BioSystems 99 (2010) 70–78

Contents lists available at ScienceDirect

BioSystems

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

Reproducibility and usability of chronic virus infection model using agent-based simulation; comparing with a mathematical model

Jun Itakura[a,*], Masayuki Kurosaki[a], Yoshie Itakura[a], Sinya Maekawa[b], Yasuhiro Asahina[a], Namiki Izumi[a], Nobuyuki Enomoto[b]

[a] Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, 1-26-1 Kyonan-cho, Musashino-shi, Tokyo 180-8610, Japan
[b] First Department of Internal Medicine, Faculty of Medicine, University of Yamanashi, 1110, Shimogatou, Chuou-shi, Yamanashi 409-3898, Japan

ARTICLE INFO

Article history:
Received 30 June 2009
Received in revised form 27 August 2009
Accepted 6 September 2009

Keywords:
Agent-based model
Virus infectious disease

ABSTRACT

We created agent-based models that visually simulate conditions of chronic viral infections using two software. The results from two models were consistent, when they have same parameters during the actual simulation. The simulation results comprise a transient phase and an equilibrium phase, and unlike the mathematical model, virus count transit smoothly to the equilibrium phase without overshooting which correlates with actual biology in vivo of certain viruses. We investigated the effects caused by varying all the parameters included in concept; increasing virus lifespan, uninfected cell lifespan, uninfected cell regeneration rate, virus production count from infected cells, and infection rate had positive effects to the virus count during the equilibrium period, whereas increasing the latent period, the lifespan-shortening ratio for infected cells, and the cell cycle speed had negative effects. Virus count at the start did not influence the equilibrium conditions, but it influenced the infection development rate. The space size had no intrinsic effect on the equilibrium period, but virus count maximized when the virus moving speed was twice the space size. These agent-based simulation models reproducibly provide a visual representation of the disease, and enable a simulation that encompasses parameters those are difficult to account for in a mathematical model.

1. Introduction

All viruses need hosts as a basis for their life. When a virus enters the host body, it invades cells and uses both its own enzymes and those of the host cells to replicate. Host cells infected by viruses launch a self-defense system known as the innate immune system (See and Wark, 2008; Naniche, 2009), which inhibits viral replication and uses the human leukocyte antigen system and cytokines to elicit an immune response. Immune cells that have received signals from host cells activate other immune cells, neutralize viruses in the serum by means of antibodies, and prevent the virus from replicating and proliferating by destroying or curing host cells. Viral infection is a disorder based on the interactions between viruses and cells.

The power relationship between these agents changes along with the progression of the disease. In the very early stages of infection, as the host defense mechanisms are immature, the virus has the ability to overwhelm the host cells, actively replicate, and proliferate. Subsequently, as the capacity of the immune system improves, the speed of viral proliferation drops and the virus count reaches a peak. Infected host cells begin to be disrupted by the immune system or virus particles, and symptoms appear as a result. If the immune system is stronger than the virus, then the viral counts decline, and, in transient viral disorders, the virus is finally eliminated and the host recovers. In chronic viral disorders, however, the power relationship between the virus and host cells reaches equilibrium, and a long-term power balance is maintained with the virus count reaching a plateau.

Mathematical models have been proposed to study the dynamics of such viral disorders, and are regarded as being of value in understanding this phenomenon (Ho et al., 1995; Nowak et al., 1996; Neumann et al., 1998). However, these models are difficult to understand for clinicians, and their applicability is somewhat limited in everyday practice. In clinical research, measurements of viral dynamics in patients for short duration have been made for human

Abbreviations: HIV, human immunodeficiency virus; HBV, hepatitis B virus; HCV, hepatitis C virus.

* Corresponding author. Tel.: +81 422 32 3111; fax: +81 422 32 9551.
E-mail address: jitakura@musashino.jrc.or.jp (J. Itakura).

0303-2647/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2009.09.001

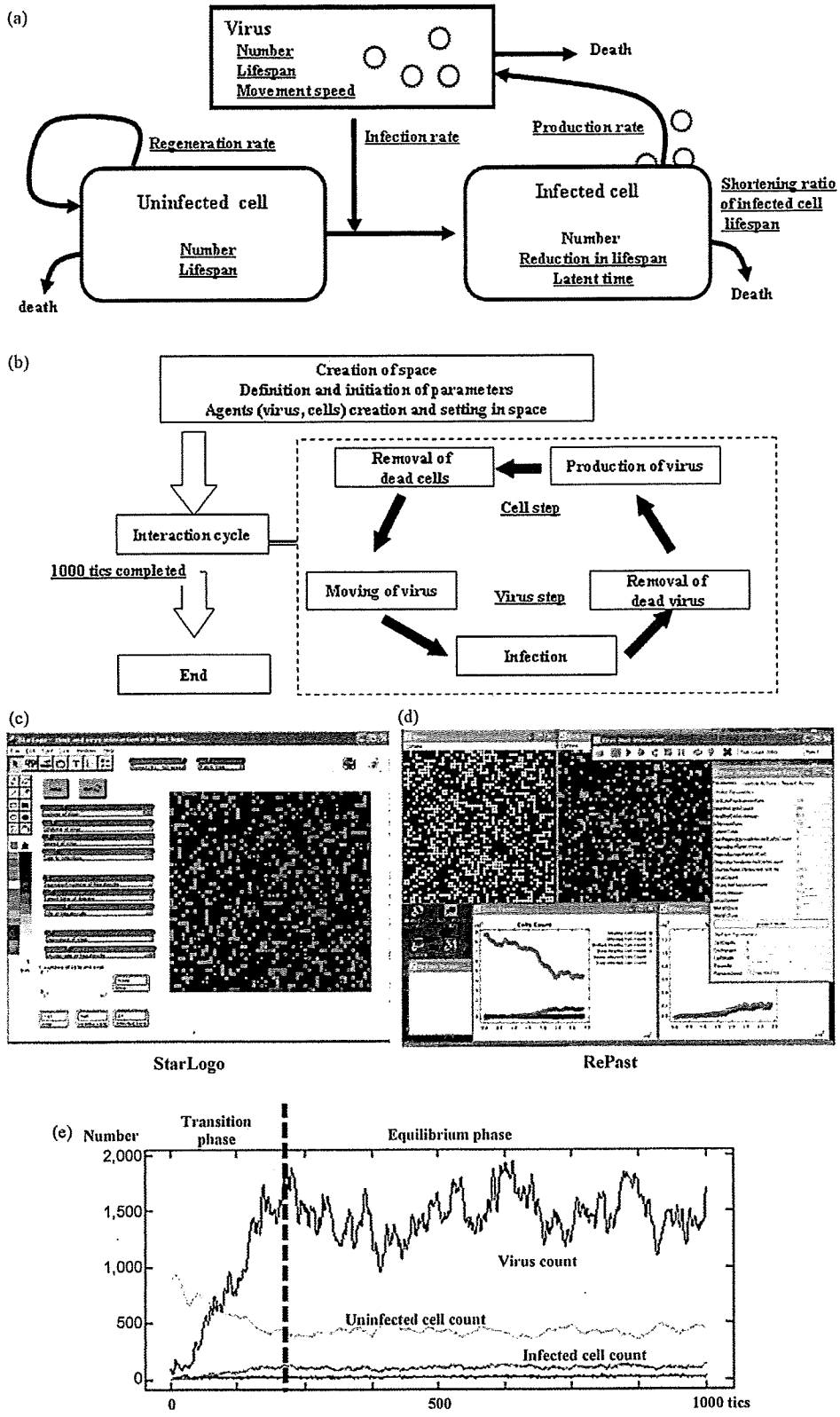

Fig. 1. Simulation design and an example of simulation results. (a) Model concept. Viruses, uninfected cells, and infected cells were treated as agents, and parameters were set for each of these and for interactions between agents (underlined). (b) Flowchart of the program. After preparing the simulation, we entered the interaction cycle, in which virus steps (such as movement) and cell steps were repeated. One cycle was counted as 1 tic, and the simulation concluded after 1000 tics. (c and d) Simulation screen using (c) StarLogo and (d) RePast. Yellow circles are viruses, green squares are uninfected cells, and orange and red indicate infected cells, with orange indicating the latent period. In StarLogo, all the agents are shown on the same screen, but in RePast, viruses and cells are shown in separate windows. (e) Example of a simulation chart in StarLogo. After the start of simulation the virus count and infected cell count increase while the uninfected cell count decreases, with equilibrium state reached after a certain number of tics.

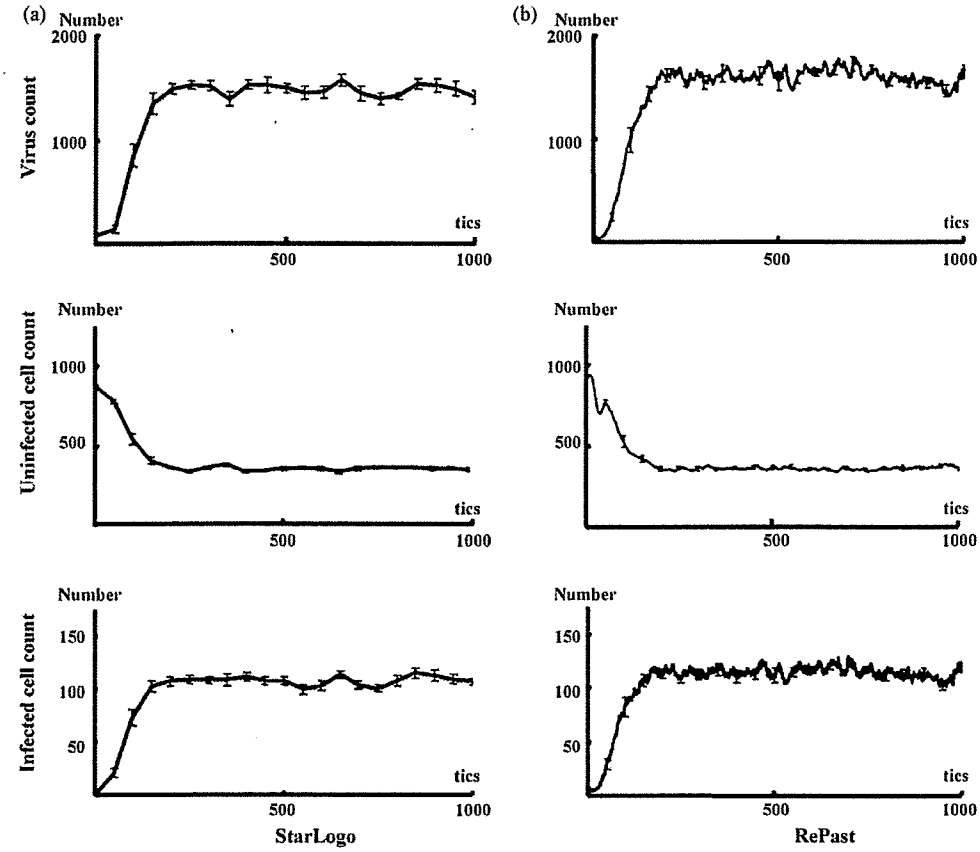

Fig. 2. Comparison of simulation results in (a) StarLogo and (b) RePast. The results were consistent when the parameters were made consistent. (Virus count [average ± SD]: StarLogo 1458.03 ± 173.1, RePast 1462.71 ± 178.8, $p = 0.94$. Uninfected cell count: 364.24 ± 30.4, 368.11 ± 33.4, $p = 0.83$. Infected cell count: 105.73 ± 13.0, 107.74 ± 13.0, $p = 0.24$. Unpaired Student's t-test.) Parameter values were set as follows: initial virus count, 100; uninfected cell count, 880; infected cell count, 0; virus speed of movement, 5 grids/tic; infection rate, 10%; uninfected cell regeneration rate, 1%; latent period, 3 tics; and virus reproduction rate, 5/cells/tic. The following parameter settings were taken from actual measurements: virus lifespan, 4.5 tics; uninfected cell lifespan, 49.8 tics; and infected cell lifespan, 6.7 tics.

immunodeficiency virus (HIV) (Ho et al., 1995), hepatitis B virus (HBV) (Nowak et al., 1996) and hepatitis C virus (HCV) (Neumann et al., 1998), and research is also underway on a range of models based on animal experiments and cell culture systems. As chronic viral disorders persist over long periods of time complete follow-up of viral dynamics is difficult. Furthermore, limitations of items that can be measured, such as the difficulty of measuring whole numbers of host cells, make it extremely difficult to investigate their consistency in mathematical models.

The recent ascend of dynamic-models owes much to advances in computers. Computer performance has improved markedly in recent years, not only in terms of their calculating capacity but also with regard to image displays, and models that offer a visual representation of viral disorders are now being reported (Gilbert and Bankes, 2002; Duca et al., 2007; Shapiro et al., 2008; Castiglione et al., 2007). One advantage of such visual models is that by providing a visual representation, they make understanding the disease status easy. Another benefit is that they enable parameters to be identified that are hidden as background noise in mathematical models. However, these models have some problems; it is difficult to prove the reproducibility of the simulation results derived from different languages or libraries, difficult to prove the validity of the model's concepts, and difficult to prove that the simulation results accurately reflect the reality. In this study, we created agent-based computer models that visually simulate the conditions of chronic viral infections using two software. The reproducibility of two agent-based computer models and the differences between agent-based models and the mathematical model were analyzed. This agent-based model enabled us to investigate how each parameter included in the concept affects the conditions of chronic viral infections.

2. Methods

2.1. Selection of Software

In this study, we used two different types of softwares: StarLogo version 2.0 (http://education.mit.edu/starlogo/) supplied by MIT Media Laboratory and Recursive Porous Agent Simulation Toolkit (RePast-3.0, http://repast.sourceforge.net/) supplied by the Argonne National Laboratory. StarLogo uses Logo, one of the simplest programming languages, and has a fixed graphical user interface. RePast is a library that uses Java, another programming language, which also has a fixed graphical user interface.

Logo is an assembly language, and StarLogo carries out processing sequentially. Java is an object-oriented language, and RePast has a faster processing speed than StarLogo. In addition, StarLogo has a number of stipulations to simplify simulations, such as parameters can only be set up to five decimal places and the simulation space is also fixed as 51 × 51 square grids. RePast, on the other hand, has fewer such restrictions. Thus, it offers a higher degree of freedom in program settings than StarLogo. Taking simulation space as an example, in spite of the restrictions imposed by the underlying operating system's image display system, any number of grids can be set and a hexagonal grid could also be chosen rather than a square one. However, users must stipulate and set all parameters themselves. This means that they must first declare the shape of the grid and the number of grids they will use to fill the simulation space. Java is also more difficult to learn than Logo, and debugging and correcting the program is also more difficult. Thus, it is difficult to judge whether or not the results agree with the planned simulation.

In effect, these two different types of softwares are polar opposites. It is simple to start a simulation in StarLogo, but producing results takes time and it is difficult to carry out more complex simulations. In RePast it is difficult to compose the program and judge whether or not the planned study has actually been achieved, but the

simulation itself takes only a short time to complete and there are lesser restrictions in the construction of a simulation model.

2.2. Concept for Modeling

We applied the basic virus–host interaction mathematical model to the agent-based simulation system with slight modifications. The mathematical model was used to describe the dynamics of HIV (Ho et al., 1995), HBV (Nowak et al., 1996), and HCV (Neumann et al., 1998) and the only agents involved were host cells and viruses, without the inclusion of immune cells. The effects of the immune system are expressed by varying parameters such as lifespan of host cells and viruses.

Fig. 1a illustrates the study concept. Viruses have the ability to infect healthy host cells (uninfected cells) and the infected cells produce new viruses. Both cells and viruses have definite lifespans, and the lifespan of infected cells is usually shorter than that of uninfected cells. Uninfected cells automatically regenerate within the space, whereas infected cells only arise due to infection of uninfected cells. Viruses also lack the ability to regenerate themselves and are only produced from infected cells.

2.3. Parameter Settings

In the present study, as the StarLogo settings are circumscribed, we limited the simulation space to 51 × 51 square grids. However, we made an exception here while investigating the effects of size of space on the simulation results. The numbers of viruses, uninfected cells, and infected cells could only be set before the start of the simulation. As described in the later, our simulation ran in cycles, with 1 cycle defined as 1 tic.

In mathematical simulation models, the death rate is required as a parameter. However, in our program we set lifespans for viruses and uninfected cells. These lifespans were not uniform, but were set to have a deviation of about 10%. The lifespan of cells was shortened by infection with ratio decided beforehand.

The infection ratio was meaningful only when an infected cell and a virus coincidentally occupied the same grid, and this was used to calculate the probability of the infection occurring in that situation. The virus production rate was set as the number of viruses produced by an infected cell during 1 tic. Infected cells could be set as a parameter indicating the latent period between the time of virus infection and the time of virus replication.

In order to emulate the tissue repair capacity, we set uninfected cell regeneration rate such that grids without any cells had a specified probability of producing uninfected cells on top of themselves. As a result, the more the cell count declined within a space the more regenerated uninfected cells were produced, whereas the number of regenerated cells declined as cell count increased.

The number of grids through which a virus could move in 1 tic was set as the speed of movement, and the direction of movement was set within a range of 90° toward the top of the simulation space. The program used a circulatory method of movement; when a virus arrived at the top of the space, it was translocated to the bottom, and moved again toward the top. Cells were fixed on the grid.

2.4. Simulation Flowchart

Fig. 1b shows a flowchart of the program. First, the simulation space was produced, after which each parameter was defined and the initial settings were made. Next the agents – viruses and uninfected and infected cells – were produced. The simulation cycle was as follows. Viruses moved to a new grid, and if an uninfected cell was present, this was infected with a probability based on the infection rate. The lifespan of the virus decreased, and viruses that had completed their lifespan and those that had caused an infection were removed from the space. Infected cells produced new viruses, the lifespans of both uninfected and infected cells decreased. Then, cells that had completed their lifespan were eliminated and a new cycle began. The program was set such that the simulation ended after this cycle had repeated 1000 times. This meant that one simulation was complete after 1000 tics.

2.5. Data Collection

The RePast model was programmed such that data for each tic was saved automatically as a text file at the end of the simulation. This text file could be opened by a database software. The StarLogo model was programmed to stop the simulation and collect data after every 50 tics.

2.6. Mathematical Model

In order to compare the results of this agent-based simulation, we used a viral infection mathematical model, which we improved as follows.

$$\frac{dT}{dt} = s[2601 - (T+I)] - dT - bVT \quad (1)$$

$$\frac{dI}{dt} = bVT - dI \quad (2)$$

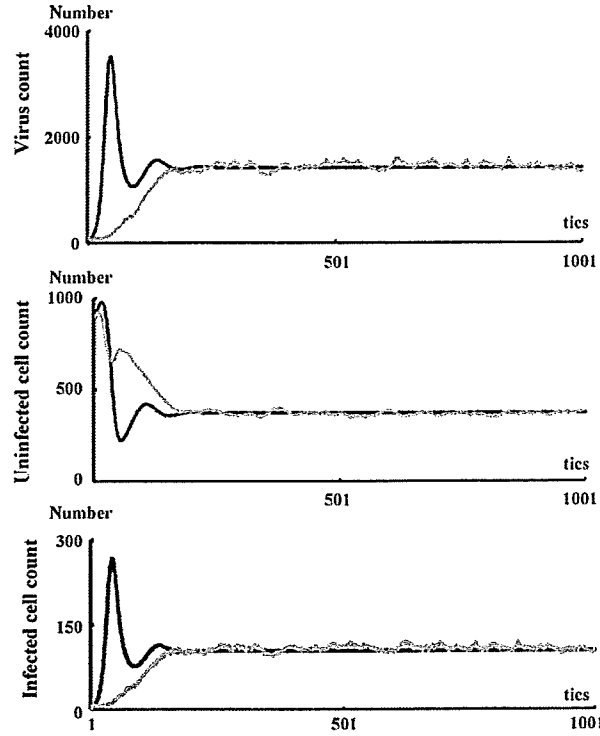

Fig. 3. Comparison of results of agent-based simulation and mathematical simulation. Both sets of results were consistent for the equilibrium phase, but differed in the shift in transition phase. Black line: mathematical model; grey line: results of simulation in RePast. Parameter values were set as follows: initial virus count, 100; uninfected cell count, 880; infected cell count, 0; virus speed of movement, 5 grids/tic; infection rate, 10%; uninfected cell regeneration rate, 1%; latent period, 3 tics; virus reproduction rate, 5/cells/tic; virus lifespan, 10 tics; uninfected cell lifespan, 50 tics; and cell lifespan-shortening ratio as a result of infection, 69%.

$$\frac{dV}{dt} = pI - cV \quad (3)$$

where, T is the uninfected cell count, I is the infected cell count, and V is the virus count. Uninfected cells are supplied to the space with a probability $s[2601 - (T+I)]$, as the number of grids in this agent-based simulation model was 2601 (51 × 51). The death rate of uninfected cells is d, the death rate of infected cells is δ, and the death rate of viruses is c. The infection rate is indicated by β. Viruses are released from infected cells at a probability p.

2.7. Statistical Analysis

Statistical analyses were performed by statistical tests using the program StatView 5.0 (SAS Institute Inc.). All tests of significance were two-tailed, with p values of <0.05 considered to be significant.

3. Results

3.1. Reproducibility of Chronic Viral Infection Disease Models Using Agent-based Simulation Methods

We constructed the chronic viral infection model with StarLogo library. Fig. 1c shows the simulation screen, and Fig. 1e shows one sample result. Immediately after the start of the simulation, the virus count temporarily dropped in accordance with the onset of an infection. Subsequently, the virus count started to increase with an increase in the infected cells and a decrease in the uninfected cells. After a certain number of tics (around 300 in this example), although the virus count, infected cell count, and uninfected cell count had some fluctuation, an equilibrium state was reached. We use the following descriptive terms in this paper: the transient phase is the period during which virus growth peaks, and the equilibrium phase is the period during which an equilibrium state is

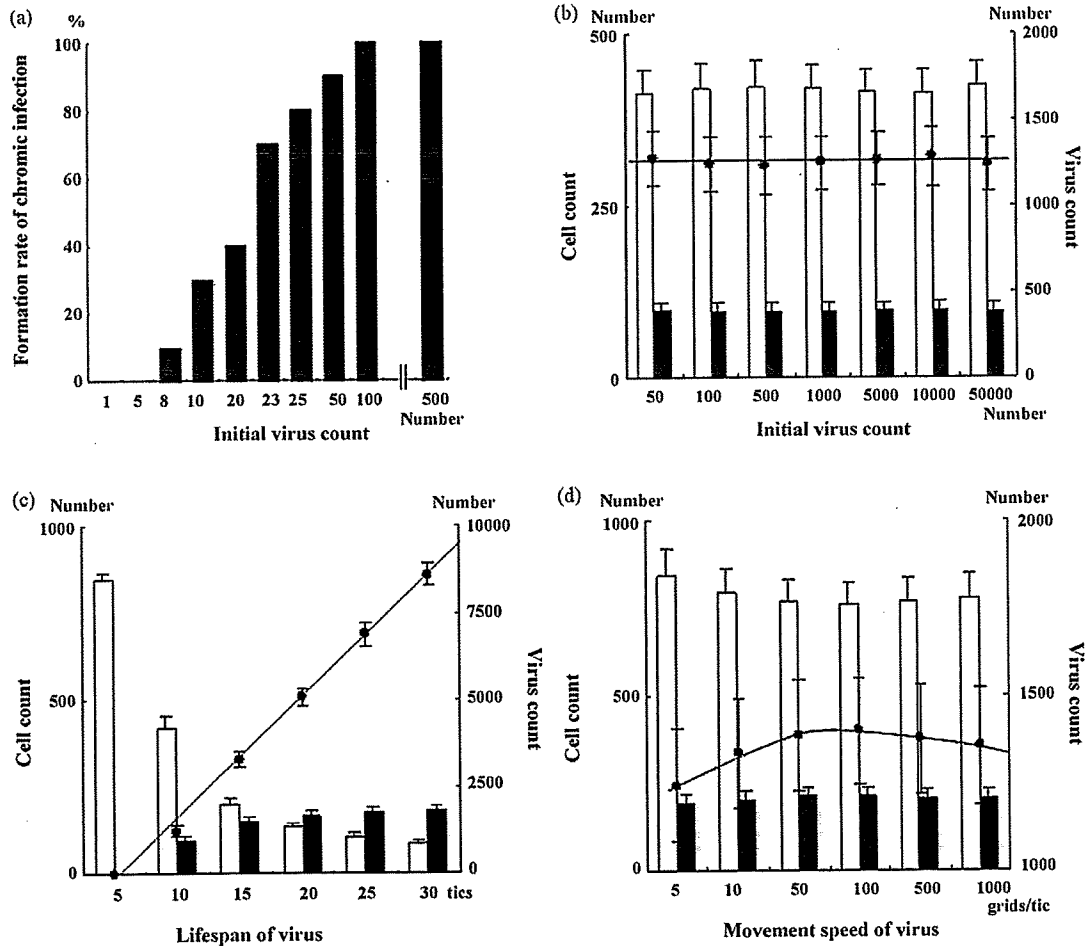

Fig. 4. Effects of changes in viral parameters. (a) The higher the initial virus count, the greater is the increase in the rate of formation of chronic infection, but (b) there was no effect on the conditions in the equilibrium phase. (c) Extending the virus lifespan increased the virus count. (d) Increasing the speed of virus movement to 100 grids/tic increased the virus count, but increasing it to 500 grids/tic had the opposite effect, with a slight declining trend. (a) Black bars: number of infections produced; (b–d) black circles: virus count; line: virus count approximation curve; white bars: uninfected cell count; black bars: infected cell count.

established. When the simulation was performed multiple times, the features described above were maintained, and the average values for virus, infected cell, and uninfected cell counts during the equilibrium state were all consistent.

Fig. 1d shows the simulation screen of the RePast. When we attempted setting all the initial parameters to the same values as those in the StarLogo, the results were not consistent. When we recalculated the parameters from the simulation results, in RePast, the parameters were largely maintained at the levels of the settings, but in StarLogo, the lifespans of both cell types became shorter than the settings while the simulation was in progress. We made the results of both simulations consistent by using the same parameters during the actual simulation (Fig. 2a and b).

3.2. Comparison Between Agent-based Simulation Models and Mathematical Simulation Model

We investigated whether the results of a chronic viral infection disease model produced by RePast would be consistent with the results of a mathematical model. For the mathematical model, we carried out an approximate integration using a four-dimensional Runge–Kutta method to ensure that the uninfected cell count and infected cell count would be in the same class. Parameters were always fixed as constant between simulations. The simulation results were consistent for the equilibrium phase, but transitions in virus count during the transient phase varied, with a shift to equilibrium state following two overshoots in the mathematical model, but a monotonic increase following a logistic curve in the agent-based model (Fig. 3). In the mathematical model, when the equilibrium condition was calculated with $dT/dt = dI/dt = dV/dt = 0$, the equilibrium-phase virus count, uninfected cell count, and infected cell count were very similar to those of the agent-based model (virus count: mathematical model 371.8/space, agent-based model 371.1 ± 32.4/space [average ± SD]; uninfected cell count: mathematical model 1605/space, agent-based model 1454 ± 194/space; infected cell count: mathematical model 115.9/space, agent-based model 108.3 ± 14.2/space).

3.3. Usability of the Models; Effect of Changing Parameters

We investigated the changes in the equilibrium phase brought about by changing each parameter. All the investigations below were carried out by using RePast, and we used the average values from ten simulations.

3.4. Viral Parameters

The lower the virus counts at the beginning of the simulation, the lower the probability of a chronic infection (Fig. 4a). However, the initial virus count did not have any effect on the equilibrium

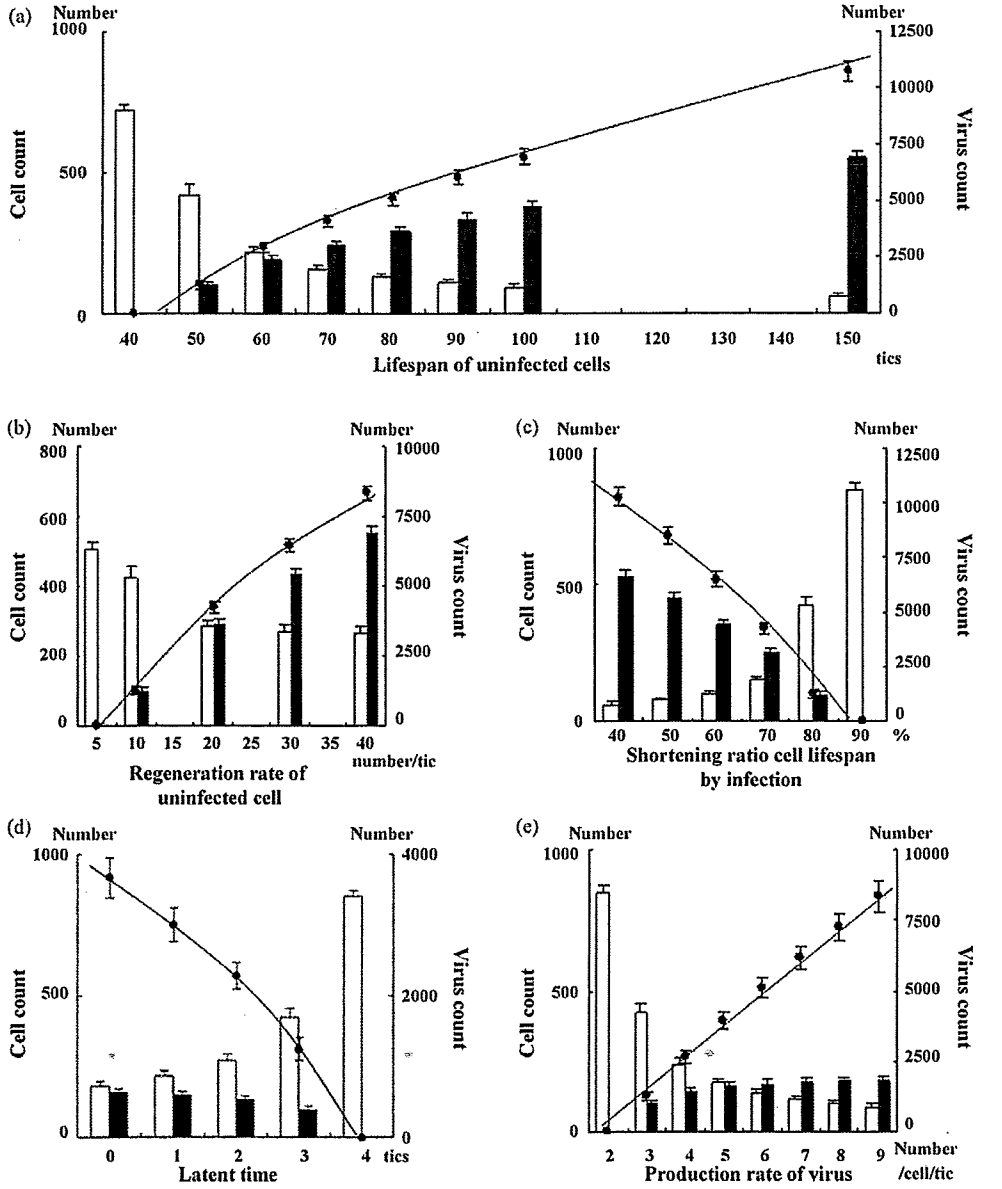

Fig. 5. Effects of changes in cell parameters. (a) Extending the uninfected cell lifespan and (b) increasing the uninfected cell regeneration rate increased the virus count. (c) Raising the lifespan-shortening ratio as a result of infection shortened the lifespan of infected cells, thereby decreasing the virus count. (d) Extending the latent period shortened the period of virus production from infected cells, thereby decreasing the virus count. (e) Increasing the virus production count resulted in a linear increase in equilibrium-phase virus count. Black circles: virus count; line: virus count approximation curve; white bars: uninfected cell count; black bars: infected cell count.

phase itself (Fig. 4b). Extending the lifespan of viruses resulted in a linear increase in equilibrium-phase virus count (Fig. 4c). Although the infected cell count increased, the rate of increase gradually declined. Changing the speed of viral movement resulted in the equilibrium-phase virus count to eventually decline after 100 grids/tic was reached, allowing movement over an area twice the size of the simulation space (Fig. 4d).

3.5. Uninfected Cell Parameters

Extending the lifespan of uninfected cells led to an increased virus count during the equilibrium phase (Fig. 5a). Increasing the uninfected cell regeneration rate also contributed to increased equilibrium-phase virus count (Fig. 5b). In both the cases, the increases in virus count and infected cell count were not linear, but showed a tendency for the rate of increase to decline gradually.

3.6. Infected Cell Parameters

We carried out an investigation of the effects of variation in the lifespan-shortening ratio on the virus count on the assumption that cell lifespan is shortened by infection. When this ratio was increased, the virus count decreased (Fig. 5c). An extended latent period was also related to a decreased virus count (Fig. 5d). However, the virus production from infected cells led to a linear increase in the virus count (Fig. 5e).

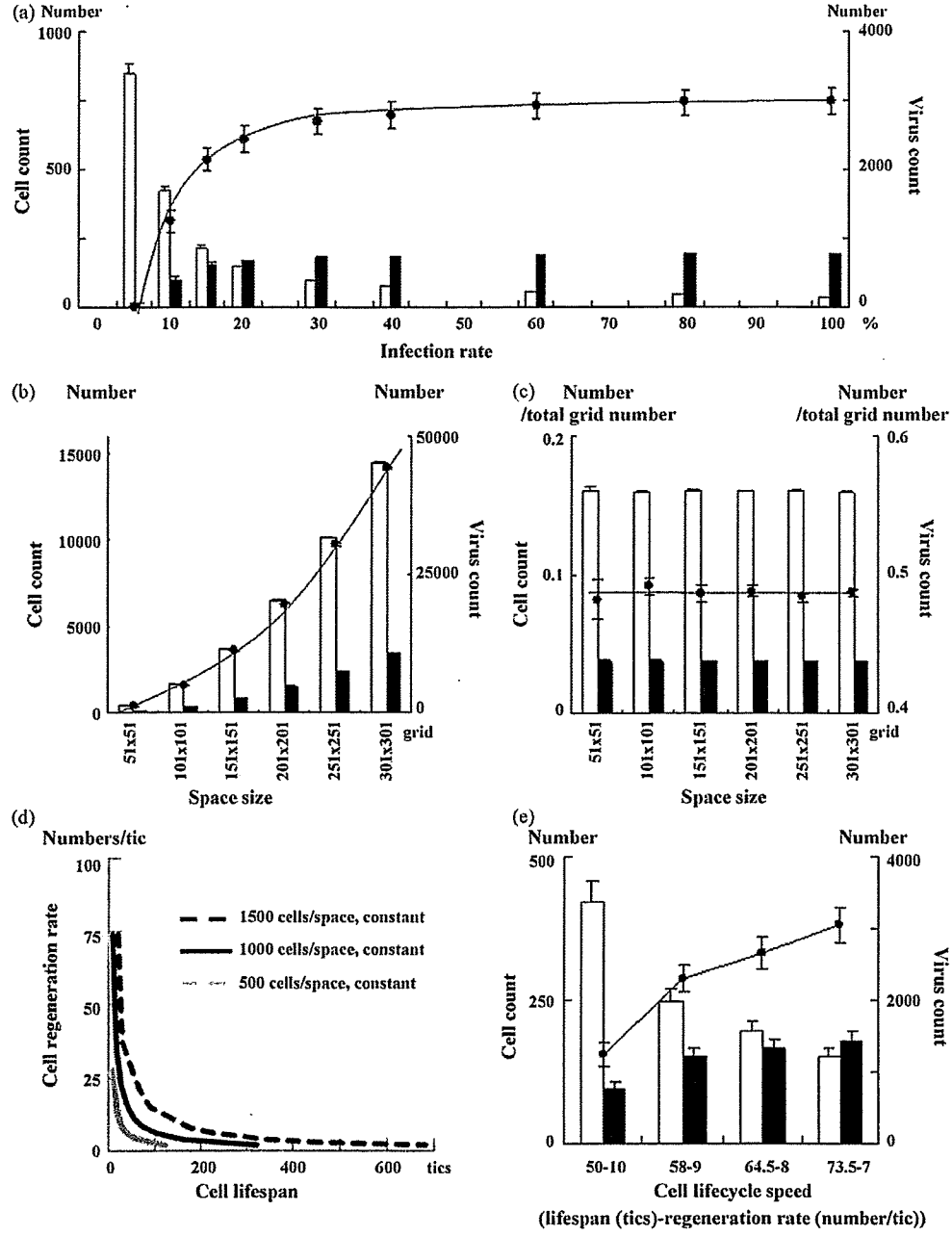

Fig. 6. (a) Increasing the infection rate increased the virus count in equilibrium periods, but the virus count did not change at infection rates of 30% or more. (b) The size of the simulation space increased not only virus count but also the cell count; however, (c) when virus and cell counts were divided by the total number of grids in the space, they were constant for all space sizes. (d) Changing the lifespan and regeneration rate of uninfected cells in opposite directions at the same time makes it possible to change only the cell cycle speed without altering the uninfected cell count. (e) When the cell cycle speed was reduced, the virus count increased toward the right of the graph. This may be because the effect of extending the lifespan of cells exceeds that of reducing their regeneration rate. (a–c and e) Black circles: virus count; line: virus count approximation curve; white bars: uninfected cell count; black bars: infected cell count.

3.7. Infection Rate and Space Size

Increasing the infection rate caused an increase in the virus count, but the change was minimal at an infection rate of 30% or more. The same results were seen for infected cell count, but a decrease in uninfected cell count resulted in a tendency for the infection rate to decrease by up to 60% (Fig. 6a).

The larger the space, higher the increase in both virus and cell counts (Fig. 6b). This increase was proportional to space size, however, when virus and cell counts were divided by the total number of grids in the space they were all constant (Fig. 6c).

3.8. Cell Cycle Speeds

Running a simulation with the initial virus count set to zero enables only the equilibrium condition for uninfected cells to be simulated. Changing the lifespan and regeneration rate of uninfected cells in opposite directions at the same time makes it possible

to change the cell cycle speed without altering the uninfected cell count (Fig. 6d). We used this technique to investigate how changing the cell cycle speed affected the equilibrium phase. Fig. 6e shows the results. Cell lifespan increases while the cell cycle speed declines. The equilibrium virus count increased in accordance with slower cell cycle speeds.

4. Discussion

In this study, we investigated the models using two agent-based simulation methods to program a simple virus–host chronic infection model. The same model written in two different programming language systems displayed the same results. The transient phase was unlike that seen in a mathematical simulation with no overshoot in virus count, but rather a smooth transition to the equilibrium phase. The virus count at the start of the simulation only had effect on the rate of infection development. Increases in virus lifespan, uninfected cell lifespan, uninfected cell regeneration rate, virus production count from infected cells, and infection rate all led to increased equilibrium-phase virus count. Rises in the infected cell lifespan-shortening ratio, latent period, and cell cycle speed decreased the equilibrium-phase virus count. The size of the space itself had no innate effect on the equilibrium phase, but a speed of movement of the virus that was twice the size of the space produced the maximum virus count.

Reproducibility is the basis for all scientific study, but there are many problems to prove it in computer simulations, such as programming bugs. As agent-based simulation deals with numerous agents individually, it requires vast amounts of calculations. Accumulation of very small change of values leads to large differences of results. In this study, we investigated two programs based on two programming languages to confirm the reproducibility of our simulation results in different programming languages. The results of two simulations were consistent, but in StarLogo, the lifespan parameters had a tendency to be lower than when they were set while simulations were actually in progress. This may be because the number of digits used in calculations was different between the two programs. RePast performs calculations to at least eight decimal places. In StarLogo, the library settings only enable settings to be made up to five decimal places. It is probable that these small differences accumulate during repeated calculations and are reflected in the simulation. Ultimately, we confirmed that the differences in results obtained by using different libraries and programming languages were not innate and by making the parameters consistent during simulation, consistent results were obtained.

Mathematical models using formulae for HIV therapy was published in 1994, the method has since been applied to HBV and HCV (Ho et al., 1995; Nowak et al., 1996; Neumann et al., 1998), and they were thought to be good reflections of the reality. In the mathematical model, viruses and cells are conceived as individuals in the concept itself, but both of them are perceived *en masse* when calculations are performed. However a feature of the agent-based simulation is that it deals with individual viruses and cells as separate agents. By moving each agent individually, it probes the factors influencing overall shifts from the micro viewpoint. When the space is viewed as a whole, it is possible to watch on the screen the collective movement of groups of agents. Recently, models that provide a visual representation of Epstein-Barr virus and HIV infection have been reported, both of which are useful for an instinctive and intuitive understanding (Duca et al., 2007; Shapiro et al., 2008; Castiglione et al., 2007).

In agent-based simulation model, virus count transit smoothly to the equilibrium phase. On the other hand, virus counts overshoot during transient phase in mathematical model. We think this difference is derived from technicality of different modeling. The difference in concepts between mathematical models and agent-based models is the space. The mathematical model has no space in concept, but agents move across the space in the agent-based model. In agent-based models, the densities of virus and cells change overtime especially in the transition phase because of the limited space. These changes of the densities of virus and cells lead to the dynamic change of the encounter rate of viruses and cells. The mathematical model does not make such concept of the density; the encounter rate is constant. This may be the reason for the difference between two models in the transition phase. Since no overshoot of virus counts in transient phase had been reported from in vivo studies of hepatitis C virus and simian immunodeficiency virus (Dahari et al., 2005; Nowak et al., 1997), agent-based model correlates with actual biology in vivo at least for these viruses. The increase of initial virus count at the start of simulation correlates with higher encounter rate of viruses and cells which make the linear increasing of infection forming rate. Mathematical model can only express the infection formation rate as "infected or not".

The importance of viral passing speed in the agent-based model is also explained by the "space". Although the virus actually moves through the blood stream in our body and virus could not decide their moving speeds by themselves, there is most appropriate speed for virus to meet the cells on the simulation space by the highest probability. The effect of cell cycle speed should be mentioned by another affection of the space. A fast cell cycle speed means that the lifespan of uninfected cells is short. Then fast cell cycle speed leads to the short lifespan of infected cells. A higher regeneration rate for uninfected cells results in a higher rate of infection among uninfected cells by viruses, but in situations where viruses and cells are dispersed around the space this is ineffective in increasing the infection rate, as the latter depends on the probability that they will encounter one another. As a result, the infected cell count decreases during the equilibrium phase, as does the virus count.

In this study, we confirmed the reproducibility and usability of agent-based models in expressing the interaction between viruses and cells. A feature of this simulation system is that it uses the concept of space as actual space, which means that the existence of the space becomes an additional controlling factor on the simulation results. This is a concept that is absent from mathematical models. The reality is that we have a spatial existence, and an advantage of the agent-based simulation system is the fact that it accounts for the space. Another feature of the simulation system is that it enables the condition to be perceived in visual terms, making it easy to understand. However it may be affected by computer performance and by the limitations of programming languages or the program itself, this system may offer a powerful tool for the future analysis of real virus–host interaction disease.

Conflict of interest

No conflicts of interest exist for all authors.

References

Castiglione, F., Pappalardo, F., Bernaschi, M., Motta, S., 2007. Optimization of HAART with genetic algorithms and agent-based models of HIV infection. Bioinformatics 23, 3350–3355, doi:10.1093/bioinformatics/btm408.

Dahari, H., Major, M., Zhang, X., Mihalik, K., Rice, C.M., Perelson, A.S., Feinstone, S.M., Neumann, A.U., 2005. Mathematical modeling of primary hepatitis c infection: noncytolytic clearance and early blockage of virion production. Gastroenterology 128, 1056–1066, doi:10.1053/j.gastro.2005.01.049.

Duca, K.A., Shapiro, M., Delgado-Eckert, E., Hadinoto, V., Jarrah, A.S., Laubenbacher, R., Lee, K., Luzuriaga, K., Polys, N.F., Thorley-Lawson, D.A., 2007. A virtual look at Epstein-Barr virus infection: biological interpretations. PLoS Pathog. 3, 1388–1400, doi:10.1371/journal.ppat.0030137.

Gilbert, N., Bankes, S., 2002. Platforms and methods for agent-based modelling. Proc. Natl. Acad. Sci. U.S.A. 99 (Suppl. 3), 7197–7198.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., Markowitz, M., 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373, 123–126, doi:10.1038/373123a0.

Naniche, D., 2009. Human immunology of measles virus infection. Curr. Top. Microbiol. Immunol. 330, 151–171.

Neumann, A.U., Lam, N.P., Dahari, H., Gretch, D.R., Wiley, T.E., Layden, T.J., Perelson, A.S., 1998. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. Science 282, 103–107, doi:10.1126/science.282.5386.103.

Nowak, M.A., Bonhoeffer, S., Hill, A.M., Boehme, R., Thomas, H.C., McDade, H., 1996. Viral dynamics in hepatitis B virus infection. Proc. Natl. Acad. Sci. U.S.A. 93, 4398–4402.

Nowak, M.A., Lloyd, A.L., Vasquez, G.M., Wiltrout, T.A., Wahl, L.M., Biscofberger, N., Williams, J., Kinter, A., Fauci, A.S., Hirsch, V.M., Lifson, J.D., 1997. Viral dynamics of primary viremia and antiretroviral therapy in simian immunodeficiency virus infection. J. Virol. 71, 7518–7525.

Shapiro, M., Duca, K.A., Lee, K., Delgado-Eckert, E., Hawkins, J., Jarrah, A.S., Laubenbacher, R., Polys, N.F., Hadinoto, V., Thorley-Lawson, D.A., 2008. A virtual look at Epstein-Barr virus infection: simulation mechanism. J. Theor. Biol. 252, 633–648, doi:10.1016/j.jtbi.2008.01.032.

See, H., Wark, P., 2008. Innate immune response to viral infection of the lungs. Paediatr. Respir. Rev. 9, 243–250, doi:10.1016/j.prrv.2008.04.001.

nature genetics

Genome-wide association of *IL28B* with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C

Yasuhito Tanaka[1,18], Nao Nishida[2,18], Masaya Sugiyama[1], Masayuki Kurosaki[3], Kentaro Matsuura[1], Naoya Sakamoto[4], Mina Nakagawa[4], Masaaki Korenaga[5], Keisuke Hino[5], Shuhei Hige[6], Yoshito Ito[7], Eiji Mita[8], Eiji Tanaka[9], Satoshi Mochida[10], Yoshikazu Murawaki[11], Masao Honda[12], Akito Sakai[12], Yoichi Hiasa[13], Shuhei Nishiguchi[14], Asako Koike[15], Isao Sakaida[16], Masatoshi Imamura[17], Kiyoaki Ito[17], Koji Yano[17], Naohiko Masaki[17], Fuminaka Sugauchi[1], Namiki Izumi[3], Katsushi Tokunaga[2] & Masashi Mizokami[1,17]

The recommended treatment for patients with chronic hepatitis C, pegylated interferon-α (PEG-IFN-α) plus ribavirin (RBV), does not provide sustained virologic response (SVR) in all patients. We report a genome-wide association study (GWAS) to null virological response (NVR) in the treatment of patients with hepatitis C virus (HCV) genotype 1 within a Japanese population. We found two SNPs near the gene *IL28B* on chromosome 19 to be strongly associated with NVR (rs12980275, $P = 1.93 \times 10^{-13}$, and rs8099917, 3.11×10^{-15}). We replicated these associations in an independent cohort (combined P values, 2.84×10^{-27} (OR = 17.7; 95% CI = 10.0–31.3) and 2.68×10^{-32} (OR = 27.1; 95% CI = 14.6–50.3), respectively). Compared to NVR, these SNPs were also associated with SVR (rs12980275, $P = 3.99 \times 10^{-24}$, and rs8099917, $P = 1.11 \times 10^{-27}$). In further fine mapping of the region, seven SNPs (rs8105790, rs11881222, rs8103142, rs28416813, rs4803219, rs8099917 and rs7248668) located in the *IL28B* region showed the most significant associations ($P = 5.52 \times 10^{-28}$–2.68×10^{-32}; OR = 22.3–27.1). Real-time quantitative PCR assays in peripheral blood mononuclear cells showed lower *IL28B* expression levels in individuals carrying the minor alleles ($P = 0.015$).

Hepatitis C is a global health problem that affects a significant proportion of the world's population. The World Health Organization estimated that in 1999, there were 170 million HCV carriers worldwide, with 3–4 million new cases appearing each year. HCV infection affects more than 4 million people in the United States, where it represents the leading cause of cirrhosis and hepatocellular carcinoma as well as the leading cause of liver transplantation[1]. The American Gastroenterological Association estimated that drugs are the largest direct costs of hepatitis C[1].

The most effective current standard of care in patients with chronic hepatitis C, a combination of PEG-IFN-α with ribavirin, does not produce SVR in all patients treated. Large-scale studies on 48-week-long PEG-IFN-α/RBV treatment in the United States and Europe showed that 42–52% of patients with HCV genotype 1 achieved SVR[2–4], and similar results were found in Japan. However, older patients (greater than 50 years of age) had a significantly lower rate of SVR due to poor adherence resulting from adverse events and laboratory-detectable abnormalities such as neutropenia and thrombocytopenia[5,6]. Specifically, various well-described side effects (such as a flu-like syndrome, hematologic abnormalities and adverse neuropsychiatric events) often necessitate dose reduction, and 10–14% of patients require premature withdrawal from interferon-based therapy[7]. To avoid these side effects in patients who will not be helped by the treatment, as well as to reduce the substantial cost of PEG-IFN-α/RBV treatment, it would be useful to be able to predict an individual's response before or early in treatment. Several viral factors, such as genotype 1, high baseline viral load, viral

[1]Department of Clinical Molecular Informative Medicine, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan. [2]Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [3]Division of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Tokyo, Japan. [4]Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, Tokyo, Japan. [5]Division of Hepatology and Pancreatology, Kawasaki Medical College, 577 Matsushima, Kurashiki, Japan. [6]Department of Internal Medicine, Hokkaido University Graduate School of Medicine, Sapporo, Japan. [7]Molecular Gastroenterology and Hepatology, Kyoto Prefectural University of Medicine, Kyoto, Japan. [8]National Hospital Organization Osaka National Hospital, Osaka, Japan. [9]Department of Medicine, Shinshu University School of Medicine, Matsumoto, Japan. [10]Division of Gastroenterology and Hepatology, Internal Medicine, Saitama Medical University, Saitama, Japan. [11]Second department of Internal Medicine, Faculty of Medicine, Tottori University, Yonago, Japan. [12]Department of Gastroenterology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan. [13]Department of Gastroenterology and Metabology, Ehime University Graduate School of Medicine, Ehime, Japan. [14]Department of Internal Medicine, Hyogo College of Medicine, Nishinomiya, Japan. [15]Central Research Laboratory, Hitachi Ltd., Kokubunji, Japan. [16]Gastroenterology and Hepatology, Yamaguchi University Graduate School of Medicine, Yamaguchi, Japan. [17]Research Center for Hepatitis and Immunology, International Medical Center of Japan Konodai Hospital, Ichikawa, Japan. [18]These authors contributed equally to this work. Correspondence should be addressed to M.M. (mmizokami@imcjk2.hosp.go.jp).

Received 29 June; accepted 21 August; published online 13 September 2009; doi:10.1038/ng.449

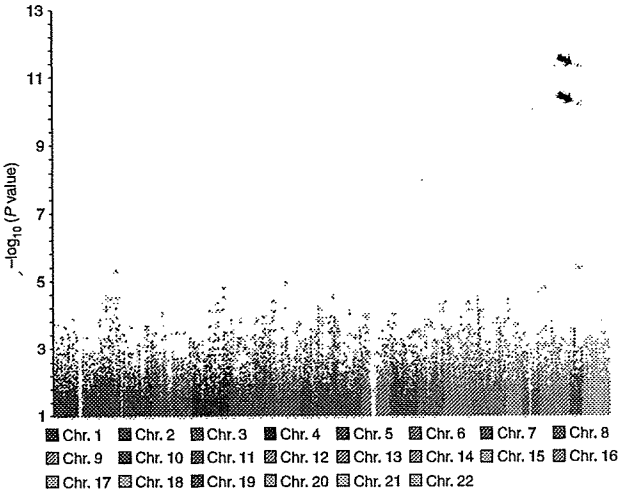

Figure 1 Genome-wide association results with PEG-IFN-α/RBV treatment in 142 Japanese patients with HCV (78 NVR and 64 VR samples). *P* values were calculated by using a χ^2 test for allele frequencies. The dots with arrows for chromosome 19 denote SNPs that showed significant genome-wide associations ($P < 8.05 \times 10^{-8}$) with response to PEG-IFN-α/RBV treatment.

kinetics during treatment, and amino acid pattern in the interferon sensitivity–determining region, have been reported to be significantly associated with the treatment outcome in a number of independent studies[8–10]. Studies have also provided strong evidence that ~20% of patients with HCV genotype 1 and 5% of patients with genotype 2 or 3 have a null response to PEG-IFN-α/RBV. No definite predictor of this resistance is currently available that make it possible to bypass the initial 12–24 weeks' treatment before deciding whether treatment should be continued. If a reliable predictor of non-response were identified for use in patients before treatment initiation, then an estimated 20%, including those who have little or no chance to achieve SVR, could be spared the side effects and cost of treatment.

Host factors, including age, sex, race, liver fibrosis and obesity, have also been reported to be associated with PEG-IFN-α/RBV therapy outcome[11,12]. However, little is known about the host genetic factors that might be associated with the response to therapy: thus far only a few candidate genes, including those encoding type I interferon receptor-1 (*IFNAR1*) and mitogen-activated protein kinase–activated protein kinase 3 (*MAPKAPK3*), have been reported to be associated with treatment response[13,14]. We describe here a GWAS for response to PEG-IFN-α/RBV treatment.

We conducted this GWAS to identify host genes associated with response to PEG-IFN-α/RBV treatment in 154 Japanese patients with HCV genotype 1 (82 with NVR and 72 with virologic response (VR), based on the selection criteria as described in Online Methods). We used the Affymetrix SNP 6.0 genome-wide SNP typing array for 900,000 SNPs. A total of 621,220 SNPs met the following criteria: (i) SNP call rate ≥95%, (ii) minor allele frequency (MAF) ≥1% and (iii) deviation from Hardy-Weinberg equilibrium (HWE) $P \geq 0.001$ in VR samples. After excluding 4 NVR and 8 VR samples that showed quality control (QC) call rates of <95%, 78 NVR and 64 VR samples were included in the association analysis. **Figure 1** shows a genome-wide view of the single-point association data based on allele frequencies. Two SNPs located close to *IL28B* on chromosome 19 showed strong associations, with a minor allele dominant model (rs12980275, $P = 1.93 \times 10^{-13}$, and rs8099917, $P = 3.11 \times 10^{-15}$, respectively), with NVR to PEG-IFN-α/RBV treatment (**Table 1**). The rs8099917 lies between *IL28B* and *IL28A*, ~8 kb downstream from *IL28B* and ~16 kb upstream from *IL28A*. These associations reached genome-wide levels of significance for both SNPs in this initial GWAS cohort (Bonferroni criterion $P < 8.05 \times 10^{-8}$ (0.05/621,220)). The frequencies of minor allele–positive patients were much higher in the NVR group than in the VR group for both SNPs (74.3% in NVR, 12.5% in VR for rs12980275; 75.6% in NVR, 9.4% in VR for rs8099917). Notably, individuals homozygous for the minor allele were observed only in the NVR group. The VR group, as compared to the NVR group, showed genotype frequencies closer to those in the healthy Japanese population[15], yet the minor allele frequencies were slightly higher in the transient virologic response (TVR) group (23.1%, 15.4%) than in the SVR group (9.8%, 7.8%) (**Table 1**). We applied the Cochrane-Armitage test on all the SNPs and found a genetic inflation factor, λ, of 1.029 for the GWAS stage (**Supplementary Fig. 1**). We also carried out principal component analysis in 142 samples for the GWAS stage together with the HapMap samples (CEU, YRI, CHB and JPT) (**Supplementary Fig. 2**); this suggested that the effect of population stratification was negligible.

Table 1 Significant association of two SNPs (rs12980275 and rs8099917) with response to PEG-IFN-α/RBV treatment

					Null responder (NVR[a], *n* = 128)			Responder (VR[a], *n* = 186)			Responder (SVR[a], *n* = 140)			NVR vs. VR		NVR vs. SVR	
dbSNP rsID	Nearest gene	MAF[b] (allele)	Allele (1/2)	Stage	11	12	22	11	12	22	11	12	22	OR (95% CI)[c]	*P* value[d]	OR (95% CI)[c]	*P* value[d]
rs12980275	*IL28B*	0.15 (G)	A/G	GWAS	20 (25.6)	54 (69.2)	4 (5.1)	56 (87.5)	8 (12.5)	0 (0.0)	46 (90.2)	5 (9.8)	0 (0.0)	20.3 (8.3–49.9)	1.93×10^{-13}	26.7 (9.3–76.5)	7.41×10^{-13}
				Replication	10 (20.0)	37 (74.0)	3 (6.0)	101 (82.8)	21 (17.2)	0 (0.0)	73 (82.0)	16 (18.0)	0 (0.0)	19.2 (8.3–44.4)	5.46×10^{-15}	18.3 (7.6–44.0)	8.37×10^{-13}
				Combined	30 (23.4)	91 (71.1)	7 (5.5)	157 (84.4)	29 (15.6)	0 (0.0)	119 (85.0)	21 (15.0)	0 (0.0)	17.7 (10.0–31.3)	2.84×10^{-27}	18.5 (10.0–34.4)	3.99×10^{-24}
rs8099917	*IL28B*	0.12 (G)	T/G	GWAS	19 (24.4)	56 (71.8)	3 (3.8)	58 (90.6)	6 (9.4)	0 (0.0)	47 (92.2)	4 (7.8)	0 (0.0)	30.0 (11.2–80.5)	3.11×10^{-15}	36.5 (11.6–114.6)	5.00×10^{-14}
				Replication	11 (22.0)	37 (74.0)	2 (4.0)	108 (88.5)	14 (11.5)	0 (0.0)	78 (87.6)	11 (12.4)	0 (0.0)	27.4 (11.5–65.3)	9.47×10^{-18}	25.1 (10.0–63.1)	1.00×10^{-14}
				Combined	30 (23.4)	93 (72.7)	5 (3.9)	166 (89.2)	20 (10.8)	0 (0.0)	125 (89.3)	15 (10.7)	0 (0.0)	27.1 (14.6–50.3)	2.68×10^{-32}	27.2 (13.9–53.4)	1.11×10^{-27}

[a]NVR, null virologic response; VR, virologic response; SVR, sustained virologic response. The 186 VRs consisted of 46 transient virologic response (TVRs) and 140 SVRs. [b]Minor allele frequency and minor allele in 184 healthy Japanese individuals[15]. The MAF of the SNPs in SVR is similar to that of TVR group, whereas that of NVR is much higher (76.6%). [c]Odds ratio for the minor allele in a dominant model. [d]*P* value by χ^2 test for the minor allele dominant model.

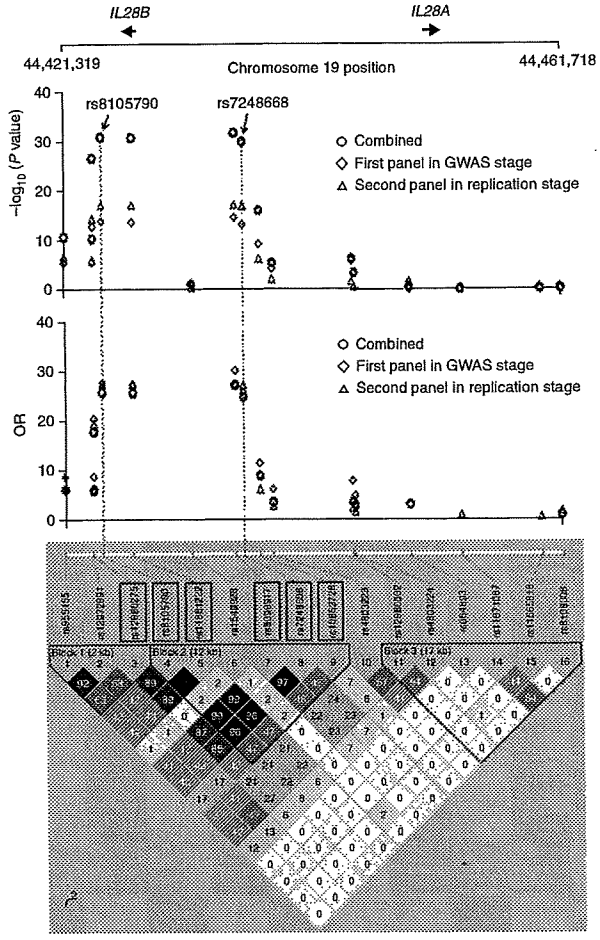

Figure 2 Genomic structure, *P* value and OR plots in association analysis and LD map around *IL28B* and *IL28A* (chr.19, nucleotide positions 44421319–44461718; build 35). *P* values by the χ^2 test for minor allele dominant effect model are shown for the first panel of 142 samples in the GWAS stage, the second panel of 172 samples in the replication stage, and the combined analysis. Below are estimates of pairwise r^2 for 16 SNPs selected in the replication study using a total of 314 Japanese patients with HCV treated with PEG-IFN-α/RBV. Boxes indicate the significantly associated SNPs with response to PEG-IFN-α/RBV treatment both in the GWAS stage and in the replication stage. Dotted lines indicate the region with the strongest associations from the positions of rs8105790 to rs7248668.

We analyzed the region of ~40 kb (chr. 19, nucleotide positions 44421319–44461718; build 35) containing the significantly associated SNPs (rs12980275 and rs8099917) using Haploview software for linkage disequilibrium (LD) and haplotype structure based on the HapMap data for individuals of Japanese ancestry. The LD blocks were analyzed using the four-gamete rule, and four blocks were observed (**Supplementary Fig. 3**). We selected 16 SNPs for both replication study and high-density association mapping, including tagging SNPs estimated on the basis of the haplotype blocks, one SNP located within *IL28B* (rs11881222) and the significantly associated SNPs from the GWAS stage (rs12980275 and rs8099917) (**Supplementary Table 1**).

To validate the results of the GWAS stage, 16 SNPs selected for the replication stage, including the original SNPs, were genotyped using the DigiTag2 assay in an independent set of 172 Japanese patients with HCV treated with PEG-IFN-α/RBV treatment (50 NVR and 122 VR samples), together with the first panel of 142 samples analyzed in the GWAS stage (**Supplementary Table 1**). The associations of the original SNPs were replicated in the replication cohort of 172 patients ($P = 5.46 \times 10^{-15}$, OR = 19.2 for rs12980275; $P = 9.47 \times 10^{-18}$, OR = 27.4 for rs8099917; **Table 1**). The combined *P* values for both stages reached 2.84×10^{-27} (OR = 17.7; 95% CI = 10.0–31.3) and 2.68×10^{-32} (OR = 27.1; 95% CI = 14.6–50.3), respectively (**Table 1**). Notably, when we compared the SVR ($n = 140$) with the NVR group ($n = 128$), the original two SNPs (rs12980275 and rs8099917) again showed strong associations: both *P* values and ORs were similar to those observed in the comparison between VR and NVR, and the combined *P* values for both stages reached 3.99×10^{-24} (OR = 18.5; 95% CI = 10.0–34.4) and 1.11×10^{-27} (OR = 27.2; 95% CI = 13.9–53.4), respectively (**Table 1**). Comparing SVR ($n = 140$) versus NVR plus TVR ($n = 174$), we again found that these SNPs were significantly associated ($P = 1.71 \times 10^{-16}$, OR = 8.8; 95% CI 5.1–15.4 for rs12980275; $P = 1.18 \times 10^{-18}$, OR = 12.1; 95% CI 6.5–22.4 for rs8099917, **Supplementary Table 2**), suggesting that these SNPs would predict NVR as well as SVR before PEG-IFN-α/RBV therapy.

Among the newly analyzed SNPs in the replication study, six (rs12980275, rs8105790, rs11881222, rs8099917, rs7248668 and rs10853728) showed significant associations both in the GWAS stage ($P < 8.05 \times 10^{-8}$) and in the replication stage ($P < 0.0031$ (0.05/16)) after Bonferroni correction. These SNPs are located within a 15.7-kb region that includes *IL28B* (**Fig. 2** and **Supplementary Table 1**). In particular, the strongest associations with NVR were observed for four SNPs, rs8105790, rs11881222, rs8099917 and rs7248668, that are located in the downstream flanking region, the third intron and the upstream flanking region of *IL28B*. The combined *P* values for these polymorphisms were 1.98×10^{-31} (OR = 25.7; 95% CI = 13.9–47.6), 2.84×10^{-31} (OR = 25.6; 95% CI = 13.8–47.3), 2.68×10^{-32} (OR = 27.1; 95% CI = 14.6–50.3) and 1.84×10^{-30} (OR = 24.7; 95% CI = 13.3–45.8), respectively (**Supplementary Table 1**). We then sequenced this region to identify further variants and found three SNPs (rs8103142, rs28416813 and rs4803219) located in the third exon, the first intron and the upstream flanking region of *IL28B*, and a few infrequent variations. These SNPs also showed strong associations in the combined dataset of 128 NVR and 186 VR samples ($P = 1.40 \times 10^{-29}$, OR = 26.6 for rs8103142; $P = 5.52 \times 10^{-28}$, OR = 22.3 for rs28416813; $P = 2.45 \times 10^{-29}$, OR = 23.3 for rs4803219; **Supplementary Table 3**). We also performed LD and haplotype analyses with seven SNPs. These SNPs were in strong LD, and the risk haplotype showed a level of association similar to those of individual SNPs ($P = 1.35 \times 10^{-25}$, OR = 11.1; 95% CI = 6.6–18.6) (**Table 2**). These results suggest that the association with NVR was primarily driven by one of these SNPs.

Table 2 Association analysis of response to treatment by *IL28B* haplotype

SNP							Frequencies			
rs8105790	rs11881222	rs8103142	rs28416813	rs4803219	rs8099917	rs7248668	NVR group	VR group	*P* value	OR (95% CI)
T	**A**	**T**	C	C	T	G	0.543	0.942	1.81×10^{-32}	0.1 (0.04–0.12)
C	**G**	**C**	G	T	G	A	0.387	0.054	1.35×10^{-25}	11.1 (6.6–18.6)

Association analysis of haplotypes consisting of seven SNPs with response to PEG-IFN-α/RBV treatment in 314 Japanese patients with HCV. Boldface letters: rs11881222 (third intron); rs8103142 (third exon).

Table 3 Factors associated with NVR by logistic regression model

Factors	Odds ratio	95% CI	*P* value
rs8099917 (G allele)	37.68	16.71–83.85	<0.0001
Age	1.02	0.98–1.07	0.292
Gender (Female)	3.32	1.49–7.39	0.003
Re-treatment[a]	1.12	0.55–2.33	0.750
Platelet count	0.93	0.87–1.01	0.080
Aminotransferase level	1.00	0.99–1.00	0.735
Fibrosis stage[20]	1.10	0.73–1.66	0.658
HCV-RNA level	1.01	0.99–1.02	0.139

[a]Re-treatment, non-response to previous treatment with interferon-α (plus RBV).

To examine the relative contribution of factors associated with NVR, we used a logistic regression model. One tagging SNP located within *IL28B* (minor allele of rs8099917) was the most significant factor for predicting NVR, followed by gender (**Table 3**). Clinically, viral factors such as HCV genotype and HCV RNA level are important for the outcome of PEG-IFN-α/RBV therapy. Indeed, mean HCV-RNA level was significantly lower in SVR (SVR versus TVR, $P = 0.002$; SVR versus NVR, $P = 0.016$; **Supplementary Table 4**). Mean platelet count and the proportion of mild fibrosis (F1–F2) were significantly higher in SVR than in NVR.

Real-time quantitative PCR assays in peripheral blood mononuclear cells revealed a significantly lower level of *IL28* mRNA expression in individuals with the minor alleles (**Fig. 3**), suggesting that variant(s) regulating *IL28* expression is associated with a response to PEG-IFN-α/RBV treatment. *IL28B* encodes a cytokine distantly related to type I (α and β) interferons and the interleukin (IL)-10 family. This gene and *IL28A* and *IL29* (encoding IL-28A and IL-29, respectively) are three closely related cytokine genes that encode proteins known as type III IFNs (IFN-λs) and that form a cytokine gene cluster at chromosomal region 19q13 (ref. 16). The three cytokines are induced by viral infection and have antiviral activity[16,17]. All three interact with a heterodimeric class II cytokine receptor that consists of IL-10 receptor beta (IL10Rβ) and IL-28 receptor alpha (IL28Rα, encoded by *IL28RA*)[16,17], and they may serve as an alternative to type I IFNs in providing immunity to viral infection.

Notably, a recent report showed that the strong antiviral activity evoked by treating mice with TLR3 or TLR9 agonists was significantly reduced in both *IL28RA*$^{-/-}$ and *IFNAR*$^{-/-}$ mice, indicating that IFN-λ is important in mediating antiviral protection by ligands for TLR3 and TRL9 (ref. 18). IFN-λ induced a steady increase in the expression of a subset of IFN-stimulated genes, whereas IFN-α induced the same genes with more rapid and transient kinetics[19]. Therefore, it is possible that IFN-λ induces a slower but more sustained response that is important for TLR-mediated antiviral protection. This might be one of the ways that a genetic variant regulating *IL28* expression influences the response to PEG-IFN-α/RBV treatment. Further research will be required to fully understand the specific mechanism by which a genotype might affect the response to treatment.

In conclusion, the strongest associations with NVR were observed for seven SNPs, rs8105790, rs11881222, rs8103142, rs28416813, rs4803219, rs8099917 and rs7248668, that are located in the downstream flanking region, the third intron, the third exon, the first intron and the upstream flanking region of *IL28B*. Further studies following our report of this robust genetic association to NVR may make it possible to develop a pre-treatment predictor of which individuals are likely to respond to PEG-IFN-α/RBV treatment. This would remove the need for the initial 12–24 weeks of treatment that is currently used as a basis for a clinical decision about whether treatment should be continued. That would allow better targeting of PEG-IFN-α/RBV treatment, avoiding the unpleasant side effects that commonly accompany the treatment where it is unlikely to be beneficial, and reduce overall treatment costs. Because of the small number of samples in this study, we plan to conduct a further prospective multicenter study to establish these SNPs as a clinically useful marker.

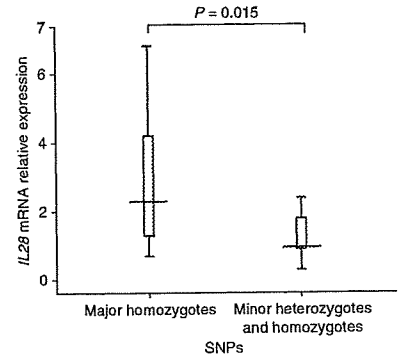

Figure 3 Quantification of *IL28* mRNA expression. The expression level of *IL28* genes was determined by real-time quantitative RT-PCR using RNA purified from peripheral blood mononuclear cells. Distribution of relative gene expression levels was compared between the individuals homozygous for major alleles ($n = 10$) and the heterozygous or homozygous individuals carrying minor alleles ($n = 10$) of rs8099917 by using the Mann-Whitney *U*-test. The bars indicate the median. All samples were obtained from HCV-infected patients before PEG-IFN-α/RBV therapy.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This study was supported by a grant-in-aid from the Ministry of Health, Labour, and Welfare of Japan (H19-kannen-013). This study is based on 15 multicenter hospitals throughout Japan, in the Hokkaido area (Hokkaido University Hospital), Kanto area (Saitama University Hospital; Konodai Hospital; Musashino Red Cross Hospital; Tokyo Medical and Dental University Hospital), Koshin area (Shinshu University Hospital; Kanazawa University Hospital), Tokai area (Nagoya City University Hospital), Kinki area (Kyoto Prefectural University of Medicine Hospital; National Hospital Organization Osaka National Hospital; Hyogo College of Medicine Hospital) and Chugoku/Shikoku area (Tottori University Hospital; Ehime University Hospital; Yamaguchi University Hospital; Kawasaki Medical College Hospital). We thank Y. Uehara-Shibata, Y. Ogasawara, Y. Ishibashi and M. Yamaoka-Sageshima (Tokyo University) for technical assistance; A. Matsumoto (Shinshu), K. Naiki (Saitama), K. Nishimura (Kyoto), H. Enomoto (Hyogo), K. Oyama (Tottori) and the Ochanomizu Liver Conference Study Group for collecting samples; M. Watanabe (Tokyo Medical and Dental University), S. Kaneko (Kanazawa University) and M. Onji (Ehime University) for their advice throughout the study; and H. Ito (Aichi Cancer Center) for conducting statistical analyses.

AUTHOR CONTRIBUTIONS

Study design and discussion: Y.T., N.N., N.M., K.T., M.M.; sample collection: Y.T., M.K., K.M., N.S., M.N., M.K., K.H., S.H., Y.I., E.M., E.T., S.M., Y.M., M.H., A.S., Y.H., S.N., I.S., M.I., K.I., K.Y., F.S., N.I.; genotyping: N.N.; statistical analysis: N.N., A.K., K.I.; quantitative RT-PCR: M.S.; manuscript writing: Y.T., N.N., K.T., M.M.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Ray Kim, W. Global epidemiology and burden of hepatitis C. *Microbes Infect.* **4**, 1219–1225 (2002).
2. Manns, M.P. *et al.* Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* **358**, 958–965 (2001).

3. Fried, M.W. *et al.* Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N. Engl. J. Med.* **347**, 975–982 (2002).
4. Hadziyannis, S.J. *et al.* Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann. Intern. Med.* **140**, 346–355 (2004).
5. Bruno, S. *et al.* Peginterferon alfa-2b plus ribavirin for naive patients with genotype 1 chronic hepatitis C: a randomized controlled trial. *J. Hepatol.* **41**, 474–481 (2004).
6. Sezaki, H. *et al.* Poor response to pegylated interferon and ribavirin in older women infected with hepatitis C virus of genotype 1b in high viral loads. *Dig. Dis. Sci.* **54**, 1317–1324 (2009).
7. Fried, M.W. Side effects of therapy of hepatitis C and their management. *Hepatology* **36**, S237–S244 (2002).
8. Pascu, M. *et al.* Sustained virological response in hepatitis C virus type 1b infected patients is predicted by the number of mutations within the NS5A-ISDR: a meta-analysis focused on geographical differences. *Gut* **53**, 1345–1351 (2004).
9. Shirakawa, H. *et al.* Pretreatment prediction of virological response to peginterferon plus ribavirin therapy in chronic hepatitis C patients using viral and host factors. *Hepatology* **48**, 1753–1760 (2008).
10. Akuta, N. *et al.* Predictive factors of early and sustained responses to peginterferon plus ribavirin combination therapy in Japanese patients infected with hepatitis C virus genotype 1b: amino acid substitutions in the core region and low-density lipoprotein cholesterol levels. *J. Hepatol.* **46**, 403–410 (2007).
11. Walsh, M.J. *et al.* Non-response to antiviral therapy is associated with obesity and increased hepatic expression of suppressor of cytokine signalling 3 (SOCS-3) in patients with chronic hepatitis C, viral genotype 1. *Gut* **55**, 529–535 (2006).
12. Gao, B., Hong, F. & Radaeva, S. Host factors and failure of interferon-alpha treatment in hepatitis C virus. *Hepatology* **39**, 880–890 (2004).
13. Matsuyama, N. *et al.* The dinucleotide microsatellite polymorphism of the IFNAR1 gene promoter correlates with responsiveness of hepatitis C patients to interferon. *Hepatol. Res.* **25**, 221–225 (2003).
14. Tsukada, H. *et al.* A polymorphism in MAPKAPK3 affects response to interferon therapy for chronic hepatitis C. *Gastroenterology* **136**, 1796–1805 (2009).
15. Nishida, N. *et al.* Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals. *BMC Genomics* **9**, 431 (2008).
16. Kotenko, S.V. *et al.* IFN-λs mediate antiviral protection through a distinct class II cytokine receptor complex. *Nat. Immunol.* **4**, 69–77 (2003).
17. Sheppard, P. *et al.* IL-28, IL-29 and their class II cytokine receptor IL-28R. *Nat. Immunol.* **4**, 63–68 (2003).
18. Ank, N. *et al.* An important role for type III interferon (IFN-lambda/IL-28) in TLR-induced antiviral activity. *J. Immunol.* **180**, 2474–2485 (2008).
19. Marcello, T. *et al.* Interferons alpha and lambda inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. *Gastroenterology* **131**, 1887–1898 (2006).
20. Desmet, V.J., Gerber, M., Hoofnagle, J.H., Manns, M. & Scheuer, P.J. Classification of chronic hepatitis: diagnosis, grading and staging. *Hepatology* **19**, 1513–1520 (1994).

J Gastroenterol (2009) 44:952–963
DOI 10.1007/s00535-009-0087-x

ORIGINAL ARTICLE—LIVER, PANCREAS, AND BILIARY TRACT

Predictive values of amino acid sequences of the core and NS5A regions in antiviral therapy for hepatitis C: a Japanese multi-center study

Takeshi Okanoue · Yoshito Itoh · Hiroaki Hashimoto · Kohichiroh Yasui · Masahito Minami · Tetsuo Takehara · Eiji Tanaka · Morikazu Onji · Joji Toyota · Kazuaki Chayama · Kentaro Yoshioka · Namiki Izumi · Norio Akuta · Hiromitsu Kumada

Received: 31 March 2009 / Accepted: 20 April 2009 / Published online: 11 June 2009

Abstract
Background Chronic hepatitis C (CHC) genotype 1b patients with high viral load are resistant to peginterferon (PEG-IFN) and ribavirin (RBV) combination therapy, especially older and female patients.
Methods To elucidate the factors affecting early and sustained viral responses (EVR and SVR), 409 genotype 1b patients CHC with high viral loads who had received 48 weeks of PEG-IFN/RBV therapy were enrolled. The amino acid (aa) sequences of the HCV core at positions 70 and 91 and of the interferon sensitivity determining region (ISDR) were analyzed. Host factors, viral factors, and treatment-related factors were subjected to multivariate analysis.
Results Male gender, low HCV RNA load, high platelet count, two or more aa mutations of ISDR, and wild type of core aa 70 were independent predictive factors for SVR. In patients with over 80% adherences to both PEG-IFN and RBV, male gender, mild fibrosis stage, and wild type of core aa 70 were independent predictors for SVR.
Conclusions Independent predictive factors for SVR were: no aa substitution at core aa 70, two or more aa mutations in the ISDR, low viral load, high values of platelet count, mild liver fibrosis and male gender.

T. Okanoue (✉)
Hepatology Center, Saiseikai Suita Hospital,
1-2 Kawazonocho, Suita 564-0013, Japan
e-mail: okanoue@suita.saiseikai.or.jp

T. Okanoue · Y. Itoh · H. Hashimoto · K. Yasui · M. Minami
Department of Gastroenterology and Hepatology,
Kyoto Prefectural University of Medicine,
Kawaramachi-Hirokoji Kamigyo-Ku,
Kyoto 602-8566, Japan

Y. Itoh
e-mail: yitoh@koto.kpu-m.ac.jp

H. Hashimoto
e-mail: road1820@yahoo.co.jp

K. Yasui
e-mail: yasuik@koto.kpu-m.ac.jp

M. Minami
e-mail: minami@koto.kpu-m.ac.jp

T. Takehara
Department of Gastroenterology and Hepatology,
Osaka University Graduate School of Medicine,
2-2 Yamadaoka, Suita 565-0871, Japan
e-mail: takehara@gh.med.osaka-u.ac.jp

E. Tanaka
Department of Medicine, Shinshu University School
of Medicine, 3-1-1 Asahi, Matsumoto 390-8621, Japan
e-mail: etanaka@shinshu-u.ac.jp

M. Onji
Department of Gastroenterology, Ehime University,
454 Shizukawa, Tohon, Ehime 791-0295, Japan
e-mail: onjimori@m.med.ehime-u.ac.jp

J. Toyota
Departmet of Gastroenterology, Sapporo-Kosei General
Hospital, Kita Sanjyo, Chuo-ku, Sapporo 060-0033, Japan
e-mail: joji.toyota@ja_hokkaidoukouseiren.or.jp

K. Chayama
Department of Medicine and Molecular Science,
Graduate School of Science, Hiroshima University,
1-2-3 Kasumi, Minami-ku, Hiroshima,
Hiroshima 734-8551, Japan
e-mail: chayama@hiroshima-u.ac.jp

K. Yoshioka
Department of Hepato-Biliary-Pancreas, Fujita Health Science,
Kutsukake-cho, Toyoake 470-1192, Japan
e-mail: kyoshiok@fujita-hu.ac.jp

Keywords Chronic hepatitis C · Peginterferon and ribavirin · Core amino acid · Interferon sensitivity determining region

Abbreviations

CHC	Chronic hepatitis C
PEG-IFN	Peginterferon
RBV	Ribavirin
RVR	Rapid viral response
cEVR	Complete early viral response
LVR	Late viral response
ETR	End of treatment response
NR	Non response
SVR	Sustained viral response
ISDR	Interferon sensitivity determining region
Aa	Amino acid
ALT	Alanine aminotransferase
PLT	Platelet
HCC	Hepatocellular carcinoma

Introduction

A combination of pegylated interferon (PEG-IFN) and ribavirin (RBV) therapy for 48 weeks achieves a sustained viral response (SVR) rate of 40–50% in chronic hepatitis C (CHC) patients with a high viral load of genotype 1 [1–4]. The dose-reduction rate and the frequency of discontinuation of this treatment are high in aged patients [5]. The SVR rate of the therapy is lower in females than males, especially in older patients in Japan [6].

Around 30% of HCV carriers have serum alanine aminotransferase (ALT) levels within the upper limit of normal ranges [7, 8] and HCV carriers with persistently normal serum ALT (PNALT) and serum platelet (PLT) counts of over $15 \times 10^4/mm^3$ show low grade hepatic fibrosis and good prognosis [9]. Before treating HCV carriers, it is very important to predict non-response to PEG-IFN plus RBV therapy because of its medical cost, adverse effects, and its impact on the long term quality of life.

N. Izumi
Department of Gastroenterology and Hepatology, Musashino Red Cross Hospital, Sakaiminamimachi, Musashino 180-8610, Japan
e-mail: nizumi@musashino.jrc.or.jp

N. Akuta · H. Kumada
Department of Hepatology, Toranomon Hospital, Kajigaya, Takatsu-ku, Kawasaki 213-8587, Japan

N. Akuta
e-mail: akuta-gi@umin.ac.jp

H. Kumada
e-mail: kumahiro@toranomon.gr.jp

There are many factors affecting response to IFN monotherapy and PEG-IFN/RBV therapy, including body mass index (BMI) [10, 11], steatosis [12, 13], insulin resistance [14], stage of liver fibrosis [15, 16], total cholesterol (T. Chol), triglyceride (TG), adherence to both PEG-IFN and RBV [17], race [18, 19], age [1, 2, 20], and viral factors including serum quantity of HCV RNA, HCV genotype and substitution of amino acids (aa) in the interferon sensitivity determining region (ISDR, 2209–2248) of the nonstructural protein 5A (NS5A) [21] and in the core protein [22, 23]. Early viral response is an important predictive factor in PEG-IFN/RBV therapy for CHC patients with genotype 1 and high viral loads [24–27].

The aim of this study was to elucidate the valuable predictive factors of SVR in Japanese patients with HCV genotype 1b high viral loads following 48 weeks of PEG-IFN/RBV therapy, focusing on the relationship between aa substitutions in the ISDR and at core aa 70 and 91 and early viral kinetics.

Patients and methods

Selection of patients

This retrospective study was conducted at 15 clinical sites in Japan which are part of the Study Group of Optimal Treatment of Viral Hepatitis supported by the Ministry of Health, Labor and Welfare, Japan. Eligible subjects were CHC patients, who (1) had received liver biopsy; (2) were genotype 1b with high viral load (≥100 KIU/ml by Cobas Amplicor Hepatitis C Virus Test, version 2.0) at the start of PEG-IFN/RBV therapy; (3) received weekly injections of PEG-IFN-α-2b (PEG-INTRON; Shering-Plough, Kenilworth, NJ) of 1.5 μg/kg bw and oral administration of RBV (Rebetol; Shering-Plough) for 48 weeks. The amount of RBV was adjusted based on the subject's body weight; (600 mg for ≤60 kg bw, 800 mg for 60–80 kg bw, 1,000 mg for > 80 kg bw); (4) were examined serially for quantitative and qualitative HCV RNA; and (5) the aa sequences at positions 70 and 91 in the core region and of the ISDR in the NS5A had been determined in pretreatment sera.

Hepatitis B virus (HBV) infection, human immunodeficiency virus (HIV) infection, autoimmune hepatitis, primary biliary cirrhosis, hemochromatosis, and Wilson's disease were excluded. Histopathological diagnosis was based on the scoring system of Desmet et al. [28]. The definition of alcohol abuse included patients having a history of more than 100 kg of total ethanol intake. Complete blood counts, liver function tests, serum lipids, serum ferritin, serum fibrosis markers, fasting plasma glucose (FPG), and immune reactive insulin (IRI) were examined in most cases. Written informed consent was obtained from all

patients before treatment, and the protocol was approved by the ethics committees in each site.

Study design

Four hundred and nine patients who completed 48 weeks of treatment and were followed for more than 24 weeks after treatment were enrolled in the first study (*Study design 1*).

To elucidate the effect of aa substitution of HCV core and in the ISDR on HCV dynamics, including a rapid viral response (RVR), complete early viral response (cEVR), a late viral response (LVR) and SVR, according to gender and age (<60 years ≥ 60 years), 201 of the 409 patients maintaining over 80% adherences to both PEG-IFN and RBV were enrolled in the second study (*Study design 2*).

Nucleotide sequencing of the core and NA5A gene

The nucleotide sequences encoding aa 1–191 (HCV core) and aa 2209–2248 (ISDR) were analyzed by direct sequencing as described by Akuta et al. [22, 27] and Enomoto et al. [21]. In brief, RNA was extracted from the sera and converted to cDNA and two nested rounds of polymerase chain reaction (PCR) were performed. Primers used in the PCR were as follows; (a) Nucleotide sequences of the core region: the first-round PCR was performed with CC11 (sense) and e14 (antisense) primers [22, 27], and the second-round PCR with CC9 (sense) and e14 (antisense) primers [22, 27]. (b) Nucleotide sequences of the ISDR in NS5A: the first-round PCR was performed with ISDR1 (sense) and ISDR2 (antisense) primers [21], and the second-round PCR with ISDR3 (sense) and ISDR4 (antisense) primers [21]. These sequences were compared with the consensus sequence of genotype 1b (HCV-J) [29]. Wild types virus encoded arginine and leucine at aa 70 and 91, respectively, and the aa substitutions were glutamine or histidine at aa 70 and methionine at aa 91.

Viral kinetic study

Serum HCV RNA levels were measured by PCR (Amplicor HCV RNA kit, version 2.0, Roche Diagnostics) using samples taken before treatment and at 4, 12, 24, and 48 weeks after the therapy. SVR was defined as HCV RNA negativity by qualitative analysis by PCR at 24 weeks after the treatment. RVR was defined as HCV RNA negativity at 4 weeks, cEVR as HCV RNA negativity at 12 weeks, LVR as HCV RNA negativity during 13–24 weeks and an end of treatment response (ETR) as HCV RNA negativity at the end of treatment. Patients who remained positive for HCV RNA at the end of the treatment and at 24 weeks after the therapy were defined as non-responders (NR).

Adherences to PEG-IFN and RBV

Adherences to PEG-IFN and RBV were assessed by separately calculating the actual doses of PEG-IFN and RBV received as percentages of the intended dosages. Adherences to PEG-IFN and RBV were divided into two groups; 80%≤ and <80%.

Statistical analysis

All data analyses were conducted using the SAS version 9.1.3 statistical analysis packages (SAS Institute, Cary, NC, USA). Individual characteristics between groups were evaluated by Mann–Whitney U test for numerical variables or Fisher's exact test for categorical variables. Variables exhibiting values of $p < 0.1$ in the univariate analysis were subjected to stepwise multivariate logistic regression analysis. The grade of steatosis and iron deposition in liver tissue, BMI, albumin (Alb), low density lipoprotein-cholesterol (LDL-C), homeostasis model assesment-insulin resistance (HOMA-IR), ferritin, and hyaluronic acid were excluded from multivariate logistic regression analysis because of the absence of those data in more than 10% of the patients. All p values of $p < 0.05$ by the two-tailed test were considered statistically significant.

Results

Study design 1

Baseline backgrounds, characteristics and adherences of peginterferon and ribavirin in males and females

The treatment outcome of PEG-IFN and RBV combination therapy depends on gender in Japanese patients, so in addition to aa substitutions in the ISDR in NS5A [21] or at HCV core 70 and 91 [22, 27], we compared the baseline characteristics according to gender (Table 1). Males were younger and the grade of hepatic inflammation was milder in males. The serum levels of LDL-C, PLT count, and aa substitutions of ISDR and at core 70 and 91 did not differ significantly different between males and females. The frequency of no alcohol abuse was significantly ($p < 0.0001$) higher in females than males (Some of them are not described in Table 1).

The rates of over 80% adherences to PEG-IFN and RBV were significantly lower ($p = 0.0066$, $p < 0.00001$, respectively) in females than males. Only in those above 60 years did the rate of over 80% adherence to PEG-IFN not differ significantly between males and females, but the rate of over 80% adherence to RBV was significantly lower ($p = 0.035$) in females than males (Table 1).

Table 1 Backgrounds and characteristics of male and female patients

Factors	Gender		*p* value
	Male	Female	
No. of patients	256 (62.6%)	153 (37.4%)	
Age			
Median (range)	53 (18–73)	59 (23–75)	0.00001
F stage			
F0–2	206 (80.5%)	119 (77.8%)	0.592
F3–4	50 (19.5%)	34 (22.2%)	
Grade (A factor)			
A0–1	163 (63. 7%)	79 (51.6%)	0.026
A2–3	93 (36.3%)	74 (48.4%)	
HCV RNA load 0 week (KIU/mL)			
Median (range)	1500 (100–5000 <)	1280 (100–5000<)	0.384
ALT 0 week (IU/L)			
Median (range)	74.5 (16–504)	59 (19–391)	0.001
BMI			
Median (range)	23.6 (17.5–31.2)	22.1 (16.1–33.9)	0.00033
Alb (g/dL)			
Median (range)	4.0 (3.0–5.2)	3.8 (3.0–4.8)	0.011
LDL-C (mg/dL)			
Median (range)	97 (30–185)	90 (34–174)	0.612
T-Chol (mg/dL)			
Median (range)	167 (85–273)	176 (114–261)	0.0016
PLT count ($\times 10^4/mm^3$)			
Median (range)	17.0 (8.0–31.9)	16.4 (8.1–39.9)	0.350
Amino acid mutation of ISDR			
0–1	200 (78.1%)	121 (79.1%)	0.608
2≤	56 (21.9%)	32 (20.9%)	
Amino acid substitution of core 70			
Wild	177 (69.1%)	114 (74.5%)	0.261
Mutant	79 (30.9%)	39 (25.5%)	
Amino acid substitution of core 91			
Wild	153 (59.8%)	98 (64.1%)	0.403
Mutant	103 (40.2%)	55 (35.9%)	
PEG-IFN adherence			
<80%	41 (17.7%)	42 (30.4%)	0.0066
80%≤	190 (82.3%)	96 (69.6%)	
Ribavirin adherence			
<80%	54 (23.6%)	73 (52.1%)	<0.00001
80%≤	175 (76.4%)	67 (47.9%)	
Age: <60 years			
PEG adherence			
<80%	30 (17.8%)	23 (31.5%)	0.027
80%≤	139 (82.2%)	50 (68.5%)	
Ribavirin adherence			
<80%	27 (16.2%)	31 (42.5%)	0.000029
80%≤	140 (83.8%)	42 (57.5%)	
Age: 60 years≤			
PEG adherence			
<80%	11 (17.7%)	19 (29.2%)	0.147
80%≤	51 (82.3%)	46 (70.8%)	
Ribavirin adherence			
<80%	27 (43.5%)	42 (62.7%)	0.035
80%≤	35 (56.5%)	25 (37.3%)	

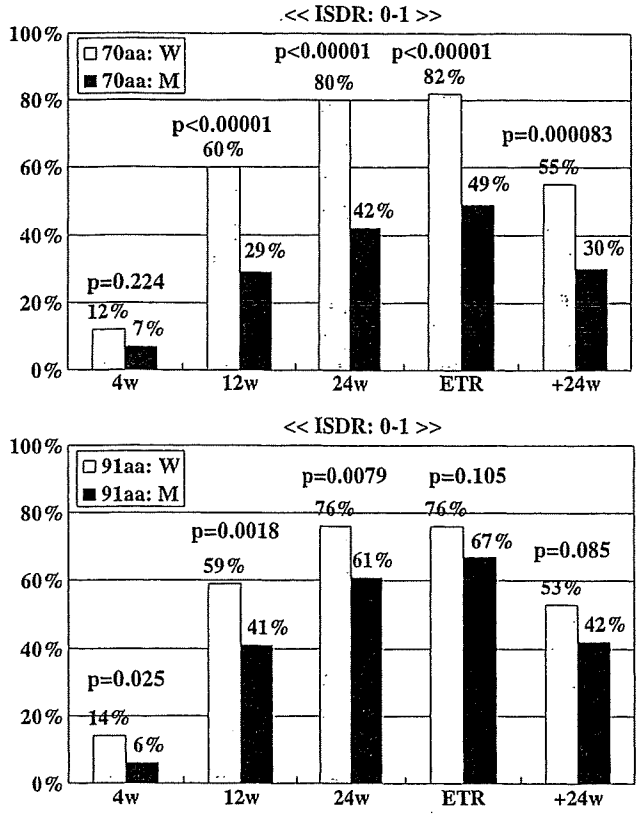

Fig. 1 Relationship between time course of serum HCV RNA negativity and amino acid substitutions in the ISDR and core amino acids 70 and 91. For cases with no or only one amino acid (aa) change in the ISDR, the rates of cEVR, LVR, ETR and SVR were significantly higher in patients with wild type core aa 70 but only the rates of RVR, cEVR, and LVR were significantly higher in patients with wild type core aa 91

Amino acid substitutions

There were no significant differences in the frequency of aa substitutions in the ISDR between males and females. Core aa substitutions at positions 70 and 91 were as follows; 291 (71.1%) were wild type and 118 (28.9%) were mutant at core aa 70, and 251 (61.4%) were wild type and 158 (38.6%) were mutant at core aa 91. There were no significant differences between males and females and between patients below and above 60 years of age.

Virological responses and aa substitutions

The rate of RVR did not differ significantly between males and females. However, more male patients showed HCV RNA negativity at 12 weeks (males vs. females; 60.7 vs. 48.4%, $p = 0.018$), 24 weeks (76.8 vs. 64.2%, $p = 0.0078$) and 48 weeks (78.2 vs. 68.6%, $p = 0.049$), and the proportion of male patients in SVR was significantly higher than that of females (61.3 vs. 37.3%, $p < 0.00001$).

RVR, cEVR and SVR rates were significantly higher in patients with two or more aa mutations in the ISDR compared to patients having no or one aa substitution in that region (20 vs. 11%, $p = 0.044$; 71 vs. 52%, $p = 0.0021$; 66 vs. 49%, $p = 0.0054$, respectively). AA substitution at core position 70 resulted in significantly lower rate of cEVR, LVR, ETR, and SVR (40 vs. 63%, $p = 0.000037$; 51 vs. 81%, $p < 0.00001$; 56 vs. 83%, 41 vs. 57%; $p < 0.00001$, $p = 0.0031$, respectively). Although the patients with the wild type aa at core 91 showed significantly higher rates of RVR and cEVR, the rate of SVR was not significantly higher in those patients ($p = 0.054$).

SVR rates were 30% for patients with no or one aa substitution in the ISDR and the core aa 70 substitution, and were significantly lower compared to those with the wild type aa core 70 (Fig. 1). These findings were not confirmed in patients with no or one aa substitution in the ISDR and the core aa 91 substitution (Fig. 1).

Factors affecting SVR by univariate analysis

Univariate analysis identified nine parameters that influenced non-SVR significantly: female gender, older age, advanced staged liver fibrosis, high viral load, low serum Alb level, low PLT count, no or one aa substitution in the ISDR, aa substitution at core aa 70, and low adherence to RBV (Table 2). The frequency of steatosis and HOMA-IR were significantly ($p = 0.0057$, $p < 0.00001$, respectively) lower in patients with SVR compared with non-SVR (data not shown). However, these factors were not entered in the multivariate analysis because of the absence of the data in many cases.

Factors affecting RVR, cEVR, and SVR by multivariate logistic regression analysis

Multivariate analysis identified four parameters that influenced RVR independently: low HCV RNA load, low serum ALT level, two or more aa mutations in the ISDR and the wild type aa at core position 91 (Table 3).

Concerning cEVR, male gender, mild fibrosis stage, low HCV RNA load, two or more aa mutations in the ISDR, and the wild type aa at core positions 70 and 91 were independent predictors (Table 3).

Concerning SVR, male gender ($p < 0.0001$), low HCV RNA load ($p = 0.013$), high PLT count ($p = 0.0019$), two or more aa mutations in the ISDR ($p = 0.024$), and wild type core aa 70 ($p = 0.0045$) were found to be independent predictors (Table 3).

The predictive values of the combination of gender, PLT count, ISDR and core aa 70 are shown in Fig. 2a. In male patients having PLT of $<15 \times 10^4/mm^3$, and, no or one aa substitution in the ISDR, the SVR rate was 68% when core 70

Table 2 Univariate analysis to identify the factors of SVR

Factors	Negative of HCV RNA after 24 weeks		*p* value
	(−)	(+)	
No. of patients	214 (52.3%)	195	
Gender			
Male	157 (61.3%)	99	<0.00001
Female	57 (37.3%)	96	
Age			
Median (range)	52.5 (18–75)	58 (20–74)	<0.00001
<60 years	155 (58.1%)	112	0.0018
60 years≤	59 (41.5%)	83	
Age: <60 years			
Male	118 (63.4%)	68	0.010
Female	37 (45.7%)	44	
Age: 60 years≤			
Male	39 (55.7%)	31	0.0011
Female	20 (27.8%)	52	
F stage			
F0–2	190 (58.5%)	135	0.000013
F3–4	25 (29.8%)	59	
Grade (A factor)			
A0–1	138 (56.8%)	104	0.130
A2–3	81 (48.5%)	86	
HCV RNA load 0 week (KIU/mL)			
Median (range)	1300 (100–5000<)	1700 (130–5000<)	0.016
ALT 0 week (IU/L)			
Median (range)	66 (16–391)	67 (19–504)	0.892
BMI			
Median (range)	23.0 (17.3–32.4)	23.25 (16.1–33.9)	0.714
Alb (g/dL)			
Median (range)	4.0 (3.2–5.2)	3.8 (3.0–4.8)	0.0088
LDL-C (mg/dL)			
Median (range)	94.5 (31–185)	97.5 (30–182)	0.611
T-Chol (mg/dL)			
Median (range)	169.5 (85–257)	170 (103–273)	0.511
PLT count ($\times 10^4$/mm^3)			
Median (range)	18.2 (8.7–39.9)	15.1 (8.0–31.9)	<0.00001
<15	54 (36.5%)	94	<0.00001
15≤	160 (61.3%)	101	
Amino acid mutation of ISDR			
0–1	156 (48.6%)	165	0.0054
2≤	58 (65.9%)	30	
Amino acid substitution of core 70			
Wild	166 (57.0%)	125	0.0031
Mutant	48 (40.7%)	70	
Amino acid substitution of core 91			
Wild	141 (56.2%)	110	0.054
Mutant	73 (46.2%)	85	
PEG-IFN adherence			
<80%	35 (42.2%)	48	0.063
80%≤	154 (53.8%)	132	
Ribavirin adherence			
<80%	55 (43.3%)	72	0.048
80%≤	132 (54.5%)	110	