Hepatitis C Virus of Subtype 2I in Marseille, Southeastern France

Sarah Aherfi^{a, b} Olga Glazunova^{a, b} Patrick Borentain^c Daniele Botta-Fridlund^c
Laurent Chiche^d Sylvie Bregigeton^e Anne Motte^b Catherine Tamalet^{a, b}
Philippe Colson^{a, b}

^aURMITE UM 63 CNRS 7278 IRD 198 INSERM U1095, Facultés de Médecine et de Pharmacie, Aix-Marseille University,

^bFondation Institut-Hospitalo-Universitaire (IHU) Méditerranée Infection, Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, Centre Hospitalo-Universitaire Timone, ^cService d'Hépatogastro-Entérologie, Centre Hospitalo-Universitaire Conception, ^dService de Médecine Interne, Centre Hospitalo-Universitaire Conception, and ^eCISIH SUD Service d'Hématologie, Département de Recherche Clinique, CHU Ste Marguerite, Marseille, France

Key Words

Hepatitis C virus · Genotype · Subtype 2I · Treatment · Polymerase

Abstract

The rate of eradication of chronic hepatitis C considerably increases with direct-acting antiviral agents, particularly hepatitis C virus (HCV) polymerase inhibitors. While implementing full-length HCV NS5B polymerase sequencing in our clinical microbiology laboratory, we identified atypical HCV sequences, classified as subtype 2I, from 2 patients. HCV-2I NS5B polymerase sequences were detected from 5 and 14 additional patients by screening our laboratory hepatitis virus sequence database and the NCBI GenBank sequence database. Phylogenetic analyses show unambiguously that all HCV-2I sequences are clustered apart from HCV 2 non-I sequences, which compose a second cluster. Mean ($\pm$ SD) nucleotide identity between near full-length NS5B fragments of subtype 2I was $93.4 \pm 0.8\%$ (range: 92.4–95.1). Of note, all

HCV-2I sequences obtained in our laboratory and in other centers were from serum samples collected in France. Analysis of the HCV-2I NS5B polymerase amino acid sequences at 30 positions critical for interaction with or resistance to HCV polymerase inhibitors showed specific patterns.

© 2015 S. Karger AG, Basel

Introduction

Hepatitis C virus (HCV) infection entered a new era with the availability of direct-acting antiviral agents to treat chronic infections in combination with pegylated interferon and ribavirin [1, 2]. Thus, clinical trials have shown considerable increases in the rates of sustained virological response. Direct-acting antiviral agents encompass HCV NS3 protease inhibitors, approved for the treatment of chronic HCV genotype 1 infections, and pangenotypic inhibitors of HCV NS5B RNA-dependent RNA polymerase and HCV NS5A.

HCV is characterized by a high level of genetic diversity [3]. Seven HCV genotypes and more than 100 subtypes, including 69 confirmed, were described. Genomes of the same subtype differ by <13% over their complete coding region sequences, and those that differ by >15% belong to different genotypes and subtypes. While implementing full-length HCV NS5B polymerase gene sequencing in our laboratory, we identified atypical HCV sequences related to recently proposed subtype 2l [3, 4], which were obtained from the serum of 2 patients. Then, we screened our laboratory hepatitis virus sequence database [5] and detected that HCV sequences of subtype 2l had been obtained from serum samples of 5 additional patients, whereas HCV-2l sequences collected from only 14 patients were detected in the NCBI GenBank nucleotide sequence database.

Materials and Methods

Conserved regions in HCV RNA-dependent RNA polymerase gene (NS5B) were identified and used to design PCR primers using the SVARAP tool [6] based on a set of aligned HCV reference genomes obtained from the Los Alamos HCV sequence database [7] (online suppl. table S1; for all online suppl. material, see www.karger.com/10.1159/000369015). HCV RNA was extracted from 200 µl of serum using the EZ1 viral nucleic acid extraction kit (Virus Mini kit version 2.0, Qiagen, Courtaboeuf, France) according to the manufacturer's recommendations. Aliquots (10 µl) of extracted RNA were amplified by reverse transcription PCR using 0.5 µl of SuperScript III/Platinum Taq polymerase (Invitrogen Life Technologies, Carlsbad, Calif., USA), 0.4 µM of each dNTP, 3.2 µM MgSO₄ and 200 nM of each primer under the following conditions: 50° for 30 min for reverse transcription, 95° for 4 min for initial denaturation, then 40 cycles including 30 s at 95°, 45 s at 56° and 2 min at 72°, and a final elongation step at 72° for 10 min. Nested PCR was performed with the FastStart Taq polymerase (Roche Applied Science, Mannheim, Germany) and included 40 cycles with 30 s at 95°, 45 s at 56° and 2 min at 72° before

a final elongation step at 72° for 10 min. Thereafter, 1,570 nucleotide-long HCV genome fragments that cover the near full-length NS5B-encoding gene were obtained using the ABI Prism 3130xl Genetic Analyzer (Applied Biosystems, Branchburg, N.J., USA) under conditions previously described [5] and were analyzed using the SeqScape v2.5 software (Applied Biosystems). Primers used for PCR amplification and sequencing are indicated in online suppl. table S1. In addition, a full-length HCV NS3 protease encoding gene and a fragment of 5'-untranslated region (5'-UTR) were obtained as previously described [5]. These sequences were aligned using Muscle software [8] with an HCV sequence reference panel [7]. Pairwise p-distances were computed using the BioEdit software (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>). Phylogenetic trees were built using MEGA v5.1 software [9] with the maximum-likelihood method based on the most appropriate substitution model, as determined by MEGA v5.1 for the sequence set, or the neighbor-joining method based on the Kimura 2-parameter model; tests of phylogeny used the bootstrap method with 1,000 resamplings of the data. BLAST searches were done for HCV sequences obtained in the present study against the NCBI GenBank nucleotide sequence database (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) and our clinical microbiology laboratory sequence database [5]. Inter-genotypic or intersubtypic recombination was excluded using the SimPlot tool (sray.med.som.jhmi.edu/SCSoftware/simplot/) through comparison with a reference set of HCV genomes, including those classified as 2l (KC197235 and KC197240).

Results

HCV sequences corresponding to the near full-length HCV NS5B polymerase gene, most similar to genotype 2l apart from other genotype 2 HCV, were first obtained from serum samples of 2 case-patients (fig. 1). These sequences exhibited 93% nucleotide identity between each other. BLAST searches against our laboratory sequence database identified 5 other sequences that belong to the same phylogenetic cluster. In addition, these 2 sequences

Fig. 1. Phylogenetic reconstruction based on near full-length HCV NS5B RNA polymerase encoding gene recovered in the present study, their best BLAST hits in our laboratory [5] and the NCBI nucleotide sequence databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) and a set of HCV reference genomes [7]. The sequences obtained in the present study are indicated by a black background and a white font. Sequences obtained in our laboratory are indicated by a black frame. Sequences with the highest BLAST score recovered from the NCBI GenBank nucleotide sequence database are indicated in boldface. Sequence alignment was performed by the MUSCLE program [8] and the tree was built with MEGA5 software [9], by using the maximum likelihood

method based on the general time reversible model. The tree with the highest log likelihood (-31776.8398) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Branches with bootstrap values >50% as obtained from 1,000 resamplings of the data are labeled on the tree. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter = 0.3910)]. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved a set of 45 nucleotide sequences with a total of 1,568 positions (nucleotides 7,611–9,179 in reference to GenBank accession No. NC004102) in the final dataset.

(For figure see next page.)

showed 93–94 and 92–95% identity, respectively, with their 3 top BLAST hits in the NCBI GenBank nucleotide sequence database, which were genomes KC197235 and KC197240 that were released in GenBank in September 2013 and classified as subtype 2l [3, 4] and sequence FJ872270 [10]. These 3 latter sequences were obtained in France. Apart from these 3 sequences, the 10 best BLAST hits in GenBank included HCV sequences of subtypes 2c, 2q or 2b, with 79–80, 79 and 79% nucleotide identity, respectively, for the first NS5B sequence recovered here, and HCV sequences of subtypes 2c or 2b with 79% nucleotide identity for the second NS5B sequence recovered here. Overall, mean ($\pm$ SD) nucleotide identity between NS5B long fragments of subtype 2l was $93.4 \pm 0.85\%$ (range: 92.4–95.1), whereas the mean nucleotide identity between sequences of subtype 2l and 2 non-l was $77.9 \pm 0.80\%$ (75.7–79.9). Therefore, sequences reported in the present work can be classified as of subtype 2l because, in addition to their phylogenetic clustering with the 2 recently described HCV-2l genomes, they exhibit far lower nucleotide differences with HCV-2l sequences than those defined as threshold for subtype classification for genotype 2 (13.1–17.6%) in the recent article on expanded classification of HCV genotypes [3]. Intergenotypic or intersubtypic recombination was excluded by using the SimPlot tool (sray.med.som.jhmi.edu/SCSoftware/simplot/) with HCV reference genomes (including HCV-2l genomes).

BLAST searches using a shorter HCV NS5B fragment (nt 8,342–8,650 in reference to HCV 2a genome accession No. AF177036) classically used for genotype identification and used in routine clinical microbiology practice in our institution detected 5 additional HCV-2l sequences in our laboratory sequence database and 9

additional HCV-2l sequences in the GenBank sequence database, 7 being labeled as 2l (fig. 2). All of these additional HCV sequences were obtained in France. The same results were obtained when incorporating into the analysis sequences from the recent article by Smith et al. [3] on expanded classification of HCV genotypes and other HCV-2 non-l sequences recovered from the GenBank sequence database (online suppl. fig. S1). Two of these sequences were branched with HCV subtype 2l sequences, apart from HCV-2 non-l subtypes, but they exhibited a mean nucleotide identity of 84.3 ± 1.4 (81.3–86.6) with HCV-2l sequences. In congruence with the HCV NS5B encoding gene, the best BLAST hits for HCV NS3 protease encoding sequences (543 nucleotides) obtained from the 2 case-patients were sequences identified either as belonging to subtype 2l (KC197235 and KC197240 [4]; with 91–92% nucleotide identity for the first sequence and 87–88% nucleotide identity for the second sequence) or as belonging to genotype 2 but unclassified at the subtype level (HQ623293 and HQ623295 [10]; with 91% nucleotide identity for the first sequence and 88–91% nucleotide identity for the second sequence; fig. 3). All of these HCV NS3 sequences found in GenBank had been recovered from serum samples collected in France. Finally, the best matches for 5'-UTR nucleotide sequences obtained from the 2 case-patients were the 2 HCV-2l genomes (KC19740 and KC197235, with 99% nucleotide identity; online suppl. fig. S2).

Overall, HCV-2l sequences represented 0.25, 0.15 and 1% of 2,788, 1,333 and 688 HCV NS5B polymerase, NS3 protease and 5'-UTR sequences, respectively, available in our laboratory. For all HCV sequences analyzed here, phylogeny reconstructions delineated with strong confidence

Fig. 2. Phylogenetic reconstruction based on a short fragment of the HCV NS5B RNA polymerase encoding gene recovered in the present study, their best BLAST hits in our laboratory [5] and the NCBI nucleotide sequence databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) and a set of HCV reference genomes [7]. The legend is the same as fig. 1, except that the tree was built using the neighbor-joining method with the Kimura 2-parameter model; test

of phylogeny used the bootstrap method with 1,000 resamplings of the data. This analysis involved a set of 58 nucleotide sequences with a total of 250 positions (nucleotides 8,344–8,591 in reference to GenBank accession No. NC004102) in the final dataset. A tree built by using the maximum likelihood method with an expanded set of HCV genotype 2 sequences is shown in online supplementary figure S1.

Fig. 3. Phylogenetic reconstruction based on HCV NS3 protease encoding gene recovered in the present study, their best BLAST hits in our laboratory [5] and the NCBI nucleotide sequence databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) and a set of HCV reference genomes [7]. The legend is the same as for fig. 1,

except that the tree was built by using the neighbor-joining method with the Kimura 2-parameter model; test of phylogeny used the bootstrap method with 1,000 resamplings of the data. The analysis involved a set of 39 nucleotide sequences with a total of 542 positions (nucleotides 3,420–3,962 in reference to GenBank accession No. NC004102) in the final dataset.

(For figures 2 and 3 see next pages.)

Table 1. Clinical and epidemiological data available for 4 of the 7 HCV-2l-infected patients followed up in our institution

Patient No.	Gender, age	Geographical origin	Possible source of infection	Treatment of hepatitis C	Date of blood collection
1	F, 55	France	Hemophilic husband chronically infected with HCV through transfusion	Pegylated interferon and ribavirin, with initial virological response, then relapse	July 27, 2012
2	M, 86	The Netherlands	Unknown	None	December 14, 2010
3	F, 60	France	Unknown	Liver transplantation in 2002, then interferon-ribavirin for HCV recurrence; lost to follow-up thereafter	November 27, 2002
4	M, 38	France	Exchange transfusion at birth	HCV treatment initiated in 2001, but patient was lost to follow-up	October 1, 2001

a cluster encompassing HCV-2l sequences apart from other genotype 2 subtypes (fig. 1–3; online suppl. fig. S1).

The same phylogenetic tree topology as the near full-length NS5B nucleotide sequences was obtained with the near full-length NS5B amino acid sequences (online suppl. fig. S3). Analysis of the NS5B RNA-dependent RNA polymerase amino acid sequence at 30 positions critical for interaction with or resistance to HCV polymerase inhibitors [11–14] showed presence of amino acid substitution C451T in all HCV-2l sequences, as in all HCV genotype 4, 6 and 7 sequences and most of the genotype 3 HCV, whereas subtypes 2 non-l HCV harbored the C451V substitution. The M414L substitution that confers resistance to HCV nonnucleoside inhibitors [15, 16], already described in most of the genotype 4 HCV [17], was found in all HCV-2l sequences, whereas HCV of subtypes 2 non-l harbor amino acid substitution M414Q that was also associated with drug resistance in replicon assays [18]. In contrast, among the amino acid substitutions identified at 7 positions in the NS3 protease as associated with decreased susceptibility to HCV protease inhibitors, only the Q80G substitution was found in HCV-2l, as in other HCV genotype 2 subtypes [13].

Clinical and epidemiological data were available for 4 of the 7 HCV-2l-infected patients followed up at our institution and are summarized in table 1.

Discussion

We describe here 7 HCV NS5B RNA polymerase sequences of subtype 2l, representing more than half the number of sequences (n = 12) previously available world-

wide. When considering any HCV genome region, only 19 HCV-2l sequences were available from GenBank, corresponding to 14 HCV strains, including only 2 full-length genomes released in 2013, and some of these sequences were un- or misclassified at the subtype level. Phylogenetic analyses based on HCV NS5B and NS3 genes show in congruence with a previous report [4] that these HCV sequences are clustered together apart from HCV 2 non-l sequences that compose a second cluster. In addition, HCV NS5b sequences recovered here can be classified as 2l based on their level of nucleotide identity with other HCV-2l genomes and with sequences classified as 2 non-l based on the recent recommendation for the classification within HCV genotype 2 subtypes [3]. It is noteworthy that all currently available HCV-2l sequences and those recovered in the present study were obtained from serum samples collected in France (in metropolitan France in all but 1 case) and the reason for this geographical distribution is unresolved. In addition, HCV-2l NS5B polymerases were found to specifically harbor 2 amino acids that may influence susceptibility to currently available HCV polymerase inhibitors, albeit the clinical relevance of these observations is unclear, particularly with respect to the tremendous potency of these drugs. Taken together, previous findings indicate that HCV-2l has been rarely detected worldwide and only from serum samples collected in France, and they highlight that HCV diversity could still be incompletely explored.

Disclosure Statement

No potential conflict of interest or financial disclosure for all authors.

References

- 1 Liang TJ, Ghany MG: Current and future therapies for hepatitis C virus infection. *N Engl J Med* 2013;368:1907–1917.
- 2 Fox AN, Jacobson IM: Recent successes and noteworthy future prospects in the treatment of chronic hepatitis C. *Clin Infect Dis* 2012; 55(suppl 1):S16–S24.
- 3 Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, Simmonds P: Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and assignment web resource. *Hepatology* 2014;59:318–327.
- 4 Jordier F, Deligny ML, Barre R, De MP, Cantaloube JF: Evidence for two phylogenetic clusters within hepatitis C virus (HCV) genotype 2 inferred from analysis of complete coding sequences of 15 HCV strains. *J Med Virol* 2013;85:1754–1764.
- 5 Colson P, Bregigeon S, Tourres C, Solas C, Poizot-Martin I, Tamalet C: Relapse of hepatitis C virus after 14 months of sustained virological response following pegylated-interferon alpha plus ribavirin therapy in a human immunodeficiency virus type 1 infected patient. *J Clin Virol* 2013;58: 309–314.
- 6 Colson P, Tamalet C, Raoult D: SVARAP and aSVARAP: simple tools for quantitative analysis of nucleotide and amino acid variability and primer selection for clinical microbiology. *BMC Microbiol* 2006;6:21.
- 7 Kuiken C, Hraber P, Thurmond J, Yusim K: The hepatitis C sequence database in Los Alamos. *Nucleic Acids Res* 2008;36:D512–D516.
- 8 Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797.
- 9 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28:2731–2739.
- 10 Legrand-Abravanel F, Henquell C, Le Guillou-Guillemette H, Balan V, Mirand A, Dubois M, Lunel-Fabiani F, Payan C, Izopet J: Naturally occurring substitutions conferring resistance to hepatitis C virus polymerase inhibitors in treatment-naïve patients infected with genotypes 1–5. *Antivir Ther* 2009;14: 723–730.
- 11 Jaspe RC, Sulbaran YF, Sulbaran MZ, Loureiro CL, Rangel HR, Pujol FH: Prevalence of amino acid mutations in hepatitis C virus core and NS5B regions among Venezuelan viral isolates and comparison with worldwide isolates. *Virol J* 2012;9:214.
- 12 Paolucci S, Fiorina L, Mariani B, Gulminetti R, Novati S, Barbarini G, Bruno R, Baldanti F: Naturally occurring resistance mutations to inhibitors of HCV NS5A region and NS5B polymerase in DAA treatment-naïve patients. *Virol J* 2013;10:355.
- 13 Sarrazin C, Zeuzem S: Resistance to direct antiviral agents in patients with hepatitis C virus infection. *Gastroenterology* 2010;138:447–462.
- 14 Bartenschlager R, Lohmann V, Penin F: The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection. *Nat Rev Microbiol* 2013;11:482–496.
- 15 Plaza Z, Soriano V, Gonzalez MM, Di Lello FA, Macias J, Labarga P, Pineda JA, Poveda E: Impact of antiretroviral therapy on the variability of the HCV NS5B polymerase in HIV/HCV co-infected patients. *J Antimicrob Chemother* 2011;66:2838–2842.
- 16 Mo H, Lu L, Pilot-Matias T, Pithawalla R, Mondal R, Masse S, Dekhtyar T, Ng T, Koev G, Stoll V, Stewart KD, Pratt J, Donner P, Rockway T, Maring C, Molla A: Mutations conferring resistance to a hepatitis C virus (HCV) RNA-dependent RNA polymerase inhibitor alone or in combination with an HCV serine protease inhibitor in vitro. *Antimicrob Agents Chemother* 2005;49:4305–4314.
- 17 Newman RM, Kuntzen T, Weiner B, Berical A, Charlebois P, Kuiken C, Murphy DG, Simmonds P, Bennett P, Lennon NJ, Birren BW, Zody MC, Allen TM, Henn MR: Whole genome pyrosequencing of rare hepatitis C virus genotypes enhances subtype classification and identification of naturally occurring drug resistance variants. *J Infect Dis* 2013;208:17–31.
- 18 Nyanguile O, Pauwels F, Van den Broeck W, Boutton CW, Quirynen L, Ivens T, van der Helm L, Vandercruyssen G, Mostmans W, Delouvrois F, Dehertogh P, Cummings MD, Bonfanti JF, Simmen KA, Raboisson P: 1,5-benzodiazepines, a novel class of hepatitis C virus polymerase nonnucleoside inhibitors. *Antimicrob Agents Chemother* 2008;52:4420–4431.