efficacy, other options, including combination therapy by pegylated IFN [such as PEG-interferon-α2b (PEG-IFN)] with antiviral drugs (such as ribavirin) have been developed. PEG-IFN plus ribavirin is significantly more effective than IFN plus ribavirin or PEG-IFN alone for the treatment of chronic hepatitis C, with sustained virological response rates across hepatitis C genotypes ranging from 48% in patients infected with genotype 1 to 88% in those infected with genotype 2 or 3 (5).

2′-5′-Oligoadenylate synthetase (2-5AS), first discovered in IFN treated cells, polymerizes ATP into 2′-5′-oligoadenylate (2-5A), which is probably related to antiviral activity and perhaps the anti-proliferative effect of IFN, now called the 2-5AS system (6, 7). A number of reports has shown the relation between the level of 2-5AS and the effect of IFN on chronic hepatitis C, signifying that the measurement of 2-5AS activity could be useful for the evaluation of IFN treatment (8–10).

This study is aimed at assessing whether 2-5AS activity is a useful clinical marker of virological response to PEG-IFN plus ribavirin therapy in patients with chronic hepatitis C.

METHODS

Patients

From December 2004 to June 2005, the 32 consecutive patients included in this study had high viral loads of serum HCV-RNA of genotype 1b and had been diagnosed with chronic hepatitis C on the basis of abnormal serum ALT for at least 6 months, and with positive HCV-RNA assessed by polymerase chain reaction (PCR). None of the patients was positive for hepatitis B surface antigen or other liver diseases (autoimmune hepatitis, alcoholic liver disease). The study protocol conformed to the ethical guideline of the 1975 Declaration of Helsinki, and was approved by the ethics committee of our hospital. All the patients gave written informed consent before treatment. All the patients received a regimen of PEG-IFNα-2b (Peg-Intron; Schering-Plough, Kenilworth, NJ, USA) (1·5μg/kg/week, subcutaneously) in combination with ribavirin (Rebetol; Schering-Plough) (600–1000 mg/day) for 48 weeks. Ribavirin was administered at a dose of 600 mg/day (three capsules) in patients weighing <60 kg, 800 mg/day (four capsules) in patients weighing <80 kg but not <60 kg, and 1000 mg/day (five capsules) in those weighing ⩾80 kg. The patients were then divided into two groups as follows: one group (effective group) demonstrated undetectable serum HCV-RNA at 24 weeks ($n = 22$); the other group (ineffective group) demonstrated persistent presence of HCV-RNA in serum at 24 weeks ($n = 10$).

The genotypes of HCV were identified by PCR using type-specific primers from the core region, as previously described (11).

Hepatitis C virus-RNA was quantified using the Amplicor HCV Monitor v 2·0 assay (Roche Molecular Systems Inc., Pleasanton, CA, USA) with a lower detection limit of 50 IU/mL. Serum HCV core antigen quantification was assessed by the immuno-radiometric assay (IRMA) (Ortho Clinical diagnostics, Tokyo, Japan) with a lower detection limit of 20 fg/mL.

Serum HCV core antigen was measured 24 h, 1 week, 2 weeks and 4 weeks before initial administration and after the combination therapy.

The 2-5AS activity in serum was assessed with the use of the RIA kit (Eiken Immunochemical Laboratory, Tokyo, Japan) with a lower detection limit of 10 pmol/dL, and was measured 2, 8 and 12 weeks before initial administration and after the combination therapy.

Statistical analysis

Differences between the groups were assessed by the χ^2-test, Student's t-test and the Mann–Whitney test. $P < 0{\cdot}05$ was considered statistically significant.

RESULTS

The baseline clinical characteristics of the patients are summarized in Table 1. There were no significant differences in any of the characteristics between the two groups.

Hepatitis C virus core antigen log reduction in viral load from the baseline at 24 h after administration was 1·08 ± 0·43 in the effective group and 0·62 ± 0·28 in the ineffective group ($P < 0{\cdot}01$); 0·89 ± 0·56 and 0·22 ± 0·23 at 1 week ($P < 0{\cdot}005$); 1·42 ± 0·74 and 0·36 ± 0·34 at 2 weeks ($P < 0{\cdot}0005$); 1·97 ± 0·64 and 0·74 ± 0·40 at 4 weeks ($P < 0{\cdot}0005$).

© 2006 The authors. Journal compilation © 2006 Blackwell Publishing Ltd,

Table 1. Baseline clinical characteristics of the patients

	Effective group	Ineffective group	P-value
Gender			
Male	12	7	
Female	10	3	NS
Age (years)	54·4 ± 11·1	61·0 ± 8·3	NS
Body weight (kg)	58·0 ± 10·5	62·5 ± 11·3	NS
Treatment-naive	8	3	
Retreatment	14	7	NS
HCV-RNA level (kIU/mL)	1417·7 ± 1156·7	1647·0 ± 1623·7	NS
HCV core antigen (fmol/L)	6025 ± 4954	13735 ± 18062	NS
2-5AS (pmol/dL)	73·4 ± 42·6	135·3 ± 85·5	NS

Data are expressed as mean ± SD.
2-5AS, 2′-5′-oligoadenylate synthetase; HCV, hepatitis C virus; NS, not significant.

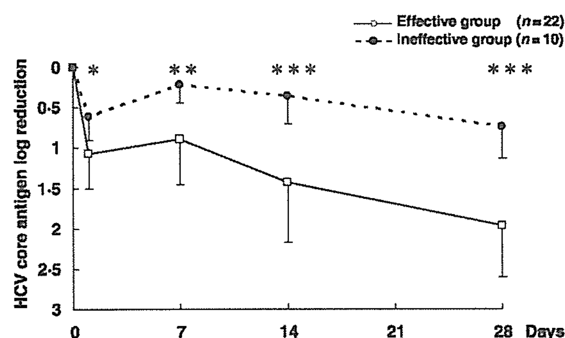

Fig. 1. Changes in hepatitis C virus (HCV) core antigen expressed as log reduction in chronic hepatitis C patients treated with PEG-IFN-α2b plus ribavirin therapy for 28 days. HCV core antigen log reduction in the effective group (open square) and in the ineffective group (solid circle). *$P < 0·01$ **$P < 0·005$ ***$P < 0·0005$ (mean ± SD).

There was a significant difference between the two groups at every assessment point (Fig. 1).

The 2-5AS levels before IFN treatment were 73·4 ± 42·6 pmol/dL in the effective group and 135·3 ± 85·5 pmol/dL in the ineffective group, with no significant difference between the two groups. The 2-5AS activity at 2 weeks after administration was 182·3 ± 51·7 pmol/dL in the effective group and 175·0 ± 105·2 pmol/dL in the ineffective group; 183·0 ± 61·8 and 158·1 ± 75·6 pmol/dL at 8 weeks; 169·2 ± 42·5 and 151·3 ± 67·6 pmol/dL at 12 weeks. There was no significant difference between the two groups (Fig. 2). The 2-5AS response ratio (measured value/measured value of baseline 2-5AS) at 2 weeks after administration was

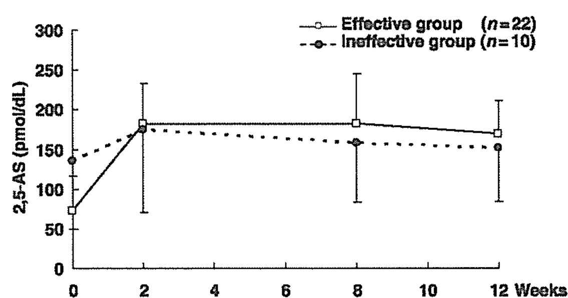

Fig. 2. Changes in 2′-5′-oligoadenylate synthetase (2-5AS) activity in chronic hepatitis C patients treated with PEG-IFN-α2b plus ribavirin therapy at 12 weeks. 2-5AS activity in the effective group (open square) and in the ineffective group (solid circle).

3·02 ± 1·45 in the effective group and 1·81 ± 1·06 in the ineffective group ($P < 0·05$); 3·12 ± 1·91 and 1·61 ± 1·41 at 8 weeks ($P < 0·005$); 2·86 ± 1·53 and 1·43 ± 0·87 at 12 weeks ($P < 0·005$). There was a significant difference between the two groups (Fig. 3).

DISCUSSION

Early loss and persistence of serum HCV-RNA predict positive and negative sustained virological response respectively to IFN-based therapy. Compared with standard and previous IFN treatment, combination therapy with IFN plus ribavirin (12) or with PEG-IFN plus ribavirin (13) enhances clearance rates of HCV-RNA.

Clearance of HCV-RNA in the early stages of treatment demonstrates IFN efficacy (14–16), and

© 2006 The authors. Journal compilation © 2006 Blackwell Publishing Ltd,

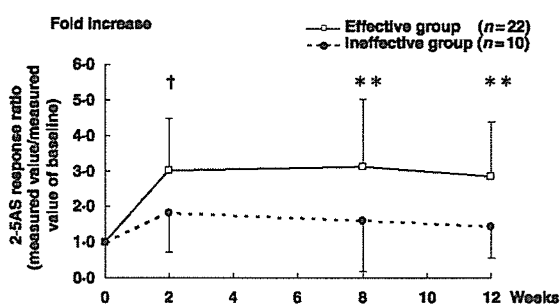

Fig. 3. Changes in 2′-5′-oligoadenylate synthetase (2-5AS) response ratio (measured value/measured value of baseline) in chronic hepatitis C patients treated with PEG-IFN-α2b plus ribavirin therapy at 12 weeks. 2-5AS response ratio in the effective group (open square) and in the ineffective group (solid circle). †$P < 0{\cdot}05$ **$P < 0{\cdot}005$ (mean ± SD).

patients who achieve complete response show early HCV-RNA clearance on IFN treatment, regardless of IFN administration schedules (17, 18). Results of combination therapy suggest that ribavirin administration in concert with IFN therapy is effective in both enhancing viral suppression in infected cells and preventing the reappearance of the virus after the end of therapy, by promoting the clearance of HCV-infected cells (19). Administration of IFN produces biphasic exponential decay of the viral load in vivo: a rapid first phase and a much slower second phase (14, 15). The first phase correlates with the clearance of free virus particles from serum through the inhibition of viral replication by the direct, non-specific action of IFN, whereas the second phase is dependent on the elimination rate of HCV-infected cells by the IFN-stimulated host immune system, and is predictive of sustained virological response (20, 21).

Monitoring early viral kinetics has recently been suggested for earlier identification of the likelihood of response in combination therapy with PEG-IFN plus ribavirin.

To achieve a sustained virological response (no detectable HCV-RNA 24 weeks after cessation of therapy), HCV-RNA negativity should be acquired at the latest 24 weeks after receiving PEG-IFN plus ribavirin (5).

Although all the mechanisms of the elimination of HCV by IFN have not been fully elucidated, the 2-5A system is one of the mechanisms of the antiviral effect of IFN, and 2-5AS is one of the antiviral enzymes induced by IFN through activation of RNase (6, 7). Therefore, the measurement of 2-5AS activity could be useful for the evaluation of IFN treatment (8–10). The 2-5AS level before IFN treatment in the ineffective group was higher than in the effective group but not significant so. The reason remains unclear. This might be associated with the endogenous IFN system, which has already been activated in patients in the ineffective group. 2-5AS activity rises with IFN treatment, and its level before IFN therapy in the effective group was higher than in the ineffective group. The difference between the two groups in the induced levels of 2-5AS probably reflects individual differences.

After stimulation by IFN, the response ratio of 2-5AS (measured value/measured value of baseline 2-5AS) increased more than 3-fold in the effective group within 8 weeks, whereas it increased only about 2-fold in the ineffective group at 12 weeks.

The differences in 2-5AS response ratio in the two groups might be strongly associated with the early phase of the clearance of virus from serum by IFN stimulation. The differences of the 2-5AS response ratio in the two groups might be strongly associated with the early phase of the clearance of virus from serum by IFN stimulation. It has been shown that the level of HCV core antigen correlates with HCV RNA concentrations and, consequently, can be used as a marker of HCV replication (22). HCV core antigen log reduction in viral load from the baseline in the effective group is significantly higher than in the ineffective group at any point (23, 24), which might reflect the therapeutic effect of IFN. Recently Marek *et al.* (25) have reported that serum TGF-β1 and TGF-β1 mRNA expression in liver biopsy are useful prognostic markers in patients undergoing antiviral therapy. We have not assessed the association between TGF -β1 and antiviral therapy in this study and further studies are needed to clarify this relationship.

Our results showed that the antiviral effect of combination therapy is related to the increase of 2-5AS activity, suggesting that the activity of 2-5AS is closely related to the antiviral effect and that the measurement of the 2-5AS response ratio might be a useful clinical parameter of virological response to PEG-IFN plus ribavirin therapy in patients with chronic hepatitis C. The activity of 2-5AS is less costly and time-consuming to measure than HCV

© 2006 The authors. Journal compilation © 2006 Blackwell Publishing Ltd,

RNA quantification. Therefore it is recommended for inclusion in standard clinical practice. 2-5AS takes 2 days to measure and costs about $24 per measurement, whereas HCV RNA costs about $43·1 and takes 4 days to measure.

REFERENCES

1. Hu KQ, Vierling JM, Redeker AG (2001) Viral, host and interferon-related factors modulating the effect of interferon therapy for hepatitis C virus infection. *Journal of Viral Hepatitis*, **8**, 1–18.
2. Shiratori Y, Kato N, Yokosuka O *et al.* (1997) Prediction of the efficacy of interferon therapy in chronic hepatitis C virus infection. Tokyo-Chiba Hepatitis Research Group. *Gastroenterology*, **113**, 558–566.
3. Nakamura H, Ito H, Ogawa H *et al.* (1999) Initial daily interferon administration can gain more eradication of HCV-RNA in patients with chronic hepatitis C, especially with serum intermediate viral load. *Hepato-Gastroenterology*, **46**, 1131–1139.
4. Chayama K, Suzuki F, Tsubota A *et al.* (2001) Evaluation of quantitative measurements of hepatitis C virus RNA to predict sustained response to interferon by genotype. *Journal of Virological Methods*, **95**, 33–45.
5. Manns MP, McHutchison JG, Gordon SC *et al.* (2001) Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet*, **358**, 958–965.
6. Schmidt A, Zilberstein A, Shulman L, Federman P, Berissi H, Revel M (1978) Interferon action: isolation of nuclease F, a translation inhibitor activated by interferon-induced (2′-5′) oligo-isoadenylate. *FEBS Letters*, **95**, 257–264.
7. Nilsen TW, McCandless S, Baglioni C (1982) 2′, 5′-oligo(A)-activated endonuclease in NIH 3T3 mouse cells chronically infected with Moloney murine leukemia virus. *Virology*, **122**, 498–502.
8. Karino Y, Hige S, Matsushima M, Miyazaki T, Toyota J (1994) Significance of 2′,5′-oligoadenylate synthetase activity and HCV genotype in IFN therapy for type C chronic hepatitis. *Hokkaido Igaku Zasshi. Hokkaido Journal of Medical Science*, **69**, 1354–1359.
9. Toda K, Kumagai N, Iwabuchi N *et al.* (1997) 2-5AS activity in serum and peripheral blood mononuclear cells for chronic active hepatitis C and the relationship to clinical outcome of interferon therapy. *Japan Journal of Clinical Immunology*, **20**, 428–436.
10. Tong WB, Zhang CY, Feng BF, Tao QM (1998) Establishment of a nonradioactive assay for 2′-5′ oligoadenylate synthetase and its application in chronic hepatitis C patients receiving interferon-alpha. *World Journal of Gastroenterology*, **4**, 70–73.
11. Okamoto H, Sugiyama Y, Okada S *et al.* (1992) Typing hepatitis C virus by polymerase chain reaction with type-specific primers: application to clinical surveys and tracing infectious sources. *Journal of General Virology*, **73**, 673–679.
12. Poynard T, Marcellin P, Lee SS *et al.* (1998) Randomized trial of interferon α 2b plus ribavirin for 48 weeks or 24 weeks versus interferon α 2b plus placebo for 48 weeksfor treatment of chronic infection with hepatitis C virus. *Lancet*, **352**, 1426–1432.
13. Glue P, Rouzeir-Panis R, Raffanel C *et al.* (2000) A dose-ranging study of pegylated interferon alpha-2b and ribavirin in chronic hepatitis C. *Hepatology*, **32**, 647–653.
14. Yasui K, Okanoue T, Murakami Y *et al.* (1998) Dynamics of hepatitis C viremia following interferon-alpha administration. *Journal of Infectious Diseases*, **177**, 1475–1479.
15. Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, Perelson AS (1998) Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science*, **282**, 103–107.
16. Ampurdanes S, Olmedo E, Maluenda MD *et al.* (1996) Permanent response to alpha-interferon therapy in chronic hepatitis C is preceded by rapid clearance of HCV-RNA from serum. *Journal of Hepatology*, **25**, 827–832.
17. Orito E, Mizokami M, Suzuki K *et al.* (1995) Loss of serum HCV-RNA at week 4 of interferon alpha therapy is associated with more favorable long-term response in patients with chronic hepatitis C. *Journal of Medical Virology*, **46**, 109–115.
18. Lee WM, Reddy R, Tong MJ *et al.* (1998) Early hepatitis C virus-RNA responses predict interferon treatment outcomes in chronic hepatitis C. *Hepatology*, **28**, 1411–1415.
19. Izumi N, Asahina Y, Kurosaki M *et al.* (2004) A comparison of the exponential decay slope between PEG-IFN alfa-2a/ribavirin and IFN alfa-2a/ribavirin combination therapy in patients with chronic hepatitis C genotype 1b infection and a high viral load. *Intervirology*, **47**, 102–107.
20. Anonymous (2002) National Institutes of Health Consensus Development Conference Statement: Management of hepatitis C. *Hepatology*, **36**, S3.
21. Anonymous (2002) Consensus conference. Treatment of Hepatitis C. Agence Nationale d'Accreditation et d'Evaluation en Sante (ANAES). *Gastroentérologie clinique et biologique*, **26**, B303.

© 2006 The authors. Journal compilation © 2006 Blackwell Publishing Ltd,

22. Buti M, Mendez C, Schaper M *et al.* (2004) Hepatitis C virus Core Antigen as a predictor of non-response in genotype 1 chronic hepatitis C patients treated with peginterferonα-2b plus ribavirin. *Journal of Hepatology*, **40**, 527–532.
23. Veillon P, Payan C, Pcchio G, Maniez-Montreiul M, Guntz P, Lunel F (2003) Comparative evaluation of the total hepatitis C virus core antigen, branched-DNA, and amplicor monitor assays in determining viremia for patients with chronic hepatitis C during interferon plus ribavirin combination therapy. *Journal of Clinical Microbiology*, **41**, 3212–3220.
24. Zanetti AR, Romano L, Brunetto M, Colombo M, Bellati G, Tackney C (2003) Total HCV core antigen assay: a new marker of hepatitis C viremia for monitoring the progress of therapy. *Journal of Medical Virology*, **70**, 27–30.
25. Marek B, Kajdaniuk D, Mazurek U *et al.* (2005) TGF-β1 mRNA expression in liver biopsy specimens and TGF-β1 serum levels in patients with chronic hepatitis C before and after antiviral therapy. *Journal of Clinical Pharmacy and Therapeutics*, **30**, 271–277.

© 2006 The authors. Journal compilation © 2006 Blackwell Publishing Ltd, *Journal of Clinical Pharmacy and Therapeutics*, **31**, 441–446

Journal of Viral Hepatitis, 2007, 14, 96–106 doi:10.1111/j.1365-2893.2006.00773.x

Is hepatitis C virus NS3 protease quasispecies heterogeneity predictive of progression from cirrhosis to hepatocellular carcinoma?

S. Vallet,[1,2] S. Gouriou,[2] G. Nkontchou,[3] H. Hotta,[7] M. Vilerio,[2] M. C. Legrand-Quillien,[1] M. Beaugrand,[3] J. C. Trinchet,[3] J. B. Nousbaum,[4] P. Dény,[5] C. Gaudy,[6] A. Goudeau,[6] B. Picard[5] and C. Payan[1,2] [1]*Microbiologie, CHU Morvan, Brest, France;* [2]*Bactériologie-Virologie EA3882, Faculté de Médecine, Brest, France;* [3]*Hépato-gastroentérologie, CH J.Verdier, Bondy, France;* [4]*Hépato-gastroentérologie, CHU La Cavale Blanche, Brest, France;* [5]*Bactériologie-Virologie-Hygiène, CH Avicenne, Bobigny, France;* [6]*Virologie, CHU Bretonneau, Tours, France;* [7]*Divisions of Microbiology, Kobe University Graduate School of Medicine, Kobe, Japan*

Received December 2005; accepted for publication February 2006

SUMMARY. We investigated whether an HCV NS3 protease quasispecies heterogeneity was associated with progression from viral cirrhosis to hepatocellular carcinoma (HCC). The NS3 protease quasispecies structure of 10 HCV-1b cirrhotic patients (controls) was compared with that of 10 paired HCV-1b cirrhotic patients who displayed progression to HCC (cases). NS3 protease genetic complexity and diversity did not differ significantly between cases and controls. Amino acid substitutions were detected at 20 (11%) and 25 (14%) sites in at least two variants of the NS3 protease in cases and controls, respectively. Significant differences in the percentage of substituted clones were observed for 10 NS3 sites. Mutations Y56F, I71V, T72I, Q86P, P89S, S101G/D, R117H, S122G/T/N, V132I and V170I were more frequently observed in the NS3 protease sequences of controls than in those of cases. Residue V107 was substituted in NS3 cases but not in controls. However, these differences did not allow the definition of a specific NS3 profile related to HCC occurrence. The NS3 secondary structure B1-1 previously identified as potentially predictive of HCC was identified with a higher frequency in cases quasispecies (84.2%) than in controls (55.9%; $P < 0.05$). Our results suggest that there may be a relationship to fibrosis progression when diversity parameters are considered together with secondary structure profiles. Further investigations are required to determine the cellular interactions of HCV NS3 protease in the context of carcinogenesis.

Keywords: cirrhosis, hepatitis C virus, hepatocellular carcinoma, NS3 protease, polymorphism, quasispecies.

INTRODUCTION

Hepatitis C virus (HCV) infection is a major public health problem worldwide, with about 3% of the world's population being infected. Three to four million people become infected each year, and there are 170 million chronic carriers at risk of developing liver cirrhosis and, possibly, liver cancer. Hepatocellular carcinoma (HCC) mostly affects patients with liver cirrhosis and is currently the most common cause of death in these patients. The annual incidence of HCC varies between 1% and 4% of cirrhotic HCV-infected patients [1]. New HCV treatments are required, because bitherapy is only partly effective, particularly for HCV genotype 1 infections, and limited by undesirable side effects.

The HCV genome encodes a polyprotein of about 3000 amino acids, which is co- and postranslationally cleaved by cellular and viral proteases. The N-terminal domain of HCV NS3, which contains a 181-amino acid chymotrypsin-like serine protease, is responsible for well-ordered proteolytic cleavage at the NS3/NS4A (in cis) and NS5A/NS5B, NS4A/NS4B and NS4B/NS5A junctions (in trans) of the viral polyprotein [2–5]. The NS3 protease is unique among serine proteases in requiring the action of a virus-encoded protein cofactor, NS4A, for cleavage at the NS3/NS4A and NS4B/NS5A junctions and for increasing cleavage efficiency at the NS4A/NS4B and NS5A/NS5B junctions [2,5]. The NS3 protease is essential for HCV replication and is therefore one of the most promising targets for specific anti-HCV therapy [6,7].

Various lines of evidence suggest that the HCV NS3 protease is involved in carcinogenesis. This protein is thought to

Abbreviations: HCC, hepatocellular carcinoma; HCV, Hepatitis C virus.

Correspondence: Dr Sophie Vallet, Département de Microbiologie, Centre Hospitalier Morvan, 2 avenue Foch, 29609 Brest Cedex, France. E-mail: sophie.vallet@univ-brest.fr

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd

affect normal cellular functions, such as cell proliferation and cell death, and to be involved, either directly or indirectly, in HCV hepatocarcinogenesis. NS3 can induce the transformation of murine fibroblasts [8,9] and NS3 protease abolishes the actinomycin-D-induced apoptosis of NIH 3T3 cells [10]. NS3 protease interacts with the p53 tumour suppressor [11]; it may also specifically repress p21 promoter activity in a dose-dependent manner [12]. It was also recently suggested that analysis of the secondary structure of HCV NS3 protease might be useful for predicting the relative risk of developing HCC [13,14]. In these studies, a particular secondary conformation of the amino-terminal 120 residues of the protease was found to be significantly associated with HCC.

In the present work, we aimed to determine whether a particular nucleotide and amino acid signature pattern of the HCV-1b NS3 protease was predictive of progression from viral cirrhosis to HCC. HCV is never found as a homogeneous population of identical RNA genomes *in vivo*; instead, it exists as a mixture of closely related genomes defined as quasispecies [15–17]. We compared the NS3 protease quasispecies structure of two groups of patients as homogeneous as possible to eliminate exogenous factors that might affect cirrhosis progression: 10 HCV-1b-infected cirrhotic patients with that of 10 paired HCV-1b-infected cirrhotic patients displaying subsequent progression to HCC.

PATIENTS AND METHODS

Patients

Twenty women with histologically proved cirrhosis related to HCV-1b genotype, without detectable HCC and prospectively screened for HCC by US scan combined to serum dosage of α-foetoprotein every 4 months, were included in this case–control study. Patients were selected according to the following criteria: Caucasian origin; absence of comorbidities as alcohol abuse, HIV or HBV coinfection; no prior antiviral treatment. Ten patients (cases) who developed HCC during the follow-up were matched with 10 patients with the same age at diagnosis of cirrhosis and who had not developed HCC. These patients have been followed in a monitoring programme over a period of time equal or higher than that of the cases (controls) (Table 1). Period of follow-up corresponded to 1991–2004 between cirrhosis and HCC or non-HCC diagnoses. The diagnosis of HCC was made according to noninvasive criteria (four patients) [18] or on histology (six patients) [19]. In all cases, serums were collected at the time of the diagnosis of cirrhosis and kept at −20 °C. A consent was obtained for all patients.

Genotyping and quantification

Hepatitis C virus genotyping was carried out by 5′ noncoding region sequencing, using the Trugene HCV method (TRUGENE 5′NC; Bayer HealthCare LLC, Berkeley, CA, USA), as previously described [20]. If HCV subtype had not been determined, we sequenced the NS5B region according to the French ANRS consensus protocol [21]. Serum HCV-RNA quantification was performed by the RT-PCR-based method (Monitor HCV RNA 2.0; Roche Diagnostics, Meylan, France) with a cutoff level at 600 IU/mL.

Hepatitis C virus NS3 RNA amplification and sequencing

RNA was extracted from 140 μL of serum (QIAmp Viral RNA Mini kit; Qiagen, Courtaboeuf, France) and eluted in 60 μL of molecular biology grade water. cDNA was synthesized at 42 °C for 50 min with 200 IU of Superscript II Reverse Transcriptase (Invitrogen, Cergy Pontoise, France) and random hexamers (Applera, Courtaboeuf, France) in a final volume of 20 μL. The complete NS3 protease gene (543 nucleotides) was amplified using previously published primers G (sense: 5′-GAGCCCGTCGTCTTCTC-3′, nucleotides 3243–3259 according to HCV-J numbering) and J (anti-sense: 5′-AGGAACTTGCCGTAGGTGGAGTA-3′, 4206–4228, HCV-J numbering) [22]. Introduction of artefactual mutations was minimized by carrying out the amplification reactions with *Pfu* DNA polymerase (Stratagene, Amsterdam, The Netherlands), a high-fidelity DNA polymerase. If amplicon yield was insufficient, a nested reaction was performed with the NS3-protease-specific primers previously described: P1b0 (sense: 5′-TCGTCTTTTCTGACATGGAG-3′, 3250–3259, HCV-J numbering) and P1b1 (anti-sense: 5′-TTGTACCCTTGGGCTGCATA-3′, 4059–4078, HCV-J numbering) [23]. Amplicons were purified (High Pure PCR Product Purification Kit; Roche Molecular Biochemicals, Meylan, France). Amplicons with 3′-A overhangs were inserted into pCR 2.1™-TOPO (TOPO TA Cloning Kit; Invitrogen). Plasmids containing protease NS3 inserts were purified and sequenced on both strands with the M13 universal and M13 reverse primers, by the dideoxy chain termination method, on a MegaBace DNA Analysis System (Amersham Biosciences, Orsay, France). Electrophoregram data were analysed with the Staden package program last updated in August 2002 [24]. Each of the 20 purified PCR products were also directly sequenced.

Sequence analysis of the protease NS3 region

Quasispecies sequences were aligned with the Clustal X program [25] and the final alignment was manually edited with the alignment editor GeneDoc [26]. A consensus sequence was determined for both the case and control groups. The number of viral variants and the genetic diversity of the HCV quasispecies were assessed. Genetic complexity, defined as the number of viral variants within HCV quasispecies, was evaluated by calculating the Shannon entropy S, according to the formula $S = -\Sigma_i(p_i \ln P_i)$, where P_i is the relative frequency of each sequence in the

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd,

Table 1 Patient characteristics. HCV infection with cirrhosis and subsequent HCC occurrence (cases, patients K1–K10) or absence of HCC (controls, patients C11–C20)

	Sex	Year of contamination	Age at the time of contamination (years)	Presumed mode of contamination	Age at the time of cirrhosis diagnosis	Child –Pugh	Viral load (IU/mL)*	ALT (IU/L)	Time between cirrhosis diagnosis and HCC or duration of HCC free follow-up (years)
Cases									
K1	F	1953	7	T	53.4	A	0.82×10^6	92	2.9
K2	F	1945	ND	ND	53.6	A	4.32×10^6	45	4.5
K3	F	1981	48	T	60.7	B	2.03×10^6	110	8.6
K4	F	1948	20	T	64.1	A	7.83×10^6	175	6
K5	F	1934	9	T	68.0	A	3.83×10^6	122	7.4
K6	F	1949	25	T	69.5	A	3.96×10^6	56	5.8
K7	F	1957	20	T	64.4	A	0.13×10^6	75	3.8
K8	F	1940	70	T	71.2	A	1.46×10^6	70	10.2
K9	F	1946	73	T	73.5	A	4.93×10^6	120	8.3
K10	F	ND	ND	T	79.3	A	1.95×10^6	146	4
Median			22.5		66.2		2.93×10^6	101	5.9
Controls									
C11	F	1955	19	SUR	58	A	0.8×10^6	94	10.5
C12	F	ND	ND	ND	49.5	A	0.07×10^6	157	7
C13	F	1971	38	T	60.5	A	1.03×10^6	80	11.1
C14	F	1970	40	ACP	63.1	A	2.15×10^6	94	10.9
C15	F	ND	ND	ND	67.1	A	6370	103	11.9
C16	F	1954	28	SUR	66.7	A	0.59×10^6	233	6.7
C17	F	1976	52	T	70.1	A	0.84×10^6	46	10.1
C18	F	1951	27	T	71.1	A	4.59×10^6	41	7.8
C19	F	ND	ND	ND	69.4	A	2.03×10^6	400	8.3
C20	F	1978	57	T	78.7	A	0.3×10^6	90	5
Median			38		66.9		0.94×10^6	94	9.2

T, transfusion; ACP, acupuncture; SUR, surgical intervention; ND, data not available; ALT, alanine aminotransferase.
*IU/mL (Cobas Amplicor Monitor; Roche Diagnostics, Meylan, France).

© 2006 The Authors

viral quasispecies. The normalized entropy S_n was calculated as $S_n = S/ \ln N$, where N is the total number of sequences analysed [27]. S_n theoretically varies from 0 (all isolates identical) to 1 (all isolates different). The normalized Shannon entropy was calculated at both nucleotide and amino acid levels.

The diversity of the NS3 protease gene was evaluated using the Mega program to determine the type of substitution (version 3; Department of Biology, Arizona State University; http://www.megasoftware.net). The number of synonymous and nonsynonymous substitutions per synonymous and nonsynonymous site (d_S) and (d_N), respectively, was calculated by the Nei–Gobojori method, with Jukes–Cantor correction to account for multiple substitutions at the same site [28]. Genetic distance was estimated for the viral quasispecies of each patient, by means of pairwise comparisons, using Kimura's two-parameter gamma model of the Mega software with a gamma parameter of 0.29 [29].

Phylogenetic analysis was performed with MEGA3 software, using the neighbour-joining method, a Kimura two-parameter model and a gamma distribution ($\alpha = 0.29$). The reliability of the various inferred clades was estimated by bootstrapping (525 replicates).

Secondary structure prediction

The secondary structure of the amino-terminal 120 residues of NS3 protease was predicted by computer-assisted Robson analysis [30] with GENETIX-MAC software (version 10.1; Software Development Co., Ltd, Tokyo, Japan), as previously described by Ogata *et al.* [13,14].

Statistical analyses

Groups were compared by means of chi-squared tests for percentages and Mann–Whitney and Student's *t*-tests for quantitative variables. $P < 0.05$ was considered significant.

Nucleotide sequence GenBank accession numbers

All nucleotide sequence data have been deposited in the GenBank sequence database under accession numbers DQ338124–DQ338432.

RESULTS

Characteristics of the patients

All 20 patients were contaminated in Europe (mostly in France), via blood transfusion in most cases (Table 1). Median age at the time of cirrhosis diagnosis was 66.2 and 66.9 years for cases and controls, respectively. The duration of HCV disease progression was longer in cases than in controls (medians of 45 *vs* 23 years). All but one of the cases had a Child–Pugh cirrhosis classification of A. Patient K3 has class B cirrhosis. Viral load was almost statistically different ($P = 0.051$) between the two groups of patients (2.93×10^6 and 0.94×10^6 IU/mL for cases and controls, respectively). Median serum alanine aminotransferase level was similar, with 2.2 and 2.1 times upper limit of normal in cases and controls, respectively. The median time between the diagnosis of cirrhosis and the diagnosis of HCC or non-HCC outcome was 5.9 and 9.2 years for cases and controls, respectively.

NS3 protease quasispecies features and subsequent progression or nonprogression to HCC

Hepatitis C virus NS3 protease region was analysed both at nucleotide and amino acid levels, corresponding to fragments of 543 bases and 181 residues, respectively. Inter- and intrasamples' variability of sequences was assessed. A median of 15 NS3 protease clones per sample was analysed in the two groups. A total of 289 NS3 protease sequences were obtained (146 sequences for cases and 143 for controls).

Consensus sequences were determined by aligning the NS3 variants in each group; 98.3% of nucleotide sites were identical when the consensus sequences of the two groups were compared. In variants from cases and controls, 206 and 205, respectively, of the 543 nucleotide sites showed at least one substitution. A stop codon was identified in two control clones from patients C12 and C16, at NS3 protease codons 85 and 86, respectively.

Complexity and diversity

The intra- and intersample complexity and diversity of NS3 protease were evaluated for both nucleotide and amino acid sequences (Table 2). Genetic complexity, estimated by Shannon entropy and defined as the number of different sequences or clusters of sequences appearing at a given time point, did not differ significantly between cases and controls. The diversity of the NS3 protease gene was evaluated as the mean genetic distance within quasispecies. The mean within-sample genetic distance calculated for the quasispecies in each of the 20 patients was 0.017. Genetic distances in the quasispecies present in samples from cases and controls varied from patient to patient but were similar between the two groups. The proportion of synonymous substitutions was significantly higher than the proportion of nonsynonymous substitutions ($P < 0.0001$), suggesting that NS3 protease mutations mostly resulted from random genetic drift rather than from positive selective pressure.

A phylogenetic tree built from an alignment of 289 nucleotide sequences showed no specific distribution of NS3 protease clones according to subsequent progression or nonprogression of cirrhosis to HCC (Fig. 1). It was actually not possible to identify a dominant intrasample

© 2006 The Authors

Table 2 HCV NS3 protease quasispecies complexity, diversity and secondary structure features

	Duration of hepatitis C infection (years)*	d_N/d_S	Number of clones analysed	Number of different AA sequences	Number of different nt sequences	Genetic distance	Shannon entropy S_n (AA)	NS3 secondary structure	(%)
Cases									
K1	45	0.015	15	4	15	0.023	0.266	B1-1	100
K2	54	0.042	13	3	13	0.008	0.209	B1-2	92.3
								A2-2	7.7
K3	11	0.048	15	6	14	0.031	0.522	B1-1	66.7
								B1-2	26.7
								C-2	6.6
K4	43	0.057	15	8	15	0.048	0.631	B1-1	80
								B1-2	13.3
								C-2	6.7
K5	59	0.167	15	5	10	0.005	0.399	B1-1	93.3
								C-2	6.7
K6	25	0.067	15	7	13	0.011	0.513	B1-1	100
K7	42	0.094	13	6	13	0.013	0.591	B1-1	100
K8	48	0.093	15	9	13	0.032	0.701	B1-1	93.3
								C-2	6.7
K9	46	0.029	15	3	14	0.024	0.326	B1-1	100
K10	ND	0.053	15	4	12	0.007	0.317	B1-1	100
Median	45	0.041	15	5.5	13.5	0.096	0.46	B1-1	84.2†
Controls									
C11	38	0.070	14	9	13	0.046	0.745	B1-1	100
C12	ND	0.020	13	5	6	0.002	0.330	B1-1	92.3
								C-2	7.7
C13	22	0.333	15	2	13	0.013	0.214	B1-1	100
C14	23	0.026	15	2	13	0.006	0.09	B1-2	100
C15	ND	0.000	13	1	4	0.001	0	B1-2	100
C16	38	0.250	15	6	13	0.011	0.433	B1-2	86.6
								A1-2	6.7
								C-4	6.7
C17	17	0.097	14	5	12	0.014	0.377	B1-1	85.7
								A1-1	14.3
C18	27	0.143	15	7	14	0.04	0.620	B1-1	100
C19	ND	0.000	15	2	7	0.002	0.09	B1-2	100
C20	21	0.429	14	7	13	0.011	0.550	B1-1	85.7
								C-2	14.3
Median	23	0.05	14.5	5.1	13	0.139	0.35	B1-1	55.9†

ND, unknown; AA, amino acidic; nt, nucleotidic; Mean intrapatient genetic distance was estimated with the Kimura two-parameter method in MEGA3 software.
*Time between supposed date of contamination and intake of serum sample.
†Percentage of B1-1 clones in each group.

sequence that clustered with those of the other HCC cases. Bootstrap resampling values outside intrasample clusters were low, probably because of the weak phylogenetic signal of the NS3 protease region [31]. However, 20 clusters had a high level of reliability (most with bootstrap values exceeding 90%), corresponding to the quasispecies of the 20 patients. The topology of the tree is otherwise consistent with estimates of intrasample quasispecies diversity based on genetic distance.

Amino acid polymorphism

Amino acid consensus sequences were 100% identical in the two groups and were 100% and 95.58% similar to the

© 2006 The Authors

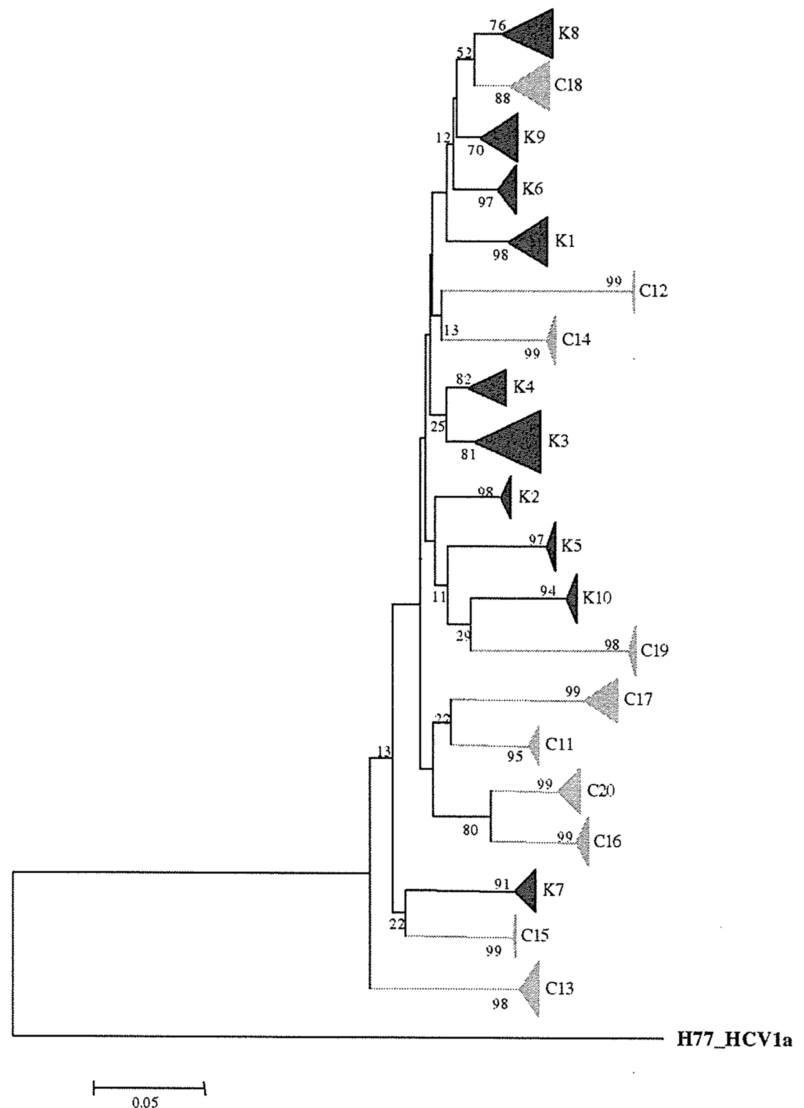

Fig. 1 Phylogenetic tree issued from nucleotide alignment of 289 NS3 protease clones analysed. Lineages of each 20 quasispecies are represented by a black (cases) or a grey (controls) solid elongated triangle whose thickness is proportional to the number of taxa condensed. HCV-1a H77 reference strain was included as outgroup. Tree was built with neighbour-joining method with a Kimura two-parameter and a gamma distribution ($\alpha = 0.29$). Bootstrap value higher than 10 are shown. Branch lengths are drawn to scale. The scale bar represents 5% of nucleotide sequence divergence.

sequences of the prototype HCV-BK [32] and HCV-J [33] strains, respectively (Fig. 2). Eight differences from the sequence of HCV-J were identified: substitutions E30D, V36L, Q86P, L94M, I114V, V132I, A150V and V170I.

Of the 181 amino acid residues in the HCV NS3 protease, 131 (72%) and 139 (77%) residues were conserved in cases and controls, respectively. Thirty and 17 sites, respectively, displayed substitution with respect to the consensus sequence in only one clone per group. NS3 protease substitutions occurring in at least two variants were detected at 20 (11%) and 25 (14%) sites in the case and control groups, respectively. Most of the mutations detected corresponded to conservative hydrophobic to hydrophobic substitutions (Fig. 2). The residues implicated in HCV resistance to NS3 protease inhibitors *in vitro* – D168, A156 and R155 [34,35] – were consistently conserved in the 289 variants analysed.

In none of the clones did polymorphism at a single NS3 position exceed 46.2% of control variants and 23.3% of case variants (Fig. 3). Neither in cases nor controls did we observe substitutions in the catalytic triad residues H57, D81 and S139 (Fig. 2), with the exception of one case, with a variant displaying a mutation in histidine 57 (H57R). Histidine 149, which holds the zinc ion on the top of the zinc finger, was substituted in one variant from several cases (H149Y). None of the major residues implicated in NS3 protease substrate binding (F154, L135, R161, K165) [36] were substituted in either of the groups, with the exception of alanine 157 from the S1 binding pocket (A157T/V), which was mutated in three and two proteases from the case and control groups, respectively.

Some identical mutations were found in the NS3 protease sequences of isolates from cases and controls, when these

© 2006 The Authors

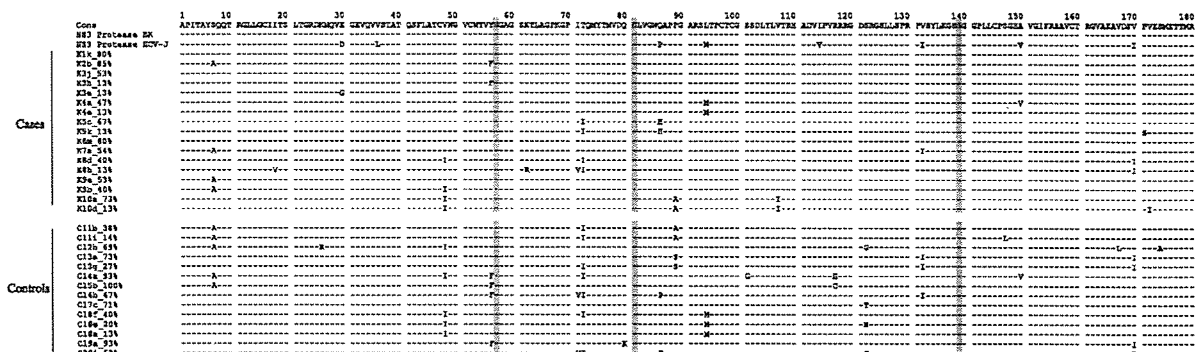

Fig. 2 NS3 protease sequence alignment for the predominant clones of each of the 20 quasispecies analysed. The consensus sequence 'cons' – obtained by aligning the sequences of every sequenced clone – the sequence of the BK strain [32] – used as the reference sequence for numbering residues – and the sequence of the HCV-J reference strain are shown for comparison. Only sequences corresponding to the predominant isolate sequences for a given quasispecies (frequency >13%) from cases and controls are aligned. The name and frequency of the predominant clone of each quasispecies is shown to the left of the sequences. The residues of the catalytic triad are shown in grey.

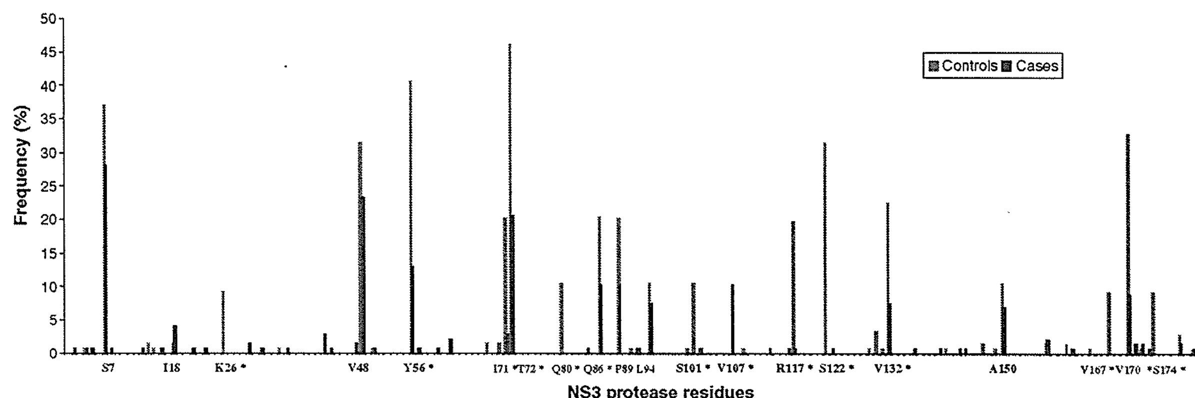

Fig. 3 Frequency of polymorphism mutations for each 181 NS3 protease residue as a function of subsequent progression from cirrhosis to HCC (cases, black bars) or absence of progression (controls, grey bars). The percentage polymorphism corresponds to differences in the amino acid sequence from the NS3 BK strain (accession number gi5542135). The proportion of substituted sequences for each 181 NS3 protease residue is shown along the *y*-axis. Asterisks indicate NS3 positions at which substitutions were significantly more prevalent in HCV isolates from one of the two groups.

sequences were compared with that of HCV-BK strain (Fig. 2). However, there were also several amino acid substitutions in control isolates that were not found in isolates from cases. Moreover, mutations affecting the consensus sequence were found exclusively in the control group at sites 26 (K26R), 80 (Q80K), 86 (Q86P), 89 (P89S), 101 (S101G), 117 (R117C), 122 (S122G/T/N), 167 (V167L) and 174 (V174A) (Table 3). Mutations at sites 26, 80, 101, 167 and 174 were restricted to all the quasispecies variants analysed from a single control patient. Polymorphism at site 122 was more widespread (31.5% of the control clones analysed), with three different amino acid substitutions (S122G/T/N). A Blast search of the GenBank^{TM} database was conducted using the amino acid sequences of the HCV-1b NS3 protease domain from the consensus sequence for all 289 clones. A total of 250 HCV-1b isolates were identified, without taking into account hepatitis stage or clonal *vs* direct sequencing, and S122G/T/N polymorphism was also observed (G/T/N for 11.6%, 10% and 4.4% of sequences, respectively; data not shown). We compared HCV NS3 protease substitutions occurring in clones from at least two different samples between the two groups (Fig. 3). Differences, in terms of the percentage of clones with substitutions, were statistically significant for 10 NS3 sites. Thus, mutations Y56F, I71V, T72I, Q86P, P89S, S101G, R117H, S122G and V170I ($P < 0.001$), S122T ($P = 0.0013$) and V132I ($P = 0.003$) were more frequently observed in control than in case NS3 protease sequences.

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd,

Table 3 HCV NS3 protease substitutions occurring exclusively in cases or in controls group variants

HCV NS3 residues	Mutations	Patients		Mutated clones (%)*
		Cases	Controls	
26	K26R	–	C12	100
80	Q80K	–	C19	100
86	Q86H	K5	–	100
86	Q86P	–	C16	100
89	P89S	–	C13	100
101	S101G	–	C14	100
107	V107I	K10	–	100
117	R117H	–	C14	100
	R117C	–	C15	100
122	S122G	–	C12, C20	100, 100
	S122T	–	C17	100
	S122N	–	C18	27
167	V167L	–	C12	100
174	V174A	–	C12	100

*Mutation occurrence in each patient analysed quasispecies.

Only one residue, V107, was substituted in case NS3 sequences but not in control sequences (V107I, 100% of the 15 clones analysed from patient K10's serum sample) (Table 3). This site was substituted in only one of the 250 NS3 sequences from GenBank; this NS3 mutation was also reported in an HCC sample of a total of 146 (77 HCC and 69 non-HCC) [13]. Q86H substitution was observed in all clones from patient K5 sample and was not present in NS3 control clones.

Classification of NS3 protease clones based on the analysis of the secondary structure of an amino-terminal portion of the protease

The secondary structure of the amino-terminal 120 residues of NS3 was predicted for each of the 289 clones sequenced and for the sequence generated by direct sequencing of each of the 20 quasispecies. NS3 protease variants were assigned to three different groups – group A (4 clones), group B (277 clones) and an indeterminate group (8 clones) – according to their secondary structure (Table 2). Each group was further divided into a number of subgroups. The classification used corresponds to that previously defined by Ogata *et al.* [13,14].

We found that in 9 of the 10 cases and 6 of the 10 controls, 66.7–100% of the NS3 protease quasispecies variants had secondary structures of the B1-1 subgroup (Table 2). The difference between cases and controls was not significant ($P > 0.05$). The proportion of NS3 protease variants assigned to subgroup B1-1, which has been shown to be associated with HCC, differed significantly between cases and controls ($P < 0.001$). The number of quasispecies belonging to this secondary structure group was higher in cases (123 variants, 84.2%) than in controls (80 variants, 55.9%). No protease sequence was classified as belonging to subgroup B2-1, which has been shown to be associated with HCC [13]. Eight control and eight case amino acid sequences obtained by direct sequencing were 100% identical to the predominant sequence of the corresponding quasispecies and consequently displayed the same secondary structure (data not shown). Of the four discordant direct sequences, only that for the sample from patient C14 had a B1-1 secondary structure, despite the population analysed being relatively homogeneous and belonging to the B1-2 group.

DISCUSSION

The HCV NS3 protease is essential for viral replication and is also thought to be involved in hepatocarcinogenesis. As this enzyme is a key target for specific anti-HCV therapy, it is important to define its diversity at different stages of liver disease. We previously described its natural polymorphism, resulting in the existence of quasispecies in untreated patients with chronic hepatitis C [37]. We investigated the relationship between the progression of hepatic disease and features of the nucleotide and amino acid sequences corresponding to the HCV-1b NS3 protease region by analysing these sequences for viral quasispecies derived from patients with cirrhosis. One of the strong points of this study is the long interval (several years) between sample analysis and outcome determination (occurrence or absence of HCC). As no particular sequence motif was identified in the case group at the time of cirrhosis diagnosis, the NS3 protease population of cases presumably differentiated later, between the diagnosis of cirrhosis and HCC. It would therefore be of interest to analyse NS3 protease sequences at the HCC stage. Our comparison of these two viral populations is probably accurate, as the patients recruited to the two groups differed exclusively in terms of subsequent HCC occurrence. The patients recruited were matched for most criteria, without reported risk factors for HCC – women, with similar presumed modes of contamination (transfusion), devoid of toxic factors such as alcohol or drug abuse [38]. None of the 20 patients studied had HCC at the time of cirrhosis diagnosis or cirrhosis with extremely rapid progression to HCC. We chose to consider patients with HCV genotype 1 infections because this genotype has a high prevalence in European HCV patients, particularly those infected for long periods of time and because this genotype does not display a complete response to pegylated interferon–ribavirin bitherapy [39]. HCV genotype 1 infections are therefore more likely to progress to HCC than are infections with other HCV genotypes. Given the small numbers of patients in the two groups, we cannot exclude the possibility of a selection bias linked to the absence of a main NS3 sequence or NS3 secondary structure profile predictive of progression from cirrhosis to HCC in our study population.

© 2006 The Authors

NS3 protease polymorphism depended on the structural and functional constraints on the enzyme, with tolerance observed for substitutions occurring outside sites essential for catalytic function and three-dimensional structure. The polymorphic residues and types of substitution observed in the 289 NS3 variants from cirrhotic patients analysed are consistent with those previously identified in chronic hepatitis patients [37]. Substitutions that do not conserve the physicochemical properties of the original residue are exposed to solvent at the surface of the protease and are probably more sensitive to external pressures. NS3 protease gene complexity and quasispecies diversity varied considerably from patient to patient and was not related to the subsequent outcome of HCV cirrhosis. Quasispecies consisted of highly homogeneous populations or a 'cloud' of more divergent variants. Polymorphic sites 26, 80, 101, 167 and 174, which were restricted to variants from the control group, could not be considered as markers of subsequent nonprogression, as substitutions at these sites were observed exclusively in quasispecies from a single control patient. Moreover, mutations at positions 26, 80, 101 and 122 have previously been found in NS3 proteases from both HCC and non-HCC patients [13,14]. Two NS3 protease variants from the control group had a premature stop codon, at sites 85 and 86. These two variants probably correspond to defective viruses, as they were also detected in previous studies [40].

Although no amino acid signature patterns specific to cirrhotic patients displaying subsequent evolution to HCC could be identified, NS3 protease polymorphism was more frequent in quasispecies from patients not displaying subsequent progression to HCC. As patients were matched for most criteria, to minimize bias, we could not exclude the possibility of different levels of diversity related to the duration of infection, which was longer for cases (45 *vs* 23 years). This lower variability may be of clinical significance for disease progression, during which the major variant is selected.

It remains unclear whether there is an association between the evolution of genetic heterogeneity and the severity of HCV liver infection. Several studies have suggested that there is a correlation between the degree of HVR1 sequence diversity and the development of severe liver disease [41–43], but these findings have not been confirmed by other studies [44]. Diversity data have also been obtained for regions implicated in hepatocarcinogenesis other than NS3, with greater variability reported for patients with HCC than in patients without HCC for core protein [45] and the NS5A-PKR-binding domain [46,47]. Conversely, we show here that patients displaying progression to HCC had lower levels of NS3 protease polymorphism than patients who did not develop cancer. Comparative analyses of the core and NS3 helicase regions showed no specific amino acid differences between isolates from cirrhotic patients with and without HCC, suggesting that specific amino acid changes in these regions have no direct oncogenic role in patients with HCV-related cirrhosis [46]. Mutations in the HCV genome have been explored over a 13-year period in a single patient [48]; mutations clustering in the HVR-1 gene were found to have accumulated over time. Conversely, the carboxy-terminal NS3 and NS5B regions were found to be the most highly conserved. To date, no study has identified amino acid signatures associated with disease severity or disease outcome, in any region of the HCV genome.

An HCV NS3 protease secondary structure conformation, B1-1, has been identified as associated with progression to HCC [13,14]. We determined the secondary structures of our NS3 protease clones by the same method, to investigate whether this conformation was also observed in variants from cirrhotic patients who subsequently developed HCC. Surprisingly, the NS3 protease sequences from most of the cases and controls adopted the B1-1 configuration. This type of secondary structure may actually be more closely linked to fibrosis than to HCC. It would therefore be of interest to carry out analyses for a third paired group of patients without cirrhosis, over a similar period of follow-up for hepatitis C infection. Our study differed from the previous study in several respects: population studied (cirrhotic patients *vs* HCC and no HCC), sequence analysis (quasispecies *vs* direct sequencing). For a more accurate comparison, we also directly sequenced NS3 protease genes from each of the case and control samples. All but one of the secondary structures of the 20 quasispecies was consistent with that of the major variants. Among case and control NS3 protease variants, non-B1-1 secondary structures (B1-2) were found in the least complex quasispecies, mostly in cirrhotic patients without HCC (K2, C14, C15, C16 and C19) (Table 2). One substitution, Y56F was found exclusively in NS3 protease clones with the B1-2 secondary structure. This mutation, observed in 15% of the 250 GenBank NS3 protease sequences, and low quasispecies complexity may be two markers of non-HCC outcome. As reported by Ogata *et al.*, the presence of a phenylalanine in position 56 is associated with the absence of the turn structure at position 57 present in the B1-1 secondary structure. Interaction between the p53 tumour suppressor protein and the NS3 region above this turn structure (amino acids 29–174) would almost certainly be compromised by the structural modification [11].

In conclusion, we assessed and compared the variability of HCV NS3 protease in two groups of viruses in samples taken from cirrhotic patients at the time of cirrhosis diagnosis. Polymorphism differences have been shown, in favour of NS3 proteases from patients who did not subsequently develop HCC compared with those from patients displaying subsequent progression from cirrhosis to HCC. However, these differences did not allow the definition of a specific NS3 profile related to HCC occurrence. Our results suggest that there may be a relationship to fibrosis progression when diversity parameters are considered together with secondary structure profiles. If these findings could be confirmed, they would demonstrate the

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd.

potential utility of quasispecies analysis in attempts to interpret disease outcome. The mechanism of hepatocarcinogenesis in patients with hepatitis C infection remains to be determined. In addition to further NS3 protease sequence analysis, the cellular interactions of this enzyme require investigation. We are currently carrying out such studies, using a subgenomic replicon model.

REFERENCES

1 Degos F, Christidis C, Ganne-Carrie N *et al.* Hepatitis C virus related cirrhosis: time to occurrence of hepatocellular carcinoma and death. *Gut* 2000; 47: 131–136.

2 Bartenschlager R, Ahlborn-Laake L, Mous J, Jacobsen H. Kinetic and structural analyses of hepatitis C virus polyprotein processing. *J Virol* 1994; 68: 5045–5055.

3 Eckart MR, Selby M, Masiarz F *et al.* The hepatitis C virus encodes a serine protease involved in processing of the putative nonstructural proteins from the viral polyprotein precursor. *Biochem Biophys Res Commun* 1993; 192: 399–406.

4 Gallinari P, Brennan D, Nardi C *et al.* Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus. *J Virol* 1998; 72: 6758–6769.

5 Failla C, Tomei L, De Francesco R. An amino-terminal domain of the hepatitis C virus NS3 protease is essential for interaction with NS4A. *J Virol* 1995; 69: 1769–1777.

6 Pawlotsky JM, McHutchison JG. Hepatitis C. Development of new drugs and clinical trials: promises and pitfalls. Summary of an AASLD hepatitis single topic conference, Chicago, IL, February 27–March 1, 2003. *Hepatology* 2004; 39: 554–567.

7 Smith RM, Wu GY. Structure-based design of hepatitis C virus inhibitors. *J Viral Hepat* 2003; 10: 405–412.

8 Sakamuro D, Furukawa T, Takegami T. Hepatitis C virus nonstructural protein NS3 transforms NIH 3T3 cells. *J Virol* 1995; 69: 3893–3896.

9 Zemel R, Gerechet S, Greif H *et al.* Cell transformation induced by hepatitis C virus NS3 serine protease. *J Viral Hepat* 2001; 8: 96–102.

10 Fujita T, Ishido S, Muramatsu S, Itoh M, Hotta H. Suppression of actinomycin D-induced apoptosis by the NS3 protein of hepatitis C virus. *Biochem Biophys Res Commun* 1996; 229: 825–831.

11 Ishido S, Hotta H. Complex formation of the nonstructural protein 3 of hepatitis C virus with the p53 tumor suppressor. *FEBS Lett* 1998; 438: 258–262.

12 Kwun HJ, Jung EY, Ahn JY, Lee MN, Jang KL. p53-dependent transcriptional repression of p21(waf1) by hepatitis C virus NS3. *J Gen Virol* 2001; 82: 2235–2241.

13 Ogata S, Florese RH, Nagano-Fujii M *et al.* Identification of hepatitis C virus (HCV) subtype 1b strains that are highly, or only weakly, associated with hepatocellular carcinoma on the basis of the secondary structure of an amino-terminal portion of the HCV NS3 protein. *J Clin Microbiol* 2003; 41: 2835–2841.

14 Ogata S, Ku Y, Yoon S, Makino S, Nagano-Fujii M, Hotta H. Correlation between secondary structure of an amino-terminal portion of the nonstructural protein 3 (NS3) of hepatitis C virus and development of hepatocellular carcinoma. *Microbiol Immunol* 2002; 46: 549–554.

15 Martell M, Esteban JI, Quer J *et al.* Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J Virol* 1992; 66: 3225–3229.

16 Smith DB, McAllister J, Casino C, Simmonds P. Virus 'quasispecies': making a mountain out of a molehill? *J Gen Virol* 1997; 78 (Pt. 7): 1511–1519.

17 Gomez J, Martell M, Quer J, Cabot B, Esteban JI. Hepatitis C viral quasispecies. *J Viral Hepat* 1999; 6: 3–16.

18 Bruix J, Sherman M, Llovet JM *et al.* Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. *J Hepatol* 2001; 35: 421–430.

19 Edmondson H, Steiner P. Primary carcinoma of the liver: a study of 100 cases among 48,900 necropsies. *Cancer* 1954; 7: 462–503.

20 Roque-Afonso AM, Ferey MP, Poveda JD, Marchadier E, Dussaix E. Performance of TRUGENE hepatitis C virus 5′ noncoding genotyping kit: a new CLIP sequencing-based assay for hepatitis C virus genotype determination. *J Viral Hepat* 2002; 9: 385–389.

21 Laperche S, Lunel F, Izopet J *et al.* Comparison of hepatitis C virus NS5b and 5′ noncoding gene sequencing methods in a multicenter study. *J Clin Microbiol* 2005; 43: 733–739.

22 Holland-Staley CA, Kovari LC, Golenberg EM, Pobursky KJ, Mayers DL. Genetic diversity and response to IFN of the NS3 protease gene from clinical strains of the hepatitis C virus. *Arch Virol* 2002; 147: 1385–1406.

23 Lodrini S, Bagaglio S, Canducci F *et al.* Sequence analysis of NS3 protease gene in clinical strains of hepatitis C virus. *J Biol Regul Homeost Agents* 2003; 17: 198–204.

24 Staden R, Beal KF, Bonfield JK. The Staden package, 1998. *Methods Mol Biol* 2000; 132: 115–130.

25 Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 1998; 23: 403–405.

26 Nicholas KB, Nicholas HB, Jr. Deerfield, DW. GeneDoc: analysis and visualization of genetic variation. *Embnew News* 1997; 4: 14.

27 Wolinsky SM, Korber BTM, Neumann AU *et al.* Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* 1996; 272: 537–542.

28 Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986; 3: 418–426.

29 Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 2004; 5: 150–163.

30 Garnier J, Osguthorpe DJ, Robson B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol* 1978; 120: 97–120.

31 Salemi M, Vandamme AM. Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. *J Mol Evol* 2002; 54: 62–70.

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd,

32 Kim JL, Morgenstern KA, Lin C *et al.* Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* 1996; 87: 343–355.
33 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.
34 Lin C, Lin K, Luong YP *et al.* In vitro resistance studies of hepatitis C virus serine protease inhibitors, VX-950 and BILN 2061: structural analysis indicates different resistance mechanisms. *J Biol Chem* 2004; 279: 17508–17514.
35 Lu L, Pilot-Matias TJ, Stewart KD *et al.* Mutations conferring resistance to a potent hepatitis C virus serine protease inhibitor in vitro. *Antimicrob Agents Chemother* 2004; 48: 2260–2266.
36 Bartenschlager R. The NS3/4A proteinase of the hepatitis C virus: unravelling structure and function of an unusual enzyme and a prime target for antiviral therapy. *J Viral Hepat* 1999; 6: 165–181.
37 Vallet S, Gouriou S, Nousbaum JB, Legrand-Quillien MC, Goudeau A, Picard B. Genetic heterogeneity of the NS3 protease gene in hepatitis C virus genotype 1 from untreated infected patients. *J Med Virol* 2005; 75: 528–537.
38 Poynard T, Mathurin P, Lai CL *et al.* A comparison of fibrosis progression in chronic liver diseases. *J Hepatol* 2003; 38: 257–265.
39 Fried MW, Shiffman ML, Reddy KR *et al.* Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; 347: 975–982.
40 Vallet S, Nousbaum JB, Gouriou S, Legrand-Quillien MC, Goudeau A, Picard B. A particular hepatitis C virus protease NS3 gene pattern in a patient not responding to interferon–ribavirin therapy. *Eur J Clin Microbiol Infect Dis* 2005; 24: 79–80.
41 Gao G, Stuver SO, Okayama A, Tsubouchi H, Mueller NE, Tabor E. The minimum number of clones necessary to sequence in order to obtain the maximum information about hepatitis C virus quasispecies: a comparison of subjects with and without liver cancer. *J Viral Hepat* 2005; 12: 46–50.
42 Koizumi K, Enomoto N, Kurosaki M *et al.* Diversity of quasispecies in various disease stages of chronic hepatitis C virus infection and its significance in interferon treatment. *Hepatology* 1995; 22: 30–35.
43 Curran R, Jameson CL, Craggs JK *et al.* Evolutionary trends of the first hypervariable region of the hepatitis C virus E2 protein in individuals with differing liver disease severity. *J Gen Virol* 2002; 83: 11–23.
44 Gretch DR, Polyak SJ, Wilson JJ, Carithers RL Jr, Perkins JD, Corey L. Tracking hepatitis C virus quasispecies major and minor variants in symptomatic and asymptomatic liver transplant recipients. *J Virol* 1996; 70: 7622–7631.
45 Ogata S, Nagano-Fujii M, Ku Y, Yoon S, Hotta H. Comparative sequence analysis of the core protein and its frameshift product, the F protein, of hepatitis C virus subtype 1b strains obtained from patients with and without hepatocellular carcinoma. *J Clin Microbiol* 2002; 40: 3625–3630.
46 Gimenez-Barcons M, Franco S, Suarez Y *et al.* High amino acid variability within the NS5A of hepatitis C virus (HCV) is associated with hepatocellular carcinoma in patients with HCV-1b-related cirrhosis. *Hepatology* 2001; 34: 158–167.
47 Franco S, Gimenez-Barcons M, Puig-Basagoiti F *et al.* Characterization and evolution of NS5A quasispecies of hepatitis C virus genotype 1b in patients with different stages of liver disease. *J Med Virol* 2003; 71: 195–204.
48 Ogata N, Alter HJ, Miller RH, Purcell RH. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc Natl Acad Sci USA* 1991; 88: 3392–3396.

© 2006 The Authors

Journal of General Virology (2007), 88, 000–000 DOI 10.1099/vir.0.82604-0

Complete genomes for hepatitis C virus subtypes 6f, 6i, 6j and 6m: viral genetic diversity among Thai blood donors and infected spouses

Ling Lu,[1] Chunhua Li,[1,2] Yongshui Fu,[3] Lakkana Thaikruea,[4] Satawat Thongswat,[5] Niwat Maneekarn,[6] Chatchawann Apichartpiyakul,[6] Hak Hotta,[7] Hiroaki Okamoto,[8] Dale Netski,[9,10] Oliver G. Pybus,[11] Donald Murphy,[12] Curt H. Hagedorn[1] and Kenrad E. Nelson[9,10]

Correspondence
Ling Lu
llu@kumc.edu

Kenrad E. Nelson
kenelson@jhsph.edu

[1]Division of Gastroenterology/Hepatology, Department of Medicine, University of Kansas Medical Center, 4035 Delp, MS 1023, Kansas City, KS 66160, USA

[2]The First People's Hospital of Yunnan Province, Kunming, China

[3]Gunagzhou Blood Center, Guangzhou City, Guangdong Province, China

[4]Department of Community Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[5]Research Institute for Health Sciences, Chiang Mai University, Chiang Mai, Thailand

[6]Department of Microbiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[7]Department of Microbiology, Kobe University School of Medicine, Chuo-ku, Kobe, Hyogo 650, Japan

[8]Division of Virology, Department of Infection and Immunity, Jichi Medical University School of Medicine, Tochigi-Ken, Japan

[9]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

[10]Department of Medicine, Johns Hopkins University, Baltimore, MD, USA

[11]Department of Zoology, Oxford University, Oxford, UK

[12]Institut National de Santé Publique du Quebec, Laboratoire de Santé Publique du Quebec, Sainte-Anne-de-Bellevue, Quebec, Canada

In this study, the first complete genome sequences for hepatitis C virus (HCV) subtypes 6f, 6i, 6j and 6m, obtained from infected blood donors in Chiang Mai, Thailand, are reported. Pairwise genome-wide nucleotide similarities between some of these isolates were higher than the 75–80 % value used previously to define different HCV subtypes. To investigate further, the entire genomes of four prototype isolates, Th602 (6i), Th553 (6j), B4/92 (6m) and D86/93 (6n), were sequenced. Pairwise comparison of these sequences gave a similar range of nucleotide similarities, thereby providing new information for HCV subtype classification. In order to study the hypothesis of interspousal HCV transmission, four additional complete HCV genome sequences were obtained from two infected Thai blood donors and their spouses, C-0044 and C-0046 (6f), and C-0192 and C-0185 (6m). Pairwise comparison of the sequences revealed that C-0044 and C-0046 share a nucleotide similarity of 98.1 %, whilst C-0185 and C-0192 have a similarity of 97.8 %. Several other studies of partial HCV sequences of different genomic regions from HCV-infected couples have shown nucleotide similarities ranging from 96.3 to 100 %. The similarities of the complete genome sequences from the two couples in the current study are consistent with HCV transmission between spouses.

Received 29 September 2006
Accepted 25 January 2007

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this study are DQ835760–DQ835770.
Supplementary tables showing PCR primers and amplification strategy and origin of the retrieved sequences reanalysed are available in JGV Online.

0008-2604 © 2007 SGM *Printed in Great Britain*

INTRODUCTION

Hepatitis C virus (HCV) is a blood-borne pathogen that infects an estimated 170 million people worldwide (3 % of the global population). HCV infection can be asymptomatic for many years, but in many patients it leads eventually to adverse effects and is a major cause of liver cirrhosis and hepatocellular carcinoma. Alone or in combination with alcoholism, HCV-related end-stage liver disease has become the leading indication for liver transplantation in most USA transplant programmes (Moreno & Berenguer, 2002). HCV is transmitted efficiently through blood transfusion and injection drug use (Murphy *et al.*, 2000). Transmission is also associated with other parenteral exposures, such as tattooing, ear or body piercing, surgery, acupuncture, haemodialysis, occupational needle-stick injuries and other health-care procedures (Alter, 1994; Alter *et al.*, 1990). Although the efficiency of HCV transmission through sexual or household contact is not yet understood entirely (Luksamijarulkul *et al.*, 2000; Romanowski *et al.*, 2003), the observation of closely related HCV sequences obtained from infected couples or sexual partners could be interpreted as evidence of HCV sexual transmission (Capelli *et al.*, 1997; Chayama *et al.*, 1995; Halfon *et al.*, 2001; Healey *et al.*, 1995; Kao *et al.*, 1992, 2000; Komine *et al.*, 1999; Nakayama *et al.*, 2005; Quer *et al.*, 2003; Rice *et al.*, 1993; Romanowski *et al.*, 2003; Ross *et al.*, 1999; Thaikruea *et al.*, 2004; Yagura *et al.*, 2002).

Currently, in most industrialized countries, HCV antibody screening and mini-pool nucleic acid amplification testing (NAT) are mandatory for virtually all blood donations collected. However, in many developing countries, HCV transmission through blood transfusion remains a critical health problem. For example, in Thailand, resources preclude the general implementation of NAT and, in some rural areas, HCV antibody screening of blood donors is not performed consistently (Luksamijarulkul *et al.*, 2004; Wiwanitkit, 2005). The seroprevalence of HCV among the general population in Thailand has been estimated to be approximately 5 % (3.5 million of the total 63 million population) (Songsivilai *et al.*, 1997; Wiwanitkit & Suyaphan, 2002). The group with the highest risk of HCV infection in Thailand is injection drug users, 95 % of whom have been found to be HCV-infected (Luksamijarulkul & Plucktaweesak, 1996). Another important risk group in Thailand is female sex workers, who have been reported to have an infection rate of 9.5 % (Luksamijarulkul & Deangbubpha, 1997). Minority hill-tribe populations in northern Thailand have also been reported to have a high HCV prevalence of approximately 8 % (Wiwanitkit & Suyaphan, 2002). The seroprevalence of HCV among Thai blood donors has increased since 1991 (Nantachit *et al.*, 2003; Songsivilai *et al.*, 1997) and a number of unique HCV subtypes have been identified in Thailand (Thaikruea *et al.*, 2004).

HCV is classified in the genus *Hepacivirus* of the family *Flaviviridae*. It has a single-stranded, positive-sense RNA genome of about 9600 nt in length. The genome contains a single open reading frame (ORF) that encompasses nearly the entire genome range. Flanked by non-coding regions (NCRs) at both 5′ and 3′ ends, this ORF encodes three structural (core, E1, E2) and seven non-structural (P7, NS2, NS3, NS4A, NS4B, NS5A and NS5B) proteins. HCV has been classified by phylogenetic analysis into six major genotypes; variants within each genotype are grouped into a number of subtypes. Recently, HCV nomenclature has been revised and 71 subtypes in total were defined. Among these subtypes, 17 have been assigned to genotype 6 (subtypes 6a–6q). Geographically, genotype 6 infections are restricted exclusively to south-eastern Asia or immigrants from this region (Lu *et al.*, 2006).

In one of our previous studies, partial sequences representing six HCV subtypes (6a, 6f, 6i, 6j, 6m and 6n) were obtained from samples of blood donors in Chiang Mai, northern Thailand. Despite the enormous importance of genetic diversity for our understanding of HCV infection and evolution, only eight of the 17 subtypes of genotype 6 have had their whole genomes sequenced to date (subtypes 6a, 6b, 6d, 6e, 6g, 6h, 6k and 6n; Li *et al.*, 2006; Lu *et al.*, 2006; Simmonds *et al.*, 2005). As a step towards a comprehensive understanding of HCV diversity, we have sequenced the entire genomes of HCV isolates of subtypes 6f, 6i, 6j and 6m, and report the data in this paper. By using phylogenetic analysis, we found that subtype 6i was genetically similar to 6j and that subtype 6m was similar to 6n. In addition to the four prototypic isolates (Th602, Th553, B4/92 and D86/93) that represent subtypes 6i, 6j, 6m and 6n, we also sequenced the complete genomes of HCV isolates obtained from two HCV-infected couples (Thaikruea *et al.*, 2004), enabling us to study the possibility of interspousal HCV transmission.

METHODS

Subjects and samples. All of the subjects studied were residents of northern Thailand. Among them, C-0046, C-0159, C-0192, C-0208 and C-0667 were male replacement blood donors. Subject C-0044 was the spouse of donor C-0046; they had lived together since 1992. Subject C-0185 was the spouse of donor C-0192; they had been married since 1978. Donor C-0192 reported a sexual history of 10 female partners and one occasion of sex with a female sex worker. Several features were shared by the other three male blood donors, C-0159, C-0208 and C-0667. None were married, but each had initiated sexual activities when 14–19 years old. They each had a history of sex with female sex workers on seven or eight occasions and had had five to ten female partners during their lifetime. These subjects were also associated with other risk factors, as listed in Table 1. From the remaining four individuals, partial sequences had been determined previously (Apichartpiyakul *et al.*, 1994; Doi *et al.*, 1996; Tokita *et al.*, 1995), which have been used for classifying HCV subtypes 6i, 6j, 6m and 6n, respectively (Simmonds *et al.*, 2005). These individuals included a healthy voluntary blood donor (B4/92), an injection drug user (D86/93) and two patients receiving kidney transplantation (Th553 and Th602).

Sequence amplification and analysis. Complete HCV genomic sequences were each amplified from 100 µl serum by using modified, previously described approaches (Li *et al.*, 2006). Briefly, RNA was extracted by using Tripure (Roche), cDNA was synthesized by using

Table 1. Epidemiological data for the five blood donors and two spouses

	Isolate						
	C-0159	C-0208	C-0667	C-0046	C-0044	C-0192	C-0185
Age (years)	24	34	23	40	39	47	42
Gender	M	M	M	M	F	M	F
Occupation	Military	Employee	Employee	Labourer	Employee	Merchant	Civil servant
Donations (*n*)	4	7	1	8	Spouse	1	Spouse
Reason for donation	Relative	Relative	Relative	Relative	–	Relative	–
Relative with hepatitis	+	–	–	+	–	–	–
History of:							
Jaundice	+	–	–	–	–	+	–
Surgery	–	+	+	–	+	–	+
Suture	+	+	+	–	–	–	+
Transfusion	–	–	+	–	+	–	+
Tattooing	+	–	+	–	–	–	–
Body piercing	+	–	+	–	+	–	+
Injection drug use	+	–	–	–	–	–	–
Shared instrument when snorting drugs	–	–	+	–	–	–	–
No. sexual partners of opposite gender	5	10	5	1	2	10	1
No. times sex with sex workers	8	7	8	0	0	1	0
Age (years) at first time of sex	14	15	19	25	25	15	20

avian myeloblastosis virus reverse transcriptase (Roche) and random primers (Promega), and overlapping fragments were amplified by using conventional PCR (Roche) with the strategies illustrated in Fig. 1 and primers listed in Supplementary Table S1 (available in JGV Online). Standard procedures were adopted to avoid nested RT-PCR false positives. These included at least a negative control, a positive control and a water blank, which were tested during each batch of RNA extraction, reverse transcription and cDNA amplification. All of these steps and the final product-resolving step were completed in different spaces with considerable distance and floor separation. After amplification, the fragments were sequenced directly by using methods described previously (Lu *et al.*, 2006). All sequence information was analysed by using GCG version 10.0 (Wisconsin Sequence Analysis Package; Genetic Computer Group), PHYML (Guindon & Gascuel, 2003) and MEGA3 (Kumar *et al.*, 2004) software. Initially, multiple alignments were performed with PILEUP and adjustments to the alignments were performed with PRETTY. Phylogenetic trees were reconstructed by using the maximum-likelihood method under the HKY+I+Γ substitution model (gamma distribution approximated by using six rate categories; Hasegawa *et al.*, 1985). The transition/transversion ratio, proportion of invariable sites and gamma-distribution shape parameter were estimated from the data. The base frequencies were adjusted to maximize the likelihood. Bootstrap resampling was performed by using 500 neighbour-joining replicates. For comparison between isolates, pairwise nucleotide similarities were calculated by using the MEGA3 software. For the detection and analysis of potential genetic recombination events among all available genotype 6 sequences, the RDP2 software (Recombination Detection Program, version 2) was used (Martin *et al.*, 2005). The program was run by using default settings with the following adjustments: (i) window size was 40 nt, (ii) the option of linear sequences was chosen, (iii) six different methods (RDP, GENECONV, MaxChi, Bootscan, Chimaera and SiScan) were run simultaneously against the multiple sequence alignment, and (iv) listing events detected by more than two methods. In addition, the complete genomes were analysed by using the bootscanning approach implemented in the hepatitis C virus subtyping tool available from http://www.bioafrica.net/virus-genotype/.

RESULTS

Characterization of the complete sequences for C-0159, C-0208 and C-0667

The complete genomic sequences of samples C-0159, C-0208 and C-0667 were amplified and sequenced by using strategies shown in Fig. 1. Their genome lengths and genomic organizations are described in Table 2. Pairwise comparisons of nucleotide sequences over the entire genome and within the ten protein-encoding regions showed that C-0159 was related closely to C-0667, whilst C-0208 was related closely to the reference strain km42. The genome-wide nucleotide similarity between C-0159 and C-0667 was 82.7 %; it was 81.3 % between C-0208 and km42 (Table 3). It has been proposed that HCV genotypes differ by 31–35 % and HCV subtypes by 20–25 % of nucleotides over the entire genome length (Simmonds *et al.*, 1994, 2005). This is consistent with our analysis, which included 189 retrievable complete HCV genomic sequences from GenBank (data not shown). We found that the highest inter-subtype nucleotide similarities were 79.7–80.0 % between subtypes 1a and 1c and 79.5 % between subtypes 6a and 6b. Based on the above criteria, C-0159 and C-0667 qualify as belonging to one subtype, whereas C-0208 and km42 qualify as belonging to another. Phylogenetic analysis was performed with complete genome sequences from 30 reference isolates that repre-

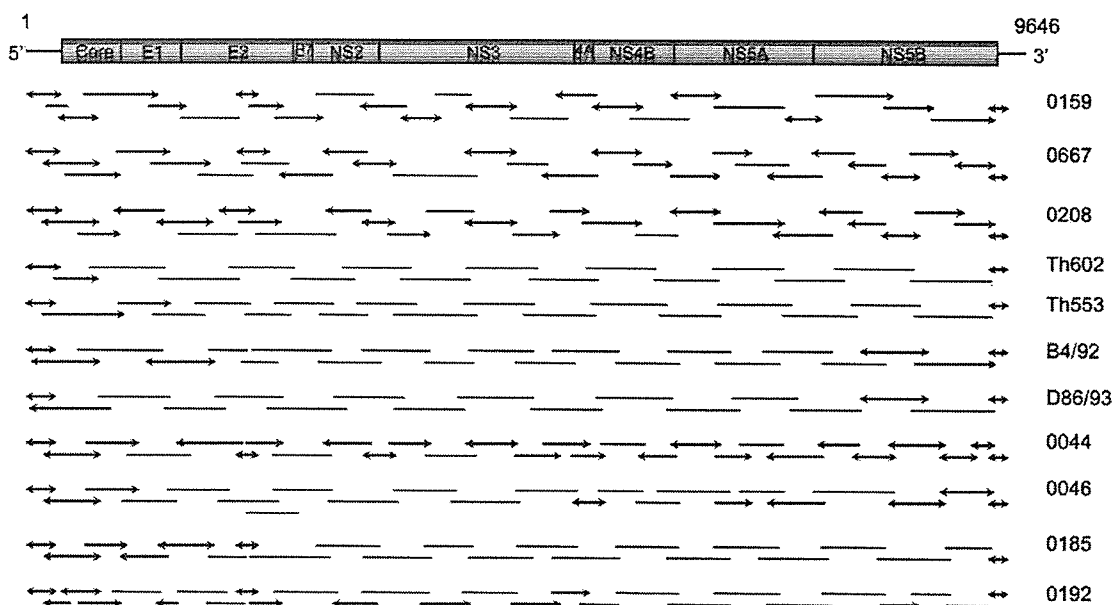

Fig. 1. Strategies used to amplify the 11 complete HCV genomic sequences. The bar at the top represents the genomic organization of HCV and shows the 10 protein-encoding regions of various lengths. Two lines attached to the bar at both sides indicate the 5′ and 3′ NCRs. The nucleotides start at 1 and end at 9646, according to the numbering of the H77 genome (GenBank accession no. NC_004102). Lines and arrows under the bar represent the overlapping fragments amplified for the HCV isolates, with their designations shown on the right. Among them, double-ended arrows identify fragments amplified by using conserved or degenerate primers, single-ended arrows indicate fragments amplified by using conserved or degenerate primers at the arrowed end and strain-specific primers at the other end, and lines without arrowed ends designate fragments amplified by using strain-specific primers.

sented various HCV genotypes and subtypes; the analysis showed that C-0159 clustered with C-0667 and that C-0208 clustered with km42, with each cluster having a bootstrap score of 100 % (Fig. 2).

Characterization of the complete sequences for Th602 and Th553

To test whether C-0159 and C-0667 represent a single HCV subtype or two different ones, complete genomic sequences were subsequently determined for Th602 and Th553 (Fig. 1). These isolates have the same ORF and encoded protein sizes as isolates C-0159 and C-0667 (Table 2). Pairwise comparison revealed that C-0159 and Th602 had a genome-wide nucleotide similarity of 95.3 %, whilst C-0667 and Th553 had a genome-wide similarity of 96.8 % (Table 3). Phylogenetic analysis demonstrated that Th602 clustered most closely with C-0159, and Th553 clustered most closely with C-0667, with each cluster having a bootstrap value of 100 % (Fig. 2). In previous studies, partial sequences of Th602 had been classified into subtype 9b and partial sequences of Th553 classified into subtype 9c (Tokita *et al.*, 1995). In the recent consensus HCV nomenclature proposal, 9b and 9c correspond to subtypes 6i and 6j, respectively (Simmonds *et al.*, 2005). With the addition of the complete genomic sequences of Th602 and Th553, it is possible to classify C-0159 definitively as subtype 6i, and C-0667 as subtype 6j. Subtypes 6i and 6j have genome-wide nucleotide similarities of 82.7–83.4 %. Although this range includes the typical range of values for a single HCV subtype, 6i and 6j still qualify as two distinct subtypes. This is because the two clusters (6i and 6j) are maintained consistently and significantly in phylogenetic analyses of different genomic regions (trees not shown), complying with the criteria for HCV subtype classification (Simmonds *et al.*, 2005). Furthermore, when 6i and 6j sequences were tested against the complete genome sequences of all available genotype 6 subtypes by using a variety of different statistical approaches, no recombination events were confirmed within the ORFs or between the isolates whose sequences were determined in this study (data not shown). This provides strong evidence that subtypes 6i and 6j are not recombinants of other genotype 6 subtypes.

Characterization of the complete sequences for B4/92 and D86/93

Similarly, C-0208 grouped with km42 and the two isolates appeared to represent another single subtype (Fig. 2). For verification, we determined the entire genomic sequences for