VIROLOGY

Detection and analysis of hepatitis C virus by a combined RT-PCR method: variation in the 5' non-coding region of the viral genome

A. KARACHRISTOS*†, S. LINARDOPOULOS†‡, M. ERGAZAKI*† and D. A. SPANDIDOS*†§

*Laboratory of Clinical Virology, Medical School, University of Crete, Heraklion, †Laboratory of Molecular Oncology and Biotechnology, National Hellenic Research Foundation, 48 Vas, Constantinou Ave., Athens, Greece and ‡The Beatson Institute for Cancer Research, Garscube Estate, Bearsden, Glasgow, Scotland

Summary. A combined reverse transcription-polymerase chain reaction (RT-PCR) method was employed for the detection of hepatitis C virus (HCV) RNA in serum from patients with chronic active hepatitis, with primers corresponding to the 5' non-coding region. The diagnosis was based on serological and biochemical methods and on liver biopsy. HCV-RNA was detected in 27 (90%) of 30 sera examined. The nucleotide sequence of PCR-amplified HCV cDNAs (256 bp) was determined from five specimens and heterogeneity varying between 0.58% and 2.89% among the clinical samples and the prototype HCV-1 was found.

Introduction

The discovery of hepatitis C virus (HCV)¹ and the development of an enzyme-linked immunoblot assay with the C100-3 HCV-antigen² revealed that HCV is the major causative agent of non-A, non-B post-transfusion hepatitis world-wide.^{2,3} Recently the complete nucleotide sequence of different HCV isolates has been determined⁴⁻⁹ and comparison of these reports led to the identification of distinct HCV genotypes. Analyses of the genomic structure of HCV have led to the conclusion that the 5' and 3' termini include non-coding regions (NC regions) that flank a single large open reading frame that encodes structural as well as non-structural proteins.

Nucleotide determination of the 5' NC region of isolates from around the world showed that this region is highly conserved among different strains;¹⁰ consequently, this region, consisting of *c.* 341 nucleotides, became the template of choice for PCR primer design. Further research showed variability of the 5' NC region and led to the definition of three major groups or types of HCV (HCV 1,2,3)^{11,12} and a fourth type¹³ that consisted of variants that cannot be classified in any of the previous types. Phylogenetic analysis of core, E1 and NS5 regions extended the classification of HCV into six major types and several subtypes.^{14,15} Sequence variability in the E1 gene among 51 isolates

collected world-wide proved the existence of at least 12 different genotypes.¹⁶ At least two more classification systems for HCV variants have been reported.^{9,17,18}

The present study investigated sequence heterogeneity of the 5' NC region, among five isolates derived from patients with chronic active hepatitis from Crete, Greece.

For reverse transcription of viral RNA and polymerase chain reaction amplification (RT-PCR), a combined assay was used in which these two steps are performed in a single tube and with the same enzyme. This method employs the DNA polymerase of the bacterium *Thermus thermophilus* (Tth polymerase) which has efficient reverse transcriptase activity in the presence of manganese.¹⁹ The use of Tth polymerase for the detection of HCV has been shown to be a convenient and specific method with increased efficiency, which additionally reduces the possibility of carry-over contamination.²⁰

Materials and methods

Serum samples

Serum samples were obtained from 30 anti-HCV-positive patients who had had elevated alanine aminotransferase levels for at least 6 months. All the patients had undergone liver biopsy that showed chronic active hepatitis. A second generation enzyme-linked immunosorbent assay (E. I. A. II, Abbot Laboratories) was

Received 7 Sept. 1994; revised version accepted 1 Nov. 1994.
§ Correspondence should be sent to Professor D. A. Spandidos.

Table. Synthetic oligonucleotide sequences*

Set-1		
External		Positions†
Sense	ACTGTCTTCACGCAGAAAGCGTCTAGCCAT	–285 to –256
Antisense	GAGACCTCCCGGGGCACTCGCAAGCACCC	–14 to –43
Internal		
Sense	ACGCAGAAAGCGTCTAGCCATGGCGTTAGT	–276 to –247
Antisense	TCCCGGGGCACTCGCAAGCACCTATCAGG	–21 to –50
Set-2		
Sense	CTGTCTTCACGCAGAAAGCG	–284 to –265
Antisense	GCACTCGCACCTATC	–28 to –47

(Set-2 oligonucleotides are a subset of Set-1 external).

*Bukh *et al.*²²

†Position 1 is the start of the genome polyprotein.

used for detection of antibody to HCV and the results were confirmed by a Western-blot assay (Matrix, Abbott Laboratories).

All sera were stored at -70°C immediately after collection and kept there until required. All the precautions proposed by Busch *et al.*²¹ for the handling and storage of sera were taken.

RNA extraction

HCV RNA was extracted by a modified RNazol-B method (Biogenesis). Briefly, 100 μl of serum were mixed with 0.9 ml of RNazol-B. After addition of 100 μl of chloroform, samples were centrifuged and the aqueous phase was precipitated with an equal volume of isopropanol. After centrifugation, the RNA pellets were washed once with ethanol 75% and resuspended in 20 μl of diethylpyrocarbonate-treated water.

cDNA synthesis and first round of amplification

The combined reverse transcription and PCR amplification was performed initially in 50- μl reaction volumes containing 10 μl of RNA solution, 400 μM of each of the four deoxyribonucleoside triphosphates (dNTPs), 50 pmol of the external antisense primer, 1 mM MnCl_2 , 2.5 U of *T. thermophilus* polymerase (Tth, Advanced Biotechnology), 20 mM Tris-HCl, pH 8.4, and 50 mM KCl. The reactions were performed in a 480 thermocycler (Perkin-Elmer) with incubation at 70°C for 15 min for the reverse transcription step. Denaturation of cDNA was at 95°C for 1 min and another 50- μl mixture was added, to a final volume of 100 μl , containing 50 pmoles of the external sense primer, 0.75 mM EGTA, 37.5 mM Tris-HCl, pH 8.8, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 1 mM MgCl_2 and Tween 0.005% w/v. The amplification consisted of 1 min at 94°C , 1 min at 45°C and 1 min at 72°C for 40 cycles. A final extension step of 5 min at 72°C followed the amplification reaction. A second PCR reaction was performed with set-2 oligonucleotides as follows: reverse transcription as described above, followed by 1 min at 94°C , 1 min at 60°C and 1 min at 72°C , for 35 cycles. The set of primers used are shown in the table.

Nested PCR and sequence determination of the 5' non-coding region of hepatitis C virus

A second round of amplification with internal primers was performed for c. 10% of cases examined in a 100- μl reaction mixture containing 50 pmoles of each internal primer, 5 μl of PCR product, 2.5 U Tth polymerase, 200 mM of each dNTP, 75 mM Tris-HCl, pH 8.8, 20 mM $(\text{NH}_4)_2\text{SO}_4$, 1.5 mM MgCl_2 and Tween 0.01% w/v. The PCR program was as previously described for 35 cycles.

The nested PCR products were analysed by electrophoresis in an agarose 2% gel, stained with ethidium bromide and visualised under UV illumination. The expected size of the band was 256 bp. After agarose gel electrophoresis, the 256-bp PCR products were purified and ligated into PCR 1000 vectors by the TA cloning system (Invitrogen; San Diego, CA, USA). Following transformation of *INVα Escherichia coli* cells, plasmids with DNA inserts (white colonies) were isolated and subject to double-stranded sequencing with T7 polymerase with the Sequenase Kit, Version 2.0 (US Biochemical Corp., Cleveland, OH, USA), according to the manufacturer's instructions.

Results

Detection of HCV-RNA by a combined RT-PCR method

HCV sequences were detected with both primer sets 1 and 2 (table) in 27 (90%) of 30 specimens examined (see fig. 1 as an example). The sets of primers used were similar and correspond to the 5' NC region. In three samples (10%) no amplified PCR product was found.

Nucleotide sequence of the 5' non-coding region from 5 HCV isolates

The nucleotide sequence of the 5' NC region of the HCV genome from a number of isolates obtained throughout the island of Crete was analysed. HCV RNA was reverse-transcribed and the resultant cDNA was PCR-amplified and sequenced to obtain the "consensus" sequence in each serum sample. The

Fig. 1. HCV amplified products with primer set 1. **A**, amplification of HCV with external primers. Lane 1, PUC18/*Hae*III (marker); 2–6, positive samples (272 bp); 5, 7–10, negative samples. **B**, amplification of the PCR products shown in **A** with internal primers. Lane 1, pUC18/*Hae*III; 2, negative control; 3, 4, 5, 7, 8, 11, positive samples (256 bp) (8, 11 shown in **A** in lanes 7, 10 as negative samples with external primers); 6, 9, 10, negative samples.

Fig. 2. Alignment of nucleotide sequences of the 5' NC region of five HCV isolates. The sequences are compared to the prototype HCV sequence (HCV-1)^{4,10} shown on the top line. Nucleotide substitutions are indicated as uppercase letters.

Fig. 3. Identification of HCV base substitutions. The sequences correspond to the 5'–3' direction of the 5' NC region of the HCV gene from top to bottom.

mutational spectrum of the 5' NC region observed in the five HCV isolates was compared to the prototype HCV-1 and is presented in fig. 2. Previously published multiple sequence alignments by others^{4,10} demonstrated that the 5' NC region is highly conserved and up to 98% homologous to the prototype isolate HCV-1. In contrast, in this study, several HCV isolates with sequences variability were identified. Three of five cases had a single substitution (A → G at position –99 for cases 9 and 14, and A → G at position –163 for case 11), one case (16) had six substitutions at positions –167 T → C, –163 A → G, –159 C → T, –122 T → C, –118 G → A and –94 T → C, and case 18 had two substitutions at positions –182 T → A and –161 G → A (fig. 3). No nucleotide deletions or insertions were detected.

Sequencing of the 5' NC region has shown that it includes 1–4 short open reading frames (ORFs).^{10,11} In this study one or two short ORFs were found in the region sequenced. In isolate 16, the base substitution C → T at position –159 forms an ATG codon, thus initiating an ORF extending from position –160 to –89 instead of the position –127 to –89 for HCV-1.¹⁰

Discussion

In this study HCV was detected by a sensitive combined RT-PCR in serum of patients with chronic active hepatitis. The need for nested PCR or a second round of amplification was 10% for the first method described and was not required for the second method. Furthermore, the use of the same enzyme for reverse transcription and PCR amplification decreases the danger of carry-over contamination and diminishes the time required for the procedure.

The nucleotide sequence of five PCR products corresponding to the 5' NC region was determined. Although this region is highly conserved among HCV isolates from around the world,¹⁰ it has been reported

that it includes variable domains that can be used to segregate HCV isolates into six major types (type 1–6).^{12–15} All types of HCV seem to have a large geographical distribution—except type 4 which has mainly been found in Zaire, Egypt and Burundi,^{11,23,24} consequently suggesting that this type may have an African distribution—and type 6 which has been found in Hong-Kong.²³ In this study, the nucleotide distance between isolate 16 and type 4 was 0.0096,²⁵ within the region sequenced, thus it clusters with this type. The other isolates are closely related to type 1.

Nucleotide variation among the isolates studied and HCV-1 varied from 0.58% to 2.89%. The greatest variability inside the 5' NC region has been reported to regions spanning positions –167 to –118 and –100 to –72.¹¹ All substitutions reported in the present study were in these variable domains, except one at position –182 (T → A), which is a highly conserved domain of the 5' NC region. This substitution has not been reported previously. Although there is a significant sequence variation in the 5' NC region, there is evidence that the secondary structure includes a conserved element¹³ that forms binding sites essential for internal ribosome entry and possibly, by this mechanism, enables the initiation of translation.^{26,27} This model is supported by the presence of several ATG codons upstream of the correct initiation codon.

Several investigators have described short ORFs in the 5' NC region.^{10,11} The number of ORFs in our isolates cannot yet be calculated because ORFs in regions that we have not sequenced have also been described. However, 1 or 2 short ORFs were found within the region analysed and isolate 16 revealed a different ORF pattern from HCV-1. Interestingly, there is a starting codon at position –160 that forms an unusual arrangement of 23 amino acids. Although it has been speculated that the short ORFs in the 5' NC region have negative control in translation of HCV,¹⁰ their function remains to be clarified.

We thank T. Liloglou for advice regarding the PCR technique and Professor J. C. Neil for critical reading of the manuscript.

References

1. Choo Q-L, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a 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 *et al.* An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* 1989; **244**: 362–364.
3. Alter HJ, Purcell RH, Shih JW *et al.* Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute chronic non-A, non-B hepatitis. *N Engl J Med* 1989; **321**: 1494–1500.
4. Choo Q-L, Richman KH, Han JH *et al.* Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci USA* 1991; **88**: 2451–2455.
5. Kato N, Hijikata M, Ootsuyama Y *et al.* 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.
6. Okamoto H, Okada S, Sugiyama Y *et al.* 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* 1991; **72**: 2697–2704.
7. Takamizawa A, Mori C, Fuke I *et al.* Structure and organization of the hepatitis C virus genome isolated from human carriers. *J Virol* 1991; **65**: 1105–1113.
8. Chen PJ, Lin MH, Tai KF, Liu PC, Lin C-J, Chen DS. The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5' termini of viral genomic and antigenomic RNA. *Virology* 1992; **188**: 102–113.
9. Okamoto H, Kurai K, Okada S *et al.* Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. *Virology* 1992; **188**: 331–341.
10. Han JH, Shyamala V, Richman KH *et al.* 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* 1991; **88**: 1711–1715.
11. Bukh J, Purcell RH, Miller RH. Sequence analysis of the 5' noncoding region of hepatitis C virus. *Proc Natl Acad Sci USA* 1991; **89**: 4942–4946.
12. Chan S-W, McOmish F, Holmes EC *et al.* Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants. *J Gen Virol* 1992; **73**: 1131–1141.
13. Simmonds P, McOmish F, Yap PL *et al.* 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* 1993; **74**: 661–668.
14. Simmonds P, Holmes EC, Cha TA *et al.* 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* 1993; **74**: 2391–2399.
15. Simmonds P, Smith DB, McOmish F *et al.* Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *J Gen Virol* 1994; **75**: 1053–1061.
16. Bukh J, Purcell R, Miller RH. At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene isolates collected worldwide. *Proc Natl Acad Sci USA* 1993; **90**: 8234–8238.
17. Houghton M, Weiner A, Han J, Kuo G, Choo Q-L. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. *Hepatology* 1991; **14**: 381–388.
18. Mori S, Kato N, Yagyu A *et al.* A new type of hepatitis C virus in patients in Thailand. *Biochem Biophys Res Commun* 1992; **183**: 334–342.
19. Myers TW, Gelfand DH. Reverse transcription and DNA amplification by a *Thermus thermophilus* DNA polymerase. *Biochemistry* 1991; **30**: 7661–7666.
20. Young KKY, Resnick RM, Myers TW. Detection of the hepatitis C virus RNA by a combined reverse transcription-polymerase chain reaction assay. *J Clin Microbiol* 1993; **31**: 882–886.
21. Busch MP, Wilber JC, Johnson P, Tobler L, Evans CS. Impact of specimen handling and storage on hepatitis C virus RNA. *Transfusion* 1992; **32**: 420–425.
22. Bukh J, Purcell RH, Miller RH. Importance of primer selection for the detection of hepatitis C virus RNA with the polymerase chain reaction assay. *Proc Natl Acad Sci USA* 1992; **89**: 187–191.
23. McOmish F, Yap PL, Dow BC *et al.* Geographical distribution of hepatitis C virus genotypes in blood donors: an international collaborative survey. *J Clin Microbiol* 1994; **32**: 884–892.
24. Stuyver L, Rossau R, Wyseur A *et al.* Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. *J Gen Virol* 1993; **74**: 1093–1102.
25. Nei M (ed). Molecular evolutionary genetics. New York, Columbia University Press. 1987: 64–110.
26. Brown EA, Zhang H, Ping L-H, Lemon SM. Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res* 1992; **20**: 5041–5045.
27. Tsukiyama-Kohara K, Iizuka N, Kohara M, Nomoto A. Internal ribosome entry site within hepatitis C virus RNA. *J Virol* 1992; **66**: 1476–1483.