RNA replication kinetics, genetic polymorphism and selection in the case of the hepatitis C virus

Michael P. H. Stumpf^{1*} and Nicole Zitzmann²

¹Department of Zoology, South Parks Road, Oxford OX1 3PS, UK

²Oxford Glycobiology Institute, South Parks Road, Oxford OX1 3QU, UK

We show in a simple theoretical quasispecies model that the replication dynamics of hepatitis C virus and a related model-system, the bovine viral diarrhoea virus, result in an effective reduction of RNA templates in infected cells. Viral fitness does not translate directly into RNA sequence replication efficiency, and hence the abundance of the viral master sequences diminishes over time. Our results suggest that genes not involved in RNA replication accumulate mutations over time because they do not undergo selection during this phase. The selection of viral RNA occurs not only during replication but also during the ensuing stages of the viral life cycle: (i) envelopment of viral RNA and (ii) successful infection of other cells, which also requires functionality of non-replicative genes. In particular, viral fitness requires the ability of the genome to encode structural proteins which do not encounter selective pressure during RNA replication. We conclude by discussing the potential value of antiviral drugs which inhibit selection on parts of the viral genome.

Keywords: RNA replication; selection; fitness; sequence evolution

1. INTRODUCTION

Since its conception more than 20 years ago, quasispecies theory (Eigen & Schuster 1977) has proved its merit as a theoretical framework for the study of primitive evolutionary processes. Theoretical work on life in the RNA world has focused almost entirely on the investigation of quasispecies; RNA sequences are rather compact and contain only tiny amounts of non-coding regions, such that the influence of neutral mutations has so far been deemed negligible or easily incorporated in generalized quasispecies models (Huynen *et al.* 1996; Schuster 1997; Schuster & Fontana 1999). For the same reason of genetic compactness, quasispecies theory has also been applied to the dynamics of the genetically compact RNA viruses (Domingo *et al.* 1978). Apart from the human immunodeficiency virus (HIV) (Nowak *et al.* 1990; Chichutek *et al.* 1992; Karlsson *et al.* 1998), the hepatitis C virus (HCV) has been the most thoroughly screened human virus for quasispecies characteristics (Martell *et al.* 1992; Kurosaki *et al.* 1994; Sullivan *et al.* 1998). Several studies have reported that RNA viruses do form a quasispecies (Domingo *et al.* 1978; Martell *et al.* 1992; Holland *et al.* 1992) and we refer to Domingo *et al.* (1999) for a recent review of viral quasispecies.

A quasispecies is the outcome of a mutation–selection process which results in a heterogeneous equilibrium distribution of genome sequences: sequences with higher fitness *a* show a higher abundance than those with lower fitness, and the distribution of sequences in sequence space is generally clustered around the sequences with the highest fitness. The sequence or degenerate set of sequences with the highest fitness is referred to as the master sequence. During replication the master sequences generate mutants with a generally lower fitness, but which, nevertheless, maintain a constant frequency in the sequence population.

Members of the Flaviviridae are small, enveloped, positive-stranded RNA viruses which encode all their proteins in a single open reading frame (ORF) (figure 1). Structural (S) proteins are located towards the N terminus and non-structural (NS) proteins are situated in the C-terminal half. Upon entrance into the cell the viral RNA is released and replicated outside the nucleus in the cytoplasm. The polycistronic viral mRNA is translated into a single polyprotein precursor which is subsequently cleaved by a combination of host and viral proteinases (Liu *et al.* 1997). The envelope proteins E1 and E2 are localized in the endoplasmic reticulum (ER), and it is thought that new RNA replicates are packed into the envelopes that bud from the ER.

The reason for focusing on HCV and its model system, bovine viral diarrhoea virus (BVDV), apart from the clinical relevance, lies in the fact that replication occurs in the cytoplasm and we are dealing effectively with replicating RNA templates (Lohmann *et al.* 1996; Liu *et al.* 1997). In principle, this can be thought of as being very similar to RNA replication in solution (Eigen & Schuster 1977). After each replication, a sequence can either remain in the template pool or it can be recruited as a viral genome; in the latter case it is packed into a viral envelope which then buds into the lumen of the ER. The virus then leaves the cell via the secretory pathway and can infect other cells. During copying of the RNA templates the sequences accumulate mutations because the RNA polymerase lacks a proofreading mechanism. The high mutation rates result in a trade-off between keeping sequences inside the cell as templates and spreading infection to other cells. The aim of this investigation is to determine the level at which fitness selection of positive-stranded RNA viruses, including HCV and BVDV, occurs. We envisage a twofold mechanism: at the first level it is solely the RNA replicates which compete for cellular resources, while at the second level viruses compete for uninfected cells. Although the latter is well described by the standard model of virus dynamics

*Author for correspondence (michael.stumpf@zoo.ox.uk).

Figure 1. Members of the *Flaviviridae* such as (a) HCV and (b) BVDV have one single ORF. The genes for structural proteins *C-p⁷* are situated towards the N terminus while genes coding for the proteins necessary for replication are found towards the C terminus of the ORF. Processing of the viral polyprotein requires both host and viral proteinases.

(Nowak & Bangham 1996), the first step has, to our knowledge, not so far received any attention.

We will simplify the theoretical model by dividing viral genes into two categories: first, genes which are essential for replication (which may include the 5' and 3' ends of the genome) and, second, genes which do not undergo selection during replication but may experience selective pressure in the ensuing lifestages; for convenience these genomic regions will be referred to as the replicative and structural parts of the genome, respectively. In the following discussion we shall approach in turn RNA replication and then virion selection dynamics. The former deals solely with the replication of virus-associated RNA sequences, while the latter step includes enveloping of suitable RNA genomes, budding from the ER and the process of infecting other target cells. As we will show, the first process shows almost paradoxical dynamics under realistic assumptions which can, however, be resolved at the level of the virion. We then discuss the interplay of viruses with associated 'replicons' (Behrens *et al.* 1998; Lohmann *et al.* 1999; Tautz *et al.* 1999; Kromykh *et al.* 2000), i.e. RNA sequences including only the parts of the genome which control replication but do not code for envelope proteins. Replicons have been created, for example, for BVDV (Behrens *et al.* 1998; Grassmann *et al.* 1999) and HCV (Thiel *et al.* 1998; Lohmann *et al.* 1999). They comprise the genes *NS2* to *5B* and can replicate inside the cytoplasm without forming virus particles.

2. RNA REPLICATION DYNAMICS AND SELECTION

Here we consider not the virus itself but the replication dynamics of its genome in the cytoplasm. In the simplest case of a single master-sequence S_1 with fitness a_1 , and a mutant tail S_2 with fitness $a_2 < a_1$, the concentrations of S_1 and S_2 can be determined from the system of ordinary differential equations:

$$\dot{S}_1 = (a_1 Q - S_1 - S_2)S_1, \quad (2.1)$$

$$\dot{S}_2 = (a_2 - S_1 - S_2)S_2 + a_1(1 - Q)S_1, \quad (2.2)$$

where $S_1(t)$ is the concentration of the sequence S_1 at time t and $Q = q^N$ is the probability for error-free replication given the probability of correct replication of a single nucleotide, q . Mutations from the mutant sequences to the master sequences are at least of order $O(1/N)$ and can therefore be neglected. The total amount of viral RNA, $S_1 + S_2$, is limited to the carrying capacity of the cell, i.e. $S_1 + S_2 = 1$. The system of equations (2.1) and

(2.2) has attracted a lot of attention and we shall only recapitulate the essentials for the following discussion and refer to Swetina & Schuster (1982) and Nowak (1992) for more detailed theoretical discussions.

Because $Q = q^N < 1$, the master sequence can only persist if $a_1 > a_2/Q$ or, given the relative fitnesses, if the error-free replication rate per nucleotide exceeds $q > (a_2/a_1)^{1/N}$ for an N -nucleotide RNA sequence. Because we only consider the cellular dynamics of the RNA quasi-species at this point, we have to reconsider the meaning of fitness of a quasispecies in this context. Generally the fitnesses a_1 and a_2 are understood to reflect the expected number of offspring in the next generation. Equally, for RNA copies the fitness is defined as the average number of sequence-replicates in the next generation; therefore, the replication rate is the fitness of the RNA sequence. In the case of HCV, RNA replication relies on virally encoded proteins: a replicase (5A), an RNA polymerase (5B) and a protease/helicase (3). Mutations in these parts of the genome can leave the copy unable to replicate, while mutations in the *S* part of the genome can occur without compromising the ability of a sequence to be replicated. Therefore, as long as mutations occur outside the crucial *NS* regions of the genome, RNA sequences have identical replication rates. In the following we denote the sequences with fully functional genomes by S_1 , those with functionality in the replicative part of the genome but faults in the *S*-part by S_2 , and those with mutations in the replicative regions by T . In the absence of deletions they are associated with replication rates (fitness) $a_1 = a_2 = a$ and $\alpha \ll a$, respectively. The fidelity of replication is given by $Q_1 = q^N$ for S_1 , $Q_2 = q^M > Q_1$ for S_2 since $M < N$ (as only part of the genome has to be replicated correctly) and unity. A ratio of $(N - M)/N$ mutations in S_1 will lead to sequences of type S_2 , while the rest give rise to mutants of type T . To a first approximation we assume that back-mutation from S_2 or T to S_1 , or from T to S_2 , is rare. For the RNA replicates (and not the virus) the simplest caricature of the dynamics is given by the system of equations (A1)–(A3) in Appendix A.

From the corresponding equilibrium conditions (equations (A4)–(A6) in Appendix A) we see that the master sequence does not survive such a scenario; sequences coding for fit virus particles are bound for extinction, while sequences that retain the ability of replication can persist. This is because at the level of RNA replication, fitness equals replication rate and does not reflect the ability of the virus to infect other cells. Selection of the structural proteins occurs only during the assembly of virions, their intracellular transport and infection of other cells. Because during replication such abilities are not needed, mutations may accumulate in the non-replicative parts of the genome (*C-p⁷* in figure 1).

We will now neglect the sequences which cannot replicate because of mutations in the replicative part of the genome and focus only on viral sequences which code for a functional replication apparatus. A simpler system (equations (A7) and (A8) in Appendix A) shows the same essential dynamics as equations (A1)–(A3); because both sequences have the same replication rate, S_1 (denoting the sequences with functional structural genome) goes extinct while S_2 (which has a corrupted structural genome) may persist as is shown in figure 2.

Figure 2. Frequencies of master sequence (black) and mutant tail sequences (grey) as calculated from equations (2.1) and (2.2) (solid lines) and equations (A7) and (A8) (dashed lines). While in the classical quasispecies picture the master sequence persists for $a_1 > a_2/Q$, its frequency in the cellular RNA pool (dashed black line) decreases due to mutations for all $Q < 1$.

We would briefly like to consider an interesting aspect of our model which relates to the phage experiments by Turner & Chao (1999). What happens if some sequences in the mutant tail can replicate themselves but do not possess the parts of the genome which code for proteins? They may even be able to enter the viral envelope and nucleocapsid, leave the cell and infect others, but not have the ability to produce the S proteins of the virus themselves. Replicons are an example of such particles: they encode the parts of the viral genome which are responsible for RNA replication but do not have the genes necessary for production of structural proteins. In the case of HCV and BVDV (cf. figure 1*a,b*), replicons which do not include the genes labelled $C-p^7$ are available (Behrens *et al.* 1998; Thiel *et al.* 1998; Grassmann *et al.* 1999; Lohmann *et al.* 1999); as they only contain replicative genes, selection acts on all parts of the replicon genome which therefore will be conserved over long periods of time.

If the sequence does not possess part of the viral structural genome, it is shorter and will therefore replicate more quickly than the master sequence with the fully functional structural genes, because RNA replication itself is largely indiscriminating against unfit sequences. Such behaviour is captured by model 2 in Appendix A. There, S_1 is the RNA sequence of the fully functional virus, S_2 denotes the sequences of the mutant tail, while S_3 could be, for example, the genome of HCV or BVDV without parts of the genes $C-p^7$. The resulting particle would have all the replicative abilities of the wild-type but lack the structural proteins. In order to leave the cell it would have to rely on 'hijacking' viral envelopes coded for by wild-type virus. S_4 is the mutant tail of S_3 and will contain sequences with decreased replication ability.

In the formulation of model 2, equations (A11) and (A12), we have made the following assumptions:

- (i) S_2 is able to replicate but does not encode the same functionality as S_1 . Only M (NS) sites have to be replicated precisely for it to retain its replicative abilities; hence $Q_2 = q^M$;

Figure 3. Concentrations of S_1 (black solid line), S_2 (grey solid line), S_3 (black dashed line) and S_4 (grey dashed line) for $r = 0.8$, $Q_1 = 0.90$, $Q_2 = 0.95$, $p = 0.0001$ and $a = 1.05$ under the initial conditions $S_1(0) = 1$ and $S_2(0) = S_3(0) = S_4(0) = 0$ against time. For $a > 1$ deletion mutants S_3 and S_4 persist and form a quasispecies while S_1 and S_2 go extinct.

- (ii) S_3 is created by a simple deletion process which occurs at a rate $p \ll (1 - Q_1)$ and is shorter than S_1 and S_2 ; hence it has a higher replication fidelity $Q_2 > Q_1$, with $Q_2 = q^M$ and $M < N$ the number of nucleotides in S_3 ; due to its shorter length it also gets replicated more quickly, at a rate $a > 1$;
- (iii) S_3 gives rise to its own mutant tail S_4 which is inefficient at replication because of errors in the replicative part of the viral genome;
- (iv) replication defective mutants of S_1 and S_2 are ignored.

In figure 3 we show the resulting concentrations of S_1 , S_2 , S_3 and S_4 for $p = 0.0001$, $Q_1 = 0.9$, $Q_2 = 0.95$ and $a = 1.05$, the fitness of S_3 sequences. While S_1 decreases rapidly, its mutant tail S_2 increases in size before giving way to the shorter sequence S_3 . For $a > 1$ the replicons S_3 and their mutant tail S_4 persist as a quasispecies, while S_1 and S_2 go extinct as shown in figure 3. If S_3 lacks a functional S genome it cannot infect other cells and will ultimately go extinct as the host cell dies. As a result, extraction of viral RNA from infected cells at different times after viral entry should show a decrease in the master sequence, while the fraction of sequences with synonymous and non-synonymous substitutions will increase over time.

Thus we expect sequences to deteriorate through mutations and especially through deletions over time. Under the constraint of a finite carrying capacity the master sequences will, over time, be replaced by their mutants; therefore persistence of fragile species, such as RNA viruses, may not occur on the cellular level. This is in contrast to what is normally expected from the quasispecies model, but follows rigorously from it.

3. VIRION PERSISTENCE AND INFECTION DYNAMICS

If, confined to a strictly cellular level, a fit virus does not stand a chance against its mutant tail, the question arises of how the virus does persist. To address this issue we propose a simple toy model of RNA virus dynamics

Figure 4. Total amount of virus released from the cell: (a) Total amount of released virus versus viral release rate b for fixed replication fidelity $Q = 0.95$ and varying initial concentration of S_1 ($S_1(0) = 0.001$ (squares), 0.1 (triangles), 0.5 (stars) or 1.0 (diamonds)). With decreasing rate concentration the optimal value of the viral release rate decreases as packing of S_1 template RNA into viruses results in decreased lifetime of the S_1 sequence inside the cell. If the initial abundance $S_1(0)$ is larger a longer persistence of S_1 sequences in the cell is possible even if a significant proportion of newly created S_1 sequences is spent on infecting other cells. (b) Total amount of virus leaving the infected cell versus b for fixed $S_1(t) = 0.1$ and varying copying fidelity $Q = 0.5$ (squares), 0.85 (triangles), 0.95 (stars) or 0.99 (diamonds). Increasing Q leads to lower absolute optimal value of b .

which includes two types of fitness: the first kind relates only to the RNA replication inside the cell and we assume that $a_1 = a_2 = 1$. The second fitness tries to capture the viability of an RNA sequence as part of a virus, i.e. the ability to infect other cells.

It is the infection of other cells that determines long-term persistence and which has to be evolutionarily optimized. Let us assume that infection is simply proportional to the number of viruses leaving the cell. We can then write down a new system of differential equations (model 3 in Appendix A), which keeps track of the number of fully functional copies of viral RNA leaving the cell in virions. We only need to consider the case where sequences in the mutant tail are unable to form functional virus and instead remain as templates in the cell body (see model 3 in Appendix A).

In figure 4a we show numerical results for the total amount of fully functional virus to leave for a variety of

initial concentrations of $S_1(0)$ and fixed virus release rate b . The influence of b for varying Q but fixed initial dose of $S_1(0)$ is presented in figure 4b. The optimal value of the viral release (or shedding) rate for fixed Q is found to increase with the initial dose. This makes intuitive sense as the 'head-start' S_1 has over S_2 increases, and more virus may be invested into the spread of the infection to other cells without compromising the short-term presence of S_1 in the cytoplasm. Similarly, the results in figure 4b indicate that with increased error-free replication rate less virus needs to be invested directly into the infection of other cells as fully functional viral RNA templates persist for longer inside infected cell.

In the present context there is a trade-off between survival of viral sequence S_1 in the cell and overall persistence of the virus in the host organism. The latter can only be achieved by releasing functional virus (i.e. diminishing the number of fit RNA copies S_1) and infecting other cells. Initially most cells will be infected by one or a small number of viruses. Therefore small values of b will tend to be preferred. The ensuing dynamics and the related competition between viral strains are described by the well known model of virus dynamics (Bonhoeffer & Nowak 1994; Nowak & Bangham 1996). At this stage, virus which has lost some of its functionality due to faulty RNA replication may be at a disadvantage compared with fully functional virus and selection of structural proteins occurs, even in the absence of an immune response.

4. DISCUSSION

We have focused only on the simplest quasispecies model of viral RNA persistence. Before turning to a discussion of the biological relevance we wish to justify the mathematical economy applied above. In reality, viruses will show a much more complicated fitness landscape than the one used here; moreover, pressure from the immune system varies with time, and both the intra-cellular and intercellular dynamics are more complicated than in the simple caricature of virus dynamics discussed here. Qualitatively, however, the validity of our discussion relies solely on the mechanism of RNA replication: if replication is indiscriminatory then our discussion applies; if replication is more efficient for RNA sequences corresponding to virus with a high fitness, then our discussion is only of mild theoretical interest. The effects of recombination need to be considered separately; in HCV recombination is, however, only of minor importance.

In the present study we have focused primarily on the dynamics of viral RNA and have neglected complicating factors such as pressure from the immune system, cytopathic effects of the virus and finite lifetime of the cell. However, all those potential factors will have a negative effect on the virus and in particular on the frequency of the master sequence S_1 . In the worst case we have established a minimal framework under which viral RNA can persist in the absence of other possible detrimental factors, i.e. it should apply to the population dynamics of viral RNA *in vitro* but not *in vivo*, but given the necessity of studying viruses and antiviral therapies *in vitro*, this is still an important point. The approach is different from the so-called 'defective interfering particles' (DI)

approach of Kirkwood & Bangham (1994). DIs show great similarity to virions but cannot replicate on their own. Due to their outward similarity to the virus they can have a profound effect on disease progression and have attracted much attention in connection with the boosting of antiviral immune response. The phenomenon we describe can be thought of as the pendant to DI dynamics: DIs can infect but not replicate, whereas we consider dynamics of replicon-like objects which can replicate RNA inside a cell but lose the ability to infect other cells.

We have argued that the viability of a quasispecies cannot depend on the cellular replication of RNA sequences. Provided that deleterious mutations do not occur in the replicative part of the viral genome it is difficult to argue that the same replication mechanism can preferably replicate a subset of sequences with identical nucleotide numbers N . Therefore, there is good reason to believe that pestiviruses, hepaciviruses and perhaps other RNA viruses are indeed plagued by the dilemma of progressing dilution after repeated replication: over time, the quality of viral RNA inside cells deteriorates to the point where non-replicative parts of the genome are rendered useless. The structural part of the RNA indeed encodes the parts of the genome where the hypervariable regions in HCV occur. It is only if virus particles leave the cell that competition and selection are reinvoked for S genes. RNA which manages to become part of a functional virus particle leaves the cell, and it is in the ensuing intercellular dynamics where it undergoes further selection even in the absence of an immune response. As a result, HCV and BVDV may not satisfy the classical quasispecies model. In its traditional definition (Eigen & Schuster 1977; Nowak 1992) a stable RNA population is the result of continuous selection acting on all parts of the genome during replication; the situation of HCV and BVDV, however, is quite different from this ideal. During the replicative phase of the viral life cycle only part of the genome is under selection, while particular structural regions of the genome are 'quasi-neutral'. It is only in the ensuing dynamics where the S genome (or parts of it) encounters selective forces. Left inside a cell, however, we expect HCV and BVDV to show largely random polymorphisms throughout their S genomes.

A direct test of our predictions would be to measure the intracellular viral RNA content. Such studies could make use of, for example, the non-cytopathic virus strain of BVDV, because with the cytopathic viruses cell death may occur before RNA sequences have significantly degraded intracellularly. The present model predicts that the composition of RNA sequences inside infected cells will barely reflect the sequences found in free virus. The number of highly polymorphic sites will increase in RNA sequences within cells, but only a fraction of these can enter the viral envelope. In particular, the ratio of non-synonymous to synonymous substitutions ought to be lower inside cells than outside cells. Incidentally, one might also expect the diversification of RNA sequences to be reflected at the virus level in a more heterogeneous virus population structure, as is indeed observed (Martell *et al.* 1992; Sullivan *et al.* 1998); at least for parts of the S genes $C-p^7$ this may occur by random genetic drift and does not necessarily reflect a quasispecies structure of the viral

RNA population. Changes in the envelope region of HCV, like those observed by Farci *et al.* (2000), can occur irrespective of immune system pressure. There is, however, the chance that a mutation in the structural part of the genome (allowing unimpeded replication) may be frequency-selected when the immune system is unable to recognize it. This may then lead to the observed changes in the abundance of viral mutants in infected host cells (Farci *et al.* 2000).

The results of this investigation suggest that extended confinement of RNA sequences to the cell body may open up a new route towards antiviral drug therapy. The RNA replication dynamics occur in a condition where structural parts of the genome do not undergo selection. If release of virus particles from infected cells could be artificially delayed for a sufficiently large number of RNA-replication periods, then the frequency of the master sequences S_1 will gradually decrease. Such a potential drug would effectively relieve the viral genome from selection pressure, eventually rendering the virus ineffective. For BVDV, for example, there is evidence that imino sugar derivatives can inhibit the formation and secretion of functional virus (Zitzmann *et al.* 1999). It is argued that the drug inhibits virus release by disrupting the folding of the viral envelope proteins by preventing interactions with calnexin, a chaperone in the ER. Therefore, viral proliferation is terminated at the final stage of assembly while replication of viral RNA seems unaffected. Given our simple mathematical treatment presented here, it is tempting to speculate that with each viral generation trapped inside the cell we see the accumulation of more mutations and a reduction in the frequency of S_1 in agreement with the idea behind Muller's ratchet (Chao 1990; Duarte *et al.* 1992). Although numerous studies have shown how selection pressure can lead to drug resistance of parasite populations, this is, to our knowledge, the first study to propose that delaying or inhibiting selection may have a beneficial therapeutic effect.

We thank R. Dwek, E. Holmes, P. Klenermann, H. Korthals-Altes and R. M. May for helpful comments on the manuscript. Support from the Wellcome Trust (M.P.H.S.), The Royal Society (N.Z.) and Linacre College, Oxford (M.P.H.S. and N.Z.), through fellowships is gratefully acknowledged.

6. APPENDIX A: THE MODELS

(a) *Model 1*

This model describes the abundance of a master sequence S_1 , a set of replication-capable mutants S_2 which may show defects in the non-replicative parts of the genome, and replication-incapable mutants T .

$$\dot{S}_1 = (aQ_1 - S_1 - S_2 - T)S_1, \quad (A1)$$

$$\dot{S}_2 = (aQ_2 - S_1 - S_2 - T)S_2 + \frac{a(N-M)}{N}(1-Q_1)S_1, \quad (A2)$$

$$\dot{T} = (\alpha - S_1 - S_2 - T)T + \frac{aM}{N}(1-Q_1)S_1 + a(1-Q_2)S_2. \quad (A3)$$

The equilibrium conditions are determined by setting the left-hand side of equations (A1)–(A3) equal to zero. Because $Q_2 > Q_1$ and $a \gg \alpha$ we find that the only stable equilibrium is given by

$$S_1 = 0, \quad (\text{A4})$$

$$S_2 = aQ_2 - \alpha, \quad (\text{A5})$$

$$T = a(1 - Q_2). \quad (\text{A6})$$

We can simplify the system of equations (A1) and (A2) but still retain the same essential RNA replication kinetics by ignoring those sequences which do not have a functional replicative genome:

$$\dot{S}_1 = (aQ_1 - S_1 - S_2)S_1, \quad (\text{A7})$$

$$\dot{S}_2 = (a - S_1 - S_2)S_2 + (1 - Q_1)S_1. \quad (\text{A8})$$

(b) Model 2

A model describing the interplay between master-sequence S_1 and its mutants S_2 on the one hand, and a sequence which is shorter but includes the uncorrupted replicative genome S_3 and its mutants S_4 , or the other, is given by

$$\dot{S}_1 = (Q_1 - S_1 - S_2 - S_3 - S_4)S_1, \quad (\text{A9})$$

$$\dot{S}_2 = (Q_2 - S_1 - S_2 - S_3 - S_4)S_2 + (1 - Q_1)S_1, \quad (\text{A10})$$

$$\dot{S}_3 = (aQ_2 - S_1 - S_2 - S_3 - S_4)S_1 + p(S_1 + S_2), \quad (\text{A11})$$

$$\dot{S}_4 = (r - S_1 - S_2 - S_3 - S_4)S_4 + (1 - Q_2)S_3. \quad (\text{A12})$$

S_3 could, for example, be the genome of HCV or BVDV without parts of the genes C – p^7 . The resulting sequence would have all the replicative abilities of the wild-type but lack the structural proteins.

(c) Model 3

Replacing equations (A1) and (A2) for the concentrations of S_1 and S_2 we can write a new system of equations that begins to take account of the increased ability of the master sequence to infect other cells:

$$\dot{S}_1 = (Q - S_1 - S_2)S_1 - bS_1, \quad (\text{A13})$$

$$\dot{S}_2 = (1 - S_1 - S_2)S_2 + a(1 - Q)S_1. \quad (\text{A14})$$

Here only the virus characterized by sequence S_1 is able to leave the cell, with a rate constant b , and infect other cells. If infection of new cells is proportional to the number of viruses with RNA sequence S_1 leaving the cell, then the integral $\int_0^\infty bS_1(t)dt$ is proportional to Fisher's basic reproductive ratio R_0 .

REFERENCES

Behrens, S., Grassmann, C., Thiel, H., Meyers, G. & Tautz, N. 1998 Characterization of an autonomous subgenomic pestivirus RNA replicon. *J. Virol.* **72**, 2364–2372.

Bonhoeffer, S. & Nowak, M. 1994 Mutation and the evolution of virulence. *Proc. R. Soc. Lond.* **B258**, 133–140.

Chao, L. 1990 Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**, 454–455.

Chichutek, K., Merget, H., Norley, S., Linde, R., Kreuz, W. & Kurth, R. 1992 Development of a quasispecies of human immunodeficiency virus type 1 *in vivo*. *Proc. Natl Acad. Sci. USA* **89**, 7365–7369.

Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. 1978 Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**, 735–744.

Domingo, E., Escarnis, C., Menéndez-Arias, L. & Holland, J. 1999 Viral quasispecies and fitness variations. In *Origin and Evolution of Viruses* (ed. E. Domingo, R. Webster & J. Holland), chapter 7, pp. 141–161. San Diego: Academic Press.

Duarte, E., Clarke, D., Moya, A., Domingo, E. & Holland, J. 1992 Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proc. Natl Acad. Sci. USA* **89**, 6015–6019.

Eigen, M. & Schuster, P. 1977 The hypercycle. *Naturwissenschaften* **64**, 541–565.

Farci, P. (and 11 others) 2000 The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339–344.

Grassmann, C., Isken, O. & Behrens, S. 1999 Assignment of the multifunctional ns3 protein of bovine viral diarrhea virus during RNA replication: an *in vivo* and *in vitro* study. *J. Virol.* **73**, 9196–9205.

Holland, J., de le Torre, J. & Steinhauer, D. 1992 RNA virus populations as quasispecies. *Curr. Topics Microbiol. Immunol.* **176**, 1–20.

Huynen, M., Stadler, P. & Fontana, W. 1996 Smoothness within ruggedness: the role of neutrality in adaption. *Proc. Natl Acad. Sci. USA* **93**, 397–401.

Karlsson, A., Lindbäck, S., Gaines, H. & Sönnernborg, A. 1998 Characterization of the viral population during primary HIV-1 infection. *AIDS* **12**, 839–847.

Kirkwood, T. & Bangham, C. 1994 Cycles, chaos and evolution in virus cultures: a model of defective interfering particles. *Proc. Natl Acad. Sci. USA* **91**, 8685–8689.

Kromykh, A., Sedlak, P. & Westaway, E. 2000 Cis- and trans-acting elements in flavivirus RNA replication. *J. Virol.* **74**, 3253–3263.

Kurosaki, M., Enomoto, N., Marumo, F. & Sato, C. 1994 Evolution and selection of hepatitis C virus variants in patients with chronic hepatitis C. *Virology* **205**, 161–169.

Liu, Q., Tackney, C., Bhat, R., Prince, A. & Zhang, P. 1997 Regulated processing of hepatitis C virus core protein is linked to subcellular localization. *J. Virol.* **71**, 657–662.

Lohmann, V., Koch, J. & Bartenschläger, R. 1996 Processing pathways of the hepatitis C virus proteins. *J. Hepatol.* **24**, 11–19.

Lohmann, V., Körner, F., Koch, J., Herian, U., Theilmann, L. & Bartenschläger, R. 1999 Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**, 110–113.

Martell, M., Esteban, J., Quer, J., Genescà, J., Weiner, A., Esteban, R., Guardia, J. & Gomez, J. 1992 Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J. Virol.* **66**, 3225–3229.

Nowak, M. 1992 What is a quasispecies? *Trends Ecol. Evol.* **7**, 118–121.

Nowak, M. & Bangham, C. 1996 Population dynamics of immune response to persistent viruses. *Science* **272**, 74–79.

Nowak, M., May, R. & Anderson, R. 1990 The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. *AIDS* **4**, 1095–1103.

Schuster, P. 1997 Genotypes with phenotypes: adventures in an RNA toy world. *Biophys. Chem.* **66**, 75–110.

- Schuster, P. & Fontana, W. 1999 Chance and necessity in evolution: lessons from RNA. *Physica D* **133**, 427–452.
- Sullivan, D., Wilson, J., Carithers, R., Perkins, J. & Gretch, D. 1998 Multigene tracking of hepatitis C virus quasispecies after liver transplantation: correlation of genetic diversification in the envelope region with asymptomatic or mild disease patterns. *J. Virol.* **72**, 10 036–10 043.
- Swetina, J. & Schuster, P. 1982 Self-replication with errors: a model for polynucleotide replication. *Biophys. Chem.* **16**, 329–345.
- Tautz, N., Harada, T., Kaiser, A., Rinck, G., Behrens, S. E. & Thiel, H. J. 1999 Establishment and characterization of cytopathogenic and noncytopathogenic pestivirus replicons. *J. Virol.* **73**, 9422–9432.
- Thiel, V., Siffell, S. & Herold, J. 1998 Replication and transcription of HCV 229e replicons. *Adv. Exp. Med. Biol.* **440**, 109–113.
- Turner, P. & Chao, L. 1999 Prisoner's dilemma in an RNA virus. *Nature* **398**, 441–443.
- Zitzmann, N., Mehta, A. S., Carrouée, S., Butters, T. D., Platt, F. M., McCauley, J., Blumberg, B. S., Dwek, R. A. & Block, T. M. 1999 Imino sugars inhibit the formation and secretion of bovine viral diarrhea virus, a pestivirus model of hepatitis C virus: implications for the development of broad spectrum anti-hepatitis virus agents. *Proc. Natl Acad. Sci. USA* **96**, 11 878–11 882.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.