mutations (including resistance unproven mutations), and clinical characteristics including HCV RNA levels and responses to PEG-IFN/RBV therapy were compared. To assess the influence of PEG-IFN/RBV therapy on NS3 mutational status, posttreatment HCV-NS3 sequences in 39 of 58 non-SVR patients were also examined.

### Statistical analysis

Statistical differences in the data, including all available patients' demographic, biochemic, hematologic, and virologic data such as sequence variation factors, were determined among the various groups by Student's $t$ test or Mann–Whitney $U$ test for numerical variables and Fisher's exact probability test for categorical variables.

## Results

### Prevalence of dominant PI-resistance-associated nonstructural 3 mutations in untreated patients

Figure 1 shows the frequency of substitutions in 261 patients for each of 181 NS3 protease amino acid residues compared to the consensus sequence. A total of 41 resistance proven mutations were detected in 35 (13.4%) patients: T54S (14 patients, 5.4%), Q80K (1 patient, 0.4%), I153V (22 patients, 8.4%), D168E (4 patients, 1.5%), T54S plus I153V double mutation (4 patients, 1.5%), and I153V plus D168E double mutation (2 patients, 0.8%). The mutation number increased to 54 in 47 (18.0%) patients when resistance unproven mutations were included: V36I (2 patients, 0.8%), I153L (11 patients, 4.2%), and I153V plus V36I double mutation (2 patients, 1.5%). Double mutations were found in 7 patients (2.7%) (Table 1). Q80L was observed in 47 (18%) patients but these were excluded from consideration because a previous study demonstrated that this mutation does not confer resistance [15]. All mutations observed in this study would confer low- to moderate-level PI resistance according to previous studies [6, 15–19]. No mutations conferring high-level resistance such as R155 or A156 [11, 17, 19–22] were observed.

### Clinical characteristics of patients with PI-resistance mutations

Table 2 presents the characteristics of patients classified according to the presence of PI-resistance mutations

**Table 1** Prevalence of PI-resistance-associated NS3 mutations

| Drug-resistance mutations described in the literature | | | | Detected resistance mutations |
|---|---|---|---|---|
| NS3 residue | Resistance mutations | Drugs | References | Genotype 1b ($N = 261$), (%) |
| V36 | A, M, L, G, C | Telaprevir, Boceprevir | [1, 3, 4, 10, 11, 19, 31, 37] | I × 2 (0.8) |
| Q41 | R | ITMN-191, Boceprevir | [19] | |
| F43 | S, C | ITMN-191, Boceprevir, Telaprevir, TMC435 | [15, 19] | |
| T54 | A, S | Telaprevir, Boceprevir, SCH900518 | [1, 3, 10, 11, 19, 20, 31, 38] | **S × 14 (5.4)** |
| V55 | A | Boceprevir | [1] | |
| Q80 | R, K | TMC435 | [6, 15] | **K × 1 (0.4)** |
| R109 | K | SCH446211 | [17] | |
| I153 | V | SCH446211 | [17] | **V × 22 (8.4)**, L ×11 (4.2) |
| R155 | K, T, I, M, G, L, S, Q | Telaprevir, Boceprevir, ITMN-191, BILN2061, TMC435 | [1, 3, 4, 6, 10, 11, 15, 19, 20] | |
| A156 | S, T, V, I, G | Telaprevir, Boceprevir, ITMN-191, BILN2061, SCH446211, TMC435, SCH900518 | [1, 3, 4, 10, 11, 15, 17, 19, 20, 38] | |
| D168 | A, V, E, N, T, H | BILN2061, ITMN-191, TMC435 | [6, 15, 20] | **E × 4 (1.5)** |
| V170 | A | Telaprevir, Boceprevir | [1, 19, 20] | |
| M175 | L | Boceprevir | [39] | |
| Total number (%) of patients with resistance proven mutations | | | | 35 (13.4) |
| Total number (%) of patients with resistance proven and unproven mutations | | | | 47 (18.0) |

Amino acid mutations conferring PI resistance in the literatures and those observed in PI-treatment-naive patients in this study are indicated. Bold indicates resistance proven mutations, and the others indicate resistance unproven mutations

Double mutations found were as follows: V36I and I153V × 1, T54S and I153V × 4, I153V and D168E × 2

(including resistance unproven mutations). Age, sex ratio, body mass index, alanine aminotransferase (ALT) levels, serum albumin, platelet count, and fibrosis stage did not differ between the NS3 mutation and wild-type groups. No significant difference was observed between the two groups in the parameters of PEG-IFN/RBV treatment response, HCV sequence variations in interferon sensitivity determining region (ISDR), Core 70, interferon plus ribavirin resistance-determining region (IRRDR), or interleukin 28B (IL28B) single nucleotide polymorphism (SNP) (rs8099917; T/G and G/G vs. T/T) [23–30]. These clinical variables were also compared between the mutation group defined as resistance proven mutations and the wild-type group, but no notable differences were observed.

### Unimpaired in vivo fitness of viral strains with resistance mutations

Because most PI-resistance mutations described till date have been associated with reduced replicative capacity of varying degrees [1, 10, 11, 13, 17, 20–22, 31, 32], we examined viral replication levels in patients with drug-resistance mutations (Fig. 2). The estimated $P$ value indicated no significant difference between the mutation (median 1,500 KIU/ml) and wild-type (median 1,800 KIU/ml) groups ($P = 0.69$). The results indicate that drug-resistant HCVs were not necessarily impaired in their ability to replicate in vivo. However, patients with double mutations ($N = 7$) tended to have low viral loads (median 1,200 KIU/ml) ($P = 0.09$).

### Resistance mutations and virologic response to PEG-IFN/RBV therapy

To determine the difference in virologic response to PEG-IFN/RBV therapy according to the PI mutation, frequency of HCV RNA levels below detection at 4 weeks (rapid viral response, RVR) and 12 weeks (complete early viral response, cEVR), and SVR rate (%) were investigated in each group. The frequency of HCV RNA levels below detection at 4 and 12 weeks was 14 and 50%, respectively, in the mutation group, and was 11 and 46%, respectively, in the wild-type group. The SVR rate was 48 and 40% in the mutation and wild-type groups, respectively ($P = 0.38$). No significant difference was observed between the two groups in any of the indexes investigated (Table 2). The time-dependent viral clearance rate during PEG-IFN/RBV therapy was estimated in 133 patients including 25 patients (19%) with PI-resistance mutations available for the analysis. Kaplan–Meier analysis demonstrated that HCV clearance did not differ between the two groups with and without resistance mutations (log-rank test, $P = 0.30$) (Fig. 3).

### Changes in nonstructural 3 amino acid sequence diversity during PEG-IFN/RBV therapy

Full-length NS3 protease sequences were determined in 39 non-SVR patients after PEG-IFN/RBV therapy. A single amino acid change at resistance-associated sites in two patients was observed. In one patient, isoleucine (Ile) at position 153 changed to valine (Val), and glutamic acid (Glu) changed to aspartic acid (Asp) at position 168 in the second (Fig. 4). At the nucleotide level, ATC (Ile) changed to GTC (Val) in I153V, and GAA (Glu) changed to GAC (Asp) in E168D. Both mutations were caused by one nucleotide exchange. No other changes were observed in the other 37 patients.

## Discussion

Here we report that in 18% (47/261) HCV genotype 1b-infected patients who had not been previously treated with NS3 PIs, the viral genome contained dominant amino acid mutations within the NS3 PI-resistance sites. Even after confining the data to established PI-resistance mutations, the mutation rate was still significant in 13.4% (35/261). No clinical differences were observed between patients

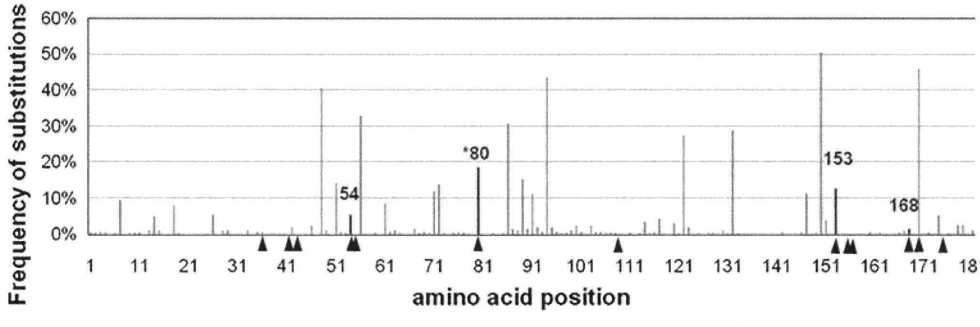



**Fig. 1** Frequency of polymorphic mutations for each of the 181 NS3 protease amino acid residues in 261 patients. *Arrowheads* indicate the sites reported to confer PI resistance. *Dark bars* denote the amino acid variations at the resistant sites in this study. **80*, we detected one resistant mutation (Q80K) and 47 (18%) non-resistant variations (Q80L) at the 80th residue

**Table 2** Characteristics of patients with or without HCV genomes harboring drug-resistance mutations

| Characteristics | Mutation type ($N = 47$) | Wild-type ($N = 214$) | $P$ value |
|---|---|---|---|
| Patients' characteristics | | | |
| Age, median (range) | 59 (46–72) | 57 (19–77) | 0.17 |
| Male, no. (%) | 26 (55) | 112 (52) | 0.70 |
| BMI, median (range) | 23.2 (15.5–31.9) | 22.8 (16.1–31.9) | 0.41 |
| ALT IU/ml | 81.3 ± 72.6[a] | 74.8 ± 51.9 | 0.93 |
| Serum albumin g/dl | 4.00 ± 0.37 | 4.01 ± 0.36 | 0.81 |
| Platelet count × $10^4$/μl | 15.8 ± 4.3 | 14.5 ± 4.8 | 0.18 |
| HCV RNA KIU/ml, median (range) | 1,500 (58–6,310) | 1800 (28–15,849) | 0.69 |
| Fibrosis, no. (%) | | | 0.97 |
| F0 | 0 (0) | 7 (3) | |
| F1 | 23 (50) | 89 (42) | |
| F2 | 9 (20) | 52 (24) | |
| F3 | 9 (20) | 40 (19) | |
| F4 | 5 (11) | 26 (12) | |
| IFN pre-treatment no. (%) | 15/40 (38)[b] | 66/172 (38) | 1.00 |
| IL28B (rs8099917) T/G or G/G no. (%) | 6/20 (30) | 19/67 (28) | 1.00 |
| Response to PEG-IFN/RBV therapy | | | |
| SVR total cases no. (%) | 22/46 (48) | 83/210 (40) | 0.38 |
| RVR in total cases no. (%) | 6/44 (14) | 22/195 (11) | 0.83 |
| cEVR in total cases no.(%) | 22/44 (50) | 92/200 (46) | 0.75 |
| SVR 48w treatment no. (%) | 16/29 (55) | 55/130 (42) | 0.29 |
| End of treatment response no. (%) | 26/41 (63) | 123/202 (61) | 0.91 |
| HCV genome sequence variation | | | |
| ISDR mutation ≤1 no. (%) | 32/46 (70) | 167/210 (80) | 0.21 |
| Core70 R no. (%) | 26/44 (59) | 136/210 (65) | 0.56 |
| IRRDR mutation >3 no. (%) | 25/38 (66) | 107/190 (56) | 0.34 |

[a] Mean ± SD

[b] Number/total number (%)

harboring viruses with and without these mutations. Moreover, no differences were observed in the responses of either group to PEG-IFN/RBV therapy.

Recent studies reported that significant number of patients who were never treated with PI possess viral sequences with PI-resistance-associated NS3 mutations. In these studies, the prevalence of PI-resistance mutations was determined to be 8.6–16.2% [13, 14], in HCV genotype 1- and 3-infected patients in European–American populations. These patients were often coinfected with HIV. Analysis of the public HCV databases (EuHCVdb and Los Alamos) also reported the presence of naturally occurring PI-resistance-associated NS3 mutations in worldwide isolates [33]. However, in vivo and in vitro studies demonstrated that most of the mutations observed conferred only low- to moderate-level PI resistance [7, 13, 14, 34, 35]. Regarding viral fitness, PI-resistant HCVs show lower fitness at varying degrees as revealed by in vitro studies [1, 10, 11, 17, 20–22, 31, 32], but HCV RNA levels in a clinical study did not differ significantly. The response to PEG-IFN/RBV therapy was almost comparable to that in HCV-infected patients without PI-resistance mutations either in HCV replicon experiments or in a clinical study of small number of treated patients [34].

The prevalence of 13.4% for PI-resistance-proven patients observed in the present study was almost comparable to the results of previous studies. Although HIV is known to increase HCV replication in coinfection with HCV [36], and HIV patients are often treated with the HIV-specific PIs, the HIV infection might not affect the natural occurrence of HCV-specific PI-resistance mutations since our studied patients were all proven to be free from coinfection with HIV infection. As shown in Table 1 and Fig. 1, I153 V (22/261, 8.4%), T54S (14/261, 5.4%), and D168E (4/261, 1.5%) were among the most prevalent PI-resistance-proven mutations in the present study. The most frequent mutation detected in our study I153V was reported to appear secondarily to the occurrence of R109K mutations in a HCV replicon system [17]. Although the role of this mutation is not understood, the I153V mutation on its own conferred SCH446211 resistance to the HCV replicon to a lesser degree [17]. Interestingly, I153V was often found in double mutations in our study, as shown in Fig. 2. This suggests analogy between in vitro and in vivo data. T54S and D168E, the other frequent mutations, have been also reported to occur as single dominant mutations in previous in vitro or in vivo studies in HCV genotype 1

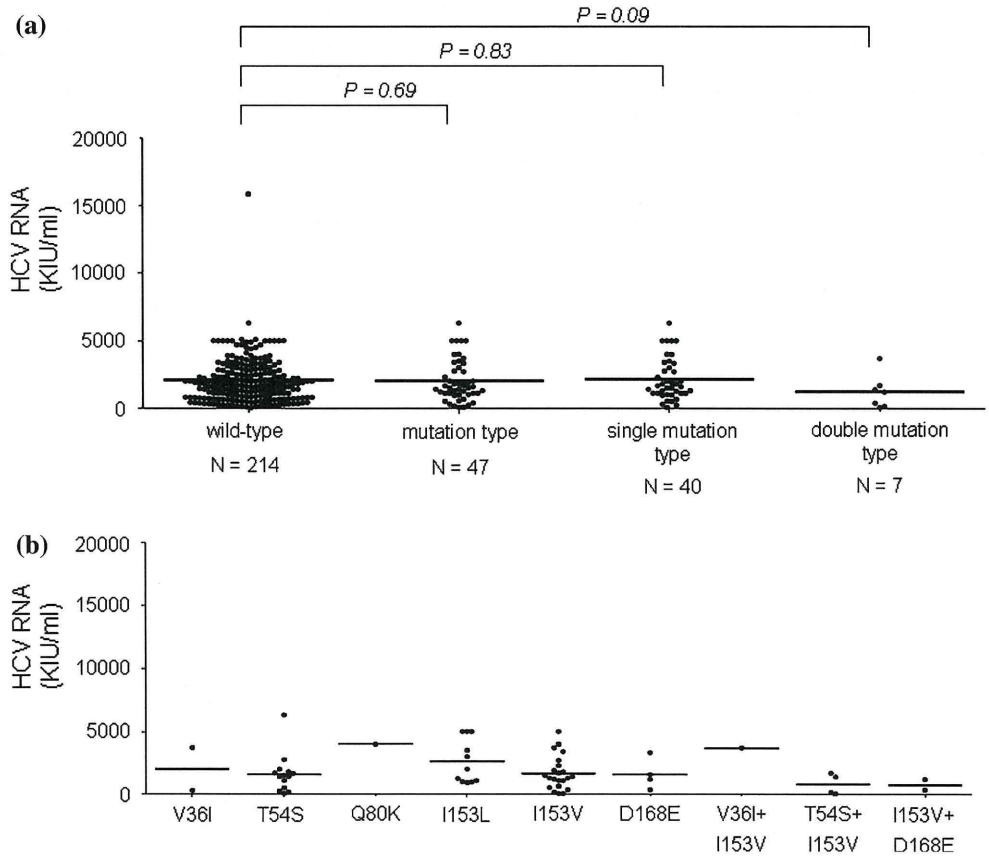



**Fig. 2** In vivo fitness of HCV with PI-resistance-associated NS3 mutations. HCV RNA levels were compared between patients with and without NS3 PI-resistance-associated mutations (**a**) and between patients with each resistance mutation (**b**). The estimated *P* value (Mann–Whitney *U* test) indicates no significant difference between the wild-type and other groups (wild-type vs. mutation type, wild-type vs. single mutation type, and wild-type vs. double mutation type). (Wild-type, $N = 214$; mutation type, $N = 47$; single mutation type, $N = 40$; double mutation type, $N = 7$; V36I, $N = 2$; T54S, $N = 14$; Q80K, $N = 1$; I153L, $N = 11$; I153V, $N = 22$; D168E, $N = 4$; E176A, $N = 1$; V36I + I153V, $N = 1$; T54S + I153V, $N = 4$, and I153V + D168E, $N = 2$)

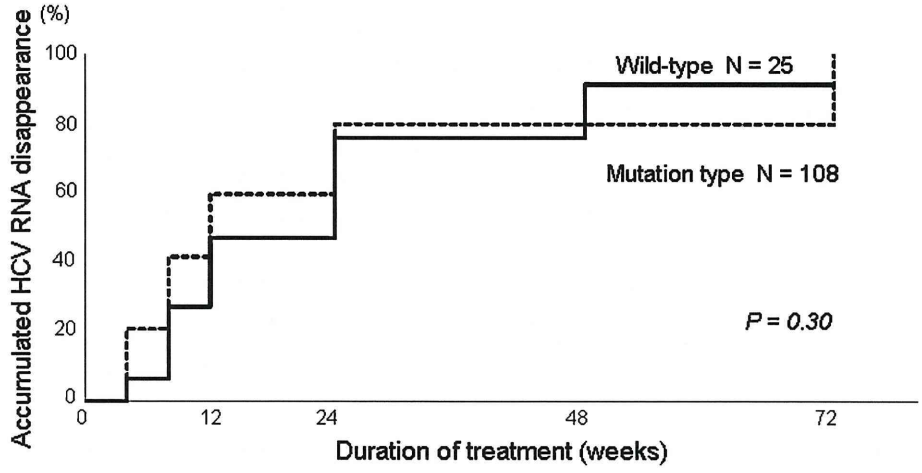



**Fig. 3** Comparison of virologic response to PEG-IFN/RBV therapy between HCV-infected patients with and without PI-resistance-associated NS3 mutations. Time-dependent HCV clearance rate analysis was based on serum HCV RNA positivity during PEG-IFN/RBV therapy for HCV isolates with resistance mutations or wild-type sequences. A total of 133 patients for whom the limit of viral genome detection could be determined were analyzed. Among this group, NS3 mutations were detected in 25 patients (19%). The estimated *P* value (log-rank test) shows no significant difference between the two groups ($P = 0.30$)

infections showing moderate degrees of resistance [16, 18, 19].

Most PI-resistance mutations described to date have been associated with varying degrees of reduced replicative capacity [10, 11, 17, 20–22, 31, 32]. In the present study, HCV RNA levels of those patients with low- to moderate-level resistance mutations were similar to those in patients in the wild-type groups, suggesting that in vitro viral fitness

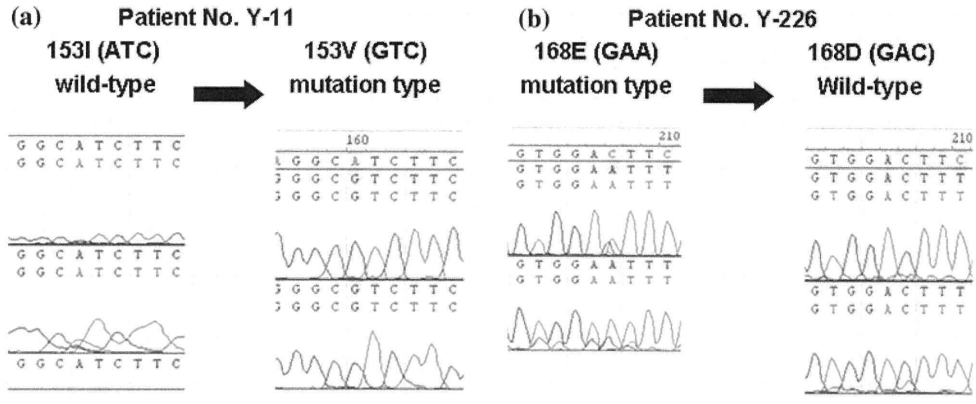



**Fig. 4** Appearance of PI-resistance-associated NS3 mutations during the PEG-IFN/RBV therapy. Chromatograms show part of the HCV NS3 sequence demonstrating PI-resistance mutations in two patients receiving therapy. **a** Site 153 isoleucine (Ile) (ATC) changed to valine (Val) (GTC), **b** Site 168 glutamic acid (Glu) (GAA) changed to aspartic acid (Asp) (GAC)

does not necessarily reflect in vivo viral fitness. This, however, does not rule out the possibility that some unknown compensatory viral mutations might have resulted in upregulation of reduced viral fitness. Interestingly, although the replicative capacity conferred by a single mutation seemed to be the same, the HCV RNA levels of double mutations were frequently low, suggesting that double mutations might weaken viral fitness.

In previous studies, clinical characteristics representing the state of liver disease other than HCV RNA levels were not studied in patients with PI-resistance mutations. In this study, we show that those clinical characteristics did not differ according to the presence of viral NS3 mutations. As shown in Table 2, age, sex ratio, fibrosis stage, ALT levels, serum albumin, platelet count, and past history of IFN pretreatment did not differ according to the presence of NS3 mutations. These results suggest that NS3 mutations occur independently of disease progression. Moreover, no evident differences were observed between viral and host factors known to affect IFN-based treatment responses. However, viral amino acid variations in the core and NS5A or the allelic frequency of IL28B SNPs, which were recently reported for the close relationship of responses to PEG-IFN/RBV therapy, did not differ between the two groups.

A significant outcome of the present study is the demonstration that PI-resistance mutations might not affect responses to PEG-IFN/RBV therapy. Previous in vitro studies demonstrated that HCV replicons harboring PI-resistance mutations were also sensitive to IFN treatment [31]. In addition, recent clinical studies also indicated that PI-resistance mutations were sensitive to the PEG-IFN/RBV [10, 34]. However, our analysis was more comprehensive because viral and host factors that contribute to treatment responses were simultaneously analyzed. A unique aspect of the present study is that we investigated the influence of the PEG-IFN/RBV treatment on the occurrence of new PI mutations by direct nucleotide sequencing, and were able to show that the PEG-IFN/RBV might not induce amino acid mutations.

Will the pre-existence of naturally occurring PI-resistance mutations have an influence on future treatment of HCV infections? Since new PIs are on the verge of clinical use, all clinicians should bear in mind the substantial numbers of HCV-infected patients with PI-resistance mutations. Although the degree of resistance is considered to be low or moderate in untreated patients, weak resistance might progress to more potent resistance with additional mutations, when PIs become widely used. Therefore, all clinicians need to be sufficiently prepared for the possibility of later onset of PI-resistance mutations that confer greater drug resistance and concomitant poorer responses to therapy. In SPRINT-1 study, the lead-in therapy was associated with a modestly lower rate of breakthrough than with no lead in [7]. Considering that PEG-IFN/RBV was equally effective for PI-resistant viruses, sufficient "lead-in" therapy before the administration of PIs could be an option in the forthcoming triple therapy modality.

In conclusion, we demonstrate here that PI-resistance-associated NS3 mutations exist in a substantial proportion of untreated HCV-1b-infected patients. Although the degree of resistance might not be strong, clinicians will need to consider this upon the introduction of triple therapy.

**Acknowledgements** This work was supported in part by a grant-in-aid scientific research fund from the Ministry of Education, Science, Sports, and Culture [grant number 20390206] and in part by a grant-in-aid from the Ministry of Health, Labor, and Welfare of Japan [grant number H19-kanen-002].

## References

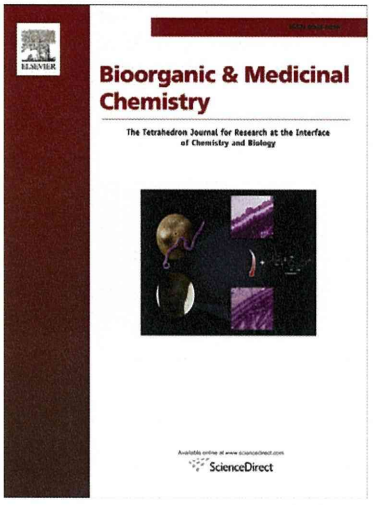
1. Susser S, Welsch C, Wang Y, et al. Characterization of resistance to the protease inhibitor boceprevir in hepatitis C virus-infected patients. Hepatology 2009;50:1709–1718
2. Hadziyannis SJ, Sette H Jr, Morgan TR, et al. Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. Ann Intern Med 2004;140:346–355

3. Hezode C, Forestier N, Dusheiko G, et al. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. N Engl J Med 2009;360:1839–1850
4. McHutchison JG, Everson GT, Gordon SC, et al. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. N Engl J Med 2009;360:1827–1838
5. McHutchison JG, Manns MP, Muir AJ, et al. Telaprevir for previously treated chronic HCV infection. N Engl J Med 2010; 362:1292–1303
6. Reesink HW, Fanning GC, Farha KA, et al. Rapid HCV-RNA decline with once daily TMC435: a phase I study in healthy volunteers and hepatitis C patients. Gastroenterology 2010;138: 913–921
7. Kwo PY, Lawitz EJ, McCone J, et al. Efficacy of boceprevir, an NS3 protease inhibitor, in combination with peginterferon alfa-2b and ribavirin in treatment-naive patients with genotype 1 hepatitis C infection (SPRINT-1): an open-label, randomised, multicentre phase 2 trial. Lancet 2010;376:705–716
8. Forestier N, Reesink HW, Weegink CJ, et al. Antiviral activity of telaprevir (VX-950) and peginterferon alfa-2a in patients with hepatitis C. Hepatology 2007;46:640–648
9. Suzuki F, Akuta N, Suzuki Y, et al. Rapid loss of hepatitis C virus genotype 1b from serum in patients receiving a triple treatment with telaprevir (MP-424), pegylated interferon and ribavirin for 12 weeks. Hepatol Res 2009;39:1056–1063
10. Kieffer TL, Sarrazin C, Miller JS, et al. Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients. Hepatology 2007;46: 631–639
11. Sarrazin C, Kieffer TL, Bartels D, et al. Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. Gastroenterology 2007;132: 1767–1777
12. Lin C, Lin K, Luong YP, et al. In vitro resistance studies of hepatitis C virus serine protease inhibitors, VX-950 and BILN 2061: structural analysis indicates different resistance mechanisms. J Biol Chem 2004;279:17508–17514
13. Kuntzen T, Timm J, Berical A, et al. Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients. Hepatology 2008;48:1769–1778
14. Gaudieri S, Rauch A, Pfafferott K, et al. Hepatitis C virus drug resistance and immune-driven adaptations: relevance to new antiviral therapy. Hepatology 2009;49:1069–1082
15. Lenz O, Verbinnen T, Lin TI, et al. (2010) In vitro resistance profile of the hepatitis C virus NS3/4A protease inhibitor TMC435. Antimicrob Agents Chemother 54:1878–1887
16. Kieffer TL, Kwong AD, Picchio GR. Viral resistance to specifically targeted antiviral therapies for hepatitis C (STAT-Cs). J Antimicrob Chemother 2010;65:202–212
17. Yi M, Tong X, Skelton A, et al. Mutations conferring resistance to SCH6, a novel hepatitis C virus NS3/4A protease inhibitor. Reduced RNA replication fitness and partial rescue by second-site mutations. J Biol Chem 2006;281:8205–8215
18. Welsch C, Domingues FS, Susser S, et al. Molecular basis of telaprevir resistance due to V36 and T54 mutations in the NS3-4A protease of the hepatitis C virus. Genome Biol 2008;9:R16
19. Tong X, Bogen S, Chase R, et al. Characterization of resistance mutations against HCV ketoamide protease inhibitors. Antiviral Res 2008;77:177–185
20. He Y, King MS, Kempf DJ, et al. Relative replication capacity and selective advantage profiles of protease inhibitor-resistant hepatitis C virus (HCV) NS3 protease mutants in the HCV genotype 1b replicon system. Antimicrob Agents Chemother 2008;52:1101–1110
21. Tong X, Chase R, Skelton A, Chen T, Wright-Minogue J, Malcolm BA. Identification and analysis of fitness of resistance mutations against the HCV protease inhibitor SCH 503034. Antiviral Res 2006;70:28–38
22. Lu L, Pilot-Matias TJ, Stewart KD, et al. Mutations conferring resistance to a potent hepatitis C virus serine protease inhibitor in vitro. Antimicrob Agents Chemother 2004;48:2260–2266
23. Akuta N, Suzuki F, Hirakawa M, et al. Amino acid substitutions in the hepatitis C virus core region of genotype 1b affect very early viral dynamics during treatment with telaprevir, peginterferon, and ribavirin. J Med Virol 2010;82:575–582
24. Ge D, Fellay J, Thompson AJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 2009;461:399–401
25. Tanaka Y, Nishida N, Sugiyama M, et al. Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. Nat Genet 2009;41: 1105–1109
26. Suppiah V, Moldovan M, Ahlenstiel G, et al. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. Nat Genet 2009;41:1100–1104
27. 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;48:38–47
28. Enomoto N, Sakuma I, Asahina Y, et al. 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; 334:77–81
29. Akuta N, Suzuki F, Kawamura Y, et al. Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. J Hepatol 2007;46:403–410
30. Akuta N, Suzuki F, Sezaki H, et al. Predictive factors of virological non-response to interferon-ribavirin combination therapy for patients infected with hepatitis C virus of genotype 1b and high viral load. J Med Virol 2006;78:83–90
31. Zhou Y, Bartels DJ, Hanzelka BL, et al. Phenotypic characterization of resistant Val36 variants of hepatitis C virus NS3-4A serine protease. Antimicrob Agents Chemother 2008;52: 110–120
32. Mo H, Lu L, Pilot-Matias T, et al. Mutations conferring resistance to a hepatitis C virus (HCV) RNA-dependent RNA polymerase inhibitor alone or in combination with an HCV serine protease inhibitor in vitro. Antimicrob Agents Chemother 2005; 49:4305–4314
33. Lopez-Labrador FX, Moya A, Gonzalez-Candelas F. Mapping natural polymorphisms of hepatitis C virus NS3/4A protease and antiviral resistance to inhibitors in worldwide isolates. Antivir Ther 2008;13:481–494
34. Bartels DJ, Zhou Y, Zhang EZ, et al. Natural prevalence of hepatitis C virus variants with decreased sensitivity to NS3.4A protease inhibitors in treatment-naive subjects. J Infect Dis 2008;198:800–807
35. Cubero M, Esteban JI, Otero T, et al. Naturally occurring NS3-protease-inhibitor resistant mutant A156T in the liver of an untreated chronic hepatitis C patient. Virology 2008;370:237–245
36. Soriano V, Puoti M, Sulkowski M, et al. Care of patients with hepatitis C and HIV co-infection. Aids 2004;18:1–12
37. Barbotte L, Ahmed-Belkacem A, Chevaliez S, et al. Characterization of V36c, a novel amino acid substitution conferring hepatitis C virus (HCV) resistance to telaprevir, a potent peptidomimetic inhibitor of HCV protease. Antimicrob Agents Chemother 2010;54:2681–2683

38. Tong X, Arasappan A, Bennett F, et al. Preclinical characterization of the antiviral activity of SCH 900518 (Narlaprevir), a novel mechanism-based inhibitor of hepatitis C virus NS3 protease. Antimicrob Agents Chemother 2010;54:2365–2370

39. Chase R, Skelton A, Xia E, et al. A novel HCV NS3 protease mutation selected by combination treatment of the protease inhibitor boceprevir and NS5B polymerase inhibitors. Antiviral Res 2009;84:178–184

# Accepted Manuscript




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

Daisuke Takaya, Atsuya Yamashita, Kazue Kamijo, Junko Gomi, Masahiko Ito, Shinya Maekawa, Nobuyuki Enomoto, Naoya Sakamoto, Yoshiaki Watanabe, Ryoichi Arai, Hideaki Umeyama, Teruki Honma, Takehisa Matsumoto, Shigeyuki Yokoyama

PII: S0968-0896(11)00745-0
DOI: 10.1016/j.bmc.2011.09.023
Reference: BMC 9507

To appear in: *Bioorganic & Medicinal Chemistry*

Received Date: 12 July 2011
Revised Date: 12 September 2011
Accepted Date: 12 September 2011

Please cite this article as: Takaya, D., Yamashita, A., Kamijo, K., Gomi, J., Ito, M., Maekawa, S., Enomoto, N., Sakamoto, N., Watanabe, Y., Arai, R., Umeyama, H., Honma, T., Matsumoto, T., Yokoyama, S., A new method for induced fit docking (GENIUS) and its application to virtual screening of novel HCV NS3-4A protease inhibitors, *Bioorganic & Medicinal Chemistry* (2011), doi: 10.1016/j.bmc.2011.09.023

ELSEVIER

Bioorganic & Medicinal Chemistry
journal homepage: www.elsevier.com

# 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
Received in revised form
Accepted
Available online

*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. (IC50 =1~10μM 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.



## 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 to 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

* Corresponding author. Tel.: +0-000-000-0000; fax: +0-000-000-0000; e-mail: yokoyama@biochem.s.u-tokyo.ac.jp

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, 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 co-workers 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 and co-workers[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 and co-workers 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 and co-workers reported the structure complexed with an inhibitor that interacts with the noncatalytic 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 and co-workers reported two structures (PDB codes 1BT7[27], 1DXW), and recently, Gallo and co-workers 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 (Figure 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

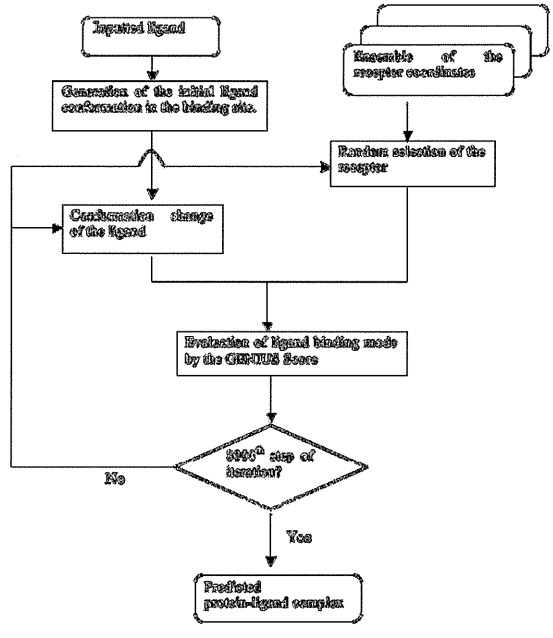



**Figure 1:** Flowchart of the GENIUS docking system.

thus discovered were modified based on the docking models, and revealed their preliminary structure-activity relationships.

## 2. Result and Discussion

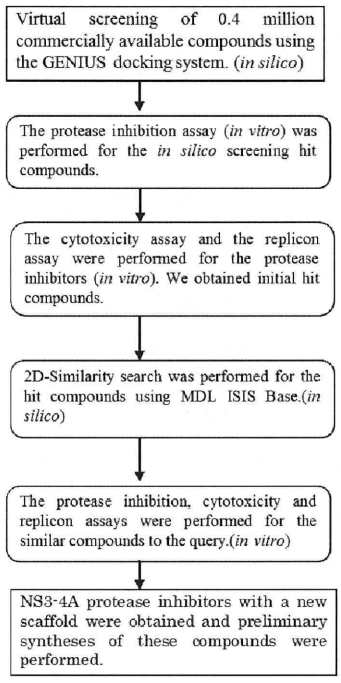


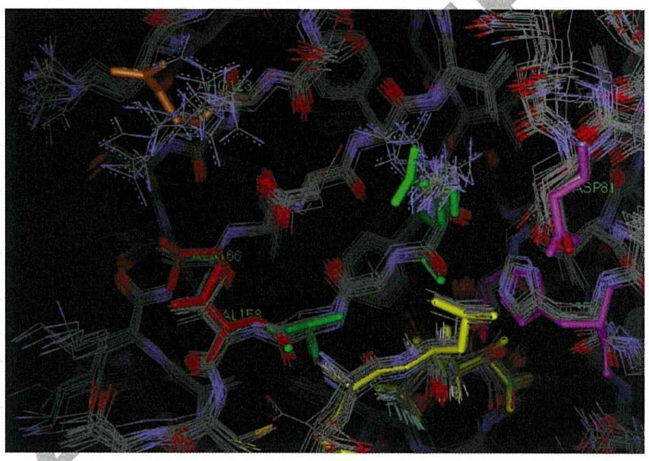

**Figure 2:** Flowchart of HCV NS3-4A *in silico* and *in vitro*

This study was performed by combining *in silico* and *in vitro* screening techniques. For the *in silico* screening part, we developed a new method for induced fit docking, called the GENIUS docking system (Figure 1 and see details in the Materials and Methods section), 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 (Figure 2, see details in the Materials and Methods section).

### *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 in the 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 (Figure 3-a). 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 (Figure 3-b).



(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-di-fluoro-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.

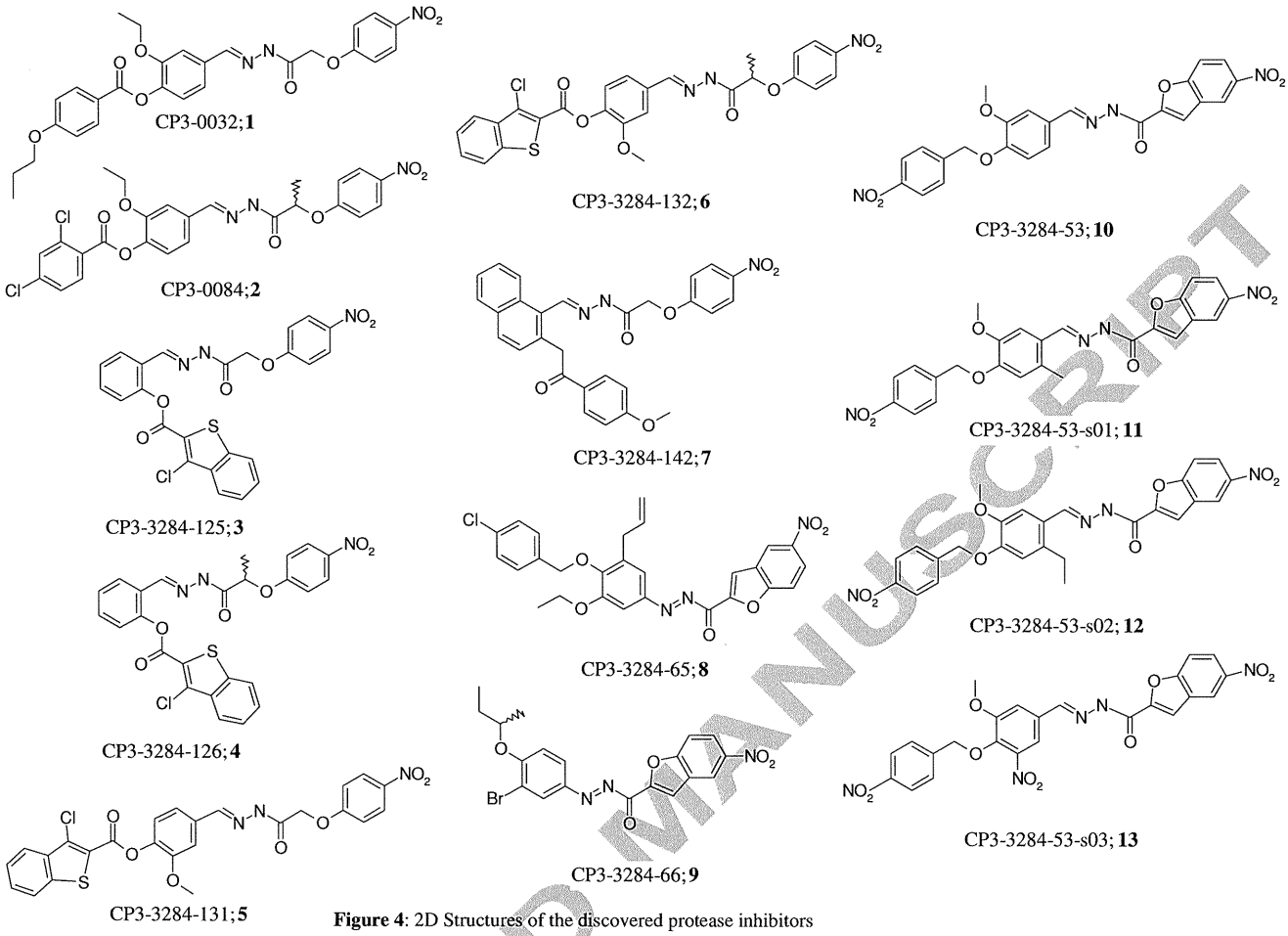



**Figure 4**: 2D Structures of the discovered protease inhibitors

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 (Figure 3-b). 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**) (Figure 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 (Figure 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 μM 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).

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

| ID; serial number | Inhibition at 100uM(%) | $IC_{50}$ | $EC_{50}$ | $CC_{50}$(MTS) | $CC_{50}$(ATP) | SI[a] | ALogP[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 $CC_{50}$ value (either $CC_{50}$(MTS) or $CC_{50}$(ATP)) to the $EC_{50}$ value.
[b] ALogP was calculated by PipeLinePilot 8.0.1(Accelrys Software Inc.).

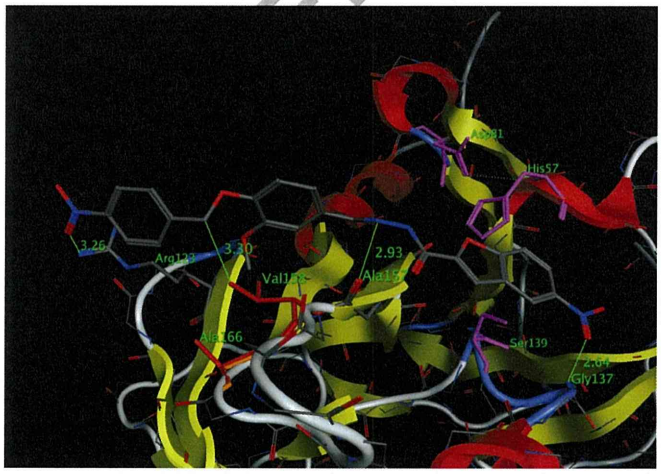
*2.4. Synthesis of Compounds 10, 11, 12 and 13*

Since the purity of compound **10** was unknown (we assumed 100% purity in the *in vitro* assay), compound **10** was synthesized (Scheme 1 in the Materials and Methods section). 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 (Figure 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.

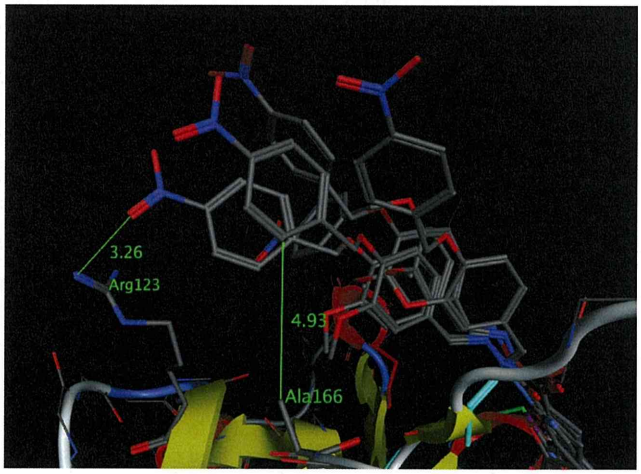
In summary, compound **1**, compound **2** and the CP3-3284 series (**3-10**) obtained in this research represent a new, unique class of non-peptide HCV NS3-4A inhibitors, because no similar HCV NS3-4A inhibitors (Tanimoto coefficient >= 0.7) are currently registered in SciFinder[36].

*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. (Figure 5-a,b, those of compounds **3**&**6** are in Figures S1 and 2).




(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 stick representation: predicted binding mode(s) of CP3-3284-53. Purple: the catalytic triad, red: hydrophobic residue on the β sheet.

**Table 2**: Ranking of the discovered compounds and the macrocyclic inhibitors and peptide-mimic inhibitors by the GENIUS docking system.

| ID | Rank (ave.) | SD. | ID | Rank (ave.) | SD. |
|---|---|---|---|---|---|
| CP3-0032 | 2163 | 636 | BILN-2061 | 23876 | 984 |
| CP3-0084 | 4410 | 1336 | ITMN-191 | 18140 | 962 |
| CP3-3284-53 | 1089 | 445 | MK-7009 | 22402 | 2966 |
| CP3-3284-65 | 1542 | 604 | TMC-435 | 19032 | 1788 |
| CP3-3284-66 | 2291 | 170 | VX-950 | N/A | N/A |
| CP3-3284-125 | 12260 | 178 | SCH-503034 | N/A | N/A |
| CP3-3284-126 | 10511 | 437 | | | |
| CP3-3284-131 | 1810 | 368 | | | |
| CP3-3284-132 | 4245 | 1254 | | | |
| CP3-3284-142 | 11047 | 441 | | | |
| CP3-3284-164 | 3056 | 478 | | | |
| CP3-3284-53-s01 | 3424 | 686 | | | |
| CP3-3284-53-s02 | 1802 | 856 | | | |
| CP3-3284-53-s03 | 3200 | 346 | | | |
| CP3-3284-53-s04 | 1415 | 369 | | | |

The predicted binding mode of compound **10** involved interactions with Val158 and Ala166, which are close to the side chain of Arg123 (Figure 5-a). 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 (Figure 5-b). 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 $NO_2$ (Figure 5-a) 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 (Figure 5-a).

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 logP 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 (Figure 6) of compound **10** was modified, by introducing methyl, ethyl, and nitro groups (Figure 4). The inhibition activity of the derivatives was not significantly changed. In compound **7**,

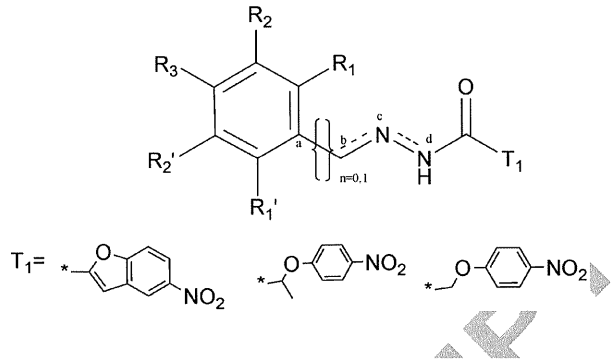



**Figure 6**: The common scaffold among CP3-3284 series. T1 means substructures in the CP3-3284 series.

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 (Figure 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 (Figure 5-a). 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.

### *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 1.06μM and 4.08μM, 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.

### *2.7. Validation of specificity for CP3-3284 series compounds and known-inhibitors by the GENIUS docking system with the EIP*

In this research, to identify the HCV NS3-4A protease inhibitors that differ from the conventional macrocyclic or peptide type inhibitors, the EIP was set to interact with not only

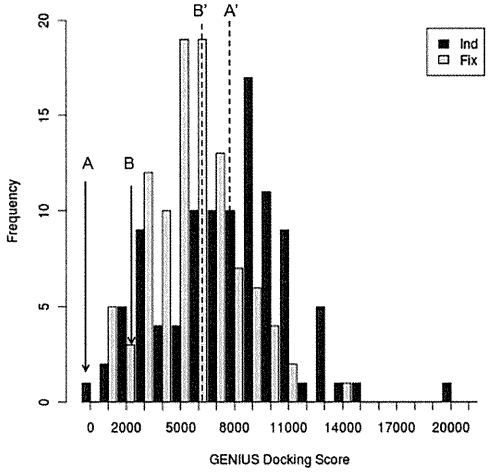



**Figure 7**: Distributions (histogram) of the docking scores of the decoy compounds and CP3-3284-53, using the induced-fit and the fixed receptor modes; "Ind" means the trial using the induced-fit receptor. "Fix" means the trial using the fixed receptor. A: The average score of CP3-3284-53 in the induced-fit receptor mode. B: The average score of CP3-3284-53 in the fixed receptor mode. A': The average score of the decoy compounds in the induced-fit receptor mode. B': The average score of the decoy compounds in the fixed receptor mode. Raw data are available in Supplementary data.

the active site but also the β sheet. We validated the effectiveness of the GENIUS docking system to detect the CP3-3284 series compounds, using the obtained EIP for the NS3-4a protease. The docking and the subsequent GENIUS score ranking of well-known protease inhibitors and 166,206 compounds (described in the Materials and Methods section) used as decoy compounds were performed. If the active compounds were ranked higher than the decoy compounds, then the *in silico* screening procedure can detect the inhibitors efficiently. The enrichment factor (EF) is one of the popular metrics for screening efficiency[39]. In this case, the EF(x) values were EF(1%)=14.3, EF(5%)=11.4, and EF(10%)=7.1 (x means the top x % of the total number of all calculated compounds). The number of active compounds is 21, including the CP3-3284 series (Figure 4) and the macro-cyclic and peptide inhibitors. Moreover, the rank orders between the CP3-3284 series and the other active compounds were compared, to validate the specificity for the CP3-3284 series of the obtained EIP for the NS3-4A protease. All of the CP3-3284 series compounds were ranked higher than the other active compounds. In addition, the Wilcoxon rank sum test indicated a significant difference between the distributions of the ranking between the CP3-3284 series and the other active compounds (p-value < 5.76e-11). This result shows that the EIP obtained for the NS3-4A protease had specificity for the CP3-3284 series compounds. Moreover, the ranking of the CP3-3284 series was higher than that of the macrocyclic compounds by the GENIUS score (Table 2). The peptide inhibitors could not be docked by the EIP because they lacked the ligand atoms specified in the obtained EIP. It was demonstrated that the GENIUS docking system, using the combination of the induced fit and the obtained EIP, had the capability to selectively detect a new class of inhibitors (CP3-3284 series compounds) that are neither peptide-type nor macrocyclic inhibitors.

*2.8. Validation of the detection capability for the CP3-3284 series compounds in terms of induced-fit and no-induced fit in the GENIUS scoring function*

Table 3. Partially-divided EIPs and Enrichment Factors for each partially-divided EIP.

| KEYATM | EIP (1) | EIP (2) | EIP (3) | EIP (4) |
|---|---|---|---|---|
| O.3 100 2.58 NE2 HISA__36 | 1 | 1 | 1 | 1 |
| O.co2 100 2.60 N GLYA_116 | 1 | 1 | 1 | 1 |
| DONOR 100 3.40 O ARGA_134 | | 1 | 1 | 1 |
| ACPTR 100 2.60 N ALAA_136 | | 1 | 1 | 1 |
| DONOR 100 2.60 O ALAA_136 | | 1 | 1 | 1 |
| C.3 100 2.60 CB ALAA_145 | | | 1 | 1 |
| C.3 100 3.80 CB VALA_137 | | | | 1 |

| | EF(1%) | EF(5%) | EF(10%) |
|---|---|---|---|
| EIP(1) | 0.0 | 0.0 | 1.3 |
| EIP(2) | 0.0 | 4.0 | 8.0 |
| EIP(3) | 6.7 | 20.0 | 10.0 |
| EIP(4) | 46.7 | 20.0 | 10.0 |

(a): The upper table; Enable KEYATMS in each partially-divided EIP. If the KEYATM was valid in the EIP, then the corresponding column bit was on. EIP (4) is the same as the EIP set up for the HCV NS3-4A *in silico* screening in this study. EIP(1): used KEYATMS only near the active-site 3 residue. EIP(2): used the hydrogen bond interactions and EIP(1); EIP(3): used part of the hydrophobic interaction and EIP(3). (b) The lower table; EF(x%) for each partially-divided EIP.

In order to clarify the effects of induced fit docking by GENIUS on the discovery of the CP3-3284 series compounds, a docking without consideration of induced fit was performed. To cancel the consideration of induced fit, the X-ray structure complexed with TMC-435 (PDB code: 3KEE) was used, instead of the receptor conformation ensemble. In addition, no collisions between ligand atoms and receptor atoms were allowed. Except for the receptor coordinates and the collision term, the docking calculation conditions were the same as those of the previous GENIUS induced fit docking calculation. Five docking calculations were performed with the receptor, and as a result, the average value of the GENIUS score with induced fit was about 3 times better than that with the fixed receptor (Figure 7). The obtained EIP contributed to the discovery of compounds that formed hydrophobic interactions with Val158 and Ala166 on the β sheet, arising from induced fit. The reason for the worse score of CP3-3284-53 (**10**) in the case of the fixed receptor is mainly due to the collision with Arg123, which was permitted in the case of an induced fit setting. We have shown that our defined EIP functions are effective with a receptor that functions by induced fitting, including side chain fluctuations. Moreover, the average score of the decoy distribution was better (6227.2) than that of the induced fit receptor mode. Additionally, in the case of the induced fit mode, the standard deviation of the score was larger than that of the fixed mode. Since the induced fit mode used multiple receptor conformations for the docking calculation, more diverse binding modes were generated, thus enlarging the standard deviation of the induced fit mode. Interestingly, the top scores of the decoy compounds for both the fixed and induced fit modes were almost the same (although the compounds with the top scores were not the same). The average score of compound **10** in the induced fit mode (965.3) and its rank (1st place) were quite improved, as compared with those in the fixed mode. Although the ranking is largely influenced by the selection of decoy compounds, the induced fit mode played very important roles in the discovery of the non-peptide inhibitor **10**.

*2.9. Validation of the effectiveness of the obtained EIP for the CP3-3284 series compounds*

To verify the effectiveness of the obtained EIP for the HCV NS3-4A protease docking, four different EIPs derived from the obtained EIP were used for *in silico* screenings. The EIPs are listed in Table 3-(a). As active compounds, 15 of the CP3-3284 series compounds were used, and as the decoy compounds, 3,000 compounds randomly selected from PubChem were used. For each decoy compound, the GENIUS score was calculated once. For each CP3-3284 series compound, the average score of five calculations was used. The EFs are listed in Table 3-(b). In the case of the EIP(1) condition, the EFs were quite poor. This result shows that it is difficult to obtain active compounds when only the active site atoms in the EIP are specified. In the case of EIP(4), the EFs were better than those of the other conditions. This result shows that EIP(4) was optimized for the CP3-3284 series. Therefore, if EIP(4) had not been used, then the CP3-3284 series compounds probably would not have been detected. However, when EIP(3) was used, the EF(5%) and EF(10%) values gave good results, and the EF(10%) value using EIP(2) was reasonable. In a future study, in order to obtain different compounds from the CP3-3284 series, we plan to perform a docking calculation with a new EIP, with KEYATM added, on the basis of EIP(2) or EIP(3). This GENIUS docking system, using these EIPs, is expected to identify a new class of HCV NS3-4A protease inhibitors that interact with the flexible region, in addition to the inhibitors detected by the conventional docking.

*2.10. Consideration of the collision term in the GENIUS docking system*

In general, it is hard to determine only one structure coordinate by NMR, because only a few restrictions, such as NOEs and the torsion angles, are available. However, NMR structures include information related to the flexibility of the protein molecule in solution. Therefore, it is likely that the flexible atoms in the NMR structure ensemble are ignored in the calculation of the collisions between the protein and the ligand. Table S3 summarizes the atoms that were judged as being flexible, based on a cluster analysis of the torsion angles, and thus were ignored in the collision term calculation. Interestingly, while all of the atoms of His57 in the active site were flexible, different atoms were flexible in Arg119, Arg123, and Arg155, because the flexible regions of the side chains were different. In Arg119, the NE, CZ, NH1, and NH2 atoms were permitted to collide. However, in Arg123 and Arg155, CG and CD were also added.

For example, in Glide, to express the induced fit of the receptor, an intermolecular collision can be relaxed by scaling each VDW radius. According to Table S3, the flexibilities of receptor atoms are dramatically different, even in the same residue. Therefore, if very small scaling coefficients were uniformly set for all of the atoms in a binding site, then most of the real inhibitors could be docked into the active site without any collision. However, many inactive compounds would also fit, and the screening efficiency would be very low. Therefore, the individual assignment of each atom, which permits a collision using the degree of torsion angle preservation derived from experimental structures (NMR or multiple X-ray structures), was effective to address the local softness of the receptor.

## 3. Conclusion

A new induced fit docking system, GENIUS, was developed, using collision term modification based on an experimentally determined protein structure ensemble and the essential interaction pair (EIP). The GENIUS system was applied to virtually screen HCV NS3-4A protease inhibitors, and a new class of non-peptide inhibitors was successfully identified. The EIPs for the induced fit of Arg123 on the β sheet and the hydrophobic interaction with the ligand in the open space were extracted by analyses of the binding site. Based on the ranking of the compounds by the GENIUS score, 97 compounds were selected and purchased. Among them, 27 compounds exhibited >50% inhibition at 100 μM in the protease inhibition assay. In the cell-based infection inhibition assay (replicon assay), two compounds showed 10 μM level potency ($EC_{50}$: 13 μM and 23 μM).

From a 2D similarity search of the chemical series, 140 compounds were obtained, and five compounds with $IC_{50}$ values lower than 10 μM were identified. In particular, compound **3** was the most potent, with an $IC_{50}$ of 1.06 μM. Unfortunately, since it exhibited cytotoxicity, this compound is not suitable as a seed molecule for drug development. Instead, compound **10**, which has 10 μM level potency ($IC_{50}$: 8.59 μM and $EC_{50}$: 12 μM) and no toxicity at >80 μM, was selected, and the preliminary structure-activity relationship was analyzed. We believe that compound **10** is promising as a seed for future synthetic development. The discovered compounds represent a new class of non-peptide HCV NS3-4A protease inhibitors. Furthermore, the new chemical series lacks an asymmetric carbon, unlike the existing inhibitor, and does not have a macrocyclic structure. Therefore, in terms of the synthetic feasibility and the ADME profile, the discovered chemical series has chemical tractability, as compared with the conventional peptide-type or macrocyclic NS3-4A inhibitors. The obtained EIP was capable of selectively identifying the CP3-3284 series, based on the validation results using both the induced fit and fixed receptor modes of GENIUS. In the validation, the score of compound **10** was greatly improved when induced fit was enabled. The rank of compound **10** over the decoy compounds and the EF of the CP3-3284 series were also superior in the induced fit mode. The effectiveness of the EIP was validated using the EF values under different EIP conditions. To improve this docking system, the collision coefficient was not set as a binary bit (0 or 1) for every atom of the receptor, but instead to a value between 0 and 1, by the clustering of the receptor conformation ensemble. It is hoped that a compound with the new skeleton identified by this research will be useful for future HCV therapies.

## 4. Materials and Methods

*4.1. in silico experiment schema*

*4.1.1. Receptor coordinates for docking calculations*

The NMR structure of the HCV NS3-4A protease complexed with an inhibitor (PDB code 1DXW [24]) was used as the receptor for this *in silico* screening. The structure was complexed with the peptide mimic inhibitor (3-amino-5,5-di-fluoro-2-keto-pentan-1-oic acid), which forms a covalent bond with Ser139 in the active site. The 20 registered structures were used for the receptor conformation ensemble. We considered the atomic coordinates in which the torsion angle is not maintained among the NMR conformations to have a low possibility for interaction with ligands in the stable conformation of the NS3-4A receptor. Thus,

the collisions between the receptor-ligand atoms in the flexible regions were tolerated. The criteria of flexibility were determined based on the preservation of the corresponding torsion angle of the receptor ensemble by clustering, as mentioned later.

*4.1.2. Clustering of the ensemble of receptor conformations*

The ensembles of the receptor conformation were clustered, in order to consider induced fit by the receptor. All of the side chain torsion angles maintained in the parent population, in the range of variation around the average angle of α % and plus or minus β degrees, were collected. The collected residues were referred to as the rigid residues. However, when the χ angle of the origin of the side chain was not maintained, it was assumed that the following atoms in the side chain were also not maintained, and these residues were referred to as the flexible residues. The side chain atoms of the flexible residues were ignored in the collision term of the GENIUS scoring function, which evaluates interactions between the receptor and the docking ligand. In the case of the NS3-4A protease, collisions between the docking ligand and the main chain atoms were not permitted. The details of the defined scoring functions are mentioned below. GENIUS (GENerating IndUced Systems)[40,41], which we encoded, implemented flexible ligand docking and induced-fit ligand docking algorithms, using the above scoring function.

*4.1.3. Introducing of EIP*

GENIUS requires three-dimensional receptor coordinate(s), ligand structures and essential interaction pairs (EIP). One EIP entry consists of an interaction pair that specifies the atom types of both the receptor and ligand atoms, the equilibrium distance, and the strength of the constraint. For example, if the CB atom of Val137 in the receptor interacts with the SP3 carbon atom in the ligand with an equilibrium distance of 3.8 A, and using the constraint value of 100, its EIP is described as follows:

KEYATM C.3 100 3.80 CB VALA_137

In the PDB format, the character string of amino acid residues is normally presented with capital letters. Therefore, it was similarly treated by the EIP. The designation of the hydrogen donor and acceptor is also possible, in addition to the character string full match of the atomic species. One or more combination(s) of the designation are available for the EIP. When at least one of the EIP criteria cannot be fulfilled, because the indicated atom type does not exist in the docking ligand, the docking calculation can be skipped.

*4.1.4. Generation of the initial interaction structure in the binding site*

First, the initial binding mode of each docking ligand was prepared. Dummy atoms were generated around the atoms of the receptor specified by the EIP(s), and the atoms between the ligand and the dummy were structurally aligned using the DALI[42]-like algorithm, while maintaining the initial ligand conformation. The formula is provided below:

$$S = \sum_{i=1}^{N}\sum_{j=1}^{N}\phi(i,j) \qquad (1)$$

$$\omega = \exp(-|d^{A}{}_{i,j} - d^{B}{}_{i,j}|)^2 \qquad (2)$$

$$\phi(i,j) = \begin{cases} \theta(1.0-\lambda) & (i=j) \\ \dfrac{\theta}{\mu}(\omega-\lambda) & (i\neq j) \end{cases} \qquad (3)$$

where $N$ is the number of joints, $d^{A}_{i,j}$ is the distance between the i-th and j-th dummy atoms, $d^{B}_{i,j}$ is the distance between the i-th and j-th ligand atoms, and μ is the average distance of $d^{A}_{i,j}$ and $d^{B}_{i,j}$. θ and λ are constants (1.535 and 0.81, respectively). To obtain the maximized $S$, the correspondence atom relationship between the dummy and the ligand was randomly generated 10,000 times.

*4.1.5. GENIUS scoring function*

A binding mode with a smaller score has an advantage in a protein-ligand interaction. To optimize the interaction of the initial ligand pose, the conformational changes of the ligand, translation and rotation, are repeated 8,000 times. In the case of using more than 2 receptor structures, one coordinate included in the ensemble of receptor conformations was randomly selected for every step. In addition, slight conformational changes (between plus or minus 1 degree) of the ligand were performed 5,000 times. The definition of the GENIUS scoring function $U_{optimum}$ is described below.

$$U_{optimum} = U_{sar} + U_{hydrogenbond} + U_{hydrophobi} + U_{stacking} + U_{collision} + U_{ligand\text{-}internal} \qquad (4)$$

The atomic radius and the distance of the interatomic interaction were determined by reference to the AMBER99[43] and MM3[44] parameters.

*4.1.5.1. EIP term*

One of the features of the GENIUS docking system is $U_{sar}$, which considers the EIP in the score function. This term is effective to make a specific ligand atom interact with a restricted binding site in the receptor. The formula is defined below:

$$U_{sar} = \sum_{i=1}^{N}\varphi_{sar}(i,j) \qquad (5)$$

$$\varphi_{sar}(i,j) = K_{sar}(R_{sar} - R)^2 - \delta \qquad (6)$$

where $R_{sar}$ is the i-th equilibrium distance, $R$ is the distance between the i-th specified atoms of the ligand and the receptor, $K_{sar}$ is the i-th strength of the restraint, and δ is a constant equal to -20.0. When the interaction distance of the binding mode is close to the specified equilibrium distance, this term judges that the interaction is favorable.

*4.1.5.2. Hydrogen bond (hb) term*

The hydrogen bonding score is calculated for the acceptor (or donor) atom of the receptor closest the donor (or acceptor) atom of the ligand. These atom types were previously defined. The formula is shown below:

$$U_{hydrogenbond} = \sum_{i=1}^{N} \varphi_{hb}(i) \quad (7)$$

$$\varphi_{hb}(i) = \begin{cases} -\dfrac{K_{hb}(i)}{|R - R_{hb}(i)| + 1.0} & (\theta \le 30.0) \\ -\dfrac{K_{hb}(i)}{(|R - R_{hb}(i)| + 1.0)\theta} & (\theta > 30.0) \end{cases} \quad (8)$$

where $N$ is the number of hydrogen bonds, and $R$ is the distance between the two atoms that formed each hydrogen bond. $R_{hb}(i)$ and $K_{hb}(i)$ are the equilibrium distance and a constant of the strength of the atom pair forming the hydrogen bond, respectively. $\theta$ is the angle of the hydrogen bond, in degrees (Figure 8). If the hydrogen bonding angle exceeds 30 degrees, then the score rapidly worsens.

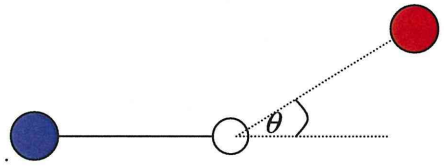



**Figure 8**: Definition of the hydrogen bond interaction. The red circle is the acceptor atom, the blue circle is the acceptor atom, and the white circle is a hydrogen atom. $\theta$ is the hydrogen bond angle.

### *4.1.5.3. Hydrophobic bond (hyd) term*

The hydrophobic score is calculated between the atoms of Ala, Cys, Phe, Ile, Leu, Met, Pro, Val, Trp, and Tyr (-OH is ignored) and the atoms of the ligand, defined as the hydrophobic atom within a fixed distance, by the following formula:

$$U_{hydrophobic} = \sum_{i=1}^{M} \sum_{j=1}^{N} \varphi_{hyd}(i, j) \quad (9)$$

$$\varphi_{hyd}(i, j) = \begin{cases} -\dfrac{K_{hyd}(i, j)}{R - R_{hyd}(i, j) + 1.0} & (R \ge R_{hyd}(i, j)) \\ -K_{hyd}(i, j) & (R < R_{hyd}(i, j)) \end{cases} \quad (10)$$

where $N$ and $M$ are the numbers of atoms that could form hydrophobic interactions in the ligand and the receptor, respectively (cutoff: 8.0 Å). $R_{hyd}(i, j)$ and $K_{hyd}(i,j)$ are the equilibrium distance and a constant defined for every interaction pair, respectively. $R$ is the distance between the i-th ligand atom and the j-th receptor atom.

### *4.1.5.4. Stacking term*

The stacking score was calculated if the distance between the i-th receptor aromatic atom and the j-th ligand aromatic atom is less than 5.0 Å. The aromatic ring center where the i-th atom belongs to the receptor side, is defined as i', the nearest aromatic atom is defined as j', from the j-th atom of the ligand, and the score was calculated by the following formula (Figure 9):

$$U_{stacking} = \sum_{i=1}^{M} \sum_{j=1}^{N} \varphi_{stacking}(i, j) \quad (11)$$

$$\varphi_{stacking} = \begin{cases} -K_{stacking}(i, j) R_{boundary} & (R_{boundary} < 0.0) \\ -K_{stacking(i,j)} \theta_{boundary} & (R_{boundary} \ge 0.0) \end{cases} \quad (12)$$

$$R_{boundary} = 1.0 - (R_{stacking}(i, j) - R)^2 \quad (13)$$

$$\theta_{boundary} = |1.0 - \Theta| \quad (14)$$

$$\Theta = \min\left\{\frac{\pi}{180.0}(\theta - 90.0)^2\right\} \quad (\theta : \theta_{i'j} or \theta_{ij'}) \quad (15)$$

where N and M are the numbers of atoms that could form stacking interactions in the ligand and the receptor, respectively. $R_{stacking}(i,j)$ and $K_{stacking}(i,j)$ are the equilibrium distance and a constant defined for every interaction pair, respectively.

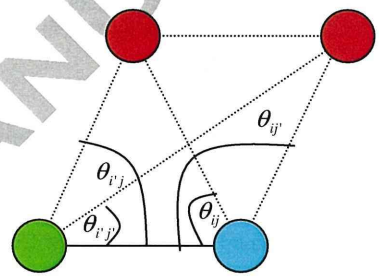



**Figure 9**: Definition of the stacking interaction. The cyan circle is the i-th atom in the aromatic ring of the receptor. The green circle means centroid of the aromatic ring including the i-th atom. The red circle is the j-th atom in the aromatic ring of the ligand.

### *4.1.5.5. Intermolecular collision term*

The intermolecular collision score was calculated for the atoms of the main chains and the rigid side chains, if the receptor ensemble was used. If the interatomic distance R between the i-th atom of the receptor and the j-th atom of the ligand is within the defined collision distance, then the following formula was applied:

$$U_{collision} = \sum_{i=1}^{M} \sum_{j=1}^{N} \varphi_{collision}(i, j) \quad (16)$$

$$\varphi_{collision}(i, j) = K_{collision} \varepsilon(i) (R_{collision}(i, j) - R)^2 \quad (17)$$

$$\varepsilon(i) = \begin{cases} 0 \\ 1 \end{cases} \quad (18)$$

where $M$ is the number of receptor atoms, $N$ is the number of ligand atoms, and $K_{collision}$ is a constant equal to 1,000.0. $R_{collision}(i,j)$ is the summation of van der Waals radii of the i-th ligand atom and the j-th receptor atom. $\varepsilon(i)$ is the collision coefficient, set to 1 or 0 for the receptor atoms through clustering

of the receptor ensembles. If the i-th atom is the ignored atom, then it is set to 0. Otherwise, it is set to 1.

#### *4.1.5.6. Internal ligand term*

In order to avoid ligand docking poses with collapsed internal structures, such as when the bond length is broken by repeating the rotation, and by intra-molecular collisions, a very strong restraint was added by the following formula:

$$U_{ligand-internal} = \sum_{i=1}^{L} \varphi_{bond-length}(i) + \sum_{i=1}^{N}\sum_{j=1}^{S} \varphi_{internal-collision}(i,j) \quad (19)$$

$$\varphi_{bond-length}(i) = K_{bond-length}\left\{(R_{bond-length}(i) - R_1)^2\right\} \quad (20)$$

$$\varphi_{internal-collision}(i,j) = K_{internal-collision}(R_{internal-collision} - R_2)^2 \quad (21)$$

where $L$ is the number of rotated bonds. $N$ is the number of atoms of the ligand. $S$ is the number of the i-th atom and the atoms that do not form a covalent bond. $K_{bond\text{-}length}$ is a constant equal to 100,000.0. $R_{bond\text{-}length}(i)$ is the bond length of the ligand in the initial structure. $K_{internal\text{-}collision}$ is a constant equal to 150.0. $R_{internal\text{-}collision}$ is a constant equal to 2.2(Å). $R_1$ is the distance for two atoms that form a covalent bond. $R_2$ is the distance between the i-th atom and an atom that does not form a covalent bond to the i-th atom.

#### *4.1.6. Setup of the EIP used in the NS3-4A protease in-silico screening*

The EIP can be automatically set up when a previously reported interaction is available from an X-ray structure or the associated literature. Nevertheless, in order to accurately dock a ligand to the important position of the receptor, the EIP should be determined manually. The EIP setting and the docking calculation were repeated until it was judged that the drug-like skeletons of compounds and appropriate binding modes were included in the ranking.

#### *4.1.7. Compound database used for in silico screening*

The MDL Available Chemical Directory 2005(ACD)[45] was used as the compound database for *in silico* screening (total 371,040 compounds). The database included the 2D structures of compounds that are commercially available. Ranking by the GENIUS score was performed for 166,206 compounds with molecular weights between 300 and 800. Generally, to reduce the number of docking compounds, drug like filter(s), such as Lipinski's Rule of five[46], were applied to the compound database. In the GENIUS docking system, the compounds without the atomic type specified in the EIP were removed from the docking calculation. For example, a ligand without a donor atom could not be docked, if the donor is specified in the EIP. This is equivalent to performing a pre-docking filtering of compounds by simple atomic species. After the *in silico* screening, the selection of the compounds that satisfy the EIP and a visual inspection of the predicted interaction status of higher ranked compounds were performed. After the NS3-4A inhibition assay, structurally similar compounds of the hit compounds were selected by a 2D-similarity search of the MDL ISIS Base[45] and were purchased.

### *4.2. In vitro experiment schema*

The protease inhibition activities of the compounds selected by *in silico* screening were measured, as the primary screen. For the hit compounds by the enzyme assay, one or both of the two cell viability test(s) described below were applied. Moreover, the concentration required for RNA generation inhibition in an HCV-infected cell was measured.

#### *4.2.1. Enzyme Assays*

The recombinant NS3 protease protein was prepared for compound screening, as an engineered single-chain NS3-protease (scNS3)[47]. The DNA sequence of the recombinant protein encoding the NS4A peptide (residues 21-33; GSVVIVGRIILSG) was genetically fused via a short linker (SGS), capable of making a beta-turn, to the N-terminus of the NS3 protease domain (residues 2-180, corresponding to 1,208-1,386 in the polypeptide). The gene encoding scNS3, with an N-terminal histidine-tag, was cloned into the pET32a(+) vector, and the protein was overexpressed in *Escherichia coli* (KRX). The scNS3 protein is reportedly soluble and fully active, with kinetic parameters virtually identical to those of the NS3/ NS4A non-covalent complex. The protein was purified by chromatography on a HisTrap HP column (GE Healthcare), a HiPrep26/60 desalting column (GE Healthcare) and then a HiTrap Q column (GE Healthcare). Finally, the purified protein was concentrated on a HiLoad Superdex75p.g. 16/60 column in 20 mM Tris-HCl buffer (pH 8.0), containing 300 mM NaCl and 2 mM dithiothreitol.

The NS3 serine protease activity was measured by the fluorogenic assay based on intramolecular fluorescence resonance energy transfer, reported previously[48]. A quenched-fluorogenic substrate, Mca-Asp-Asp-Ile-Val-Pro-Cys-Ser-Met-Lys(Dnp)-Arg-Arg (QF-2), derived from the NS5A/5B junction of the HCV polyprotein, was synthesized by Toray Research Center (Kamakura, Japan). The purity of the synthetic peptide was more than 95%, based on an HPLC analysis.

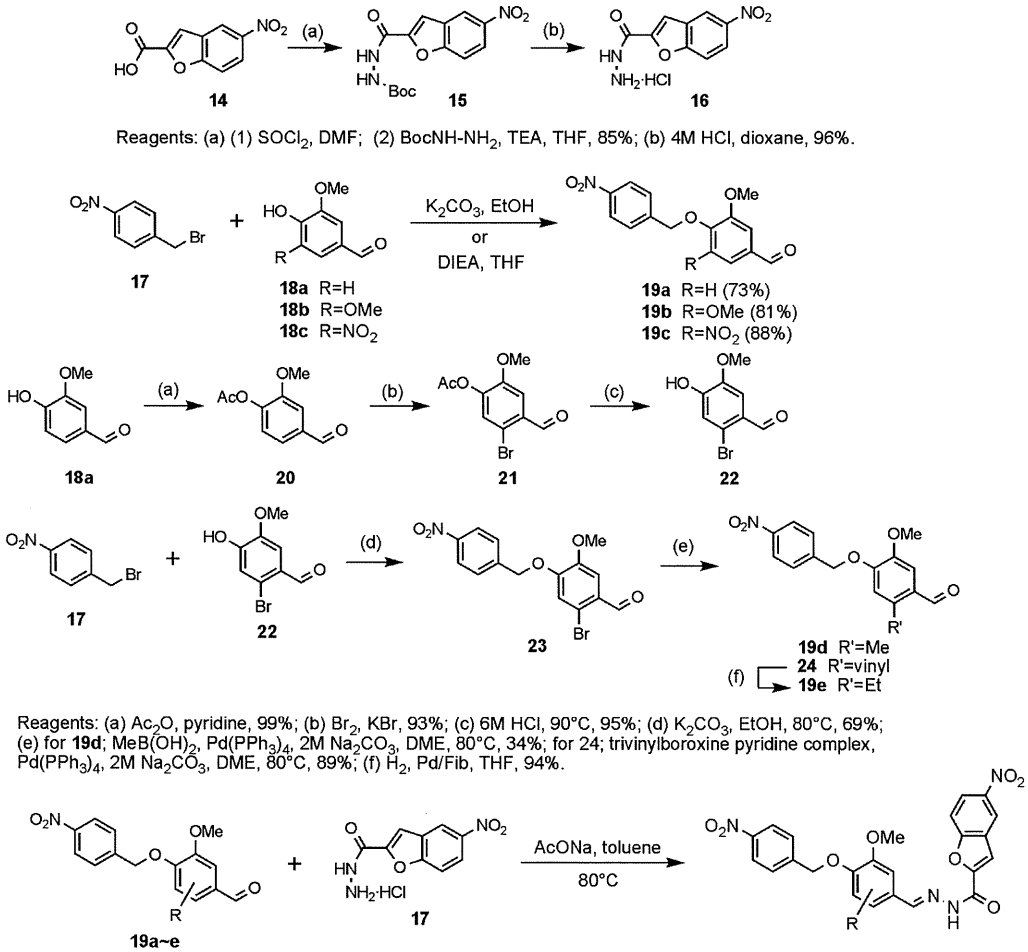
The enzyme was pre-incubated with or without chemical compounds dissolved in dimethylsulfoxide (DMSO), in a reaction mixture containing 50 mM Tris HCl (pH 7.8), 30 mM NaCl, 5 mM $CaCl_2$ and 10 mM dithiothreitol, at 37°C for 30 min, and then the reaction was started by the addition of QF-2 at a final concentration of 26 μM. The enzyme reaction was incubated at 37°C. The progress of the enzyme reaction was detected in a 96-well black plate with a Twinkle LB970 multiwell plate reader (Berthold Technologies GmbhH & Co, Bad Wildbad / Germany), using F340 and F440 filters for excitation and emission, respectively.

#### *4.2.2. Replicon assay*

##### *4.2.2.1. Cell culture*

An HCV replicon harboring cell line, Huh7/Rep-Feo[49,50], which expressed a chimeric gene encoding firefly luciferase and neomycin phosphotransferase, was used for the *in vitro* replication assay. Huh7/Rep-Feo cells were maintained in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal calf serum and 250 μg/ml of G418.

##### *4.2.2.2. Anti-HCV Assay in Huh7/Rep-Feo cells*




Scheme1: Synthetic routes of Compounds 10-15

Huh7/Rep-Feo cells were seeded in a 48-well plate at a density of 2 $X10^4$ cells per well. Two-fold serial dilutions of the test compounds in culture medium were added. After 72 h of culture, the expression levels of the HCV replicon were measured, using the luciferase assay system (Promega, Madison, WI, USA), and a JNR AB-2100 detector (Atto, Tokyo, Japan). The 50% effective concentration ($EC_{50}$) was defined as the concentration of compound that reduced the luciferase signal by 50%.

*4.2.2.3. Cell viability assays*

Huh7/Rep-Feo cells were seeded in a 96 well plate, at a density of 1 $X10^4$ cells per well, and were incubated in the presence of various compounds. The 50% cytotoxicity concentration ($CC_{50}$) was determined 72h after compound addition, using the cell titer 96 aqueous one solution cell proliferation assay (Promega, USA) (represented as "CC50 MTS" in this paper) or the cell titer-glo luminescent cell viability assay (Promega, USA) (represented as "CC50 ATP" in this paper), according to the manufacturer's protocol.

*4.3. Compounds*

*4.3.1. Commercially available compounds 1-10*

Compound **1** was SALOR-INT L29,866-2, compound **2** was SALOR-INT L39,343-6, compound **3** was R941689, and compound **4** was R942251. All were purchased from SALOR-INT. Compound **5** was R985147, compound **6** was R988529, and compound **7** was L268399, all purchased from SALOR-INT. Compound **8** was NATR212614, compound **9** was NATR206554, and compound **10** was NATR206692, all purchased from Vitas-M. The purity of these compounds was unknown, and thus 100% purity was assumed in the enzyme assay and the replicon assay.

*4.3.2. Synthesis of compounds 11-24*

*4.3.2.1. Purity analysis of the synthesized compounds*

A Waters 996 PDA (254nm) was used for detection. The column was a GL Science Inertsil ODS-3 (4.6 x 75 mm). The mobile phase gradient was a mixture of $H_2O$ and $CH_3CN$ (80:20, 0 min), (0:100, 5 min) with formic acid (0.1%). $^1H$ NMR spectra were obtained on a JEOL JNM ECP300 FT NMR system. Liquid chromatograph mass spectra (LC-MS) were detected in the ES positive mode.

*4.3.2.2. Compound 15 (tert-butyl N-[(5-nitro-1-benzofuran-2-carbonyl)amino]carbamate )*

5-Nitroglycerine benzofuran 2-carboxylic acid (**14**) (1.0 g, 4.8 mmol) was dissolved in methylene chloride (10 mL). Thionyl chloride (1.15 mL, 1.58 mmol) and N,N-dimethylformamide (30 uL) were added, and the resultant solution was stirred at 40 °C for 3.5 hours. After cooling in air, the mixture was concentrated