with reference sequences (2b.HC-J8.D10988, 2.JP.MD2b9-2, and 2a.JP.JFH-1.AB047639 obtained from the Los Alamos HCV Database as representative sequences for genotype 2b and genotype 2a HCV) and constructed a phylogenetic tree (Fig. 1). As demonstrated in the tree, no evident clustering was apparent according to the difference of responses.

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

Next, we 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. 2, left). These differences were especially significant in E1, p7 and NS5A (Fig. 2, 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 in those regions in E1, p7, and NS5A.

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

Each amino acid position in the HCV ORF was compared to detect any differences between the SVR and non-SVR patients. In Fig. 3a, differences in amino acid resides at each position are shown as dots demonstrating −logP values. As shown in Table 2, four points were extracted: amino acid (aa) 404 in the E2 region ($p = 0.008$), aa 530 in the E2 region ($p = 0.008$), aa 2359 in the NS5A region ($p = 0.002$) and aa 2631 in the NS5B region ($p = 0.012$). Among them, the residue at aa 2359 in the NS5A region differed most frequently between the SVR and non-SVR patients. Amino acids 4 and 110 in the Core region, residues that have been reported to vary according to the virological responses in genotype 2a infection [22,23], did not differ significantly in this genotype 2b HCV study. Meanwhile, amino acids 70 and 91, which have been reported to vary according to virological response to PEG-IFN/RBV therapy in genotype 1b infection, were conserved irrespective of the outcome (Fig. 3b).

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

Fig. 4a and Table 3 shows the result of sliding window analysis. 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, six regions were associated with the final outcome (p-values less than 1/20): aa 400–408 in the E2 region ($p = 0.006$), aa 723–770 in the E2 and the N-terminus of p7 region ($p = 0.001$), aa 879–893 in the NS2 region ($p = 0.01$), aa 2045–2051 in the NS5A region ($p = 0.0002$), aa 2224–2242 in the NS5A region ($p = 0.001$) and aa 2379–2405 in the NS5A region ($p = 0.03$). Interestingly, aa 2224–2242 in the NS5A was located in the interferon sensitivity determining region (ISDR). Fig. 4b shows the aligned sequences of amino acids around 2213–2274 of HCV NS5A. Among these 6 regions, aa 723–770, aa 879–893, aa 2224–2242, and aa 2379–2405 were correlated with the final outcome in an incremental manner according to the number of substitutions in those regions (Table S2, S3, S4, S5). The number of substitutions in the ISDR was also correlated to the final outcome in an incremental step-up manner (data not shown).

Multivariate analysis to detect independent predictive factors contributing to the SVR

Next, multivariate analysis was undertaken to identify pretreatment variables correlated with the final outcome. To evaluate the optimal threshold of amino acid variations for SVR prediction in each viral region extracted, a receiver operating characteristic curve was constructed and the most optimal cut off value was determined for each region. E2 aa404–408 was excluded from the analysis because we considered that the region was unlikely to be truly associated to the outcome as it is located in the hypervariable region, the region of the highest mutation rate in the HCV genome as a result of host's immune attack. E2 aa 723–770 was excluded from the analysis because all the patients above the cut-off value in the region achieved SVR and an odds calculation was not possible. The ISDR was also excluded because NS5A aa2224–2242 was completely contained in the ISDR. In addition, variables of EVR and RVR were excluded because they were post treatment variables. The multivariate analysis revealed that only NS5A aa 2224–2242 (odds ratio 11.0, $p = 0.039$) was finally identified as the independent variable predicting the final outcome (Table 4).

Biological relevance of variation in NS5A in this study group

Because NS5A aa 2224–2242 is located within the ISDR, for which the amino acid substitution numbers have been reported to be correlated with the HCV RNA titer in genotype 1 and 2a HCV infection [13], we analyzed the relationship between amino acid variations in that region and pretreatment HCV RNA titers. Contrary to our expectation, no evident relationship was found between variations in the NS5A region aa 2224–2242 and HCV RNA titer (Fig. 5). On the other hand, as shown in Table 5, although the initial viral responses (RVR or EVR) did not show evident association with the amino acid variations in the region, treatment relapse was significantly correlated with the amino acid variations in the region. In addition to NS5A aa 2224–2242, there was no evident relationship between HCV RNA level and variations in the other regions found in this study (data not shown).

Discussion

In this study, we showed that genotype 2b HCV sequences from Japanese patients who achieved SVR were more diverse than the sequences from patients with non-SVR. The result that SVR patients were more diverse in their HCV sequences than non-SVR patients is in accordance with previous studies of genotype 1 HCV infection, although the diverse viral genes varied according to genotype [18,19]. We found that these diversities were primarily found in E1, p7 and NS5A.

In systemic searching for single amino acid positions or consecutive amino acid regions in the HCV ORF associated with the treatment outcome, several regions were extracted in E2, p7, NS2, NS5A and NS5B. Among those identified regions, E2 aa 723–770, NS2 aa 879–893, NS5A aa2224–2242, and NS5A aa2379–2405 were correlated with the final outcome in an incremental manner according to the number of amino acid substitutions. Specifically, the sequences of those regions in non-SVR patients were almost homogeneous, while the sequences of the region in SVR patients were significantly diverse and multiple amino acid substitutions were found compared to the consensus sequence. Interestingly, among those regions, aa 2224–2242 was completely included in the ISDR, in which the number of amino acid substitutions is known to show significant correlation with

the treatment response to IFN-based therapy in genotype 1b, and also in genotype 2 [21,24].

In recent studies of genotype 1b infection, amino acid variation of residues 70 and 91 in the Core were reported to be associated with the treatment response to IFN-based therapy. The correlation of amino acid variation in the Core (residues 4 and 110) with the response to PEG-IFN/RBV therapy was also identified in genotype 2a infection [22,23]. In genotype 2b infection, however, we could not find such associations between amino acid variation in the core region and the response to PEG-IFN/RBV therapy (Fig. 3b). Amino acid residues of aa 70 and 91 were conserved irrespective of differences in the PEG-IFN/RBV responses. On the other hand, although amino acid variations were also sometimes found at residues 4 and 110 in genotype 2b HCV, their frequency was low, and no evident association between the variation and the treatment response was found. Although the reason of the lack of association between the Core and the PEG-IFN/RBV treatment response in genotype-2b HCV infection is unknown, it suggests that a different mechanism affecting the treatment response might exist, depending on genotype-specific viral features.

In genotype 1 HCV, variations within the PKR-binding region of NS5A, including those within the ISDR, were reported to disrupt the NS5A-PKR interaction, possibly rendering HCV sensitive to the antiviral effects of interferon [25]. Clinically, the number of substitutions within the ISDR has been reported to correlate with the serum HCV RNA level in genotype 1 and 2a infections [13]. In addition, a recent study reported that mutations in the ISDR also show the correlation with the relapse in the PEG-IFN/RBV therapy in genotype 1b infection [26]. Because NS5A aa2224–2242, part of ISDR, was extracted as one of those regions related to the treatment response in genotype 2b infection, we undertook further analysis to investigate the correlation between amino acid variation numbers and serum HCV RNA level. Though the reason is unknown, we could not find evidence of a relationship between variation in the NS5A aa 2224–2242 and HCV RNA titer in genotype 2b infection, unlike genotypes 1 and 2a. Of note, a high SVR rate in genotype1 and genotype 2a infection is known to be closely correlated with a low HCV RNA level and multiple substitutions in ISDR. However, in genotype 2b infection in our study, there was no significant difference in the HCV RNA level between SVR and non-SVR patients, as shown in Table 1. Previously, the role of the ISDR in the contribution to SVR in genotype 1 and 2a has been discussed in detail in the context of serum HCV RNA level, and multiple substitutions in the ISDR are related to a low HCV RNA level and high SVR rate. However, it is not known which of these two factors is directly associated with viral clearance. Consideration of this three-sided relationship of ISDR, HCV RNA level and SVR rate in genotype-2b infection leads to the suggestion that amino acid variation in ISDR to be more direct contributor for SVR.

In spite of these findings, there were still limitations in our study. First, because genotype 2b infection only accounts for 10% of all HCV infection in Japan, the number of studied patients was rather small, especially non-SVR patients. In addition, because genotype 2b HCV contains as many as 3033 amino acids, it is possible that incorrect amino acids or regions were judged as significant in the complete HCV ORF comparison study as a result of type I errors. Therefore, if more patients were available for the analysis, the statistical power detecting the meaningful differences would be greater. Secondly, we could not include the IL28B SNP analysis in this study. If we could have combined the information of IL28B SNPs with the full HCV ORF information, a more comprehensive analysis would have been achieved.

In conclusion, we have shown that viral sequences were more diverse in SVR patients infected with genotype 2b HCV. Through systematic comparison between SVR and non-SVR patients, we have also shown that several localized regions were extracted as hot spots whose amino acid substitutions were closely related to the final outcome by affecting the relapse rate in the PEG-IFN/RBV therapy.

Supporting Information

Table S1 GenBank Accession Numbers. Obtained GenBank accession numbers for 60 genotype-2b HCV full open reading frame sequences are listed.
(DOC)

Table S2 Substitutions in NS5A aa 2224–2242 Amino Acid Regions and SVR rate. SVR rate increased with the number of substitutions in this region.
(DOC)

Table S3 Substitutions in NS5A aa 2379–2405 Amino Acid Regions and SVR rate. SVR rate increased with the number of substitutions in this region.
(DOC)

Table S4 Substitutions in NS2 aa 879–893 Amino Acid Regions and SVR rate. SVR rate increased with the number of substitutions in this region.
(DOC)

Table S5 Substitutions in E2 aa 723–770 Amino Acid Regions and SVR rate. SVR rate increased with the number of substitutions in this region.
(DOC)

Table S6 PCR Primer List. Primers designed to perform two-step nested PCR for this study are listed. Dominant genotype-2b HCV full open reading frame sequences was determined by the 24 partially overlapping amplicons amplified by these primers.
(XLS)

Acknowledgments

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

Author Contributions

Conceived and designed the experiments: MK SM NE. Performed the experiments: MK. Analyzed the data: MK SM NE. Contributed reagents/materials/analysis tools: RS MM HS KK. Wrote the paper: MK SM NE. Critical revision of the manuscript for important intellectual content: FA TU TI MS MN NS MW.

References

1. Ghany MG, Strader DB, Thomas DL, Seeff LB (2009) Diagnosis, management, and treatment of hepatitis C: an update. Hepatology 49: 1335–1374.
2. Di Bisceglie AM, Martin P, Kassianides C, Lisker-Melman M, Murray L, et al. (1989) Recombinant interferon alfa therapy for chronic hepatitis C. A randomized, double-blind, placebo-controlled trial. N Engl J Med 321: 1506–1510.
3. Haydon GH, Jarvis LM, Blair CS, Simmonds P, Harrison DJ, et al. (1998) Clinical significance of intrahepatic hepatitis C virus levels in patients with chronic HCV infection. Gut 42: 570–575.
4. Simmonds P (1997) Clinical relevance of hepatitis C virus genotypes. Gut 40: 291–293.

5. Hadziyannis SJ, Sette H, Jr., Morgan TR, Balan V, Diago M, et al. (2004) Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. Ann Intern Med 140: 346–355.
6. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, 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.
7. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, et al. (2009) Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 461: 399–401.
8. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, et al. (2009) Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 41: 1105–1109.
9. Rauch A, Kutalik Z, Descombes P, Cai T, di Iulio J, et al. (2010) Genetic variation in IL28B Is Associated with Chronic Hepatitis C and Treatment Failure - A Genome-Wide Association Study. Gastroenterology.
10. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, et al. (1996) Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N Engl J Med 334: 77–81.
11. El-Shamy A, Nagano-Fujii M, Sasase N, Imoto S, Kim SR, et al. (2008) Sequence variation in hepatitis C virus nonstructural protein 5A predicts clinical outcome of pegylated interferon/ribavirin combination therapy. Hepatology 48: 38–47.
12. Hamano K, Sakamoto N, Enomoto N, Izumi N, Asahina Y, et al. (2005) Mutations in the NS5B region of the hepatitis C virus genome correlate with clinical outcomes of interferon-alpha plus ribavirin combination therapy. J Gastroenterol Hepatol 20: 1401–1409.
13. Chayama K, Suzuki F, Tsubota A, Kobayashi M, Arase Y, et al. (2000) Association of amino acid sequence in the PKR-eIF2 phosphorylation homology domain and response to interferon therapy. Hepatology 32: 1138–1144.
14. Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, et al. (2005) Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. Intervirology 48: 372–380.
15. Duverlie G, Khorsi H, Castelain S, Jaillon O, Izopet J, et al. (1998) Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity. J Gen Virol 79(Pt 6): 1373–1381.
16. Hofgartner WT, Polyak SJ, Sullivan DG, Carithers RL, Jr., Gretch DR (1997) Mutations in the NS5A gene of hepatitis C virus in North American patients infected with HCV genotype 1a or 1b. J Med Virol 53: 118–126.
17. Zeuzem S, Lee JH, Roth WK (1997) Mutations in the nonstructural 5A gene of European hepatitis C virus isolates and response to interferon alfa. Hepatology 25: 740–744.
18. Donlin MJ, Cannon NA, Aurora R, Li J, Wahed AS, et al. (2010) Contribution of genome-wide HCV genetic differences to outcome of interferon-based therapy in Caucasian American and African American patients. PLoS One 5: e9032.
19. Donlin MJ, Cannon NA, Yao E, Li J, Wahed A, et al. (2007) Pretreatment sequence diversity differences in the full-length hepatitis C virus open reading frame correlate with early response to therapy. J Virol 81: 8211–8224.
20. Kobayashi M, Watanabe K, Ishigami M, Murase K, Ito H, et al. (2002) Amino acid substitutions in the nonstructural region 5A of hepatitis C virus genotypes 2a and 2b and its relation to viral load and response to interferon. Am J Gastroenterol 97: 988–998.
21. Murakami T, Enomoto N, Kurosaki M, Izumi N, Marumo F, et al. (1999) Mutations in nonstructural protein 5A gene and response to interferon in hepatitis C virus genotype 2 infection. Hepatology 30: 1045–1053.
22. Kadokura M, Mackawa S, Sueki R, Miura M, Komase K, et al. (2011) 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. Hepatol Int.In Press.
23. Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, et al. (2009) Association of Amino Acid Substitution Pattern in Core Protein of Hepatitis C Virus Genotype 2a High Viral Load and Virological Response to Interferon-Ribavirin Combination Therapy. Intervirology 52: 301–309.
24. Hayashi K, Katano Y, Honda T, Ishigami M, Itoh A, et al. (2009) Mutations in the interferon sensitivity-determining region of hepatitis C virus genotype 2a correlate with response to pegylated-interferon-alpha 2a monotherapy. J Med Virol 81: 459–466.
25. Gale M, Jr., Blakely CM, Kwieciszewski B, Tan SL, Dossett M, et al. (1998) Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. Mol Cell Biol 18: 5208–5218.
26. Kurosaki M, Tanaka Y, Nishida N, Sakamoto N, Enomoto N, et al. (2010) Pre-treatment prediction of response to pegylated-interferon plus ribavirin for chronic hepatitis C using genetic polymorphism in IL28B and viral factors. J Hepatol.

Hepatol Int (2011) 5:789–799
DOI 10.1007/s12072-011-9267-x

ORIGINAL ARTICLE

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 · Shinya Maekawa · Ryota Sueki · Mika Miura · Kazuki Komase · Hiroko Shindo · Fumitake Amemiya · Tomoyoshi Uetake · Taisuke Inoue · Minoru Sakamoto · Mina Nakagawa · Naoya Sakamoto · Mamoru Watanabe · Nobuyuki Enomoto

Received: 25 November 2010 / Accepted: 21 February 2011 / Published online: 20 March 2011
© Asian Pacific Association for the Study of the Liver 2011

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 RBV 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 number of amino acid variations within the NS5A aa 2258–2306 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 RBV 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.

Abbreviations

EVR	Early virological response
IFN	Interferon
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

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

M. Kadokura · S. Maekawa (✉) · R. Sueki · M. Miura · K. Komase · H. Shindo · F. Amemiya · T. Uetake · T. Inoue · M. Sakamoto · N. Enomoto
First Department of Internal Medicine, Faculty of Medicine, University of Yamanashi, 1110, Shimokato, Chuo, Yamanashi 409-3898, Japan
e-mail: maekawa@yamanashi.ac.jp

M. Nakagawa · N. Sakamoto · M. Watanabe
Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, 1-5-45, Yushima, Bunkyo, Tokyo 113-8510, Japan

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-interferon sensitivity determinant region (ISDR) [9], NS5A-interferon ribavirin resistance determinant region (IRRDR) [10], NS5B [11], and 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 (ORF) analysis for genotype 1 also has 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 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 and 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 the University of Yamanashi, Tokyo Medical and Dental University, and related institutions were first included in the study. They all fulfilled the 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; and (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 to exclude false positives (type I errors) which might arise in successive HCV-ORF study. Since genotype-2a HCV contains as many as 3,033 amino acids, it was possible that incorrect amino acids can 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 deoxyribonucleic acid (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 s, annealing at 55°C for 15 s, and extension at 72°C for 45 s with BD Advantage^{TM} 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 was 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 1/1,000 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, were 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 <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 RBV 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 relapsers, while two patients were null responders.

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.

Table 1 Baseline characteristics of all patients (groups 1 and 2)

Characteristic	SVR ($n = 58$)			Non-SVR ($n = 20$)			p value[△]
	Group 1 ($n = 36$)	Group 2 ($n = 22$)	Combined ($n = 58$)	Group 1 ($n = 7$)	Group 2 ($n = 13$)	Combined ($n = 20$)	
Gender (male/female)	20/16	9/13	29/29	4/3	5/8	9/11	0.80[†]
Age (years)	50.0 ± 12.5*	57.3 ± 10.0	52.4 ± 12.1	55.0 ± 9.7	59.8 ± 6.4	58.1 ± 7.8	0.058[‡]
ALT (IU/l)	86.6 ± 86.6	71.2 ± 50.4	80.5 ± 74.2	52.9 ± 29.3	88.1 ± 90.1	75.8 ± 75.5	0.81[‡]
Platelet ($\times 10^4/mm^3$)	20.8 ± 6.2	19.0 ± 5.2	20.1 ± 5.8	14.7 ± 7.1	19.1 ± 4.9	17.6 ± 6.0	0.11[‡]
Fibrosis score (0–2/≥3)[§]	34/1	19/2	53/3	4/3	11/2	15/5	0.049[†]
HCV RNA (KIU/ml)	760 (2–3,100)**	340 (54–3,600)	550 (12–3,600)	1,300 (350–30,000)	1,400 (180–5,000)	1,300 (180–30,000)	0.002[‖]
IFN dose (≥80%/60–80%)[¶]	28/4	21/1	49/5	4/3	11/2	15/5	0.12[†]
Ribavirin dose (≥80%/60–80%)[¶]	27/5	17/5	44/10	4/3	5/8	9/11	0.003[†]
RVR rate (%)	87.5	54.5	74.1	33.3	46.1	42.1	0.022[†]
EVR rate (%)	100	100	100	66.7	100	89.4	0.07[†]

* Mean ± SD; ** median (range); [†] Fisher's exact probability test; [‡] Student t test; [‖] Mann–Whitney's U test; [△] p values between all SVR ($n = 58$) versus all non-SVR ($n = 20$)

Several clinical characteristics listed above were unavailable in some patients. § SVR: $n = 56$ (35 in group 1, 21 in group2), non-SVR: $n = 17$ (7 in group 1, 10 in group 2); [¶] SVR: $n = 54$ (32 in group 1, 22 in group 2)

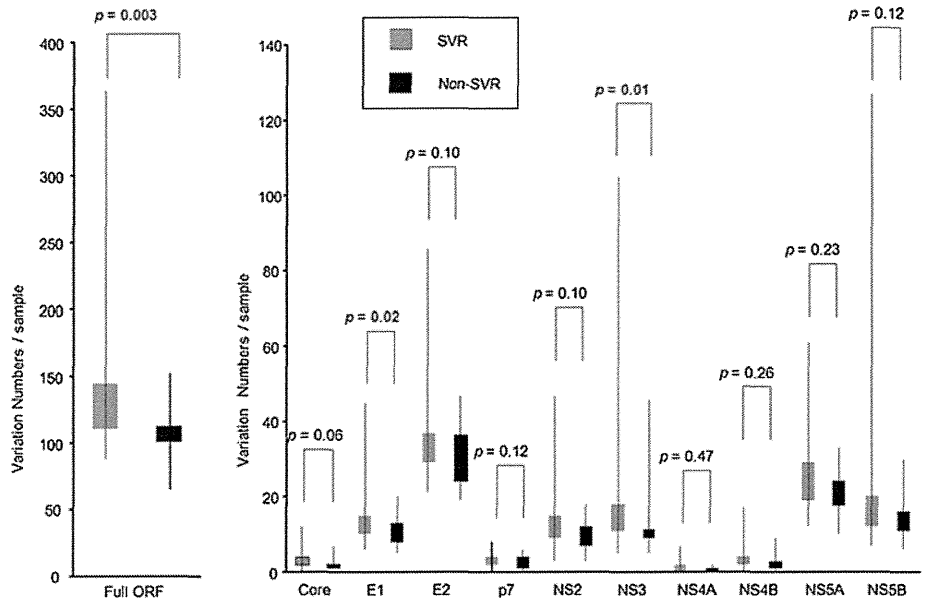


Fig. 1 Number of amino acid substitutions per sample in the sustained viral responders (SVR) and the non-sustained viral responders (non-SVR) group. The numbers of variations, relative to a population consensus, that were unique to either SVR or non-SVR patients are shown for the full ORF (*left*) and for each HCV protein (*right*)

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 –log P 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 2), 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 2). Figure 2b shows the aligned sequences of amino acids 1–120 of the core region. Substitutions at aa 110 from

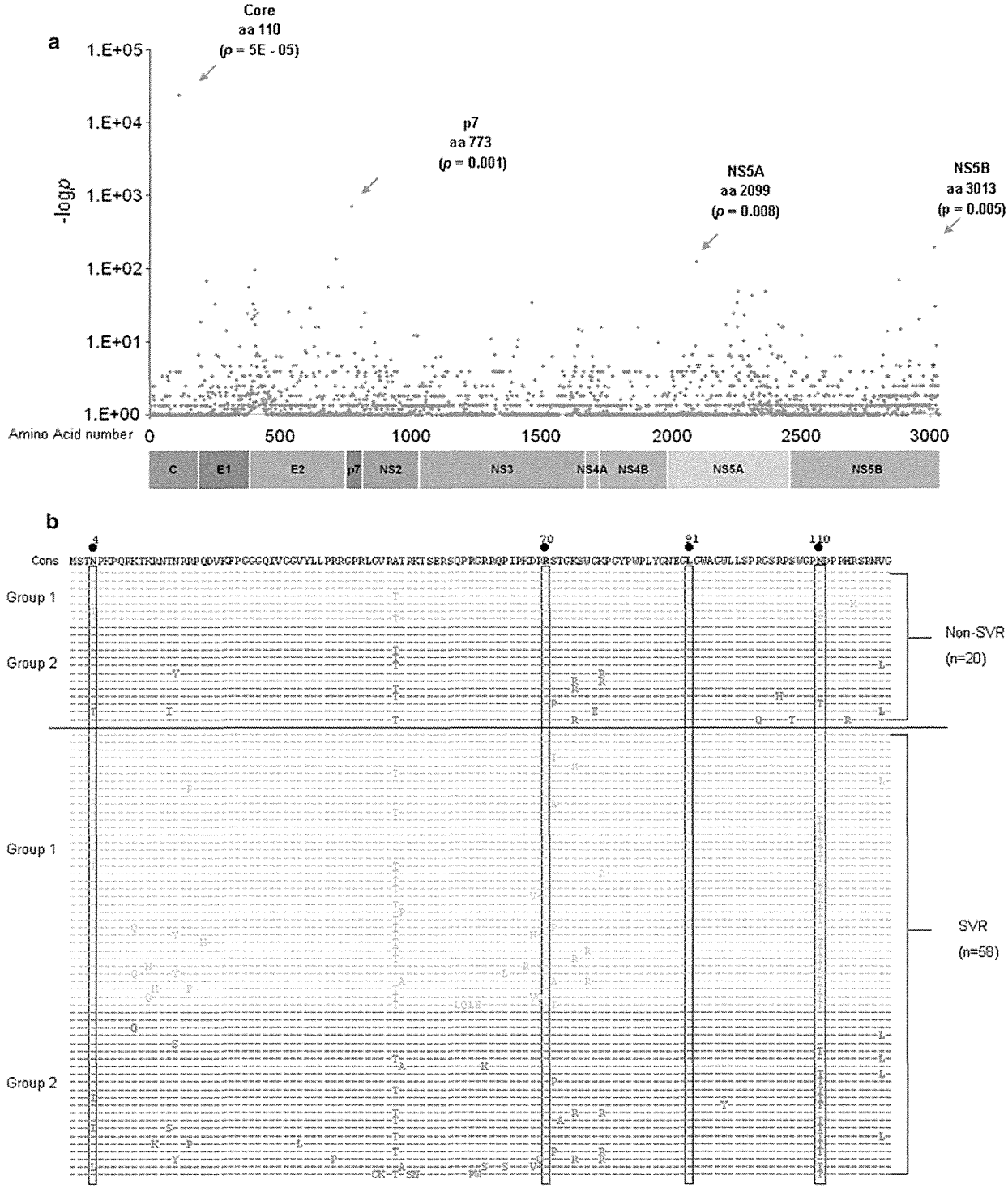


Fig. 2 **a** Different amino acid usages at each viral amino acid position between the sustained viral responders (SVR) and the non-sustained viral responders (non-SVR) patients. Amino acid variation was determined between SVR and non-SVR patients by Fisher's exact probability test. The longitudinal axis shows the –log *P* value. **b** Sequence alignment in the core region. *Dashes* indicate amino acids identical to the consensus sequence and substituted amino acids are shown by standard single letter codes. **c** Sliding window analysis. Viral regions affecting treatment outcomes are shown in *dark spots*. There are four hot spots: at core amino acid 110, amino acids 400–403 (i.e., the hypervariable region) in Envelope 2 (*E2*) region, amino acids 724–743 in E2, and amino acids 2258–2306 in the nonstructural (*NS*) 5A. **d** Sequence alignment amino acids in the nonstructural (*NS*) 5A around amino acids 2258–2306. *Dashes* indicate amino acids identical to the consensus sequence and substituted amino acids are shown by standard *single letter* codes

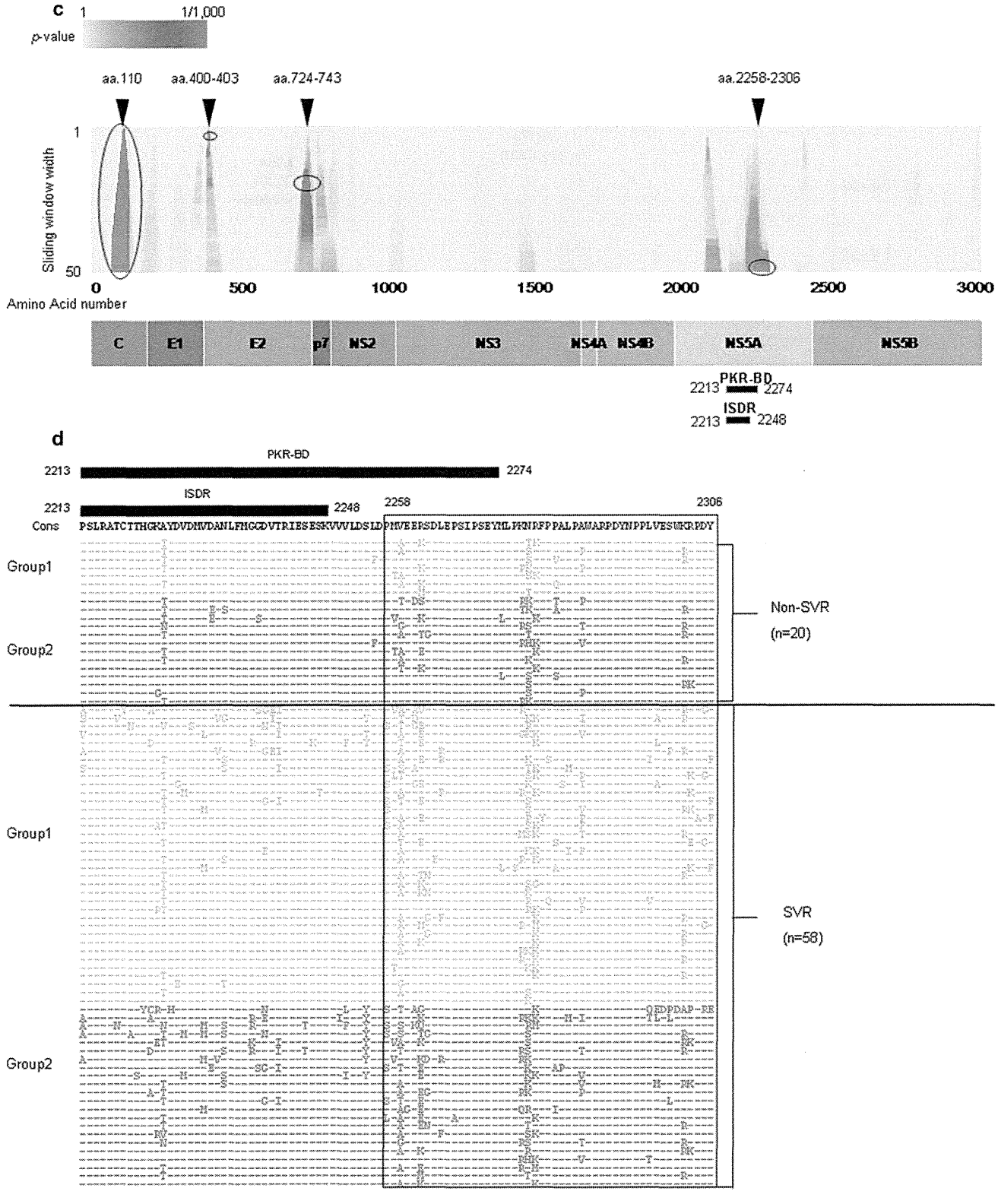


Fig. 2 continued

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

Table 2 Variations in each amino acid position and SVR rate

Position	Group 1 ($n = 43$)	p value	Group 2 ($n = 35$)	p value	Combined ($n = 78$)	p value
Core aa 110						
T	100% (19/19)	0.01	92.9% (13/14)	0.004	97% (32/33)	5E−05
Non T	70.8% (17/24)		42.9% (9/21)		57.8% (26/45)	
p7 aa 773						
V	77.4% (24/31)	0.16	53.6% (15/28)	0.03	66.1% (39/59)	0.002
Non V	100% (12/12)		100% (7/7)		100% (19/19)	
NS5A aa 2099						
R	92.9% (13/14)	0.40	91.7% (11/12)	0.01	92.3% (24/26)	0.01
Non R	79.3% (23/29)		47.8% (11/23)		65.4% (34/52)	
NS5B aa 3013						
L	78.9% (26/33)	0.17	47.8% (11/23)	0.01	66.1% (37/56)	0.008
Non L	100% (10/10)		91.7% (11/12)		95.5% (21/22)	

Table 3 Number of amino acid substitutions in each region and SVR rate

Region	Group 1 ($n = 43$)	p value	Group 2 ($n = 35$)	p value	Combined ($n = 78$)	p value
E2 aa 400–403						
Mutation ≥ 2	89.3% (25/28)	0.22	100% (11/11)	0.002	92.3% (36/39)	0.0005
Mutation 0–1	73.3% (11/15)		45.8% (11/24)		56.4% (22/39)	
E2 aa 724–743						
Mutation ≥ 1	100% (28/28)	0.0002	72% (18/25)	0.12	86.8% (46/53)	0.0006
No mutation	53.3% (8/15)		40% (4/10)		48% (12/25)	
ISDR(aa 2213–2248)						
Mutation ≥ 2	100% (15/15)	0.08	86.7% (13/15)	0.02	93.3% (28/30)	0.003
Mutation 0–1	75% (21/28)		45% (9/20)		62.5% (30/48)	
NS5A aa 2258–2306						
Mutation ≥ 5	100% (19/19)	0.01	84.2% (16/19)	0.006	92.1% (35/38)	0.0006
Mutation 0–4	70.8% (17/24)		37.5% (6/16)		57.5% (23/40)	

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

Figure 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 1/1,000). 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 3). Interestingly, the NS5A region overlapped the PKR-binding domain, which includes the IFN sensitivity determining region (ISDR). Figure 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 RBV dose $\geq$80% were finally

Table 4 Multivariate logistic regression analysis

Factor	Odds (95% CI)	*p* value
Age	1.01 (0.91–1.13)	0.85
HCV RNA	1.00 (1.00–1.00)	0.09
Fibrosis score ≥3/0–2	2.37 (0.21–26.7)	0.48
RVR achievement	3.46 (0.54–22.1)	0.19
Ribavirin dose ≥80%	16.0 (1.66–153)	0.02
Core aa 110 T	24.7 (1.72–353)	0.02
NS5A aa 2258–2306 mutations 0–4/≥5	11.5 (1.23–108)	0.03

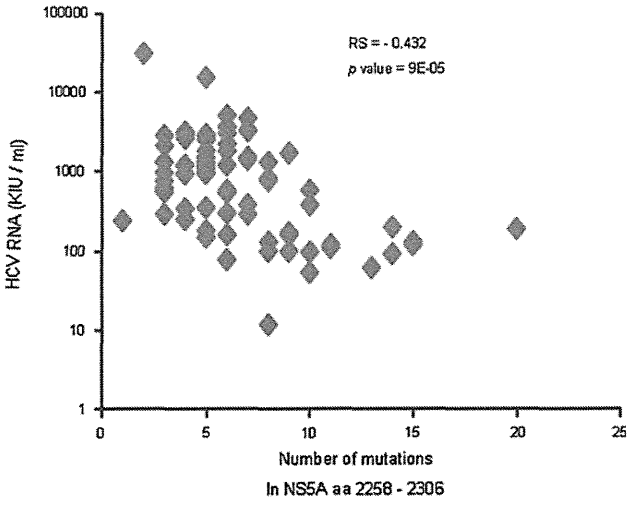


Fig. 3 Correlation between pretreatment HCV RNA levels and the number of substitutions in the NS5A region aa 2258–2306. Spearman's correlation coefficient by rank test is demonstrated

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 4).

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; Table 5). Interestingly, variation of the core aa110 was significantly associated with the patients' age ($p = 0.03$, Table 6).

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 effect 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 relapsers, 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 a 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

Table 5 Baseline characteristics of patients with NS5A aa 2258–2306 mutations 0–4 or ≥5 (groups 1 and 2)

Characteristic	Mutation 0–4 ($n = 40$)	Mutation ≥ 5 ($n = 38$)	p value
Gender (male/female)	22/18	16/22	NS[†]
Age (years)	54.3 ± 11.4*	53.5 ± 11.5	NS[‡]
ALT (IU/l)	73.8 ± 70.3	85.3 ± 78.7	NS[‡]
Platelet ($\times 10^4/mm^3$)	18.0 ± 5.9	21.0 ± 5.7	0.03[‡]
Fibrosis score (0–2/≥3)[§]	33/5	33/2	NS[†]
HCV RNA (KIU/ml)	1,100 (99–30,000)**	380 (12–5,000)	0.02[‖]
IFN dose (≥80%/60–80%)[¶]	31/8	33/2	NS[†]
Ribavirin dose (≥80%/60–80%)[¶]	25/14	28/7	NS[†]
RVR rate (%)	65.8	62.9	NS[†]
EVR rate (%)	94.7	100	NS[†]
Relapse rate (%)	35.9	7.9	0.002[†]
SVR rate (%)	57.5	92.1	0.0006[†]

* Mean ± SD; [†] Fisher's exact probability test; [‡] Student t test; [§] mutation 0–4 $n = 38$, mutation ≥5: $n = 35$; ** median (range); [‖] Mann–Whitney's U test; [¶] mutation 0–4: $n = 39$, mutation ≥5: $n = 35$

Table 6 Baseline characteristics of patients with core 110 T or N/S (groups 1 and 2)

Characteristic	Core 110 T ($n = 33$)	Core 110 N/S ($n = 45$)	p value
Gender (male/female)	18/15	20/25	NS[†]
Age (years)	50.4 ± 13.0*	56.4 ± 9.5	0.032[‡]
ALT (IU/l)	64.5 ± 48.2	88.8 ± 86.2	NS[‡]
Platelet ($\times 10^4/mm^3$)	19.3 ± 4.9	19.5 ± 6.6	NS[‡]
Fibrosis score (0–2/≥3)[§]	30/1	36/6	NS[†]
HCV RNA (KIU/ml)	580 (54–3,600)**	980 (12–30,000)	NS[‖]
IFN dose (≥80%/60–80%)[¶]	26/3	38/7	NS[†]
Ribavirin dose (≥80%/60–80%)[¶]	23/6	30/15	NS[†]
RVR rate (%)	72.4	59.1	NS[†]
EVR rate (%)	100	95.5	NS[†]
Relapse rate (%)	3.0	38.6	9E−05[†]
SVR rate (%)	97.0	57.8	5E−05[†]

* Mean ± SD; [†] Fisher's exact probability test; [‡] Student t test; [§] core 110 T: $n = 31$, core 110 N/S: $n = 42$; ** median (range), [‖] Mann–Whitney's U test; [¶] core 110 T: $n = 29$

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 clear association with age. Younger patients with core aa 110T showed favorable responses, while 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 IFN [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 RBV dose ≥80% were independent variables associated with the final outcome (Table 4). The association of RBV 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 was 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.

Acknowledgements 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). 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. Hepatology 2009;49(4):1335–1374
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 Med 1989;321(22):1506–1510
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. Gut 1998;42(4):570–575
4. Simmonds P. Clinical relevance of hepatitis C virus genotypes. Gut 1997;40(3):291–293
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 Med 2004;140(5):346–355
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. Lancet 2001;358(9286):958–965
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. Hepatology 2004;40(6):1260–1265
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 Med 2005;352(25):2609–2617
9. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, Yamamoto C, Ogura Y, Izumi N, Marumo F, Sato C. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N Engl J Med 1996;334(2):77–81
10. El-Shamy A, Nagano-Fujii M, Sasase N, Imoto S, Kim SR, Hotta H. Sequence variation in hepatitis C virus nonstructural protein 5A predicts clinical outcome of pegylated interferon/ribavirin combination therapy. Hepatology 2008;48(1):38–47
11. Hamano K, Sakamoto N, Enomoto N, Izumi N, Asahina Y, Kurosaki M, Ueda E, Tanabe Y, Maekawa S, Itakura J, Watanabe H, Kakinuma S, Watanabe M. Mutations in the NS5B region of the hepatitis C virus genome correlate with clinical outcomes of interferon-alpha plus ribavirin combination therapy. J Gastroenterol Hepatol 2005;20(9):1401–1409
12. Chayama K, Suzuki F, Tsubota A, Kobayashi M, Arase Y, Saitoh S, Suzuki Y, Murashima N, Ikeda K, Takahashi N, Kinoshita M, Kumada H. Association of amino acid sequence in the PKR-eIF2 phosphorylation homology domain and response to interferon therapy. Hepatology 2000;32(5):1138–1144
13. Akuta N, Suzuki F, Sezaki H, Suzuki Y, Hosaka T, Someya T, Kobayashi M, Saitoh S, Watahiki S, Sato J, Matsuda M, Arase Y, Ikeda K, Kumada H. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon–ribavirin combination therapy. Intervirology 2005;48(6):372–380
14. Toyoda H, Kumada T, Tada T, Arakawa T, Hayashi K, Honda T, Katano Y, Goto H. Association between HCV amino acid substitutions and outcome of peginterferon and ribavirin combination therapy in HCV genotype 1b and high viral load. J Gastroenterol Hepatol 2010;25(6):1072–1078
15. Donlin MJ, Cannon NA, Aurora R, Li J, Wahed AS, Di Bisceglie AM, Tavis JE. Contribution of genome-wide HCV genetic differences to outcome of interferon-based therapy in Caucasian American and African American patients. PLoS One 2010;5(2): e9032
16. Donlin MJ, Cannon NA, Yao E, Li J, Wahed A, Taylor MW, Belle SH, Di Bisceglie AM, Aurora R, Tavis JE. Pretreatment sequence diversity differences in the full-length hepatitis C virus open reading frame correlate with early response to therapy. J Virol 2007;81(15):8211–8224
17. Murakami T, Enomoto N, Kurosaki M, Izumi N, Marumo F, Sato C. Mutations in nonstructural protein 5A gene and response to interferon in hepatitis C virus genotype 2 infection. Hepatology 1999;30(4):1045–1053
18. Hayashi K, Katano Y, Honda T, Ishigami M, Itoh A, Hirooka Y, Nakano I, Urano F, Yoshioka K, Toyoda H, Kumada T, Goto H. Mutations in the interferon sensitivity-determining region of hepatitis C virus genotype 2a correlate with response to pegylated-interferon-alpha 2a monotherapy. J Med Virol 2009;81(3): 459–466
19. Kobayashi M, Watanabe K, Ishigami M, Murase K, Ito H, Ukai K, Yano M, Takagi K, Hattori M, Kakumu S, Yoshioka K. Amino acid substitutions in the nonstructural region 5A of hepatitis C virus genotypes 2a and 2b and its relation to viral

load and response to interferon. Am J Gastroenterol 2002;97(4): 988–998
20. Akuta N, Suzuki F, Hirakawa M, Kawamura Y, Yatsuji H, Sezaki H, Suzuki Y, Hosaka T, Kobayashi M, Saitoh S, Arase Y, Ikeda K, Kumada H. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 2a high viral load and virological response to interferon–ribavirin combination therapy. Intervirology 2009;52(6):301–309
21. Saito T, Ito T, Ishiko H, Yonaha M, Morikawa K, Miyokawa A, Mitamura K. Sequence analysis of PePHD within HCV E2 region and correlation with resistance of interferon therapy in Japanese patients infected with HCV genotypes 2a and 2b. Am J Gastroenterol 2003;98(6):1377–1383
22. Watanabe H, Nagayama K, Enomoto N, Itakura J, Tanabe Y, Sato C, Izumi N, Watanabe M. Amino acid substitutions in PKR-eIF2 phosphorylation homology domain (PePHD) of hepatitis C virus E2 protein in genotype 2a/2b and 1b in Japan and interferon efficacy. Hepatol Res 2003;26(4):268–274
23. Gale M Jr, Blakely CM, Kwieciszewski B, Tan SL, Dossett M, Tang NM, Korth MJ, Polyak SJ, Gretch DR, Katze MG. Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. Mol Cell Biol 1998;18(9):5208–5218
24. Hiramatsu N, Oze T, Yakushijin T, Inoue Y, Igura T, Mochizuki K, Imanaka K, Kaneko A, Oshita M, Hagiwara H, Mita E, Nagase T, Ito T, Inui Y, Hijioka T, Katayama K, Tamura S, Yoshihara H, Imai Y, Kato M, Yoshida Y, Tatsumi T, Ohkawa K, Kiso S, Kanto T, Kasahara A, Takehara T, Hayashi N. Ribavirin dose reduction raises relapse rate dose-dependently in genotype 1 patients with hepatitis C responding to pegylated interferon alpha-2b plus ribavirin. J Viral Hepat 2009;16(8):586–594
25. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, Heinzen EL, Qiu P, Bertelsen AH, Muir AJ, Sulkowski M, McHutchison JG, Goldstein DB. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 2009;461(7262):399–401
26. Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, Nakagawa M, Korenaga M, Hino K, Hige S, Ito Y, Mita E, Tanaka E, Mochida S, Murawaki Y, Honda M, Sakai A, Hiasa Y, Nishiguchi S, Koike A, Sakaida I, Imamura M, Ito K, Yano K, Masaki N, Sugauchi F, Izumi N, Tokunaga K, Mizokami M. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 2009;41(10):1105–1109
27. Rauch A, Kutalik Z, Descombes P, Cai T, di Iulio J, Mueller T, Bochud M, Battegay M, Bernasconi E, Borovicka J, Colombo S, Cerny A, Dufour JF, Furrer H, Gunthard HF, Heim M, Hirschel B, Malinverni R, Moradpour D, Mullhaupt B, Witteck A, Beckmann JS, Berg T, Bergmann S, Negro F, Telenti A, Bochud PY. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure—a genome-wide association study. Gastroenterology 2010 Jan 7

Bioorganic & Medicinal Chemistry 19 (2011) 6892–6905

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

A new method for induced fit docking (GENIUS) and its application to virtual screening of novel HCV NS3-4A protease inhibitors

Daisuke Takaya [a], Atsuya Yamashita [b], Kazue Kamijo [a], Junko Gomi [a], Masahiko Ito [b], Shinya Maekawa [c], Nobuyuki Enomoto [c], Naoya Sakamoto [d,e], Yoshiaki Watanabe [f], Ryoichi Arai [f], Hideaki Umeyama [f], Teruki Honma [a], Takehisa Matsumoto [a], Shigeyuki Yokoyama [a,g,*]

[a] *RIKEN Systems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan*
[b] *Department of Microbiology, Division of Medicine, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi 409-3898, Japan*
[c] *First Department of Internal Medicine, Faculty of Medicine, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi 409-3898, Japan*
[d] *Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8519, Japan*
[e] *Department for Hepatitis Control, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8519, Japan*
[f] *School of Pharmacy, Kitasato University 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan*
[g] *Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

ARTICLE INFO

Article history:
Received 12 July 2011
Revised 12 September 2011
Accepted 12 September 2011
Available online 16 September 2011

Keywords:
HCV
NS 3-4A protease
Structure based drug design
Ligand docking
Virtual screening

ABSTRACT

Hepatitis C virus (HCV) is an etiologic agent of chronic liver disease, and approximately 170 million people worldwide are infected with the virus. HCV NS3-4A serine protease is essential for the replication of this virus, and thus has been investigated as an attractive target for anti-HCV drugs. In this study, we developed our new induced-fit docking program (GENIUS), and applied it to the discovery of a new class of NS3-4A protease inhibitors (IC_{50} = 1–10 μM including high selectivity index). The new inhibitors thus identified were modified, based on the docking models, and revealed preliminary structure–activity relationships. Moreover, the GENIUS in silico screening performance was validated by using an enrichment factor. We believe our designed scaffold could contribute to the improvement of HCV chemotherapy.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Hepatitis C virus (HCV) is an etiologic agent of chronic liver disease,[1,2] and approximately 170 million people worldwide are infected with the virus.[3] Chronic hepatitis C can lead to severe liver diseases, including fibrosis, cirrhosis, and hepatocellular carcinoma.[4] The current standard therapy for chronic hepatitis C consists of pegylated interferon in combination with ribavirin.[5] Unfortunately, this therapy results in sustained antiviral activity in only about 50–60% of the patients, and is associated with serious side effects. Thus, the development of alternative and more effective anti-HCV agents has been eagerly anticipated.

HCV NS3-4A serine protease is essential for the replication of this virus, and has been investigated as an attractive target for anti-HCV drugs. Several three-dimensional structures of HCV NS3-4A protease have been deposited in the Protein Data Bank (PDB).[6] Therefore, Structure Based Drug Design (SBDD) is a promising approach for the discovery of new NS3-4A protease inhibitors. The NS3-4A protease has the catalytic triad with the anion hole, commonly found among serine protease family members. The NS3-4A protease consists of two domains: a protease domain of 180 residues and a helicase domain of 420 residues.[7] The protease domain contains the protease activity, and thus it is appropriate to use only this domain as the receptor coordinates for SBDD.[8] On the other hand, docking calculations to a complex with a helicase domain have also been performed.[9] Different receptor structures were used in the docking calculations, because no experimentally determined full-length NS3-4A protease structures complexed with small molecule inhibitors were available, as of 2011.

In recent years, many peptide or peptide-mimic inhibitors that inhibit HCV NS3-4A protease have been developed, including SCH-503034,[10] VX-950,[11] BILN-2061,[12] TMC-435,[13] ITMN-191[14] and MK-7009,[15] as specifically targeted anti-viral agents for HCV (STAT-C).[16] These compounds, which competitively inhibit the protease activity, were roughly classified into two types: the mimic type inhibitors (SCH-503034, VX-950), which have a peptide bond, and the macrocyclic compounds (BILN-2061, TMC-435350, ITMN-191, MK-7009), which have a macrocyclic ring. Recently,

* Corresponding author.
E-mail address: yokoyama@biochem.s.u-tokyo.ac.jp (S. Yokoyama).

0968-0896/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2011.09.023

ACH-806[17] was reported as an HCV NS3-4A non-peptide inhibitor, and it works in harmony with the NS3-4A protease inhibitor or the NS5B polymerase. Clinical trials (Phase III) of SCH503034 and VX-950 have been performed.[18] However, cardiotoxicity in monkeys was reported for BILN-2061, one of the macrocyclic compounds, and thus its clinical development has been interrupted.[19] Moreover, macrocyclic compounds also have a problematic ADME profile, mainly due to their large molecular weights, and the synthetic optimization of the inhibitors is difficult. In addition, various mutations, especially A156T in the active site,[20] confer resistance to these protease inhibitors, such as SCH503034, BILN-2061 and VX-950[21] Since drug-resistant viruses have readily appeared in monotherapy, a multiple drug regimen has been widely applied for anti-HCV therapy. Therefore, good ADME properties are important for the next generation of HCV NS3-4A inhibitors.[18] Generally, since peptide inhibitors lack chemical stability in relation to racemization, peptide compounds are not being pursued in the development of more effective anti-HCV drugs. Thus, a new class of non-peptide inhibitors is still expected, and an inhibitor of this protease, designed by SBDD, would be valuable for anti-HCV chemotherapy. For example, in recent years, Ismail and Hattori designed a new inhibitor with an indole skelton by a molecular modeling approach,[22] based on the structure complexed with an inhibitor bearing an indole skeleton (PDB code: 1W3C) reported by Ontoria et al.[23]

From a three-dimensional point of view, many HCV NS3-4A protease structures have been reported. In the PDB, 53 BLAST hits (*E*-value <10.0) on a query sequence obtained from the NS3-4A protease (PDB code 1DXW.A[24]) were found, as of January 2011. Almost all of the structures were determined by X-ray analyses. For example, Cummings et al. determined the complex structure of TMC-435 with the protease, and reported that the protease inhibitor interacts with the protease domain by forming non-covalent bonds (PDB code 3KEE).[25] Moreover, Hangel et al. reported the structure complexed with an inhibitor that interacts with the non-catalytic cysteine of the protease.[26] However, the structures of some HCV NS3-4A proteases have also been determined by NMR analyses. Among the BLAST hits, 3 structures determined by NMR were found. Barbato et al. reported two structures (PDB codes 1BT7,[27] 1DXW), and recently, Gallo et al. reported that of the NS3 protease, in the absence of the NS-4A co-factor, complexed with a non-covalent inhibitor (PDB code 2K1Q[28]).

Many programs are available to predict the binding modes of small molecules. Docking programs, such as AUTODOCK,[29] DOCK[30] and GOLD,[31] dock a ligand by changing their conformations to a fixed coordinate receptor and evaluating the fit by various experiential energy functions (i.e., Flexible Ligand Docking). These docking programs are useful for relatively non-flexible proteins; however, the conformations of many proteins are changed by different ligand molecules (induced fit). In such cases, conventional flexible ligand docking is not suitable for the prediction of the binding mode. To solve this induced fit problem, there are many docking programs and protocols in which the dock changes not only the conformations of the ligand but also the coordinates of the receptor, to consider the flexibility of the receptor (Flexible *Receptor* Docking or Induced-Fit Docking).

The induced-fit ligand docking methods are mainly classified into two groups,[32] soft-docking and ensemble docking. In soft-docking, the flexibility of a receptor is considered by changing the repulsion term of the protein ligand interaction in scoring functions, such as the Lennard-Jones potential term. In ensemble docking, one ligand is docked to multiple receptor conformation groups. For example, the soft-docking program Glide[33] enables scaling of the VDW radii, to relax the repulsion of the protein–ligand atoms. As an ensemble docking method, RosettaLigand considers the induced fit of the side chain, using a Backbone-dependent Rotamer Library.[34,35] Moreover, to release the volume occupied by the side chain, Glide performs an alanine substitution of the side chain in contact with the docking ligand. The open space is used for the binding pocket in the first docking, for predicting tentative binding modes. After the ligand is docked, the removed side chain is reconstructed by homology modeling using Prime, and the ligands are re-docked into the constructed protein models. These induced fit docking programs make it possible to predict interactions in difficult predictions, by only using the coordinates of one fixed receptor structure.

In this study, we developed our new induced-fit docking program (GENIUS), and applied it to the discovery of a new class of HCV NS3-4A protease inhibitors. In our program, the induced fit of protein side chains was considered by incorporating the dynamic information in solution. Among the available experimental coordinates of the NS3-4A protease, the NMR structure (PDB code 1DXW) was chosen as the receptor ensemble for docking. The collision tolerance was set for each atom of the receptor, based on the degree of preservation of the side chain torsion angle in the ensemble. Moreover, Essential Interaction Pairs (EIPs) were newly defined to interact with not only the active site but also the hydrophobic atoms on the planar beta sheet of the protease, as a constraint for ligands. The GENIUS docking system enables induced fit docking (Fig. 1), and combines ensemble docking to use the conformation cluster of the receptors, and soft-docking to set the coefficient for every atom of the receptor and to relax the collisions between protein and ligand atoms. The GENIUS docking system using EIPs was employed for the in silico screening of the NS3-4A protease inhibitors, and the selected compounds were evaluated by HCV NS3-4A enzymatic and cell-based assays. The new inhibitors thus discovered were modified based on the docking models, and revealed their preliminary structure–activity relationships.

2. Result and discussion

This study was performed by combining in silico and in vitro screening techniques. For the in silico screening part, we developed

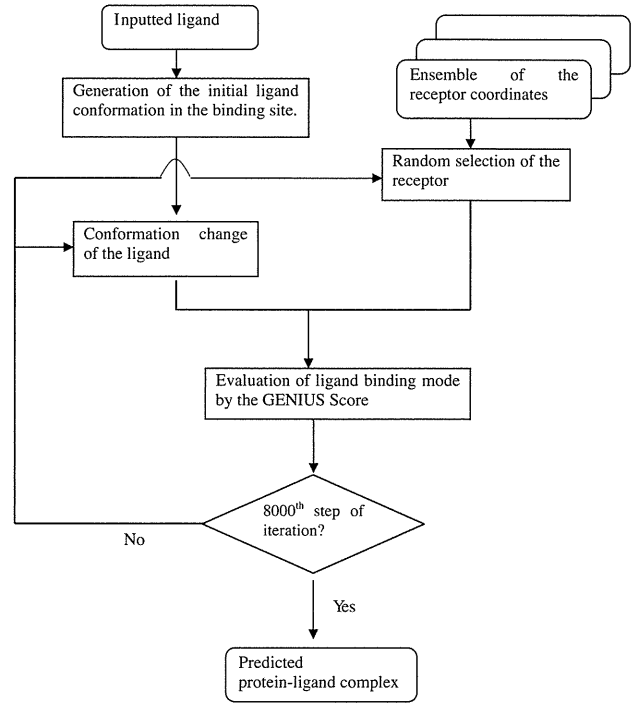


Figure 1. Flowchart of the GENIUS docking system.

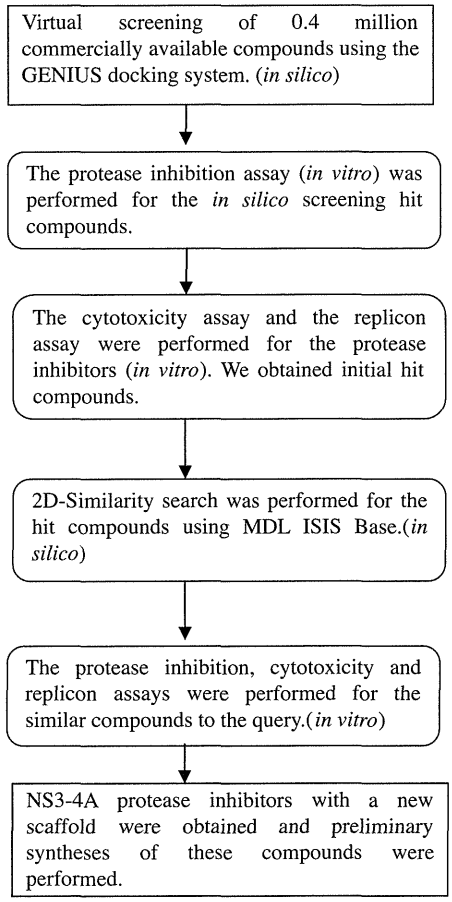
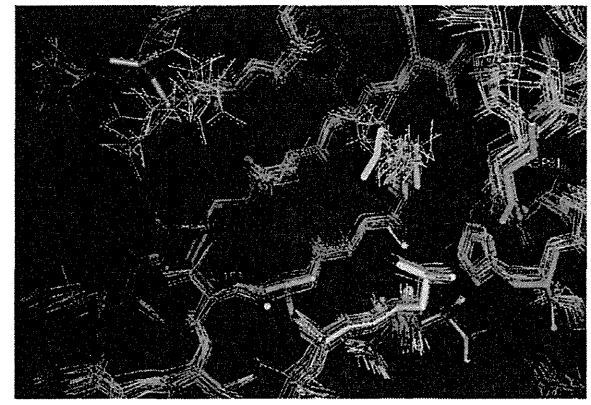


Figure 2. Flowchart of HCV NS3-4A in silico and in vitro.



(a): 20 NMR structures (PDB code: 1DXW)

```
# anion hole, anchor zone
KEYATM O.3  100 2.58 NE2 HISA_57
KEYATM O.co2 100 2.60 N  GLYA_137

# hydrogen-bond interactions
KEYATM DONOR 100 3.40  O  ARGA_155
KEYATM ACPTR 100 2.60  N  ALAA_157
KEYATM DONOR 100 2.60  O  ALAA_157

# hydrophobic interactions on the beta-sheet
KEYATM C.3  100 2.60 CB ALAA_166
KEYATM C.3  100 3.80 CB VALA_158
```

(b) The indicated interaction points on the NS3-4A protease.

Figure 3. (a) The line representations are the ensemble of 20 NMR structures of the NS3-4A protease domain. All hydrogen atoms were removed. The fraction residue on the beta sheet, Arg 123, is shown as an orange stick representation. The catalytic triad residues, His57, Asp81, and Ser139, are shown as a pink stick representation. The hydrophobic residues on the beta sheet, Val158 and Ala166, are shown as a red stick representation. The residue involved in the anion hole formation, Gly 137, is shown as a blue stick representation. The residues involved in the hydrogen-bond interaction, Arg155 and Ala157, are colored green on the beta sheet. Val158 and Ala166 are shown as red stick representations. The inhibitor, 3-amino-5,5-difluoro- 2-keto-pentan-1-oic acid, which forms a covalent bond with Ser139, is shown as a yellow stick representation. (b) The line that begins with"KEYATM" means one of the EIPs. The second column string, such as O.3, O.co2, means the designated atom type that the docking ligand must have in the docking calculation. The third column means the constraint value for the EIP term in the GENIUS scoring function. The fourth column means the equivalent distance of pairwise atoms between receptor and ligand. The 5-th and 6-th columns mean the atom type and the amino acid involved in the protein–ligand interaction on the receptor, respectively.

a new method for induced fit docking, called the GENIUS docking system (Fig. 1 and see details in the Section 4), and the system was utilized for HCV NS3-4A protease in silico screening, based on NMR structural ensembles. Subsequently, the EIP for HCV NS3-4A protease inhibitors was set. For the in vitro assays, the protease inhibition activity and efficacy in HCV infected cells (i.e., the replicon) were assessed for the compounds selected in the in silico screen. Finally, preliminary syntheses to analyze the structure–activity relationships for the effective compounds in the in-vitro assays were performed (Fig. 2, see details in the Section 4).

2.1. Setting of ligand binding site and EIP for in silico screening

Since this research commenced before the structure complexed with a non-covalent inhibitor was reported (PDB code 2K1Q), 1DXW.A was used for the docking receptor. In the GENIUS docking system, the definition of a binding site was required, as in other docking algorithms. The binding site for docking was defined at 16.0 Å around every atom of the ligand (3-amino-5 and 5-di-fluoro-2-keto-pentan-1-oic acid) contained in the coordinates. The ensemble of receptor coordinates was clustered, in order to analyze the induced fit of the receptor. The atoms conserved in the average torsion angle range from −18 to 18 degrees in 99% of the population were collected from the binding site, and were ignored in the calculation of the collision term (Table S1 Supplementary data).

Next, the EIP used in this study was set up for the docking conditions of GENIUS. The receptor conformation group revealed that the active site residues displayed minimal fluctuations between the NMR structures. On the other hand, for Arg123 on the β sheet, the fluctuation between each coordinate was large (Fig. 3a). The hydrophobic residues (Val158, Ala166) on the β sheet are exposed, as a result of the motion of Arg123. The EIP was then prepared, by reference to the interactions generated as a result of the dynamics (Fig. 3b).

The final EIP is described below. Since the HCV NS3-4A protease is a serine protease, it contains the catalytic triad and the oxyanion hole that cleave the peptide bond of the substrate, as in other serine proteases. In order to obtain the interaction with the oxyanion hole, Gly137 was assigned to the EIP setting. Furthermore, we set an EIP with a hydrophobic interaction between the atoms on the β sheet (Val158, Ala166) and the atoms of the ligand (Fig. 3b). This EIP was used for the docking conditions.

2.2. Docking by GENIUS docking system

In silico screening by GENIUS, using the obtained EIP, was performed for 166,206 compounds. Based on their GENIUS docking scores, 42,504 compounds were ranked, because compounds lacking the atoms specified in the EIP could not be docked. The ranked compounds were also verified by visual inspection from top to bottom, because the EIP term is not always valid for all docking compounds. Finally, 97 compounds were selected, based on their high scores in the docking calculation, as meeting the criteria specified in the obtained EIP and the visual inspection.

2.3. In vitro evaluation of the selected compounds

Among the 97 compounds, 27 compounds showed more than 50% protease inhibition activity at 100 μM. In addition, compounds CP3-0032 (**1**) and CP3-0084 (**2**) (Fig. 4) exhibited HCV growth inhibitory activity at 13 and 23 μM in the replicon assay, respectively, and lacked toxicity (CC_{50}(MTS) >125 μM). (Table 1) Compounds **1** and **2** have a common skeleton, featuring an acyl diazene (–N=N–) and a biarylester (Fig. 4). To clarify the structure–activity relationship of this chemical series, similar compounds were selected from commercially available compounds. In total, 140 compounds were selected as derivatives with the common substructure and a similar skeleton by a 2D-similarity search, and the protease inhibition assay was performed. Among the similar compounds, eight compounds (**3**–**10**) exhibited protease inhibition activities ranging from 1.01 to 64.3 μM of the IC_{50} values. The IC_{50}, EC_{50}, CC_{50} and selectivity index values for these compounds are summarized in Table 1. Among these compounds, CP3-3284-125 (**3**) and CP3-3284-126 (**4**) exhibited strong inhibition of the protease activity at IC_{50} = 1.06 and 1.01 μM, respectively. Moreover, in the replicon assay, their EC_{50} values were 19.5 and 12.5 μM, respectively (Table 1). However, these compounds showed relatively strong toxicity in the ATP assays. In contrast, CP3-3284-53 (**10**) exhibited moderate protease inhibition activity (IC_{50} = 8.59 μM), as compared with compounds **3** and **4**; however, in the cell-based assays, the EC_{50} was 12.0 μM with a high selectivity index (>9.3).

2.4. Synthesis of compounds 10, 11, 12 and 13

Since the purity of compound **10** was unknown (we assumed 100% purity in the in vitro assay), compound **10** was synthesized (Scheme 1 and see details in the Section 4). In addition, compounds **11**, **12** and **13** were synthesized, and a preliminary synthetic modification was performed for compound **10**, based on the predicted binding mode. First, to enhance the hydrophobic interaction between these compounds and the receptor, a methyl (compound **11**) or ethyl (compound **12**) group was introduced to the central benzene ring. Moreover, this compound contained multiple nitro groups (Fig. 4). Next, the effect of introducing a nitro group to compound **10** was examined (compound **13**). However, the inhibition activity was not significantly different (Table 1). Generally, since a nitro group is disadvantageous from the viewpoint of solubility, this functional group is removed or converted to an amino group, which can form a hydrogen bond to the receptor atoms.

In summary, compound **1**, compound **2** and the CP3-3284 series (**3**–**10**) obtained in this research represent a new, unique class of non-peptide HCV NS3-4A inhibitors, because no similar HCV

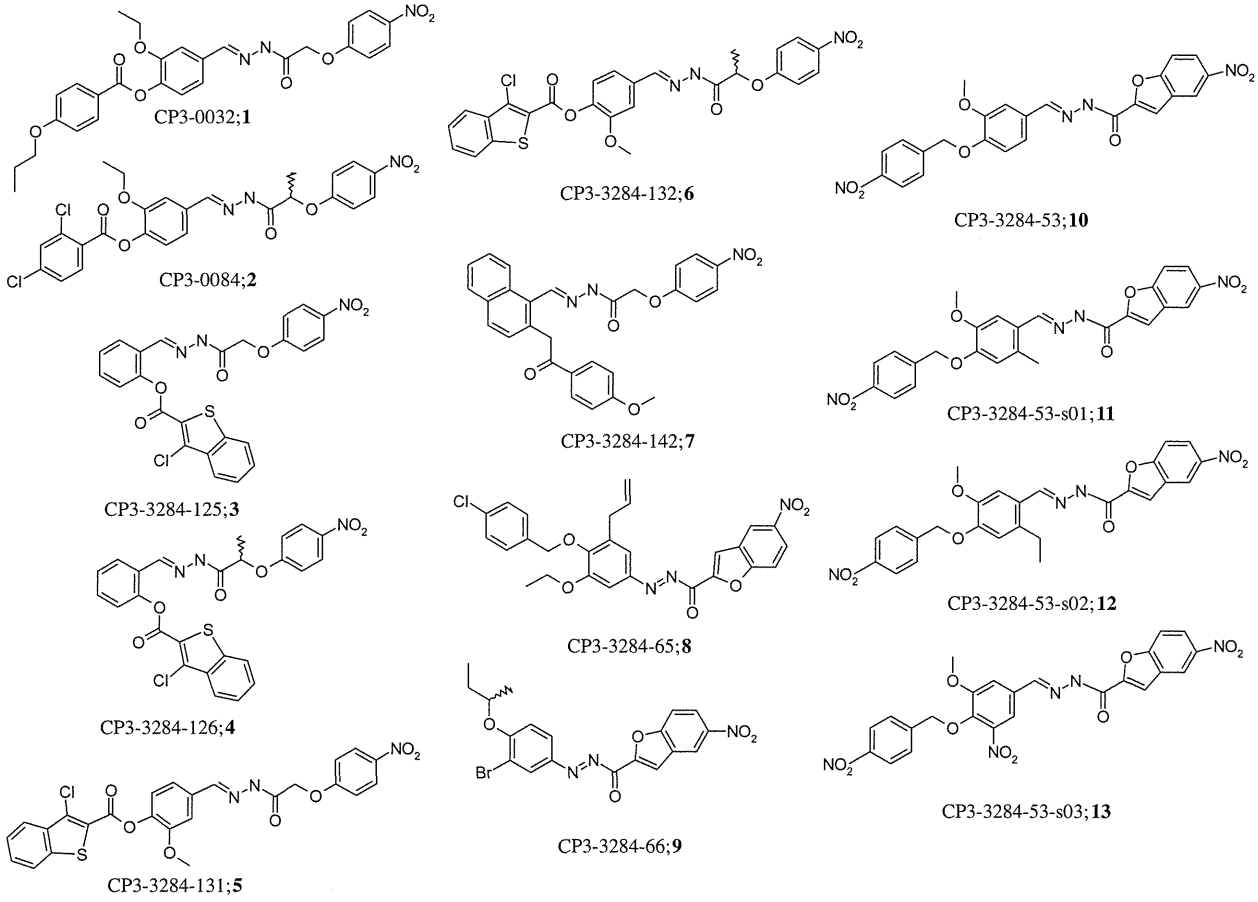
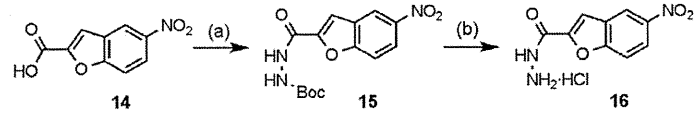


Figure 4. 2D Structures of the discovered protease inhibitors.

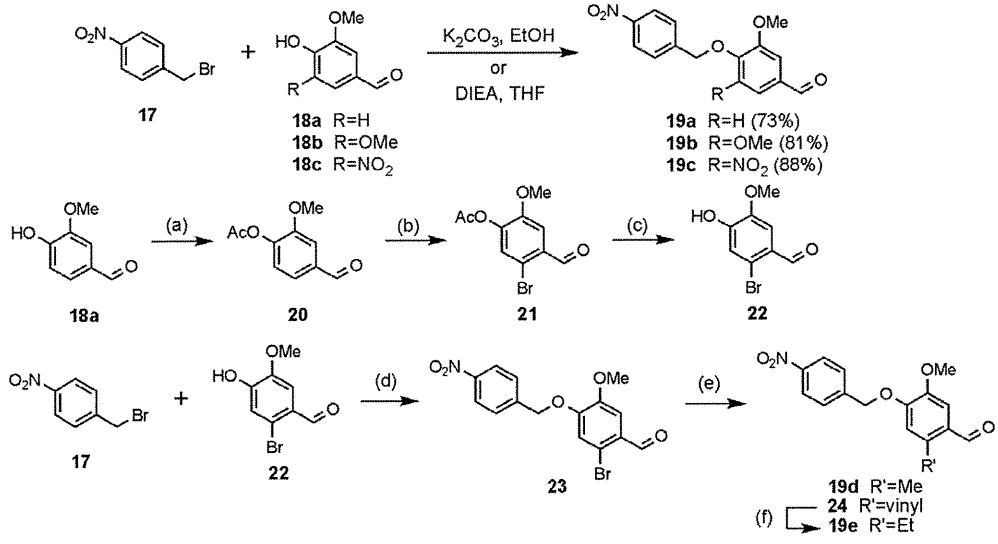
Table 1
In-vitro assay data of the discovered protease inhibitors

ID; serial number	Inhibition at 100 μM(%)	IC_{50}	EC_{50}	CC_{50} (MTS)	CC_{50} (ATP)	SI[a]	*A*Log*P*[b]
CP3-0032;**1**	38		13	>125		>9.6	5.63
CP3-0084;**2**	42.9		23	>125		>5.4	6.58
CP3-3284-125;**3**		1.06	19.5	>125	40	2.1	6.25
CP3-3284-126;**4**		1.01	12.5	>125	19	1.5	6.74
CP3-3284-131;**5**		12.3	93	>125			6.24
CP3-3284-132;**6**		4.08	121	>125			6.72
CP3-3284-142;**7**		64.3	8.5	>125	9	1.1	5.34
CP3-3284-65;**8**		8.07	>125	>125			8.72
CP3-3284-66;**9**		22.7	13.5	57	36	2.6	7.13
CP3-3284-53;**10**		8.59	12	>125	>80	>9.3	5.80
CP3-3284-53-s01;**11**		17.1					6.29
CP3-3284-53-s02;**12**		11.9					6.74
CP3-3284-53-s03;**13**		8.34					6.29

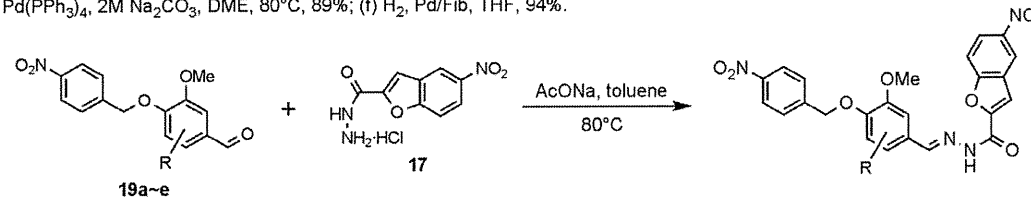
[a] The selectivity index (SI) is the ratio of the smaller CC50 value (either CC50(MTS) or CC50(ATP)) to the EC50 value.
[b] ALogP was calculated by PipeLinePilot 8.0.1(Accelrys Software Inc.).



Reagents: (a) (1) $SOCl_2$, DMF; (2) BocNH-NH_2, TEA, THF, 85%; (b) 4M HCl, dioxane, 96%.



Reagents: (a) Ac_2O, pyridine, 99%; (b) Br_2, KBr, 93%; (c) 6M HCl, 90°C, 95%; (d) K_2CO_3, EtOH, 80°C, 69%; (e) for **19d**; $MeB(OH)_2$, $Pd(PPh_3)_4$, 2M Na_2CO_3, DME, 80°C, 34%; for 24; trivinylboroxine pyridine complex, $Pd(PPh_3)_4$, 2M Na_2CO_3, DME, 80°C, 89%; (f) H_2, Pd/Fib, THF, 94%.



Scheme 1. Synthetic routes of Compounds **10–15**.

NS3-4A inhibitors (Tanimoto coefficient >= 0.7) are currently registered in SciFinder.[36]

2.5. Features of the hit compounds

The CP3-3284 series compounds have a skeleton with a diazene in common. In addition to the diazene, compounds **3**, **4**, **5** and **6** have a benzothiophene ring, and their predicted binding modes with the NS3-4A protease were almost the same, involving a hydrophobic interaction between the skeleton and various residues, such as Val158 or Ala166. (Fig. 5a,b, those of compounds **3** & **6** are in Figs. S1 and S2).

The predicted binding mode of compound **10** involved interactions with Val158 and Ala166, which are close to the side chain of Arg123 (Fig. 5a). One of the reasons why the predicted binding mode was not stable is that the side chain of Arg123 is also not

stabilized, since it is influenced by the multiple side chain conformations in the receptor ensembles (Fig. 5b). Moreover, most of the side chain atoms of Arg123 were ignored in the collision term of the GENIUS docking system (Table S1). Therefore, an undesirable angle in the hydrogen bond between the N atom of Arg123 and the O atom of NO_2 (Fig. 5a) would be observed in the flexible region of the receptor. The diazene moiety of the identified inhibitors formed a hydrogen bond with the oxygen atom of the main chain of Ala157, and the carbonyl group of the inhibitors also formed a hydrogen bond with the nitrogen atom of the main chain of Ala157 (Fig. 5a).

The IC_{50} values of compounds **3** and **4** were 4- to 12-fold lower than those of compounds **5** and **6**. In compounds **3** and **4**, the diazene and benzothiophene are ortho-substituted on the central benzene ring, while they are para-substituted in compounds **5** and **6**. The ortho-substituted benzothiophene moiety is predicted to interact more tightly with a hydrophobic surface.

In the replicon assay, the EC_{50} values of the four compounds (**3**–**6**) were approximately 10-fold larger than their IC_{50} values. In terms of hydrophobicity, very high calculated log*P* values (6.24–6.74) were observed for these compounds. Generally, hydrophobic compounds demonstrate good cell permeability. However, strong hydrophobicity also causes non-specific binding to the cell membrane. Therefore, these compounds would be less potent in the cell-based assay, as compared to the enzyme assay. In terms of cell toxicity in the ATP assay, compounds **3**, **4**, **7**, and **9** were more toxic than compound **10**. To clarify the preliminary structure–activity relationship, the R1 or R2 part (Fig. 6) of compound **10** was modified, by introducing methyl, ethyl, and nitro groups (Fig. 4). The inhibition activity of the derivatives was not significantly changed. In compound **7**, which has a naphthalene ring instead of the central benzene ring, the inhibition activity was decreased to 64.3 μM, because the atomic collision increased due to the larger volume of the sub-structure, extended by changing the substituent from benzene to naphthalene. The nitrobenzene group was commonly found at the T1 position of the active compounds (Fig. 6). The nitrobenzene group is an electron-poor aromatic ring, and is suitable to tightly bind to the electron-rich aromatic ring of His57. The nitro group also formed a weak hydrogen bond with Arg123 (Fig. 5a). In a future study, we will generate new compounds by introducing other electron-deficient substituents to interact with His57 and more powerful H-bonding acceptors to interact with Arg123, based on these structure–activity relationships.

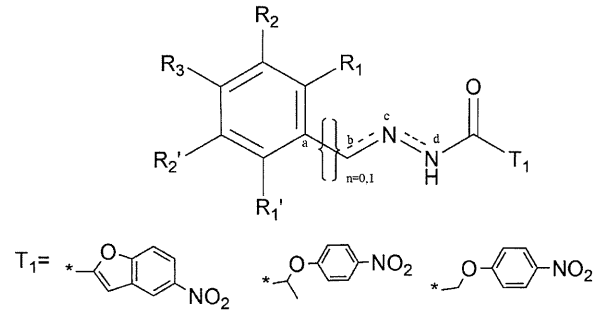
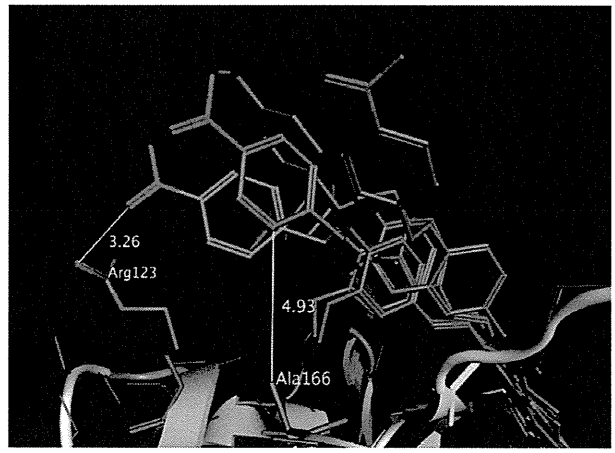


Figure 6. The common scaffold among CP3-3284 series. T1 means substructures in the CP3-3284 series.

2.6. Consideration of the predicted binding modes of the hit compounds

Since the CP3-3284 series compounds inhibited the protease activity and the cell viability, these compounds were considered to be promising as competitive inhibitors of the HCV NS3-4A protease. In a recent study, the interactions around the catalytic triad have been regarded as being important in NS3-4A protease inhibitor design.[10,11] Since the NS3-4A protease involves four connections of the HCV protein precursors, such as NS3-NS4A, NS4A-NS4B, NS4B-NS5A and NS5A-NS5B,[37] it is likely to identify peptide-type inhibitors. Generally, docking software emphasizes hydrophilic interactions, such as H-bonds, as compared with hydrophobic interactions, such as the interaction on the planar β sheet. To evaluate that kind of interaction and to identify the compounds that interact with the planar β sheet more accurately, it is necessary to determine the residues that interact with the ligand.[38] To overcome the problems with the conventional docking software, we set the hydrophobic interactions with the planar β sheet (Val158 and Ala166). Since the potent compounds **3** and **6** (IC_{50} values 1.06 and 4.08 μM, respectively) were discovered to form hydrophobic interactions between the 3-chlorobenzothiophene ring and the β sheet (the predicted binding modes are included in the Supplementary data), our pharmacophore constraints (that is, the EIPs) were effective to detect a new class of non-peptide inhibitors that interact with the planar β sheet.

(a) One of the predicted binding modes of CP3-3284-53



(b) All of the predicted binding modes of CP3-3284-53

Figure 5. Predicted binding modes of CP3-3284-53(10); Ribbon representation: one of the conformations of the NS3-4A protease. Thick stickrepresentation: predicted binding mode(s) of CP3-3284-53. Purple: the catalytic triad, red: hydrophobic residue on the β sheet.