

Sequence Analysis of the 5' Untranslated Region in Isolates of at Least Four Genotypes of Hepatitis C Virus in The Netherlands

G. E. M. KLETER,^{1*} L.-J. VAN DOORN,² J. T. BROUWER,³ S. W. SCHALM,³
R. A. HEIJTINK,¹ AND W. G. V. QUINT^{1,2}

Department of Virology, Erasmus University Rotterdam,¹ and Department of Internal Medicine II, Dijkzigt Hospital Rotterdam,³ 3015 GD Rotterdam, and Department of Molecular Biology, Diagnostic Center SSDZ, 2600 GA Delft,² The Netherlands

Received 29 April 1993/Returned for modification 15 June 1993/Accepted 26 October 1993

The RNAs of hepatitis C virus (HCV) isolates from 62 patients with chronic HCV infection were analyzed by direct sequencing of the 5' untranslated region. Two important sequence motifs were recognized: one between positions -170 and -155 and the other between positions -132 and -117. These motifs are partly complementary. All three previously published genotypes were observed; 34 (55%) isolates were classified as type 1 (including prototype [from the United States] and HCV-BK [from Japan] sequences), 11 (18%) were classified as type 2 (including HC-J6 and HC-J8), and 12 (19%) were classified as type 3 (including EB1); one patient was infected with genotypes 1 and 2. Four (6%) isolates showed aberrant sequences and were therefore provisionally classified as genotype 4. These results indicate the significance of sequence variation among the 5' untranslated regions of different HCV genotypes and indicate that this region could possibly be used for consistent genotyping of HCV isolates.

Since the discovery of hepatitis C virus (HCV), a flavivirus with a positive-sense, single-stranded RNA genome of approximately 9,400 nucleotides (6), several full-length sequences have been obtained from various isolates (7, 14, 21, 22, 26) and investigators have proposed (3, 4, 9, 12, 18, 21) that HCV isolates be classified into different types or subtypes. On the basis of all available sequence information, Chan et al. (4) distinguished three HCV genotypes. Type 1 isolates include the prototype strain HCV-1 (7) and strains HCV-H (19), HCV-K1 (9), HCV-J (14), and HCV-BK (26); type 2 includes strains HCV-K2 (9), HC-J6 (22), and HC-J8 (21); type 3 includes strains HCV E-b1 (4), HCV-T (18), and HCV1196 (17).

HCV genotyping is of interest in viral transmission studies and HCV epidemiology. Furthermore, the success of interferon treatment (25, 28) may be type or subtype related.

There are several reasons to choose the 5' untranslated region (5' UTR) for genotyping. (i) Analysis of a large number of HCV isolates results in similar phylogenetic trees for the 5' UTR, the core, NS-3, and NS-5 regions (4); (ii) the observed mutation rate of the 5' UTR is extremely low (19, 20); (iii) sequence variation within the conserved 5' UTR is mainly limited to specific regions; and (iv) the putative secondary structure of the 5' UTR, as established from biochemical and phylogenetic data (1), suggests functional conservation of this region.

In the study described here, the RNAs of 62 HCV isolates from patients in The Netherlands with chronic HCV infections were investigated to determine whether HCV genotyping by sequence analysis of PCR products derived from the 5' UTR could be performed.

MATERIALS AND METHODS

Patients. Sixty-two patients from The Netherlands between the ages of 26 and 74 years with elevated alanine aminotransferase levels, a biopsy-proven chronic non-A, non-B hepatitis, antibodies to HCV, and no recent history of infection with hepatitis B virus, hepatitis A virus, cytomegalovirus, or Epstein-Barr virus were analyzed.

Anti-HCV. Antibodies to HCV were tested by a second-generation enzyme immunoassay (Abbott, North Chicago, Ill.) and were confirmed by the recombinant immunoblot assay (RIBA-4; Ortho Diagnostics, Raritan, N.J.) according to the instructions of the manufacturer.

Blood plasma. For HCV RNA detection, EDTA-blood was collected by venipuncture, and plasma was prepared within 2 h after sampling. One-milliliter aliquots were quickly frozen in liquid nitrogen and were stored at -70°C until use.

HCV RNA PCR. HCV RNA was isolated from 100 µl of plasma by a modified version of the guanidinium method described previously (5). cDNA synthesis was performed on one-third of the isolated RNA in a 25-µl reaction volume (15) by using 20 pmol of antisense primer HCV19 (GTGCACG GTCTACGAGACCT; positions -1 to -20), 200 U of Moloney murine leukemia virus reverse transcriptase (GIBCO-Bethesda Research Laboratories, Gaithersburg, Md.), 30 U of RNasin (Promega, Madison, Wis.), and 0.5 mM (each) deoxyribonucleotide (Boehringer, Mannheim, Germany) at 42°C for 30 min after brief denaturation at 80°C. PCR was performed (40 cycles of 1 min at 94°C, 2 min at 48°C, and 3 min at 74°C) with the antisense primer HCV19 and the sense primer HCV18 (GGCGACTCCACCATAGAT; positions -304 to -324) or the sense primer HCV35 (TTGGCGGC CGCACTCCACCATGAATCACTCCCC; positions -296 to -318; underlined sequences are not complementary to the HCV sequence). For diagnosis, the first-round PCR products were analyzed by Southern blot hybridization with probe HCV17 (GAGTAGTGTGGGTCGCGAA; positions

* Corresponding author. Mailing address: Department of Virology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Phone: +31-10-463-5942. Fax: +31-10-436-5916.

–86 to –67); this was followed by washing at low stringency.

Direct sequencing of PCR products. For direct sequencing of PCR products (13), a second round of PCR (40 cycles of 1 min at 94°C, 2 min at 48°C, and 3 min at 74°C) was performed with the sense primer NCR3 (GGGGCGGCCGCCACCA TRRATCACTCCCCTGTGAGG; positions –288 to –314) and the antisense primer LD58 (5'-bio-GGCCGGGGCGGC CGCCAAGCACCTATCAGGCAGTACCACAAGGC; positions –37 to –64). LD58 is biotinylated at the 5' end. Biotinylated PCR products (estimated by agarose gel electrophoresis at approximately 100 ng) were captured on streptavidin-coated paramagnetic particles (Dynabeads M-280; Dynal, Oslo, Norway). Single-stranded DNA was prepared by denaturation of the captured amplification product by alkaline treatment according to the instructions of Dynal. Separate strands were sequenced by using the T7 DNA sequencing kit (Pharmacia, Uppsala, Sweden) and [³²P]dATP (Amersham, Buckinghamshire, United Kingdom). NCR3 served as a sense primer on the minus strand captured on the beads, and NCR4 (CACTCTCGAGCAC CCTATCAGGCAGTACC; positions –57 to –29) was used as the antisense primer on the plus strand in the supernatant. DNA sequences were read manually from autoradiographs and were analyzed with the PC/Gene computer program (Intelligenetics Inc., Mountain View, Calif.).

Nucleotide sequence accession number. The nucleotide sequences reported here have been deposited in the EMBL data library (accession numbers X58937 to X58953).

RESULTS

Sixty-two patients with chronic HCV infection were anti-HCV positive by enzyme immunoassay which was confirmed by RIBA-4, and were HCV RNA positive by reverse transcription-PCR aimed at the 5' UTR. The sense primer HCV18, which was based on the first published HCV sequences (10, 23), was initially used for the diagnosis of HCV viremia by PCR. In the case of low yields of PCR products, the sense primer HCV35 improved the PCR results considerably (data not shown), and in these cases, PCR products obtained by HCV35 were used for sequence analysis.

Nested PCR products were sequenced directly, and the results are presented in Fig. 1. If identical sequences were obtained from a number of isolates, only one representative sequence is shown. Several sequences have been reported previously and are identified by their original names. Limited sequence variations were observed essentially in two motifs: motif 1 is located between nucleotides (nt) –170 and –155, and motif 2 is located between nt –132 and –117 (Fig. 1). On the basis of these motifs, 58 of 62 isolates could be classified into the three genotypes as proposed previously. Thirty-four (55%) isolates were classified as type 1; this included the prototype HCV sequence HCV-1. Isolate HC1-N8 showed a single insertion at nt –138. In one patient (isolate HC1/2), a double infection with types 1 and 2 was observed, as deduced from bands with identical mobilities in two different lanes on the gel (Fig. 2). Type 2 sequences, including the sequence of HC-J6, were observed in 11 (18%) patients. These isolates were also recognized by the presence of a T at position –72 and a C at position –80. HC2-N2 contained a mutation at nt –127 in motif 2. Within genotype 2, sequence heterogeneity was observed at position –119, showing either a T or C. Type 3 sequences were detected in 12 (19%) patients. In addition to the type-specific sequence

motifs, type 3 isolates could also be identified by a TCA sequence at positions –93 to –95. HC3-N2 contained a point mutation at position –118 in motif 2. Four (6%) isolates were provisionally classified as type 4. These isolates showed additional sequence heterogeneity between positions –238 and –235 in comparison with the sequences of strains of types 1, 2, and 3.

Comparison between the sequences obtained in the present study (nt –262 to –66) and all previously reported 5' UTR sequences (2–4, 11) revealed a number of new 5' UTR mutations within all genotypes (Fig. 1).

The overall sequence heterogeneity among the four genotypes in the present study ranged from 4% between types 1 and 4 to 11% between types 2 and 3, whereas heterogeneity within each type was less than 2.5% (Table 1). Mutations were not randomly distributed along the 5' UTR but were clustered in motifs 1 and 2. Comparison of these motifs among the four genotypes revealed strong conservation within each genotype and showed significant differences between the genotypes (Table 1). Further analysis of the sequence variation in motifs 1 and 2 revealed the presence of covariants, i.e., compensatory mutations in each motif, maintaining the postulated secondary structure of the 5' UTR genomic RNA (1). The *P* values for the occurrence of one, two, or three covariant mutations in motifs 1 and 2 to have arisen by chance were very low (*P* = 0.06, *P* = 0.004, and *P* = 0.0002, respectively). Covariance occurred at positions –164, –163, –161, and –155 in motif 1 and positions –132, –124, –122, and –121 in motif 2. This covariance was consistently observed in the sequences of all 62 isolates, indicating the importance of this phenomenon.

DISCUSSION

Sequence analysis of the 5' UTR of HCV isolates from 62 patients with chronic HCV infection allowed consistent and efficient genotyping. Fifty-eight (94%) isolates could be classified into the three different genotypes proposed by Chan et al. (4). Identification of the genotypes is essentially based on the sequence variation in the defined motifs 1 and 2.

All three genotypes reported so far were observed in the HCV-infected patient population in The Netherlands. The coexistence of HCV genotypes in several geographic regions has been indicated earlier (3, 4). However, little is known about the distribution of HCV genotypes in Europe. The majority of published isolates belong to type 1 (8, 16). Types 2 and 3 were detected in European isolates only recently (2, 4).

Isolates belonging to a new genotype, tentatively designated genotype 4, were observed in The Netherlands, although at a low frequency. Similar sequences were first found in isolates from South Africa (3) and Denmark (2). To justify classification of these isolates as a new genotype 4, sequence analysis of the coding regions is necessary. Preliminary results (data not shown) indicate significant nucleotide sequence differences between the core region from type 4 isolates and the corresponding sequences from type 1, 2, and 3 isolates. One sequence (SA1), classified here as type 4, showed minor differences with the other type 4 sequences and might be classified as a separate genotype, genotype 5 (25a). All genotypes probably have a worldwide distribution, but the relative abundances per geographic region may differ considerably.

A double infection involving genotypes 1 and 2 was found in one patient. This was in accordance with the reported

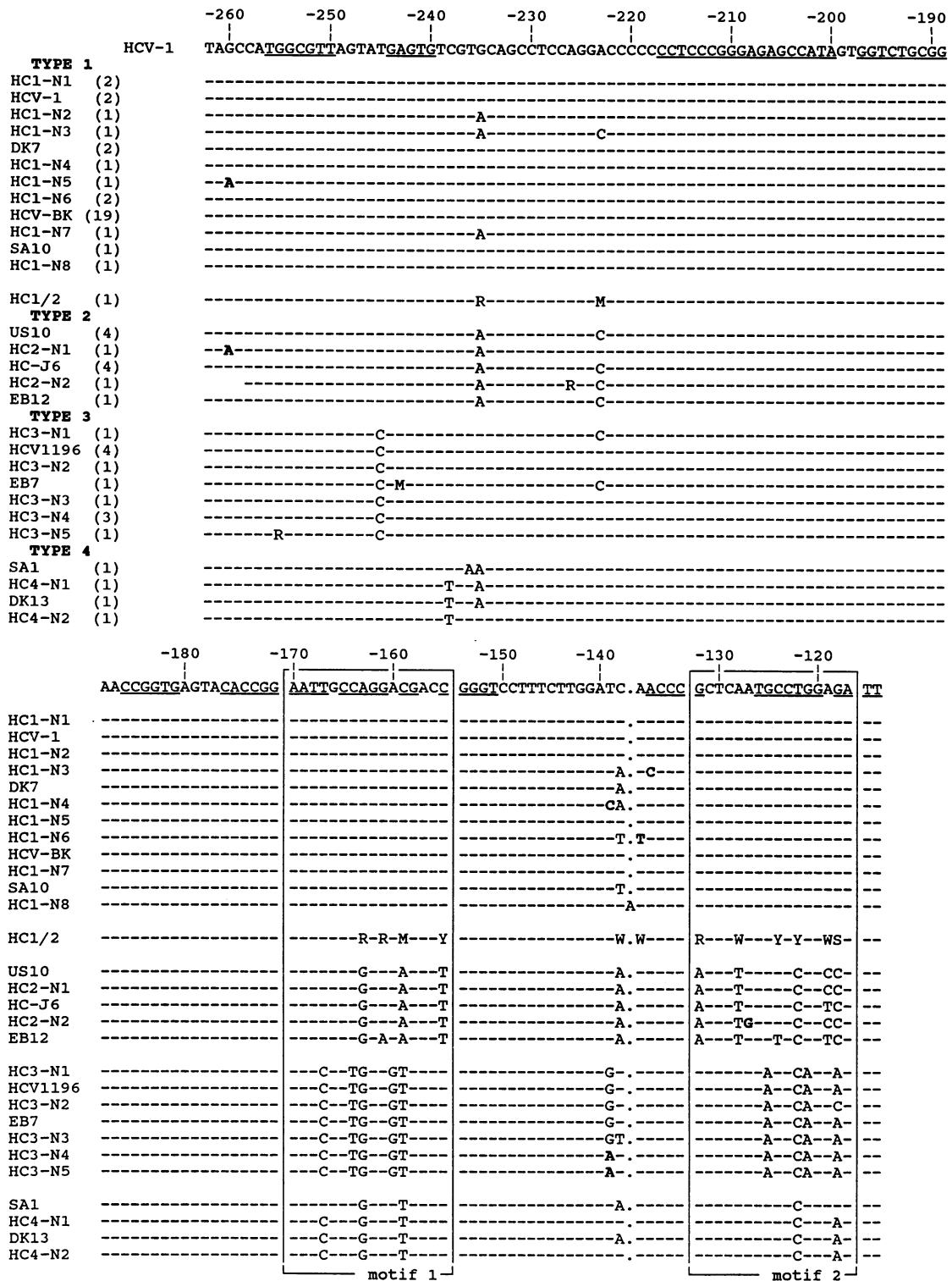


FIG. 1. Alignment of 5' UTR sequences from 62 patients. Sequences between positions -262 and -66 were classified into genotypes 1, 2, and 3 (as proposed by Chan et al. [4]) and genotype 4. Previously published sequences are identified by their original names: HCV-1 (7); DK7, SA10, US10, SA1, and DK13 (2); HCV-BK (26); HC-J6 (22); EB-12 and EB-7 (4); and HCV1196 (17). Numbers in parentheses indicate the number of isolates with that sequence. Hyphens indicate the presence of nucleotides identical to those of the prototype strains. Characters in boldface type are newly obtained point mutations. Sequence motifs 1 and 2 are boxed. H underlined nucleotides in the HCV-1 sequence are involved in the putative double-stranded RNA stem structure (1). Abbreviations: M, A or C; R, A or G; Y, T or C; W, A or T; and S, G or C.

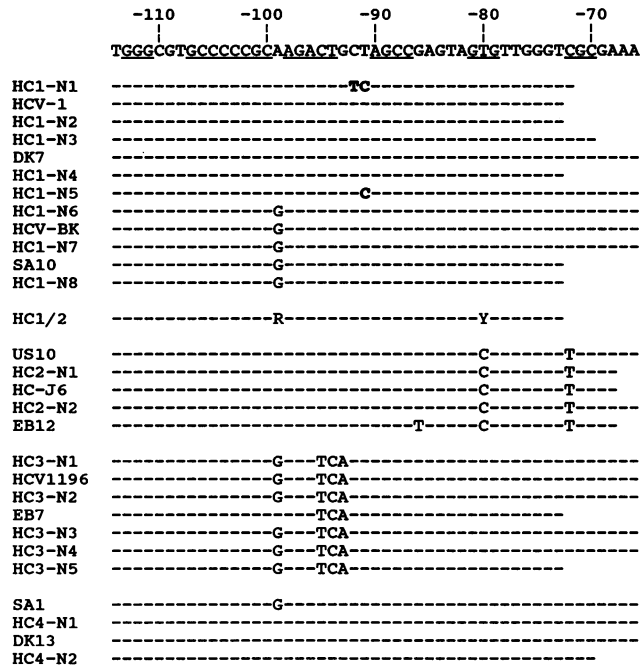


FIG. 1—Continued.

frequency of double infections which has been observed by others (24, 28) by using type-specific primers or probes.

New 5' UTR sequences were observed in 21 of the 62 isolates analyzed. The distribution of the sequence variation was not random. Recently, the putative secondary structure and possible functional elements of the 5' UTR of the HCV genome were postulated (1, 27, 29). The defined sequence motifs 1 and 2 show partial complementarity and are able to form a stable stem-loop structure; e.g., nt A at position -170 is complementary to nt T at position -115, and nt C (or T for type 2) at position -155 is complementary to nt G (or A for type 2) at position -132. Covariance in motifs 1 and 2 consistently preserves the secondary structure. Therefore, mutations are more likely to be tolerated in single-stranded regions. This is confirmed by the relatively high mutation rate in the single-stranded RNA loop between positions

TABLE 1. Mean nucleotide variation among four HCV genotypes in the entire 197-bp 5' UTR fragment and motifs 1 and 2

Fragment and genotype	No. of isolates ^a	% Nucleotide variation			
		1	2	3	4
197-bp 5' UTR					
1	34	2.5			
2	11	7.1	2.0		
3	12	7.1	11.2	1.0	
4	4	4.1	6.1	6.1	2.0
Motifs 1 and 2					
1	34	0			
2	11	28.0	2.5		
3	12	28.0	35.5	0.3	
4	4	12.5	23.5	15.8	1.5

^a Sixty-one HCV isolates were analyzed for sequence variation in 197-bp fragments from the 5' UTR (nt -262 to -66).

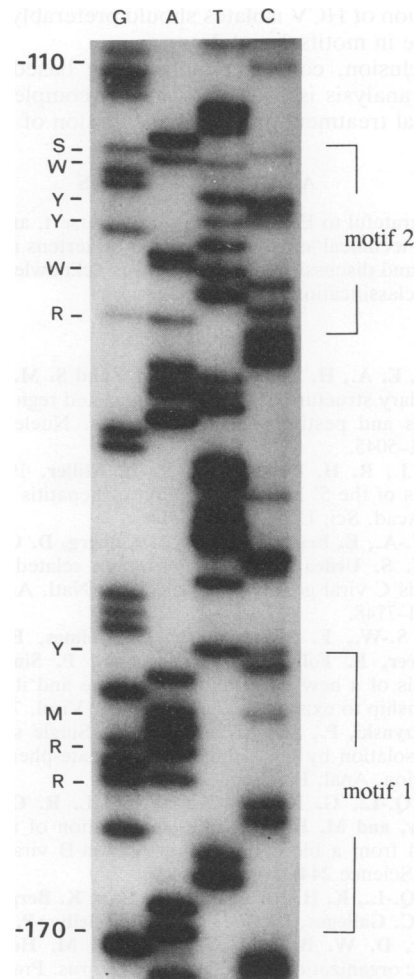


FIG. 2. Direct sequence analysis of the 5' UTR from isolate HC1/2 with the sense primer NCR3. Abbreviations: S, G or C; W, A or T; Y, T or C; R, A or G; and M, A or C.

-136 and -151. Most of the point mutations are located outside of the defined sequence motifs and do not affect classification. Only two isolates showed single point mutations within the two motifs: HC2-N2 contained a G at position -127, and HC3-N2 contained a C at position -118. Despite these variations, classification of HC2-N2 and HC3-N2 as types 2 and 3, respectively, was obvious.

The overall sequence variation between the 197-bp 5' UTR fragments was not statistically significant, because only 6 to 21 mutations were observed among the four genotypes. Because the mutations were not randomly distributed along the 5' UTR, it is impossible to apply regular statistical methods to this problem, because these require a random distribution of events. The secondary structure, i.e., the stem-loop structure formed by the partial complementarity of motifs 1 and 2, must be maintained (with a sufficiently low free energy) to function properly as an internal ribosomal entry site (27). Therefore, mutations in these motifs have functional restrictions. Comparison of the 5' UTR sequences described in this report reveals the existence of covariants. These complementary mutations in motifs 1 and 2 indeed completely conserve the secondary structure, indicating the significance of covariance. On the basis of these findings,

classification of HCV isolates should preferably be based on covariance in motifs 1 and 2.

In conclusion, consistent genotyping based on 5' UTR sequence analysis is possible and may complement studies on antiviral treatment and the transmission of HCV.

ACKNOWLEDGMENTS

We are grateful to E. Schuphof, G. de Mutsert, and P. Schrijnemakers for technical assistance and to G. Maertens and L. Stuyver for advice and discussion. P. Simmonds is acknowledged for information on classification.

REFERENCES

- Brown, E. A., H. Zhang, L.-H. Ping, and S. M. Lemon. 1992. Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res.* **20**:5041-5045.
- Bukh, J., R. H. Purcell, and R. H. Miller. 1992. Sequence analysis of the 5' noncoding region of hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **89**:4942-4946.
- Cha, T.-A., E. Beall, B. Irvine, J. Kolberg, D. Chien, G. Kuo, and M. S. Urdea. 1992. At least five related, but distinct, hepatitis C viral genotypes exist. *Proc. Natl. Acad. Sci. USA* **89**:7144-7148.
- Chan, S.-W., F. McOmish, E. C. Holmes, B. Dow, J. F. Peutherer, E. Follett, P. L. Yap, and P. Simmonds. 1992. Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants. *J. Gen. Virol.* **73**:1131-1141.
- Chomczynski, P., and N. Sacchi. 1987. Single step method of RNA isolation by guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**:152-159.
- Choo, Q.-L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**:359-362.
- Choo, Q.-L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, A. Medina-Selby, P. J. Barr, A. J. Weiner, D. W. Bradley, G. Kuo, and M. Houghton. 1991. Genetic organization of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **88**:2451-2455.
- Cuyppers, H. T. M., I. N. Winkel, C. L. van der Poel, H. W. Reesink, P. N. Lelie, M. Houghton, and A. Weiner. 1991. Analysis of genomic variability of hepatitis C virus. *J. Hepat.* **13**(Suppl. 4):S15-S19.
- Enomoto, N., A. Takada, T. Nakao, and T. Date. 1990. There are two major types of hepatitis C virus in Japan. *Biochem. Biophys. Res. Commun.* **170**:1021-1025.
- Garson, J. A., C. Ring, P. Tuke, and R. S. Tedder. 1990. Enhanced detection by PCR of hepatitis C virus RNA. *Lancet* **336**:878-879.
- Han, J. H., V. Shyamala, K. H. Richman, M. J. Brauer, B. Irvine, M. S. Urdea, P. Tekamp-Olsen, G. Kuo, Q.-L. Choo, and M. Houghton. 1991. Characterization of the terminal regions of hepatitis C viral RNA: identification of conserved sequences in the 5' untranslated region and poly(A) tails at the 3' end. *Proc. Natl. Acad. Sci. USA* **88**:1711-1715.
- Houghton, M., A. Weiner, J. Han, G. Kuo, and Q.-L. Choo. 1991. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. *Hepatology* **14**:381-387.
- Hultman, T., S. Bergh, T. Moks, and M. Uhlén. 1991. Bidirectional solid-phase sequencing of in vitro-amplified plasmid DNA. *BioTechniques* **10**:84-93.
- Kato, N., M. Hijikata, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi, T. Sugimura, and K. Shimotohno. 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.
- Kleter, G. E. M., J. T. Brouwer, R. A. Heijtkink, S. W. Schalm, and W. G. Quint. 1993. Detection of hepatitis C virus RNA in patients with chronic hepatitis C virus infections during and after therapy with alpha-interferon. *Antimicrob. Agents Chemother.* **37**:595-597.
- Kremsdorf, D., C. Porchon, and C. Brechot. 1991. Hepatitis C virus (HCV)-RNA in non-A, non-B chronic hepatitis in France. *J. Hepat.* **13**(Suppl. 4):S24-S32.
- Lee, C.-H., C. Cheng, J. Wang, and L. Lumeng. 1992. Identification of hepatitis C viruses with a nonconserved sequence of the 5' untranslated region. *J. Clin. Microbiol.* **30**:1602-1604.
- Mori, S., N. Kato, A. Yagyo, T. Tanaka, Y. Ikeda, B. Petchclai, P. Chiewsilp, T. Kurimura, and K. Shimotohno. 1992. A new type of hepatitis C virus in patients in Thailand. *Biochem. Biophys. Res. Commun.* **183**:334-342.
- Ogata, N., H. J. Alter, R. H. Miller, and R. H. Purcell. 1991. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **88**:3392-3396.
- Okamoto, H., M. Kojima, S.-I. Okada, H. Yoshizawa, H. Lizuka, T. Tanaka, E. E. Muchmore, D. A. Peterson, Y. Ito, and S. Mishiroy. 1992. Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability. *Virology* **190**:894-899.
- Okamoto, H., K. Kurai, S.-I. Okada, K. Yamamoto, H. Lizuka, T. Tanaka, S. Fukuda, F. Tsuda, and S. Mishiroy. 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.
- Okamoto, H., S. Okada, Y. Sugiyama, K. Kurai, H. Lizuka, A. Machida, Y. Miyakawa, and M. Mayumi. 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**:2697-2704.
- Okamoto, H., S. Okada, Y. Sugiyama, S. Yotsumoto, T. Tanaka, H. Yoshizawa, F. Tsuda, Y. Miyakawa, and M. Mayumi. 1990. The 5' terminal sequence of the hepatitis C virus genome. *Jpn. J. Exp. Med.* **60**:167-177.
- Okamoto, H., Y. Sugiyama, S. Okada, K. Kurai, Y. Akahane, Y. Sugai, T. Tanaka, K. Sato, F. Tsuda, Y. Miyakawa, and M. Mayumi. 1992. Typing hepatitis C virus by polymerase chain reaction with type-specific primers: application to clinical surveys and tracing infectious sources. *J. Gen. Virol.* **73**:673-679.
- Pozatto, G., M. Moretti, F. Franzin, L. S. Crocè, C. Tiribelli, T. Masayu, S. Kaneko, M. Unoura, and K. Kobayashi. 1991. Severity of liver disease with different hepatitis C viral clones. *Lancet* **338**:509.
- 25a. Simmonds, P., et al. Unpublished data.
- Takamizawa, A., C. Mori, I. Fuke, S. Manabe, S. Murakami, J. Fujita, E. Onishi, T. Andoh, I. Yoshida, and H. Okayama. 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. *J. Virol.* **65**:1105-1113.
- Tsukiyama-Kohara, K., N. Lizuka, M. Kohara, and A. Nomoto. 1992. Internal ribosomal entry site within hepatitis C virus RNA. *J. Virol.* **66**:1476-1483.
- Yoshioka, K., K. Kakumu, T. Wakita, T. Ishikawa, Y. Itoh, M. Takayanagi, Y. Higashi, M. Shibata, and T. Morishima. 1992. Detection of hepatitis C virus by polymerase chain reaction and response to interferon- α therapy: relationship to genotypes of hepatitis C virus. *Hepatology* **16**:293-299.
- Yoo, B. J., R. R. Spaete, A. P. Geballe, M. Selby, M. Houghton, and J. H. Han. 1992. 5' end-dependent translation initiation of hepatitis C virus RNA and the presence of putative positive and negative translational control elements within the 5' untranslated region. *Virology* **191**:889-899.