under reduced pressure. The residue was dissolved in THF (10 mL). Triethylamine (1.0 mL, 7.2 mmol) and tert-butyl-carbazate (0.77 g, 5.8 mmol) were then added, and the resultant solution was stirred at 0 °C for 1 hour. Water was added to the reaction mixture, followed by ethyl acetate. The organic layer was dried with anhydrous sodium sulfate.

The combined organic solutions were filtered, and concentrated under reduced pressure. The resultant solid was washed with ethyl acetate/hexane to give compound **15** (1.3 g, 4.1mmol, 85% yield).

^{1}H NMR (300 MHz, DMSO-d6) δ 10.69 (s, 1H), 9.12 (s, 1H), 8.83 (d, J = 2.4 Hz, 1H), 8.36 (dd, J = 2.4 Hz, J = 9.3 Hz, 1H), 7.95 (d, J = 9.3 Hz, 1H), 7.86 (s, 1H), 1.45 (s, 9H).

4.3.2.3. Compound 16 (5-nitro-1-benzofuran-2-carbohydrazide)

Compound **2** (300 mg, 0.93 mmol) was mixed with hydrochloric acid (3 mL, 4 mol/L in dioxane) at 0 °C, and then stirred for 2 hours at the same temperature. The solution was concentrated under reduced pressure, and the resultant solid was washed with isopropyl ether to give compound **16** (230 mg, 0.89 mmol, 96% yield).

^{1}H NMR (300 MHz, DMSO-d6) δ 8.86 (d, J = 2.4 Hz, 1H), 8.39 (dd, J = 2.4 Hz, J = 9.3 Hz, 1H), 8.06 (s, 1H), 7.99 (d, J = 9.3 Hz, 1H).

4.3.2.4. Compound 19a (3-methoxy-4-[(4-nitrophenyl)methoxy]benzaldehyde)

Vanillin (**18a**) (0.50 g, 3.3 mmol) and 4-nitrobenzyl bromide (**17**) (0.71 g, 3.3 mmol) were suspended in ethanol (2.5 mL), and potassium carbonate (0.23 g, 1.7 mmol) was added. The resultant solution was stirred at 80 °C for 15 hours. After cooling in air, the resultant solid was filtered and washed with ethanol, water, and ethanol to give compound **19a** (0.68 g, 2.4 mmol, 73% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ 9.87 (s, 1H), 8.26 (d, J = 8.7 Hz, 2H), 7.64 (d, J = 8.7 Hz, 2H), 7.46 (d, J = 1.5 Hz, 1H), 7.42 (dd, J = 1.5 Hz, J = 8.1 Hz, 1H), 6.97 (d, J = 8.1 Hz, 1H), 5.34 (s, 2H), 3.98 (s, 3H).

4.3.2.5. Compound 19b (3,5-dimethoxy-4-[(4-nitrophenyl)methoxy]benzaldehyde)

Syringaldehyde (**18b**) (0.50 g, 2.7 mmol) and 4-nitrobenzyl bromide (**17**) (0.59 g, 2.7 mmol) were suspended in ethanol (3 mL), and potassium carbonate (0.38 g, 2.7 mmol) was added. The resultant solution was stirred at 80 °C for 3 hours. After cooling in air, the resultant solid was filtered and washed with ethanol, water, and ethanol to give compound (**19b**) (0.70 g, 2.2 mmol, 81% yield) ^{1}H NMR (300 MHz, $CDCl_3$) δ 9.89 (s, 1H), 8.23 (d, J = 8.4 Hz, 2H), 7.68 (d, J = 8.4 Hz, 2H), 7.14 (s, 2H), 5.22 (s, 2H), 3.93 (s, 6H).

4.3.2.6. Compound 19c (3-methoxy-5-nitro-4-[(4-nitrophenyl)methoxy]benzaldehyde)

5-Nitrovanillin (**18c**) (0.50 g, 2.5 mmol) and 4-nitrobenzyl bromide (4) (0.55 g, 2.5 mmol) were suspended in THF (3 mL), and N, N-diisopropylethylamine was added (0.47 mL, 2.8 mmol). The resultant solution was stirred at room temperature for 12 hours and at 65 °C for 2 hours. After cooling in air, the insoluble matter was filtered, and the filtrate was concentrated under reduced pressure. The resultant solid was washed with diethyl ether to give compound **19c** (0.73 g, 2.2 mmol, 88% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ 9.95 (s, 1H), 8.26 (d, J = 8.4 Hz, 2H), 7.88 (d, J = 1.8 Hz, 1H), 7.68 (d, J = 1.8 Hz, 1H), 8.65 (d, J = 8.4 Hz, 2H), 5.38 (s, 2H), 4.03 (s, 3H).

4.3.2.7. Compound 20 ((4-formyl-2-methoxyphenyl) acetate)

Vanillin (**18a**) (9.99 g, 65.6 mmol) was suspended in methylene chloride (50 mL), and then acetic anhydride (7.5 mL, 79 mmol) and pyridine (6.4 mL, 79 mmol) were added. The resultant solution was stirred at room temperature for 18 hours. Water was added to the reaction mixture, and ethyl acetate was then added. The organic layer was washed with 1N HCl, a saturated sodium hydrogen carbonate aqueous solution, and saturated brine, and was dried with anhydrous sodium sulfate. The organic layer was filtered, and concentrated under reduced pressure to give pure compound **20** (12.6 g, 64.9 mmol, 99% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ 9.95 (s, 1H), 7.51 (d, J = 2.1 Hz, 1H), 7.48 (dd, J = 2.1 Hz, J = 7.8 Hz, 1H), 7.23 (d, J = 7.8 Hz, 1H), 3.91 (s, 3H), 2.35 (s, 3H).

4.3.2.8. Compound 21 ((5-bromo-4-formyl-2-methoxyphenyl) acetate)

Compound **20** (1.0 g, 5.1 mmol) and potassium bromide (2.0 g, 17 mmol) were suspended in water (10 mL), and then bromine (0.29 mL, 5.7 mmol) was added at 0 °C. The solution was stirred at room temperature for 15 hours. The resultant solid was filtered, washed with water and dried to give pure compound **21** (1.29 g, 4.72 mmol, 93% yield).

^{1}H NMR (300 MHz, DMSO-d6) δ 10.15 (s, 1H), 7.68 (s, 1H), 7.52 (s, 1H), 3.86 (s, 3H), 2.30 (s, 3H).

4.3.2.9. Compound 22 (2-bromo-4-hydroxy-5-methoxybenzaldehyde)

HCL (25 mL, 6 mol/L) was added to compound **21** (1.0 g, 3.7 mmol), and stirred at 90 °C for 4 hours. After cooling in air, the resultant solid was filtered, washed with water and dried to give compound **22** (0.80 g, 3.5 mmol, 95% yield).

^{1}H NMR (300 MHz, DMSO-d6) δ 10.01 (s, 1H), 7.34 (s, 1H), 7.11 (s, 1H), 3.83 (s, 3H).

4.3.2.10. Compound 23 (2-bromo-5-methoxy-4-[(4-nitrophenyl)methoxy]benzaldehyde)

Compound **22** (0.30 g, 1.3 mmol) and 4-nitrobenzyl bromide (**17**) (0.28 g, 1.3 mmol) were suspended in ethanol (3 mL), and then potassium carbonate was added. The resultant solution was stirred at 80 °C for 5 hours. After cooling in air, water was added to the solution. The resultant solid was filtered, washed with water, followed by ethanol, and dried to give compound **23** (0.33g, 0.90 mmol, 69% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ 10.19 (s, 1H), 8.28 (d, *J* = 8.4 Hz, 2H), 7.63 (d, *J* = 8.4 Hz, 2H), 7.47 (s, 1H), 7.07 (s, 1H), 5.29 (s, 2H), 3.94 (s, 3H).

4.3.2.11. Compound 19d (5-methoxy-2-methyl-4-[(4-nitrophenyl)methoxy]benzaldehyde)

Compound **23** (1.0g, 2.7mmol) and methylboronic acid (245 mg, 4.09 mmol) were dissolved in dimethoxyethane (10mL), and tetrakistriphenyl phosphine palladium (0.16 g, 0.14 mmol) and 2mol/L sodium carbonate aqueous solution (4.1 mL, 8.2 mmol) were added to the solution. The resultant solution was stirred at 80 °C for 18 hours. Methylboronic acid (245 mg, 4.09 mmol) and tetrakistriphenyl phosphine palladium (0.16 g, 0.14 mmol) were added to the solution, which was stirred at the same temperature for 24 hours. After cooling in air, ethyl acetate was added, and the organic layer was washed with water and saturated brine, and dried with anhydrous sodium sulfate. The resultant mixture was filtered and concentrated under reduced pressure. The residue was fractionated by silica gel column chromatography to give compound **19d** (0.28 g, 0.93 mmol, 34% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ 10.22 (s, 1H), 8.26 (d, *J* = 9.0 Hz, 2H), 7.63 (d, *J* = 9.0 Hz, 2H), 7.40 (s, 1H), 6.68 (s, 1H), 5.31 (s, 2H), 3.94 (s, 3H), 2.59 (s, 3H).

4.3.2.12. Compound 24 (2-ethenyl-5-methoxy-4-[(4-nitrophenyl)methoxy]benzaldehyde)

Compound 23 (0.30 g, 0.82 mmol) and tri-vinyl boroxine pyridine complex (99 mg, 0.41 mmol) were dissolved in dimethoxyethane (3 mL), and tetrakistriphenyl phosphine palladium (47 mg, 0.041 mmol) and 2 mol/L sodium carbonate aqueous solution were added. The resultant solution was stirred at 80 °C for 7 hours. After cooling in air, ethyl acetate was added to the mixture. The organic layer was washed water and saturated brine, and was dried with anhydrous sodium sulfate. The resultant mixture was filtered and concentrated under reduced pressure. The residue was fractionated by silica gel column chromatography to give compound 24 (228 mg, 0.728 mmol, 89% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ 10.26 (s, 1H), 8.27 (d, *J* = 8.7 Hz, 2H), 7.65 (d, *J* = 8.7 Hz, 2H), 7.42 (s, 1H), 7.40 (dd, *J* = 17.4 Hz, *J* = 10.8 Hz, 1H), 6.97 (s, 1H), 5.53 (dd, *J* = 17.4 Hz, *J* = 0.9 Hz, 1H), 5.47 (dd, *J* = 10.8 Hz, *J* = 0.9 Hz, 1H), 5.34 (s, 2H), 3.97 (s, 3H).

4.3.2.13. Compound 19e (2-ethyl-5-methoxy-4-[(4-nitrophenyl)methoxy]benzaldehyde)

Compound **24** (228 mg, 0.728 mmol) was dissolved in THF (2.3 mL), and palladium-fibroin (26 mg) was added. The resultant solution was stirred in a hydrogen atmosphere at room temperature for 20 hours. Palladium-fibroin (22 mg) was added to the mixture, which was stirred in a hydrogen atmosphere at room temperature for 20 hours. The resultant mixture was filtered and concentrated under reduced pressure. The residue was fractionated by silica gel column chromatography to give compound **19e** (215 mg, 0.682 mmol, 94 % yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ 10.22 (s, 1H), 8.26 (d, *J* = 8.7 Hz, 2H), 7.63 (d, *J* = 8.7 Hz, 2H), 7.42 (s, 1H), 6.71 (s, 1H), 5.31 (s, 2H), 3.94 (s, 3H), 2.96 (q, *J* = 7.5 Hz, 2H), 1.23 (t, *J* = 7.5 Hz, 3H).

4.3.2.14. Compound 10 (N-[(E)-[3-methoxy-4-[(4-nitrophenyl)methoxy]phenyl]methylideneamino]-5-nitro-1-benzofuran-2-carboxamide)

Compound **16** (100 mg, 0.388 mmol) was suspended in toluene, and compound **19a** (122 mg, 0.425 mmol) and sodium acetate (36 mg, 0.44 mmol) were added. The resultant solution was stirred at 80 °C for 20 hours. After cooling in air, the solution was filtered, and washed with water and toluene to give compound **10** (114 mg, 0.232 mmol, 60% yield, 100% purity).

^{1}H NMR (300 MHz, DMSO-d6) δ 8.83 (d, *J* = 2.4 Hz, 1H), 8.44 (s, 1H), 8.35 (dd, *J* = 9.0 Hz, *J* = 2.4 Hz, 1H), 7.27 (d, *J* = 8.7 Hz, 2H), 7.95 (d, *J* = 9.0 Hz, 1H), 7.91 (s, 1H), 7.73 (d, *J* = 8.7 Hz, 2H), 7.40 (s, 1H), 7.26 ~ 7.10 (m, 2H), 5.34 (s, 2H), 3.87 (s, 3H).

MS calcd. for $C_{24}H_{18}N_4O_8$ $(M+H)^+$ 491.11, found 490.8.

4.3.2.15. Compound 11 (N-[(E)-[5-methoxy-2-methyl-4-[(4-nitrophenyl)methoxy]phenyl]methylideneamino]-5-nitro-1-benzofuran-2-carboxamide)

Compound **16** (116 mg, 0.450 mmol) was suspended in toluene (3 mL), and compound **19d** (149 mg, 0.495 mmol) and sodium acetate (41 mg, 0.50 mmol) were added. The resultant solution was stirred at room temperature for 1 hour, at 80 °C for 5 hours and at 100 °C for 4 hours. After cooling in air, the solution was filtered and suspended in dimethoxyethane at 80 °C for 1 hour. The mixture was then cooled in air, filtered and washed with dimethoxyethane to give compound **11** (122 mg, 0.242 mmol, 54 % yield, 95.5 % purity).

^{1}H NMR (300 MHz, DMSO-d6) δ 8.83 (s, 1H), 8.76 (s, 1H), 8.36 (d, *J* = 9.0 Hz, 1H), 8.27 (d, *J* = 7.8 Hz, 2H), 7.95 (d, *J* = 9.0 Hz, 1H), 7.90 (s, 1H), 7.72 (d, *J* = 7.8 Hz, 2H), 7.43 (s, 1H), 6.98 (s, 1H), 5.31 (s, 2H), 3.83 (s, 3H), 2.56 ~ 2.37 (s, 3H).

MS calcd. for $C_{25}H_{20}N_4O_8$ $(M+H)^+$ 505.13, found 504.8.

4.3.2.16. Compound 12 (N-[(E)-[2-ethyl-5-methoxy-4-[(4-nitrophenyl)methoxy]phenyl]methylideneamino]-5-nitro-1-benzofuran-2-carboxamide)

Compound **16** (100 mg, 0.388 mmol) was suspended in toluene, and compound **19e** (134 mg, 0.425 mmol) and sodium acetate (35 mg, 0.43 mmol) were added. The resultant solution was stirred at 80 °C for 16 hours. After cooling in air, the obtained solid was filtered and suspended in dimethoxyethane (4 mL). The resultant solution was stirred at 80 °C for 16 hours. After cooling in air, the mixture was filtered and washed with dimethoxyethane to give compound **12** (128 mg, 0.247 mmol, 64% yield, 100 % purity).

^{1}H NMR (300 MHz, DMSO-d6) δ 8.82 (s, 1H), 8.80 (s, 1H), 8.36 (d, *J* = 9.3 Hz, 1H), 8.27 (d, *J* = 7.8 Hz, 2H), 7.95 (d, *J* = 9.0 Hz, 1H), 7.90 (s, 1H), 7.73 (d, *J* = 7.8 Hz, 2H), 7.45 (s, 1H), 6.97 (s, 1H), 5.32 (s, 2H), 3.83 (s, 3H), 2.71 (q, *J* = 7.5 Hz, 2H), 1.61 (t, *J* = 7.5 Hz, 3H). MS calcd. for $C_{26}H_{22}N_4O_8$ $(M+H)^+$ 519.14, found 518.7.

4.3.2.17. Compound 13 (N-[(E)-[3-methoxy-5-nitro-4-[(4-nitrophenyl)methoxy]phenyl]methylideneamino]-5-nitro-1-benzofuran-2-carboxamide)

Compound **16** (100 mg, 0.388mmol) was suspended in toluene (2 ml), and compound **19c** (142 mg, 0.427 mmol) and sodium acetate (35 mg, 0.43 mmol) were added. The resultant solution was stirred at room temperature for 15 hours and at 80 °C for 5 hours. After cooling in air, the obtained solution was filtered and washed with water and toluene to give compound **13** (148 mg, 0.276 mmol, 71 % yield, 95.9 % purity).

^{1}H NMR (300 MHz, DMSO-d6) δ 8.80 (d, J = 2.1 Hz, 1H), 8.53 (s, 1H), 8.33 (dd, J = 9.0 Hz, J = 2.1 Hz, 1H), 8.26 (d, J = 9.0 Hz, 2H), 7.94 (d, J = 9.0 Hz, 1H), 7.92 (s, 1H), 7.78 (s, 1H), 7.73 (s, 1H), 7.70 (d, J = 9.0 Hz, 2H), 5.31 (s, 2H), 4.00 (s, 3H).

MS calcd. for $C_{24}H_{17}N_5O_{10}$ $(M+H)^+$ 536.1, found 535.8.

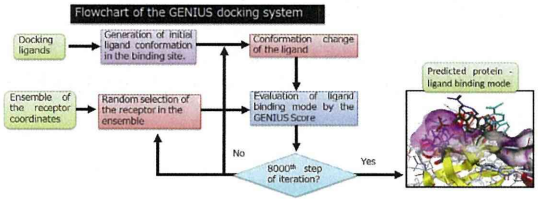
Acknowledgments

This work was supported by Health and Labor Sciences Research Grants for Research on Hepatitis, as well as by Grants-in-Aid for research on hepatitis and BSE, from the Ministry of Health, Labor and Welfare and by a contract research fund for the Program of Funding Research Centers for Emerging and Reemerging Infectious Diseases, from the Ministries of Education, Culture, Sports, Science, and Technology, Japan. This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, and the Targeted Proteins Research Program (TPRP), the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. We thank Dr. Shunta Sasaki, from Structure Based Drug Design Research, for the MS data analysis.

References and notes

1. Alter, H. J.; Purcell, R. H.; Shih, J. W.; Melpolder, J. C.; Houghton, M.; Choo, Q.-L.; Kuo, G. *N. Engl. J. Med.* **1989**, *321*, 1494-1500.
2. Choo, Q. L.; Kuo, G.; Weiner, A. J.; Overby, L. R.; Bradley, D. W.; Houghton, M. *Science* **1989**, *244*, 359-62
3. Qureshi, S.; Qureshi, H.; Hameed, A. *Eur. J. Clin. Microbiol. Infect. Dis.* **2009**, *28*, 1409-1413
4. Alter, M. J. *Hepatology* **1997**, *26*, 62S-5S
5. Fried, M. W.; Shiffman, M. L.; Reddy, K. R.; Smith, C.; Marinos, G.; Gonçales, F. L. Jr.; Häussinger, D.; Diago, M.; Carosi, G.; Dhumeaux, D.; Craxi, A.; Lin, A.; Hoffman, J.; Yu, J. *N. Engl. J. Med.* **2002**, *347(13)*, 975-82
6. Berman, H. M. *Acta Crystallogr., Sect. A.* **2008**, 64, 88-95.
7. Yao, N.; Reichert, P.; Taremi, S. S.; Prosise, W. W.; Weber, P. C. *Structure* **1999**, *15*;7(11), 1353-63.
8. Thibeault, D.; Massariol, M.-J.; Zhao, S.; Welchner, E.; Goudreau, N.; Gingras, R.; Llinàs-Brunet, M.; White, P. W. *Biochemistry* **2009**, *48*, 744-753
9. Liverton, N. J.; Holloway, M. K.; McCauley, J. A.; Rudd, M. T.; Butcher, J. W.; Carroll, S. S.; DiMuzio, J.; Fandozzi, C.; Gilbert, K. F.; Mao, S.-S.; McIntyre, C. J.; Nguyen, K. T.; Romano, J. J.; Stahlhut, M.; Wan, B.-L.; Olsen, D. B.; Vacca, J. P. *J. Am. Chem. Soc.* **2008**, *130*, 4607-4609
10. Berman, K.; Kwo P. Y. *Clin. Liver. Dis.* **2009**, *13*(3), 429-39.
11. Perni, R. B.; Almquist, S. J.; Byrn, R. A.; Chandorkar, G.; Chaturvedi, P. R.; Courtney, L. F.; Decker, C. J.; Dinehart, K.; Gates, C. A.; Harbeson, S. L.; Heiser, A.; Kalkeri, G.; Kolaczkowski, E.; Lin, K.; Luong, Y.-P.; Rao, B. G.; Taylor, W. P.; Thomson, J. A.; Tung, R. D.; Wei, Y.; Kwong, A. D.; Lin, C. *Antimicrob. Agents Chemother.*, **2006**, *50*(3), 899–90
12. Lamarre, D.; Anderson, P. C.; Bailey, M.; Beaulieu, P.; Bolger, G.; Bonneau, P.; Bös, M.; Cameron, D. R.; Cartier, M.; Cordingley, M. G.; Faucher, A.-M.; Goudreau, N.; Kawai, S. H.; Kukolj, G.; Lagacé, L.; LaPlante, S. R.; Narjes, H.; Poupart, M.-A.; Rancourt, J.; Sentjens, R. E.; St George, R.; Simoneau, B.; Steinmann, G.; Thibeault, D.; Tsantrizos, Y. S.; Weldon, S. M.; Yong, C. L.; Llinas-Brunet M. *Nature* **2003**, *13*;426(6963), 186-9
13. Lin, T.-I.; Lenz, O.; Fanning, G.; Verbinnen, T.; Delouvroy, F.; Scholliers, A.; Vermeiren, K.; Rosenquist, A.; Edlund, M.; Samuelsson, B.; Vrang, L.; de Kock, H.; Wigerinck, P.; Raboisson, P.; Simmen, K. *Antimicrob. Agents Chemother.* **2009**, *53*(4), 1377–1385
14. Rajagopalan, R.; Misialek, S.; Stevens, S. K.; Myszka, D. G.; Brandhuber, B. J.; Ballard, J. A.; Andrews, S. W.; Seiwert, S. D.; Kossen, K. *Biochemistry* **2009**, *48*, 2559-2568
15. McCauley, J. A.; McIntyre, C. J.; Rudd, M. T.; Nguyen, K. T.; Romano, J. J.; Butcher, J. W.; Gilbert, K. F.; Bush, K. J.; Holloway, M. K.; Swestock, J.; Wan, B. L.; Carroll, S. S.; DiMuzio, J. M.; Graham, D. J.; Ludmerer, S. W.; Mao, S. S.; Stahlhut, M. W.; Fandozzi, C. M.; Trainor, N.; Olsen, D. B.; Vacca, J. P.; Liverton, N. J. *J. Med. Chem.* **2010**, *25*;53(6), 2443-63.
16. Thompson, A. J. V.; McHutchison, J. G. *J. Viral Hepat.* **2009**, *16*, 377-387
17. Wyles, D. L.; Kaihara, K. A.; Schooley, R. T. *Antimicrob. Agents Chemother.* **2008**, *52*(5) 1862-1864
18. Thompson, C. A. J. V.; McHutchison J. G. *Aliment. Pharmacol. Ther.* **2009**, *29*, 689-705.
19. http://www.natap.org/2004/HCV/113004_01.htm
20. Cubero, M.; Esteban, J. I.; Otero, T.; Sauleda, S.; Bes, M.; Esteban, R.; Guardia, J.; Quer, J. *Virology* **2008**, *370*, 237-245
21. Yi, M.; Tong, X.; Skelton, A.; Chase, R.; Chen, T.; Prongay, A.; Bogen, S. L.; Saksena, A. K.; Njoroge, F. G.; Veselenak, R. L.; Pyles, R. B.; Bourne, N.; Malcolm, B. A.; Lemon, S. M. *J. Biol Chem.* **2006**, 281, 12, 8205-8215
22. Ismail, N. S. M.; Hattori, M. *Bioorg. Med. Chem.* **2011**, *19*, 374-383
23. Ontoria, J. M., Di Marco, S., Conte, I., Di Francesco, M. E., Gardelli, C., Koch, U., Matassa, V. G., Poma, M., Steinkühler, C., Volpari, C.; Harper, S. *J. Med. Chem.* **2004**, *47*, 6443-6446
24. Barbato, G.; Cicero, D. O.; Cordier, F.; Narjes, F.; Gerlach, B.; Sambucini, S.; Grzesiek, S.; Matassa, V. G.; De Francesco, R.; Bazzo, R. *EMBO J.* **2000**, *19*, 6, 1195-1206
25. Cummings, M. D.; Lindberg, J.; Lin, T.-I.; de Kock, H.; Lenz, O.; Lilja, E.; Felländer, S.; Baraznenok, V.; Nyström, S.; Nilsson, M.; Vrang, L.; Edlund, M.; Rosenquist, A.; Samuelsson, B.; Raboisson, P.; Simmen, K. *Angew. Chem., Int. Ed.* **2010** *22*;49(9):1652-5.
26. Hagel, M.; Niu, D.; Martin, T. S.; Sheets, M. P.; Qiao, L.; Bernard, H.; Karp, R. M.; Zhu, Z.; Labenski, M. T.; Chaturvedi, P.; Nacht, M.; Westlin, W. F.; Petter, R. C.; Singh, J. *Nat. Chem. Biol.* **2011**, **7**, 22-24.
27. Barbato, G.; Cicero, D. O.; Nardi, M. C.; Steinkühler, C.; Cortese, R.; De Francesco, R.; Bazzo, R. *J. Mol. Biol.* **1999**, 289, 371-384.
28. Gallo, M.; Pennestri, M.; Bottomley, M. J.; Barbato, G.; Eliseo, T.; Paci, M.; Narjes, F.; De Francesco, R.; Summa, V.; Koch, U.; Bazzo, R.; Cicero, D. O. *J. Mol. Biol.* **2009**, *385*, 1142-1155.
29. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639-1662
30. Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411-428
31. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.***1997**, *267*, 727-748
32. Koska, J.; Spassov, V. Z.; Maynard, A. J.; Yan, L.; Austin, N.; Flook, P. K.; Venkatachalam, C. M. *J. Chem. Inf. Model.* **2008**, *48*, 1965-1973
33. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739-1749
34. Dunbrack, R. L. Jr; Karplus, M. *J. Mol. Biol.* **1993**, *230*, 543–574
35. Meiler, J.; Baker, D. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 538–548
36. https://scifinder.cas.org.
37. Chevaliez, S.; Jean-Michel Pawlotsky, J.-M.; Tan, S.-L., Ed.; Norfolk (UK): Horizon Bioscience; 2006. Chapter 1. HCV Genome and Life Cycle; http://www.ncbi.nlm.nih.gov/books/NBK1613/
38. Schneider, G. *Nat. Rev. Drug Discovery* **2010**, *9*, 273-276
39. Bender, A.; Glen, R. C. *J. Chem. Inf. Model.* **2005**, 45, 1369-1375

40. Umeyama, H.; Watanabe, Y.; Arai, R., Japan Patent 4314128, 2009
41. Umeyama, H.; Watanabe, Y.; Arai, R., Japan Patent 4314206, 2009;
42. Holm, L.; Sander, C. *J. Mol. Biol.* **1993** *5; 233(1)*, 123-38.
43. Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.*, **2000**, *21;12,* 1049-1074.
44. Ma, B.; Lii, J.-H.; Allinger, N. L. *J. Comput. Chem.* **2000**, *21;10,* 813-825
45. Symyx Technologies, Inc. Corporate Address: 3100 Central Expressway, Santa Clara, CA 95051
46. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **2001**, *46 (1-3)*, 3-26.
47. Dimasi, N.; Pasquo, A.; Martin, F.; Di Marco, S.; Steinkühler, C.; Cortese, R.; Sollazzo, M. *Protein Eng., Des. Sel.* **1998**, *11(12)*, 1257-65.
48. Kakiuchi, N.; Nishikawa, S.; Hattori, M.; Shimotohno, K. *J. Virol. Methods.* **1999**, *80*, 77–84
49. Tanabe, Y.; Sakamoto, N.; Enomoto, N.; Kurosaki, M.; Ueda, E.; Maekawa, S.; Yamashiro, T.; Nakagawa, M.; Chen, C.-H.; Kanazawa, N.; Kakinuma, S. ; Watanabe, M. *J. Infect. Dis.* **2004**, *189*, 1129-39
50. Yokota, T.; Sakamoto, N.; Enomoto, N.; Tanabe, Y.; Miyagishi, M.; Maekawa, S.; Yi, L.; Kurosaki, M.; Taira, K.; Watanabe, M.; Mizusawa, H. *EMBO Rep.* 2003, **4,** 602-8.



Will There Be an HCV Meeting in 2020? Summary of the 17th International Meeting on Hepatitis C Virus and Related Viruses

TAKAJI WAKITA,[*] TETSURO SUZUKI,[‡] MATTHEW J. EVANS,[§] KUNITADA SHIMOTOHNO,[‖] KAZUAKI CHAYAMA,[¶] YOSHIHARU MATSUURA,[#] MAKOTO HIJIKATA,[**] KOHJI MORIISHI,[‡‡] TSUKASA SEYA,[§§] NOBUYUKI ENOMOTO,[¶¶] KAZUHIKO KOIKE,[##] NOBUYUKI KATO,[***] TATSUYA KANTO,[#] and HAK HOTTA[‡‡‡]

National Institute of Infectious Diseases, Japan; [‡]Hamamatsu University School of Medicine, Japan; [§]Mount Sinai School of Medicine, New York; [‖]Chiba Institute of Technology, Japan; [¶]Hiroshima University, Japan; [#]Osaka University, Japan; [**]Kyoto University, Japan; [¶¶]University of Yamanashi, Japan; [§§]Hokkaido University, Japan; [##]The University of Tokyo, Japan; [***]Okayama University, Japan; [‡‡‡]Kobe University, Japan

Hepatitis C virus (HCV), which was discovered in 1989, is a major etiologic agent in human liver disease. Approximately 130 million people, or 2% of the population, worldwide are infected. The 17th International Meeting on Hepatitis C Virus and Related Viruses was held September 10–14, 2010, in Yokohama, Japan. The meeting was attended by almost 700 scientists from all over the world who are interested in the fundamental aspects of the molecular virology, immunology, pathogenesis, prevention, and treatment of HCV infection. Two special opening lectures given by Masaaki Komatsu and Takashi Gojobori focused attention on the related research fields of autophagy and genome biology, respectively. In the subsequent sessions, the latest research, original studies, and controversies were presented in 9 keynote lectures, 82 oral presentations, and 329 poster presentations.

Viral Entry

The opening scientific session of this meeting focused on the viral host cell entry processes. Thomas Baumert presented the keynote lecture, which included an overview of the HCV cell entry process and recent advances at his laboratory. These included the finding that HCV variants that reinfect the liver after transplantation demonstrate more efficient cell entry and are less susceptible to neutralization by host antibodies. He also described the isolation of monoclonal antibodies against claudin-1 that do not inhibit either extracellular or direct cell-to-cell HCV transfer.

Alexander Ploss described the establishment of a mouse model for studying HCV cell entry. They utilized an HCV cell culture virus (HCVcc) expressing recombinase and transgenic mice bearing a recombinase-activatable fluorescent protein. Bioluminescent imaging indicated that only mice transduced with CD81 and occludin supported HCVcc entry. The presence of an intact immune system in these animals makes it particularly important for the testing of HCV vaccine candidates. Danyelle N. Martin described a role for transferrin receptor 1 (TfR1) in mediating HCV cell entry. The inhibition of HCV entry with TfR1 antibodies and silencing, suggest this factor should be added to the growing list of cellular proteins required for HCV cell entry. Joachim Lupberger presented results from a study showing an essential role for the epidermal growth factor receptor (EGFR) in HCV cell entry. He found that EGFR is required for both mediating the interactions between two other entry factors, CD81 and CLDN1, and catalyzing the fusion activity of viral glycoproteins.

Translation/Replication

Volker Lohmann began the session by describing what is known of the functions of viral nonstructural proteins and their associated host cellular factors in viral translation and replication. He included an overview of viral isolates and model systems currently used, and presented data addressing the mechanisms for efficient replication of the JFH-1 isolate.

Several reports have focused on the molecular basis of the architecture and composition of membrane-associated sites for HCV replication, which often induce membrane alterations, such as the so-called membranous web. Brenno Wolk demonstrated that NS4B is sufficient to direct all nonstructural proteins into the viral replication complex compartment, and that intragenotype-specific interactions are required for NS4B-dependent recruitment of NS5A. Ines Romero-Brey showed that the membranous web predominantly contains double-membrane vesicles with various diameters. These vesicle structures were connected to the endoplasmic reticulum (ER) through funnel-like structures.

Several DDX DEAD-box RNA helicases were identified as host factors associated with HCV replication. Yasuo Ariumi presented the cross-talk of HCV with DDX proteins and the role of distinct DDX proteins in viral replication. Tetsuro Shimakami and Selena M. Sagan reported the importance of miR-122 to not only enhance IRES-mediated translation, but stabilize positive-strand HCV RNA by binding to its 5′ extremity. Enzymatic activity of host phosphatidyl-inositol-4 kinase III alpha was shown to be critically involved in HCV replication and the activity is regulated by HCV NS5A (Simon Reiss). Nam-Joon Cho reconstituted a functionally active full-

0016-5085/$36.00
doi:10.1053/j.gastro.2011.05.027

length HCV polymerase on a biomimetic membrane platform. Deborah Harrus found that guanosine triphosphate specifically stimulates the initial step of de novo initiation by stimulating transition of newly formed linker primer.

Assembly and Release

In the keynote lecture, Guangxiang G. Luo presented an overview of particle assembly and release, and the impact of apolipoprotein (Apo) E in the entry and assembly of HCV. He demonstrated the inhibition of HCVcc entry by treatment with anti-ApoE antibody and the direct interaction of ApoE with NS5A.

Ann L. Wozniak showed an important role for p7 in the production of infectious particles. Their data suggest that p7 stimulates virus production through the alkalization of intracellular vesicles. Ophelia Granio showed that both p7 and NS2 are required for the recruitment of core from lipid droplets (LDs) to ER. Costin-Ioan I. Popescu showed that NS2 accumulated in dotted structures in the ER in juxtaposition with Core and LDs. They concluded that cross-talk among Core, E1, E2, p7, and NS2 was essential for virion assembly. Vlastimil Jirasko demonstrated point mutations in the transmembrane regions of NS2 impaired the particle production and suggested that NS2 serves as a platform of viral and cellular proteins that coordinates HCV assembly. Qisheng Li identified the proviral function of IKKα by genome wide siRNA screening. IKKα regulates lipid metabolism and biogenesis of LDs and may enhance production of virus particles. The very low-density lipoproteins are secreted via a Golgi-dependent pathway. Bryan R. Bishe demonstrated the important role of phosphatidylinositol-4-phosphate and its interacting protein GOLPH3 in HCV secretion in the trans-Golgi network. Roland Remenyi showed 3-dimensional visualization of the HCV life cycle in cultured cells by electron tomography. They detected virus-like particles at various cytoplasmic locations. Viral particles in the proximity of LDs and within sponge-like inclusion were observed. These results provide ultrastructural visualization of putative assembly sites close to LDs.

Host Factors

In the invited lecture, Sara Cherry presented an overview of high-throughput screening toward the identification of host factors required for viral infection.

The contribution of autophagy to the HCV life cycle was also presented in this section, most notably, host factors linked with lipids. Tsubasa Munakata showed that the fatty acid synthase is required for efficient HCV replication. They also suggested the importance of palmitate for HCV replication. Samantha L. Blackham presented both the thioredoxin-interacting protein and the peroxisome proliferator activated receptor-α have significant effects on HCV replication. The host factors functioning on infectious HCV particle production were also reported. Takayuki Hishiki demonstrated the isoform dependent binding affinities of ApoE for low-density lipoprotein receptors and they affect infectivity of HCV. Laurent Chatel-Chaix found that Y-box binding protein interacted with HCV NS3 protein and viral RNA and was relocalized from nucleocytoplasmics site to the core-containing surface of LDs. Mohsan Saeed reported that the ER-associated degradation pathway was activated by HCV infection in a viral envelope protein-dependent manner. Po-Yuan Ke showed that HCV infection induces the unfolded protein response and activates the autophagic pathway. They proposed that autophagy contributes to the suppression of HCV in an autolysosome formation-dependent manner. Hiroto Kambara did not find any effects on HCV replication by inhibition of autophagosome formation in replicon cells. They proposed a role for autophagy induced by HCV infection to avoid the generation of vacuolation harmful to cell survival. Qisheng Li reported the network map of cellular pathways and machineries that are associated with HCV life cycle.

Very low-density lipoprotein is now considered to be one of a component of HCV particles. LDs are composed of fatty acid, triglyceride, and cholesterol, surrounded by several types of lipoproteins. In addition, Daniel J. Felmlee reported that chylomicron-associated viruses may be generated by virion association while in the vascular compartment. Francois Jean showed that the serine protease inhibitor protein Spn4A was modified to be directed to Site-1 protease specifically and was introduced into adenovirus vector to inhibit cholesterol and fatty acid syntheses for down-regulation of HCV propagation. The modified serpin could suppress Site-1 protease activity, reduce the LD, and block HCVcc infection. Nicolas Menzel tried to identify novel cellular factors involved in HCV assembly and release and found ERK inhibitor and cytosolic phospholipase A2 (cPLA2) inhibitor reduce viral production. cPLA2 inhibitor also reduced the amount of LD-associated core and supernatant ApoB/E. cPLA2 may be crucial for assembly of infectious HCV particles, possibly through participating in the formation of lipoproteins. Kohji Moriishi reported that the proteasome activator PA28γ participates in HCV propagation. PA28γ may participate in the propagation of HCV by regulating the degradation of Core in both ubiquitin-dependent and -independent manners. NS5A is regulated by phosphorylation of several host protein kinases. Takahiro Masaki identified 79 serine threonine protein kinases that were tightly associated with NS5A. Two of these may regulate the production of viral particles and/or viral replication.

Innate Immunity

The early phase of host defense against viral infection has largely been delineated based on recent advances in innate immunity. In the invited lecture, Manoj N. Krishnan introduced his comprehensive study on the Toll-like receptor 3–TRIF (TICAM-1) pathway. Using RNAi and polyI:C, he screened the genes specifically up-regulated via the TRIF (TICAM-1) pathway. He expected that some viral infections are selectively blocked by the IPS-1 pathway, while others are blocked by the TRIF pathway.

Michael Gale, Jr., identified IFITM1 inhibits HCV infection. IFITM1 assembles with CD81 and tranlocates to the tight junction. This translocation of CD81 hampers the receptor function of CD81. They also discovered a novel pathway for ISGF3 activation. A non-receptor type tyrosine kinase-1 triggers activation of ISGF3 independent from the classical IFNAR pathway. IP-10 is a chemokine and is a negative predictor for pegylated interferon (IFN)/ribavirin therapy. Matthew L. Albert indicated that there is a 2-amino-acid–deleted form of IP-10 that serves as an antagonist for intact IP-10, and this form abrogates an early virologic response. As this IP-10 truncation is mediated by dipeptidylpeptidase IV, they believed that dipeptidylpeptidase IV is a novel therapeutic target for HCV patients during IFN therapy. Joo Chun Yoon suggested that activation of natural killer cells is inhibited by HCV-infected hepatocytes. They claimed that the early phases of HCV infection may be established through the failure of virus-inducible natural killer cell activation. Shin-ichiro Nakagawa reported that polyI:C induces both type I IFN and IFN-λ in human hepatocytes. The antiviral effect appears to parallel the induction of IFN-λ. This, together with the report by Emmanuel Thomas, suggests that the IFN-λ system is activated in HCV infected hepatocytes.

Adaptive Immunity

In a keynote lecture, Robert Thimme summarized the mechanisms of HCV-induced T-cell dysfunction. Multifaceted factors contribute to the hyporesponsiveness of T cells, including viral mutations, primary T-cell failure, lack of support from dendritic cells, expression of inhibitory molecules on T cells, and abundance of regulatory T cells (Tregs). Whether the ability of HCV-specific CTLs is comparable with that of CTLs having other specificities remains controversial. Bianca Seigel showed that HCV-specific CTLs are functionally impaired when compared with other CTLs, irrespective of their expression of inhibitory receptors or differentiation stages. CD161 is a C-type lectin that is expressed in HCV-specific $CD8^+$ T cells with tissue homing phenotype. Vicki M. Fleming found that $CD4^+CD161^+$ T cells produce large amounts of inflammatory cytokines and accumulate in the liver, where they are thought to exert pro-inflammatory roles. Naruyasu Kakita reported that certain adaptive Tregs, known as interleukin (IL)-10–producing type 1 Tregs, are increased in HCV-positive hepatocellular carcinoma patients, and their significance in hepatocellular carcinoma was greater than that of natural Tregs. Even in patients who have attained a sustained virologic response, trace amounts of HCV RNA are sporadically detectable in plasma. Barbara Rehermann reported the inoculation studies of such plasma. Residual HCV RNA in patients was able to infect chimpanzees and induced broad, HCV-specific T-cell responses. HCV RNA levels continued to be high when T-cell responses declined, suggesting that such HCV remains transmissible as hepatotropic pathogens.

Pathogenesis

In the invited lecture, Michael Diamond presented new mechanisms for West Nile virus immune evasion via 2'O methylation of viral RNA to subvert host innate immunity.

Genome-wide analysis of quantitative data (transcriptomics, proteomics, and metabolomics) facilitates systems biology analysis of HCV infection. Deborah L. Diamond analyzed the pathways involved in the progression of chronic hepatitis, namely, fibrosis and carcinogenesis, and found that molecules relating to cell metabolism including fatty acid oxidation enzymes and antioxidant systems may be master regulators of liver disease progression in HCV infection. HCV core protein has been shown to play a key role in the development of steatosis in HCV infected liver, especially in patients with genotype 3a HCV infection. Sophie Clement-Leboube showed that PTEN expression was down-regulated in the HCV infected liver. Analysis of lipo-viral-particle from hepatitis C patients by Olivier Diaz revealed that empty lipo-viral-particle lacking HCV RNA outnumbers those with RNA. The presence of virus-modified lipoproteins in HCV-infected patients may play a role in the pathogenesis of hepatitis C. Massimiliano Pagani used serum miRNA signatures to monitor liver disease in HCV infection and found miRNome candidates that are specific for HCV disease progression. Shuhei Taguwa showed that Con1 replicon induces incomplete autophagy through the dysfunction of autolysosomal acidification, which results in the secretion of immature cathepsin B in cells. Because the secretion of the protein is enhanced in many types of tumors, this observation may be associated with the pathogenesis of liver tumorigenesis in HCV infection.

The existence of extrahepatic manifestations is another issue of interest. Essential mixed cryoglobulinemia, membranoproliferative glomerulonephritis, and Sjögren syndrome are conditions that have been shown to correlate

with HCV infection. Nicola A. Fletcher reported that brain microvascular endothelial cells express all the recognized entry factors for HCV, and brain microvascular endothelial cells actually support infection by HCVpp and HCVcc. This suggests potential disorders of the central nervous system in HCV infection.

Treatment

In the keynote lecture, Masashi Mizokami presented "Genome-wide association study and its application for HCV treatment." He emphasized that the functional relevance of IL-28B single nucleotide polymorphisms should be elucidated to further advance the progress of research on the mechanisms of chronic HCV infection and treatment.

Yasuhiro Asahina presented that genetic variation in IL-28B is associated with gene expression involving innate immunity. Minor alleles of IL-28B, as well as a higher RIG-I/IPS-1 ratio are associated with null viral response. Martin Laggins correlated IL-28B genetic variation with pretreatment levels of IP-10 and HCV RNA throughout therapy. The favorable genetic variation of IL-28B single nucleotide polymorphisms (major allele) was significantly associated with lower baseline IP-10. Masao Honda revealed that hepatic IFN-stimulated genes (ISGs) are associated with genetic variation in IL-28B and the outcome of IFN therapy for chronic hepatitis C using microarray gene expression profiling of the biopsied liver samples. Multivariate logistic regression analysis showed that ISGs, fibrosis stage, and ISDR mutations were strongly associated with viral response. Hepatic ISGs were associated with the IL-28B polymorphism and expression was significantly higher in patients with the minor genotype than in those with the major genotype. Takashi Motomura also analyzed ISG expression using liver transplantation samples. Expression of ISGs in recipients' liver carrying the minor allele of IL-28B was significantly up-regulated when compared with the major allele. Surprisingly, IFN sensitivity for recurrent hepatitis C after liver transplantation is influenced by IL-28B genetic variation not only in recipients, but also in donors.

Drug Development

This session opened with a keynote lecture by Raffaele De Francesco describing the current state of drug development for patients with chronic hepatitis C. Because of the rapid development of NS3/4A, NS5A, and NS5B inhibitors, he finally presented the hopeful message "Will there be an HCV meeting in 2020?".

Lotte Coelmont characterized an NS5A D320E variant showing low-level resistance to DEB025, a cyclophilin (Cyp)-binding molecule. This study suggests that DEB025 presents a high barrier to resistance, and that D320E confers low-level resistance to DEB025 by reducing the need for CypA-dependent isomerization of NS5A. Paul Targett-Adams reported that NS5A inhibitors stimulated redistribution of NS5A from the ER to ring-like structures in the cytoplasm, and disrupted colocalization with NS5B. This study suggests that NS5A inhibitors perturb formation of new replication complexes rather than acting on preformed complexes. Luis M. Schang developed a family of small synthetic rigid amphiphiles with large hydrophilic heads and small, planar and rigid hydrophobic tails, called RAFIs (rigid amphipathic fusion inhibitors), which inhibit the infectivity of enveloped virions including HCV. Emmanuel Thomas screened host genes involving the anti-HCV activity of ribavirin. Among 64 host genes, several candidate genes were identified as host factors involving ribavirin's anti-HCV activity. Interestingly, silencing of the *ITPA* gene increased the anti-HCV activity of ribavirin. Pablo Gastaminza identified a novel family of 1,2-diamines as an anti-HCV reagent from a chemical library. The analysis of ~300 derivatives identified several compounds with enhanced potency and low cytotoxicity.

Vaccines/Epidemiology

HCV therapeutic vaccines are aimed to induce effective T-cell responses. Marianne Mikkelsen reported that vaccination of mice with recombinant adenovirus expressing HCV NS3 fused to the MHC class II chaperon protein invariant chain significantly enhanced NS3 specific $CD8^+$ T-cell responses, and protected mice against NS3-expressing vaccinia virus challenge. This vaccination induced polyfunctional $CD8^+$ memory T cells. Lars Frelin aimed to restore immunologic function through vaccination in a transgenic mouse model with impaired HCV-specific T-cell responses owing to a persistent presence of hepatic HCV NS3/4A antigens. They found that heterologous sequences improved activation and expansion of NS3/4A-specific T cells in a wild-type host, as well as in a tolerant NS3/4A-transgenic mouse model. The authors also suggested an important role for Tregs in the impaired HCV-specific T-cell responses.

Livia M.G. Rossi examined antibody cross-immunoreactivity against different HVR1 variants to identify antigens with a possible application of HCV vaccine development. The authors identified a small set of HVR1 variants that cross-immunoreacted with a large number of HVR1 peptides, thus suggesting their potential use in the development of HCV vaccine candidates.

Conclusion

HCV2010 in Yokohama was successful and contributed to the progress of research in the field. HCV infection remains one of the most serious worldwide health problems. The goals of this symposium were to

increase the scientific understanding of this virus and gain insights applicable to future efforts to control its infection. From this point of view, we gained further fundamental understanding about HCV at the meeting. The discovery of IL-28B as a new host factor involved in HCV treatment and pathogenesis had a major impact on HCV research. New treatment advances have been made in recent years and will continue in the near future. We would like to conclude that this meeting was successful in providing opportunities for exchanging up-to-date information and international collaboration. The next meeting will take place in Seattle, Washington, from September 8–12, 2011 (http://www.hcv2011.org/).

Reprint requests

Address requests for reprints to: Takaji Wakita, MD, PhD, Department of Virology II, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan. e-mail: wakita@nih.go.jp; fax 81-3-5285-1161.

Conflicts of interest

The authors disclose no conflicts.

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.

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

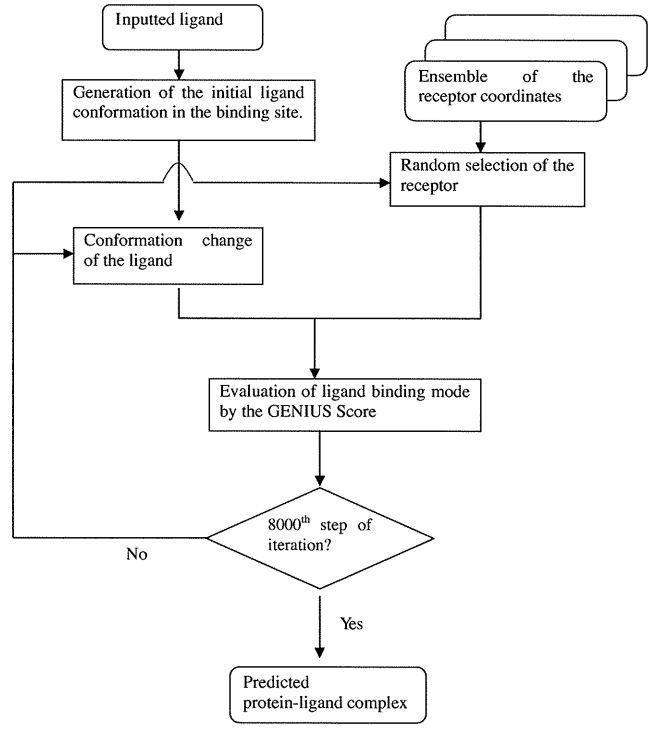


Figure 1. Flowchart of the GENIUS docking system.

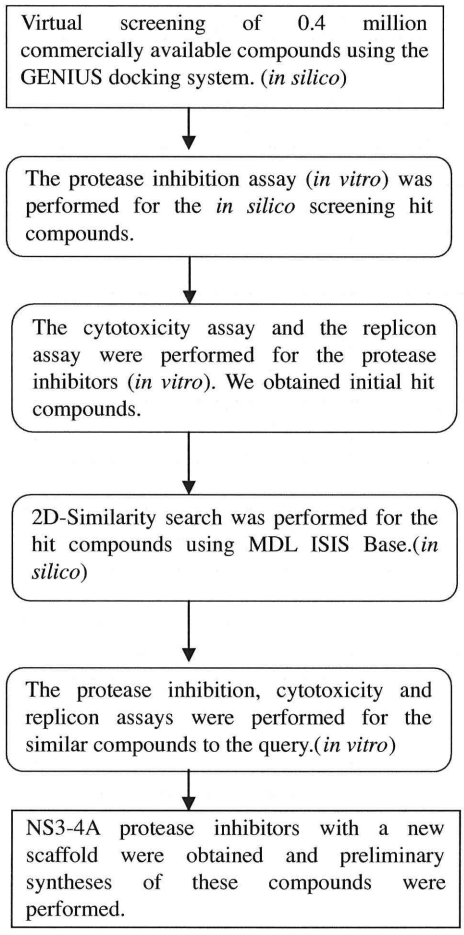
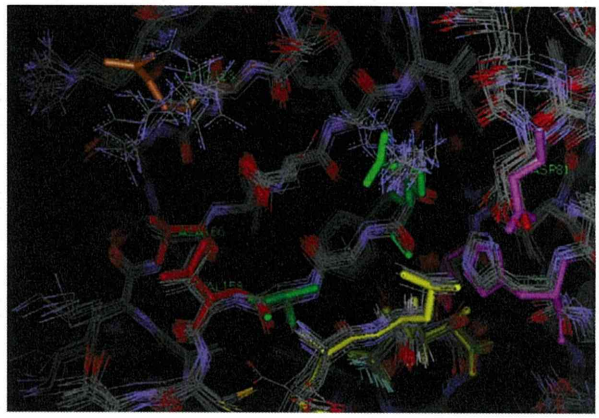


Figure 2. Flowchart of HCV NS3-4A in silico and in vitro.



(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-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.

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

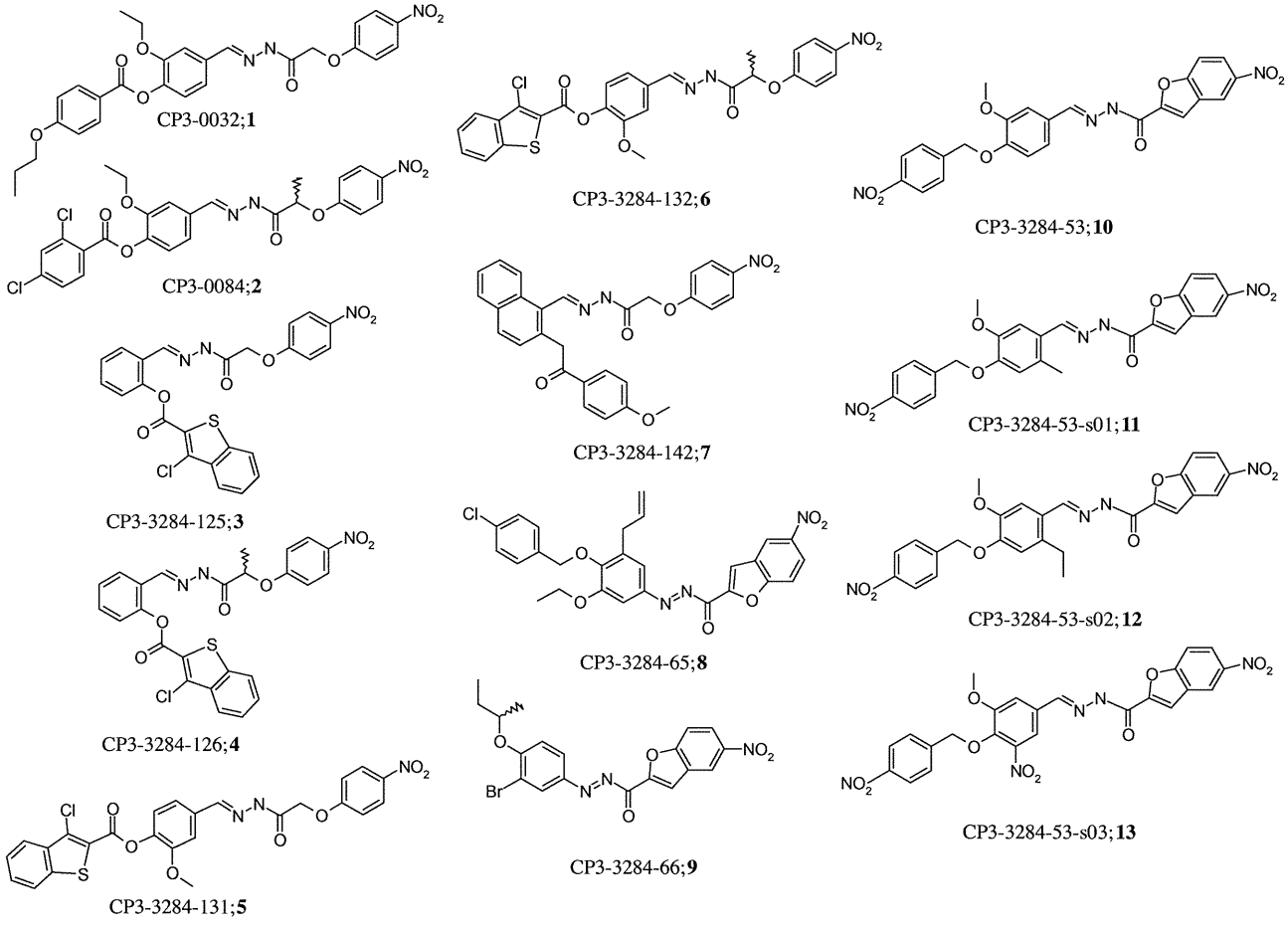
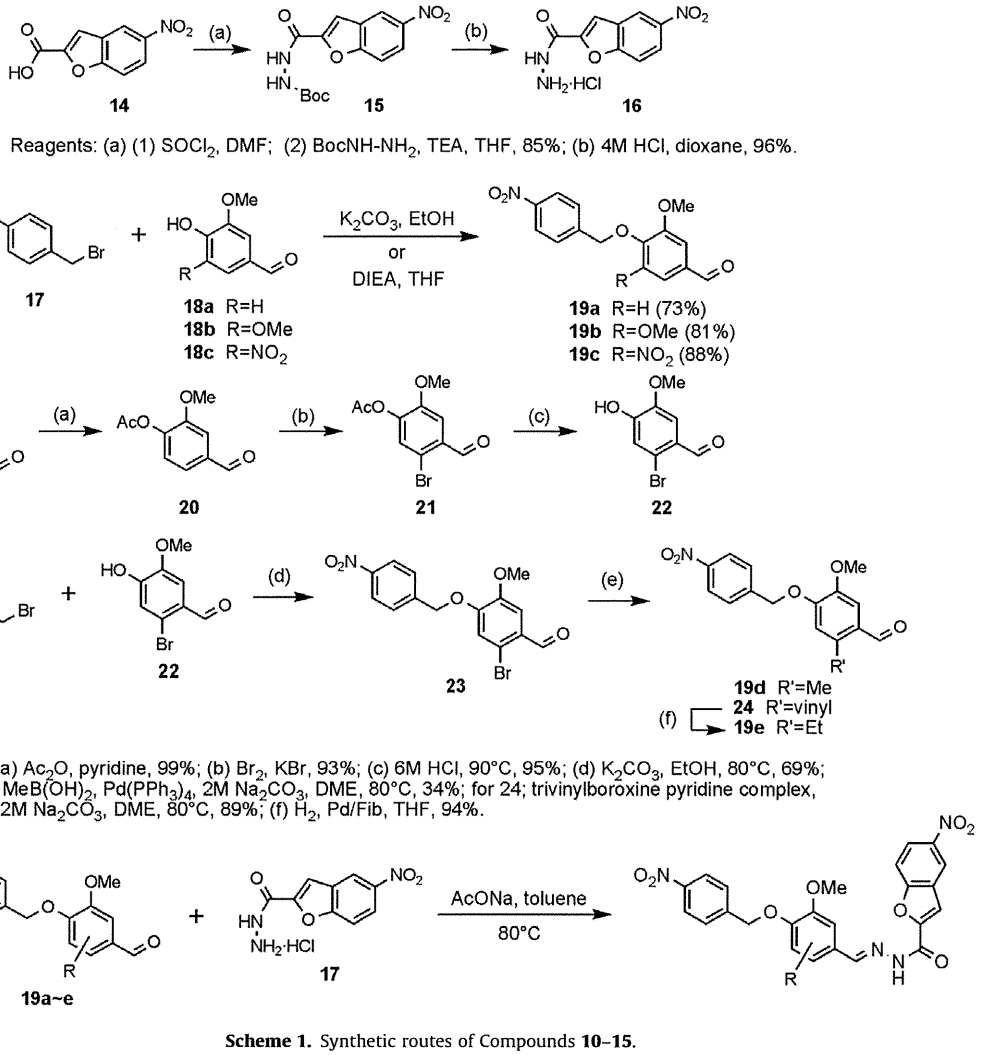


Figure 4. 2D Structures of the discovered protease inhibitors.

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.).



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.

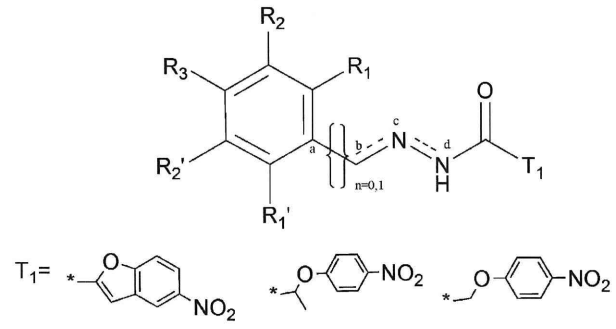
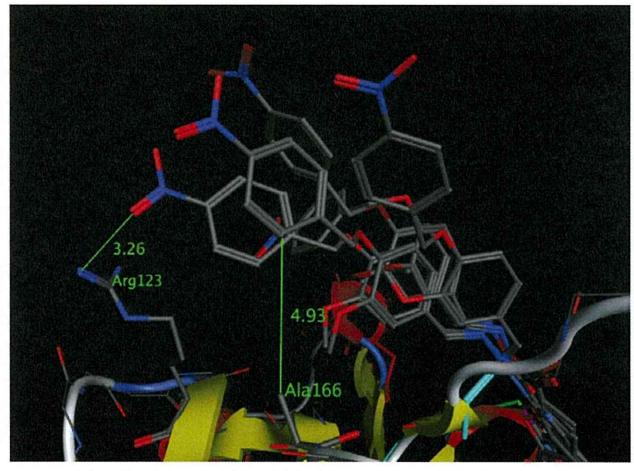


Figure 6. The common scaffold among CP3-3284 series. T1 means substructures in the CP3-3284 series.

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.

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 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 Section 4) 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 (Fig. 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

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 three times better than that with the fixed receptor (Fig. 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**.

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

ID	Rank	SD.
CP3-0032	2163	636
CP3-0084	4410	1336
CP3-3284-53	1089	445
CP3-3284-65	1542	604
CP3-3284-66	2291	170
CP3-3284-125	12260	178
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
BILN-2061	23876	984
ITMN-191	18140	962
MK-7009	22402	3966
TMC-435	19032	1788
VX-950	N/A	N/A
SCH-503034	N/A	N/A

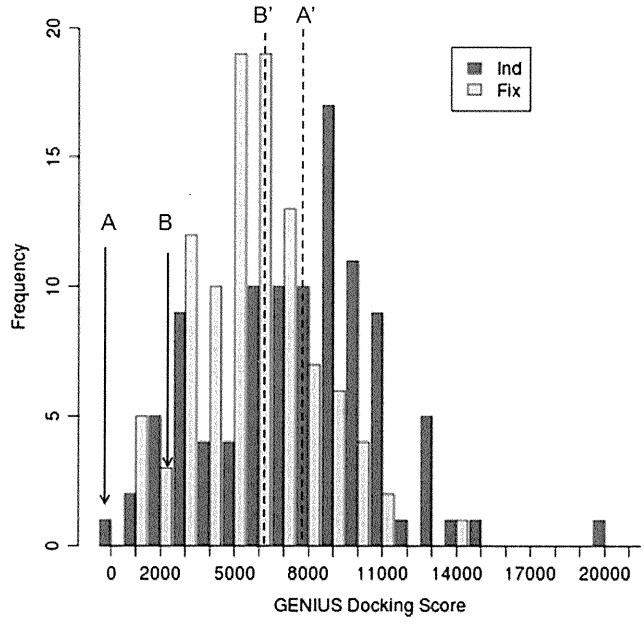


Figure 7. Distributions (histogram) of the docking scores of the decoy compounds and CP3-3284-53, using the induced-fit and the fixed receptor modes; 'Ind' means the trial using the induced-fit receptor. 'Fix' means the trial using the fixed receptor. A: The average score of CP3-3284-53 in the induced-fit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

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 3a. 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 3b. 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.

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.

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 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 Å, 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:

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

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

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

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 two 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.

$$U_{optimum} = U_{sar} + U_{hydrogenbond} + U_{hydrophobic} + U_{stacking} + U_{collision} + U_{ligand-internal} \tag{4}$$

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:

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

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

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:

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

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

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. θ is the angle of the hydrogen bond, in degrees (Fig. 8). If the hydrogen bonding angle exceeds 30 degrees, then the score rapidly worsens.

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:

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

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

where N and M are the numbers of atoms that could form hydrophobic interactions in the ligand and the receptor, respectively (cut-off: 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 (Fig. 9):

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

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

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

$$\theta_{boundary} = |1.0 - \Theta| \tag{14}$$

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

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.

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:

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

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

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

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. ε (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.

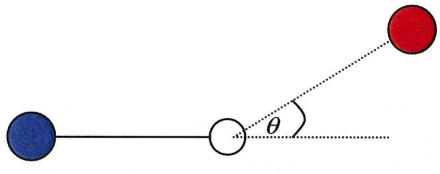


Figure 8. Definition of the hydrogen bond interaction. The red circle is the acceptor atom, the blue circle is the acceptor atom, and the white circle is a hydrogen atom. θ is the hydrogen bond angle.

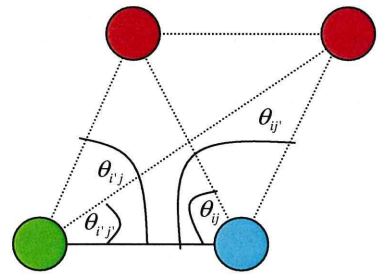


Figure 9. Definition of the stacking interaction. The cyan circle is the i-th atom in the aromatic ring of the receptor. The green circle means centroid of the aromatic ring including the i-th atom. The red circle is the j-th atom in the aromatic ring of the ligand.

4.1.5.6. 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 intramolecular collisions, a very strong restraint was added by the following formula:

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

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

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

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