Analysis of the Complete Open Reading Frame of Hepatitis C Virus in Genotype 2a Infection Reveals Critical Sites Influencing the Response to Peginterferon and Ribavirin Therapy.

Makoto Kadokura[1)], Shinya Maekawa[1)], Ryota Sueki[1)], Mika Miura[1)], Kazuki Komase[1)], Hiroko Shindo[1)], Fumitake Amemiya[1)], Tomoyoshi Uetake[1)], Taisuke Inoue[1)], Minoru Sakamoto[1)], Mina Nakagawa[2)], Naoya Sakamoto[2)], Mamoru Watanabe[2)], Nobuyuki Enomoto[1)]

1) First Department of Internal Medicine, Faculty of Medicine, University of Yamanash; 1110, Shimokato, Chuo, Yamanashi 409-3898, Japan.

2) Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University; 1-5-45, Yushima, Bunkyo, Tokyo, 113-8510, Japan

Short title: PEG-IFN/RBV response in HCV-2a

This study was supported in part by a grant-in-aid scientific research fund of the Ministry of Education, Science, Sports and Culture number 20390206 and in part by a grant-in-aid from the Ministry of Health, Labour, and Welfare of Japan (H19-kanen-002).

Correspondence : Shinya Maekawa M.D./Ph.D.

First Department of Internal Medicine, Faculty of Medicine, University of Yamanashi

1110, Shimokato, Chuo, Yamanashi 409-3898, Japan.

Tel: +81-5-5273-9584

Fax: +81-5-5273-6748

E-mail: maekawa@yamanashi.ac.jp

Abbreviations

EVR: Early Virological response

IFN: Interferon

IRF-1: Interferon Regulatory Factor 1

IRRDR: Interferon Ribavirin Resistance Determinant Region

ISDR: Interferon Sensitivity Determinant Region

ORF: Open Reading Frame

PEG-IFN: Pegylated-Interferon

PePHD: PKR -eIF2 Phosphorylation Homology Domain

PKR-BD: Double-Stranded RNA-activated Protein Kinase Binding Domain

RBV: Ribavirin

RVR: Rapid Virological Response

SVR: Sustained Virological Response

ABSTRACT

Purpose: A proportion of patients infected with genotype 2a hepatitis C virus (HCV) cannot achieve a sustained virological response (SVR) to pegylated-interferon plus ribavirin therapy (PEG-IFN/RBV) but the reason remains unclear. The present study aimed to clarify the possible correlation between viral sequence variations and final outcome.

Methods: The pretreatment complete open reading frame (ORF) sequences of genotype 2a HCV were determined by direct sequencing for two independent groups of patients (43 patients as test; group 1 and 35 as validation; group 2), and the correlation with the final outcome was explored.

Results: Patients with SVR (n=58) and with non-SVR (n=20) differed significantly in pretreatment HCV RNA level ($p=0.002$), fibrosis score ($p=0.047$), and cumulative ribavirin dosage ($p=0.003$). By comparison of all amino acid positions in the complete HCV ORFs, threonine at amino acid (aa) 110 in the core region was remarkably frequent in SVR ($p=0.01$ for group 1, $p=0.004$ for group 2, and $p=5E-05$ for combined). A sliding window analysis revealed that the total numbers of amino acid variations within the NS5A aa 2258 to 2306 region were significantly high in SVR compared to non-SVR patients ($p=0.01$ for group 1, $p=0.006$ for group 2, and $p=0.0006$ for combined). Multivariate analyses revealed that core aa 110 ($p=0.02$), NS5A aa 2258-2306 ($p=0.03$), and cumulative ribavirin dosage ($p=0.02$) were identified as independent variables associated with the final outcome.

Conclusions: The outcome of PEG-IFN/RBV therapy is significantly influenced by variation in the core and NS5A regions in genotype 2a HCV infection.

INTRODUCTION

Worldwide, 180 million of people are estimated to be infected with hepatitis C virus (HCV), and HCV is a major cause of chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (1). In HCV-infected patients with chronic hepatitis, treatment with interferon (IFN) can result in viral clearance and biochemical and histological improvements(2). The response to the therapy varies according to HCV genotype and pretreatment HCV RNA level (3-4).

The currently recommended treatment for patients infected genotype 2a HCV with high viral load is pegylated-interferon (PEG-IFN) plus ribavirin (RBV) for 24 weeks (1). Approximately 80% of patients infected with genotype 2a HCV can achieve a sustained virological response (SVR) with this regimen (5-6), although much lower percentages of patients infected with other genotypes can achieve SVR, especially with genotype 1(1). Because of its high response rate, shorter treatment duration was suggested by some studies, although an agreement has not been reached yet (7-8). On the other hand, about 20% of patients infected with this genotype cannot achieve SVR and it remains elusive which patients show poor responses.

Previous studies have reported that amino acid variations in the NS5A-ISDR (9), NS5A-IRRDR(10), NS5B(11), PKR-eIF2 phosphorylation homology domain (PePHD) of E2(12), and core (13-14) correlate with clinical outcome of IFN-based therapy, including PEG-IFN/RBV therapy in patients infected with genotype 1b HCV. Recent full HCV open reading frame analysis for genotype 1 also have reported that core, NS3, and NS5A were associated with early viral response and the outcome in PEG-IFN/RBV therapy (15-16). However, in genotype 2a infection, only a few studies have investigated the association between HCV sequence variation and treatment

response (17-19), and the role of viral factors has not been established yet, especially in the era of PEG-IFN/RBV therapy. Moreover, these previous studies investigated only several isolated HCV genomic regions, and comprehensive analysis of the full HCV open reading frame (ORF) has not been undertaken so far.

In the present study, to assess comprehensively the influence of viral variations on response to the PEG-IFN/RBV therapy in genotype 2a HCV infection, we determined the complete pretreatment HCV ORFs from Japanese patients and investigated viral amino acid variation and their correlation with the response to the combination therapy of PEG-IFN plus RBV.

PATIENTS & METHODS

Study Population

A total of 103 adult Japanese patients infected with genotype 2a HCV, who received the combination therapy with PEG-IFN (PEGINTRON®, Schering-Plough, Tokyo, Japan) plus RBV (REBETOL®, Schering-Plough) between 2005 and 2008 at University of Yamanashi, Tokyo Medical and Dental University, and related institutions were first included in the study. They all fulfilled following criteria: (1) negative for hepatitis B surface antigen, (2) high viral load (≥100 KIU/ml), (3) absence of hepatocellular carcinoma, (4) no other form of hepatitis, such as primary biliary cirrhosis, autoimmune liver disease, or alcoholic liver disease, (5) free of co-infection with human immunodeficiency virus. Informed consent was obtained from each patient. The study was approved by the ethics committees of all the participating universities and hospitals. The therapy was performed according to the standard treatment protocol of PEG-IFN/RBV therapy for Japanese patients established by a hepatitis study group of the Ministry of Health, Labour, and Welfare, Japan (PEG-IFNα-2b 1.5 μg/kg body weight, once weekly subcutaneously, and RBV 600-800 mg daily per os for 24 weeks). To clearly disclose the non-SVR viral characteristics, we have considered those patients who achieved total drug administration of 60% or more for both PEG-IFN and RBV, with the completion of the standard treatment duration. Moreover, although we excluded the patients with extended therapy to make the studied population uniform, we have included non-SVR patients with extended therapy to clarify the specific characteristics of non-SVR patients, a minor population group. As a result, 25 patients were excluded for the following reasons: 4 patients received insufficient dose, 8 patients were discontinued from the therapy within 12 weeks, and 13 SVR patients received

extended therapy. Finally, 78 patients were considered as eligible for the study. During the combination therapy, blood samples were obtained at least once every month before, during and after treatment and were analyzed for blood count, ALT and HCV RNA levels. Liver biopsy specimens were obtained from most of the patients.

The 78 patients belonging to the different institutions were separately analyzed: 43 patients registered in Y-PERS (Yamanashi Pegintron Ribavirin Study Group) were included in group 1 (test group), and the 35 patients from Tokyo Medical and Dental University and related institutions (Ochanomizu Liver Conference Group) were included in group 2 (validation group). We divided the patients into these two groups in order to exclude the false positives (type I errors) which might arise in successive HCV-ORF study. Since genotype-2a HCV contains as many as 3033 amino acids, it was possible that incorrect amino acids to be judged as significant in full HCV-ORF comparison study as a result of type I errors. Therefore, to guard against false positives, HCV-ORF comparison study was undertaken in group 1, group 2, and combined group.

Complete HCV-ORF Sequence Determination by Direct Sequencing from Pretreatment Sera

HCV RNA was extracted from pretreatment serum samples by the AGPC method using Isogen (Wako, Osaka, Japan) according to the manufacturer's protocol. Complementary DNA was synthesized with Superscript II (Invitrogen, Tokyo, Japan) using random primers (Invitrogen) and then amplified by two-step nested PCR using the primers newly designed for this study. All samples were initially denatured at 95°C for 7 min., followed by 40 cycles with denaturation at 95°C for 15 seconds, annealing at 55°C

for 15 seconds, and extension at 72°C for 45 seconds with BD Advantage™ 2 PCR Enzyme System (BD Biosciences Clontech, CA, USA).

PCR amplicons were sequenced directly by Big Dye Terminator Version 3.1 (ABI, Tokyo, Japan) with universal M13 forward / M13 reverse primers using an ABI prism 3130 sequencer (ABI). Generated sequence files were assembled using Vector NTI software (Invitrogen) and base-calling errors were corrected following inspection of the chromatogram.

Sliding Window Analysis

A sliding window analysis was introduced to search through HCV amino acid "regions", rather than single amino acid positions, related to the final outcome of PEG-IFN/RBV therapy. Briefly, the total number of amino acid substitutions compared to the consensus sequence within a given amino acid length were counted in each amino acid position in each HCV sequence. Then the relation of substitution numbers and the final outcome was compared statistically between the SVR and non-SVR groups by Mann-Whitney's U test for each amino acid position. In this study, we changed the window length from 1 to 50 to search for those HCV regions. To visualize the result, significantly lower p-values were colored in red and non-significant p-values were colored in green to generate a "heat map" appearance using Microsoft Excel software. In the present study, p-value of 1/1000 or lower was colored in the maximum red.

Statistical Analysis

Statistical differences in the parameters, including all available patients'

demographic, biochemical, hematological, and virological data such as sequence variation factors, was determined between the various groups by Student t test or Mann-Whitney's U test for numerical variables and Fisher's exact probability test for categorical variables. To evaluate the optimal threshold of variations for SVR prediction, the receiver operating characteristic curve was constructed. Variables that achieved statistical significance ($p<0.05$) in univariate analysis were entered into multiple logistic regression analysis to identify significant independent factors. We also calculated the odds ratios and 95% confidence intervals. All p values of <0.05 by the two-tailed test were considered significant.

RESULTS

Characteristics of the patients studied

Of the patients analyzed, the SVR rate was 78.3% (58/74) with the standard therapy (four non-SVR patients received an extended therapy). The baseline characteristics of the patients (group 1, group2, and combined) classified according to SVR achievement are shown in Table 1. Fibrosis score (*p=0.047*) and HCV RNA levels (*p=0.002*) were significantly higher in non-SVR patients, but the cumulative ribavirin dose ≥80% (*p=0.003*) and rapid virological response (RVR) rate (*p=0.011*) were significantly higher in SVR patients. In addition, patients with non-SVR had a tendency to be older (*p=0.058*). Achievement of RVR reached 61.5% when all patients were included, and this rate was extremely high compared to achievement of RVR in patients with genotype 1b infection (~10%) observed in Yamanashi University Hospital (data not shown). The early virological response (EVR) rate was equally high in the SVR (100%) and non-SVR (89%) groups, showing that relapse to be the characteristic feature of the non-SVR patients with genotype 2a HCV. Actually, 18 patients in non-SVR were relapser, while two patients were null responder.

Comparison of amino acid variations between the SVR and non-SVR in the complete HCV polyprotein and each HCV protein

To determine whether the sequence variations differed between the SVR and non-SVR groups, we first compared amino acid variations that were unique, relative to a population consensus, to either the SVR or non-SVR patients for the complete HCV polyprotein and each HCV protein. The number of amino acid variations in the

sequences from the SVR patients was significantly higher than in those from the non-SVR patients, when the entire HCV polyprotein was analyzed (Fig.1, left). These differences were especially significant in E1 and NS3 (Fig.1, right). This result demonstrated that HCV sequences from patients with SVR comprised a heterogeneous population, while HCV sequences from patients with non-SVR comprised a rather homogeneous population, indicating the existence of unique non-responsive HCV sequences.

Comparison of HCV sequence variation between the SVR and non-SVR patients at each amino acid position

Next, each amino acid position in the HCV ORF was compared to detect any differences between the SVR and non-SVR patients after determination of the consensus sequence from all 78 patients. In Fig.2a, the final differences of the two independent studies combined are shown as dots demonstrating – logP values. As shown in the figure, amino acid usage at amino acid 110 in the core region differed strikingly between the two groups (*p=5E-05*). The site was detected in group 1 (*p=0.01*) and was validated in group 2 *(p=0.004)* (Table 2a), and the final p-value became remarkably high, making the p-value at this site most significantly low. Variations of aa 773 in p7, aa 2099 in the NS5A, and aa 3013 in NS5B were also shown to differ significantly between the SVR and the non-SVR patients when the two studies were combined; however, they were not confirmed by one of the studies (Table 2a). Fig.2b shows the aligned sequences of amino acids 1-120 of the core region. Substitutions at aa 110 from non-T (N/S) to T were significantly more frequent in SVR (32/58, 55.2%) than in non-SVR (1/20, 3.6%, *p=5E-05*). Amino acid 4, the site reported recently to vary

according to the viral response in genotype 2a infection, did not differ significantly in our study. Amino acid 70 and 91, which have been reported to vary according to viral response to PEG-IFN/RBV therapy in genotype 1b infection, were conserved irrespective of the outcome.

Comparison of amino acid variation between the SVR and non-SVR patients across HCV "regions" using sliding window analysis

Fig.2c shows the combined result of sliding window analysis for study groups 1 and 2, this approach was used to detect differing HCV amino acid "regions", rather than single amino acid positions, between the SVR and the non-SVR patients. According to the result, four regions were notably associated with the final outcome (p-values less than 1/1000). Core aa 110, detected as a single amino acid position discriminating between the SVR and the non-SVR patients, was also identified as one of these regions. Because core aa 110 was already known for its strong correlation with the response as above, the region was excluded from further analysis. Among the other three regions, only NS5A aa 2258-2306 showed significant differences in the two independent study groups (Table 2b). Interestingly, the NS5A region overlapped the PKR-binding domain, which includes the interferon sensitivity determining region (ISDR). Fig.2d shows the aligned sequences of amino acids around 2258-2306 of HCV NS5A. As with previous studies, variations in the ISDR were also significantly more frequent in SVR patients.

Multivariate analysis to detect independent factors contributing to the SVR

Multivariate analysis revealed that variation of core aa 110, the total number of

substitutions within NS5A aa 2258-2306, and total ribavirin dose ≥80% were finally identified as the independent variables influencing the final outcome (odds ratio 24.7, 11.5 and 16.0; $p = 0.02, 0.03$ *and* 0.02, Table 3).

Biological relevance of variation in core and NS5A in this study group

To determine biological relevance of core aa110 and NS5A aa2258-2306, we investigated their relationship with clinical background factors. Multiple variations in the NS5A region aa 2258-2306 were significantly related to pretreatment HCV RNA titer ($p=9E-05$, Fig. 3 and Table 4a). Interestingly, variation of the core aa110 was significantly associated with the patients' age ($p=0.03$, Table 4b).

DISCUSSION

In this study, based on analysis of complete HCV ORF sequences and comparison of SVR and non-SVR patients in two independent study groups, we have shown that amino acid variations in the core and NS5A correlate most significantly with the final outcome in the treatment for genotype 2a chronic hepatitis C. The study is unique in that the patients studied were all Japanese, excluding any affect of racial differences and providing a clearer analysis of the viral differences.

From the analysis of the characteristics of patients infected with genotype 2a HCV, it was clear that most non-SVR patients responded to the PEG-IFN/RBV therapy at least transiently, given that most of these non-SVR patients (89%) achieved EVR. This result demonstrated that most non-SVR patients were relapser, but were not null-responders as observed frequently among genotype 1b patients treated with PEG-IFN/RBV therapy. Therefore, we compared the different viral responses according to the final outcome of SVR or non-SVR.

Variation of core aa 110 was identified as the single amino acid residue most significantly related to the final outcome ($p=5E-05$). In recent studies of treatment of genotype 1b infection with PEG-IFN/RBV, amino acid variation in the core region was reported to be associated with response. It is interesting that the core region was also identified as an HCV gene associated with the response to PEG-IFN/RBV therapy of genotype 2a infection, although the amino acid residues of core in genotype 1b were different, being aa 70 and aa 91. It is also interesting that amino acids aa 70 and aa 91 are conserved as arginine and leucine, respectively, in genotype 2a, as reported to be associated with favorable PEG-IFN/RBV responses in genotype 1b infection, consistent with the association with a high SVR rate in genotype 2a infection. Very recently, a

correlation was reported between amino acid variations in the core region and viral responses of genotype 2a HCV infection (20). Though the result seems discrepant from our study, we suspect the inconsistent results were at least partially attributable to the different groups used in comparison: we compared the difference between non-SVR patients and SVR patients while they compared the difference between non-SVR and RVR patients.

In systemic searching for the viral "regions" associated with the treatment outcome, NS5A aa2258-2306 was identified by two independent studies. Interestingly, the region overlaps the PKR-binding domain (PKR-BD), including the ISDR, in which the number of amino acid substitutions is known to be related to the response to IFN-based therapy in genotype 1b, and also in genotype 2a (17-18). Therefore, we also confirmed that total number of substitutions in the ISDR and PKR-BD is significantly associated with the final outcome in this group of patients when the two studies were combined.

Some viral regions other than core and NS5A also showed the potential association with the final outcome. Viral single amino acid substitutions of aa 773 in p7, aa 2099 in the NS5A, and aa 3013 in NS5B, or viral regions in E1 aa 400-403 and in E2 aa 724-744 were more frequent in SVR. However, because these were not extracted as significant in one of the two studies when analyzed separately, additional studies are needed to confirm the association with the final outcome. On the other hand, we could not find an association with the final outcome and the PePHD or IRRDR, including the V3 regions (data not shown) reported 1b HCV infection (21-22).

It is interesting that the variation of the core region showed the clear association with age. The younger patients with core aa 110T showing favorable

responses while the older patients with core aa 110 non-T showed unfavorable responses. It is possible that different response rates according to the patients' ages in genotype 2a infection might have been related to the core substitutions, although further study is needed. In NS5A, it was reported that the variations within the PKR-binding region, including those within the ISDR, can disrupt the NS5A-PKR interaction, possibly rendering HCV sensitive to the antiviral effects of interferon (23). Clinically, the number of substitutions within the region has been reported to correlate with the serum HCV RNA level (12). We also confirmed that the number of substitutions within the NS5A aa 2258-2306 was significantly associated with the pretreatment HCV RNA titers.

Multivariate analysis of the combined group of patients showed that variation of core aa 110, NS5A aa 2258-2306, and total ribavirin dose ≥80% were independent variables associated with the final outcome (Table. 3). The association of ribavirin dose and HCV relapse rate was reported previously (24) and that result was confirmed in this study. On the other hand, the total PEG-IFN dosage was not identified when it was administered at greater than 60% of the initially scheduled amount. Indeed, when the drug dosage is excluded, the strongest association was seen in the viral elements of core and NS5A, revealing the importance of these two regions in the treatment of genotype 2a HCV infection with PEG-IFN/RBV therapy.

On the other hand, our study still has some limitations. In recent studies, IL28B single nucleotide polymorphisms were reported to be correlated significantly with the treatment response in genotype 1b HCV infections (25-26). In genotype 2a HCV infection, a correlation was also reported to exist between the IL28B SNP and the treatment response (27). However, we could not investigate the association of the IL28B

single nucleotide polymorphisms in the treatment response in genotype 2a HCV infections. In addition, the number of analyzed patients was rather small, especially in non-SVR patients.

In conclusion, by comprehensive investigation of the complete HCV ORF in patients showing different responses to PEG-IFN/RBV therapy, we have demonstrated that amino acid variation in the core and NS5A are significantly associated with the final outcome of treatment of genotype 2a chronic hepatitis C. Considering this result, determination of those HCV regions before treatment might provide further benefits for the patients infected with genotype 2a HCV.

ACKNOWLEDGMENT

We are grateful to Yamanashi Pegintron Ribavirin Study Group and Ochanomizu Liver Conference Group for their cooperation and advices. Especially, we thank Asuka Kanayama and Takako Ohmori for their technical assistance, and Takatoshi Kitamura and Shunichi Okada for their cooperation and advices.

REFERENCES

1. Ghany MG, Strader DB, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C: an update. Hepatology2009 Apr;49(4):1335-74.

2. Di Bisceglie AM, Martin P, Kassianides C, Lisker-Melman M, Murray L, Waggoner J, Goodman Z, Banks SM, Hoofnagle JH. Recombinant interferon alfa therapy for chronic hepatitis C. A randomized, double-blind, placebo-controlled trial. N Engl J Med1989 Nov 30;321(22):1506-10.

3. Haydon GH, Jarvis LM, Blair CS, Simmonds P, Harrison DJ, Simpson KJ, Hayes PC. Clinical significance of intrahepatic hepatitis C virus levels in patients with chronic HCV infection. Gut1998 Apr;42(4):570-5.

4. Simmonds P. Clinical relevance of hepatitis C virus genotypes. Gut1997 Mar;40(3):291-3.

5. Hadziyannis SJ, Sette H, Jr., Morgan TR, Balan V, Diago M, Marcellin P, Ramadori G, Bodenheimer H, Jr., Bernstein D, Rizzetto M, Zeuzem S, Pockros PJ, Lin A, Ackrill AM. Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. Ann Intern Med2004 Mar 2;140(5):346-55.

6. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, Goodman ZD, Koury K, Ling M, Albrecht JK. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet2001 Sep 22;358(9286):958-65.

7. Dalgard O, Bjoro K, Hellum KB, Myrvang B, Ritland S, Skaug K, Raknerud N, Bell H. Treatment with pegylated interferon and ribavarin in HCV infection with genotype 2 or 3 for 14 weeks: a pilot study. Hepatology2004 Dec;40(6):1260-5.

8. Mangia A, Santoro R, Minerva N, Ricci GL, Carretta V, Persico M, Vinelli F, Scotto G, Bacca D, Annese M, Romano M, Zechini F, Sogari F, Spirito F, Andriulli A. Peginterferon alfa-2b and ribavirin for 12 vs. 24 weeks in HCV genotype 2 or 3. N Engl J Med2005 Jun