: 5477–5481
30. Stuyver S, Rossau R, Wyseur A, Duhamel M, Vanderborght B, Van Heuverswyn H, Maertens G (1993) Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. J Gen Virol 74: 1093–1102
31. Takamizawa A, Mori C, Fuke I, Manabe S, Fujita J, Onishi E, Andoh T, Yoshida I, Okayama H (1991) Structure and organization of the hepatitis C virus genome isolated from human carriers. J Virol 65: 1105–1113

32. Taniguchi S, Okamoto H, Sakamoto M, Kojima M, Tsud F, Tanaka T, Munekata E, Muchimore EE, Peterson DA, Mishiro S (1993) A structurally flexible and antigenically variable N-terminal domain of the hepatitis C virus E2/NS1 protein: implication for an escape from antibody. Virology 195: 297–301
33. Weiner AJ, Christopherson C, Hall JE, Bonino F, Sarace G, Brunetto HR, Crawford K, Marion CD, Crawford KA, Venkatakrishna S, Miyamura T, McHutchinson J, Cuypers T, Houghton M (1991) Sequence variation in hepatitis C virus isolates. J Hepatol 13 [Suppl 4]: S 6–S 14
34. Weiner AJ, Brauer MJ, Rosenblatt J, Richman KH, Tung J, Crawford K, Bonino F, Saracco G, Choo QL, Houghton M, Han JH (1991) Variable and hypervariable domains are found in the regions of HCV corresponding to the Flavivirus envelope and NS1 proteins and Pestivirus envelope glycoproteins. Virology 180: 842–848
35. Weiner AJ, Geysen HM, Christopherson C, Hall JE, Mason TJ, Saracco G, Bonino F, Chrawford K, Marion CD, Crawford KA, Brunetto M, Barr PJ, Miyamura T, McHutchinson J, Houghton M (1992) Evidence for immune selection of hepatitis C virus putative envelope glycoprotein variants: potential role in chronic HCV infections. Proc Natl Acad Sci USA 89: 3468–3472
36. Weiner AJ, Thaler MM, Crawford K, Ching C, Kansopon J, Chien DY, Hall JE, Hu F, Houghton M (1993) A unique, predominant hepatitis C virus variant found in an infant born to a mother with multiple variants. J Virol 67: 4365–4368

Authors' address: Dr. E. Schreier, Britzer Strasse 1–3, D-12439 Berlin, Federal Republic of Germany.

Received October 24, 1993