Selective advantage for conservative viruses

Yisroel Brumer and Eugene I. Shakhnovich
Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA
(Received 16 January 2004; revised manuscript received 23 December 2004; published 8 March 2005)

In this article we study the full semiconservative treatment of a model for the coevolution of a virus and an adaptive immune system. Regions of viability are calculated for both conservatively and semiconservatively replicating viruses interacting with a realistic semiconservatively replicating immune system. The conservative virus is found to have a selective advantage in the form of an ability to survive in regions with a wider range of mutation rates than its semiconservative counterpart, as well as an increased replication rate where both species can survive. This may help explain the existence of a rich range of viruses with conservatively replicating genomes, a trait that is found nowhere else in nature.

DOI: 10.1103/PhysRevE.71.031903 PACS number(s): 87.14.Gg, 87.10.+e, 87.23.Kg

Deoxyribonucleic acid (DNA) is often called the molecule of life. The vast majority of organisms in nature store their genetic information in the form of double stranded DNA, which provides a number of benefits over its close relative and likely predecessor, ribonucleic acid (RNA) [1]. These benefits include a predictable secondary structure and a resistance to autocatalytic cleavage and hence a longer half-life. However, viruses stand out as a notable exception to the "DNA world," employing a variety of methods for genetic storage including single stranded RNA (e.g., tobacco mosaic virus), double stranded RNA (such as reovirus), linear and circular single stranded DNA (including parvovirus and bacteriophage ϕX174, respectively), and a variety of double stranded DNA types (examples of which include bacteriophage T4, polyoma virus, and poxvirus). Double stranded DNA replicates semiconservatively, wherein a double stranded genome is unzipped and the single strands are individually replicated to produce new complementary strands. Many viruses, on the other hand, replicate conservatively, wherein multiple copies of a single strand are made, and the original strand is conserved. This suggests that either the benefits of DNA, such as the facilitation of proofreading, must be less for viral species (possibly due to a shorter genome length or unusual life cycles) and/or a selective advantage must exist for conservatively replicating viruses, such as lowering the cost of replication. In this paper we use the quasispecies model to demonstrate such an advantage for a conservative virus coevolving with an adaptive immune system. Although this does not explain the prevalence of conservative viruses infecting simple organisms such as bacteria, it does shed light on the interaction between viruses and complex organisms, as well as the constant battle waged by an immune system against disease. Lastly, the model can be easily adapted to represent a rapidly changing environment, independent of the existence of an adaptive immune system.

While viruses are highly diverse in nature and architecture, viral replication can generally be thought of as one step in the infectious cycle, usually characterized as host cell entry, genome replication, and transcription/translation (sometimes accompanied by integrating into the host genome), assembly of progeny virus, and host cell exit. From a biochemical perspective, viral replication is a complex, virus-specific, matter, incorporating numerous enzymes produced by the cell or coded for by the virus and we refer the interested reader to the vast literature on the subject ranging from general overviews [1,2] to detailed work on aspects such as strand asymmetry in RNA viral replication [3]. The biochemistry of the immune system represents an equally complex field [4]. It is fascinating that, despite the complexities of these processes, the quasispecies model, described below, has been very successful in providing detailed, experimentally verified, predictions on the nature and behavior of these systems.

Introduced by Eigen in 1971, the quasispecies model [5,6] has been used to study various characteristics of conservatively replicating organisms ranging from equilibrium data to punctuated evolution [7–17]. The model consists of a population of independently replicating genomes ϕ, each of which is made up of a set of "letters" $s_1 s_2 \cdots s_n$ chosen from an "alphabet" of size S. S is usually chosen to be 2 for simplicity or 4, as in this paper, to model the nucleotides found in nature, adenine, cytosine, thymine/uracil, and guanine (denoted A, C, T, and G, respectively). Each possible genome is assigned a fitness that dictates its fecundity. This mapping of fitness to genotype can be represented by a unique "fitness landscape." The process of replication includes a probability of point mutation per base pair ϵ that is generally assumed to be genome independent. By associating phenotype with genotype and assuming a first-order dependence of the growth rate on concentration, a set of differential equations can be solved to describe the competition between various genotypes [5,6]. Although the model incorporates numerous approximations, it is well suited to describing small RNA genomes and viruses and many of its predictions have been experimentally verified. One of the major successes of the model lies in recent work on novel antiviral therapies [18,19].

The quasispecies model has recently been extended to the coevolution of hosts and parasites and the particular case of an adaptive immune system interacting with a virus [20,21]. Viruses make detrimental use of host biochemical processes while the immune system expends enormous effort to keep viral concentrations as low as possible. As the immune system develops new defenses, the virus must adapt to defeat them. The immune system must then evolve to destroy the newly resistant strains, and a nonlinear coevolving feedback

loop is created. To model this behavior, the immune system and virus are both assumed to evolve on a single fitness peak landscape, where the fitness of all genomes is equal with the sole exception of a single master sequence of far greater fitness, or

$$A(\phi) = \begin{cases} \eta, & \phi \neq \phi_0, \\ \sigma \gg \eta, & \phi = \phi_0, \end{cases} \tag{1}$$

where $A(\phi)$ represents the fitness of genome ϕ and $\{\sigma, \eta\} \equiv \{\sigma_v, \eta_v\}$ for the virus and $\{\sigma, \eta\} \equiv \{\sigma_{is}, \eta_{is}\}$ for the immune system. This landscape is dynamic in that the fitness peak is allowed to shift from one genome to another at specified intervals. To model the interspecies interaction, the master sequence for the immune system is assumed to impose a death rate δ on the corresponding viral sequence. When this coincides with the viral master sequence, the viral fitness peak shifts out of self-preservation. The new viral master sequence regenerates on a time scale τ_v, defined as the time required for the new master sequence to outnumber the old. At this point, the immune fitness peak shifts to match the viral master sequence and will regenerate on a similarly defined time scale τ_{is} after which the viral peak shifts again. These steps are iterated so that the virus traverses genome space with the immune system in hot pursuit. Using recent results on the dynamics of a quasispecies on time-dependent landscapes [7], Kamp and Bornholdt [20,21] found expressions for the long-term survival of a *conservatively* replicating virus and immune system [20,21] by considering the behavior of each species on a dynamic fitness landscape. In essence, every time the fitness peak shifts, the concentration of the master sequence drops dramatically and begins to regrow, while the large concentration of the old master sequence drains away. Rigorously, the dynamics of a quasispecies on a single fitness peak landscape follow a set of differential equations defining the evolution of the various genomes in terms of their Hamming distance $HD(\phi, \phi')$. This is defined as the smallest number of point mutations that lead from ϕ to ϕ'. Utilizing the radial symmetry of the fitness landscape, the quasispecies equations become

$$\frac{dw_{l,x}}{dt} = \sum_{l'=0}^{l} \frac{(n_x - l')!}{(n_x - l)!} A_x(l')(\epsilon_x)^{l-l'}(1-\epsilon_x)^{n_x-(l-l')} w_{l',x} - f_x(t) w_{l,x}, \tag{2}$$

where $x \in \{v, is\}$, w_l is the concentration of all sequences of Hamming distance l from the master sequence, $f_{is}(t) = \Sigma_l A_{is}(l) w_{l,is} = (\sigma_{is} - \eta_{is}) w_{0,is} + \eta_{is}$ for the immune system, $f_v(t) = \delta$ for the viral sequence that coincides with the immune master sequence, and $f_v(t) = 0$ otherwise. ϵ_x represents the point mutation probability (more complex mutations such as insertions and deletions, as well as the possibility of recombination, are ignored).

Building on recent work regarding the dynamics of a quasispecies on mobile fitness landscapes [7], a number of qualitative features have come to light. Besides the standard (albeit modified by the dynamic nature of the landscape) error catastrophe at high ϵ_v, the virus is only viable above a given minimum mutation rate, below which it is unable to keep up with the moving landscape (dubbed the "adaptability" catastrophe). At these low mutation rates, each shift of the landscape is followed by a period of time wherein the new master sequence attempts to build up an equilibrium distribution. However, before the new master sequence can rebuild to the levels reached by the old master before the shift, the fitness peak moves again. As this process repeats, any initial quasispecies will disappear and genomes become stochastically distributed. Further, an optimal mutation rate for the immune system can be found that minimizes the range of viral mutation rates that allow persistence of the viral quasispecies. This optimal rate has been determined and found to be independent of the parameters of the model and the properties of the viral species as well as comparing admirably with the rates of somatic hypermutation in *B* cells [20]. The agreement suggests that this model, although approximate in nature, captures the robust features of the coevolution of viruses with an adaptive immune system.

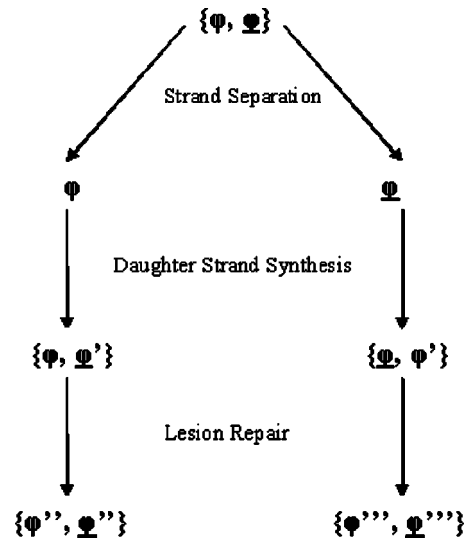

FIG. 1. A schematic model of DNA replication. Adapted from Tannenbaum *et al.* [22].

Although the host-parasite model has yielded impressive successes, work has been restricted to conservatively replicating systems, which differ greatly from the true semiconservative systems that dominate nature. In a conservative system, multiple, possibly error-prone, copies of an original strand are produced without harming or changing the original strand. Thus, the original quasispecies model is ideally suited for the study of RNA viruses or *in vitro* RNA evolution experiments. Semiconservative replication follows a different route shown schematically in Fig. 1. DNA exists as a tightly bound double helix structure, where each strand ϕ is connected to a complementary strand $\underline{\phi}$, where $\underline{\phi}$ represents the complement of strand ϕ, and can be written as $\underline{s}_1 \underline{s}_2 \cdots \underline{s}_n$ where we assign the nucleotides $A \equiv 1$, $G \equiv 2$, $T \equiv 3$, $C \equiv 4$, and $\underline{s}_i = (s_i + S/2) \bmod S$. In order to undergo replication, the double helix unzips to free two single strands ϕ and $\underline{\phi}$. Each of these is replicated to produce an error-prone complement, yielding $\{\phi, \underline{\phi}'\}$ and $\{\underline{\phi}, \phi'\}$. Each error must result in a base mismatch, which can be recognized and selectively repaired by enzymes in the cell. These enzymes can recognize the

new strand through the methylation process and ensure that the mismatch is repaired by replacing the new mismatched base, keeping the effective error rate ϵ low. This amounts to defining the error rate as the number of errors that occur per replication *after* the methyl-directed mismatch repair and methylation processes are complete. This sidesteps the complexities associated with the methylation process. In the final stage, the strands become indistinguishable and various maintenance enzymes repair the remaining errors with a 50% probability of correcting the mismatch in either strand. Thus, the final result is two new pairs, each consisting of two new strands, $\{\phi'', \phi''\}$ and $\{\phi''', \phi'''\}$. Each new genome is destroyed when it replicates, thus yielding a reproductive system with significantly different dynamics, even for viruses with large burst sizes. Recent work by Tannenbaum *et al.* [22] has extended the quasispecies model to the case of semiconservative replication, which was found to display significantly different behavior in the infinite time limit on a static landscape. Recently, the semiconservative quasispecies equations were combined with the coevolution model [23]. The behavior of the semiconservative adaptability catastrophe was investigated and a plausible explanation for the lower mutation rates found in semiconservative viruses was posited. Here, we further apply this combination to demonstrate that conservative viruses enjoy a significant selective advantage over their semiconservative counterparts when battling an adaptive immune system.

To properly treat a semiconservative quasispecies model on a single fitness landscape, ignoring back mutations, Eq. (2) must be recast as [22]

$$\frac{dw_{l,x}}{dt} = 2\sum_{l'=0}^{l} A_x(l')(\epsilon_x/2)^{l-l'}\left(1-\frac{\epsilon_x}{2}\right)^{n_x-l-l'} w_{l',x} - [A_x(l)+f(t)]w_{l,x}, \tag{3}$$

where $f(t)$ is defined above. Here, we examine the dynamics of the semiconservative equations within the confines of Kamp and Bornholdt's model of coevolution. This study will focus on the dynamical aspects of Eq. (3), as opposed to the equilibrium effects studied by Tannenbaum *et al.* [22]. Following Kamp and Bornholdt, the condition for the viability of the quasispecies is

$$\kappa_x \equiv \frac{w_{1,x}(\tau)}{n_x(S-1)e^{\eta_x\tau}w_{0,x}(0)} \geqslant 1, \tag{4}$$

where κ represents the ratio of master sequence concentrations at the beginning and end of an entire cycle of landscape shifts in an unconstrained system compared to the equivalent growth of a random sequence far from the peak. Note that Eq. (4) differs from Kamp and Bornholdt's result by a factor of $(S-1)n_x$, as $w_{1,x}$ includes the $n_x(S-1)$ possible sequences that differ from the master sequence by one error, as opposed to only a single such sequence. Obviously, if the master sequence outgrows the random sequence over this period, even with the concentration losses incurred by the peak shift, it will survive for all times. If the master sequence is outgrown by the random sequence, i.e., if $\kappa<1$, the master sequence will diminish and disappear at long times.

After a fair bit of work, the condition for viability of the immune genome can be expressed as [23]

$$\kappa_{is} = \left(\frac{\sigma_{is}\epsilon_{is}(1-\epsilon_{is}/2)^{n_{is}-1}}{(S-1)(\sigma_{is}-\eta_{is})[2(1-\epsilon_{is}/2)^{n_{is}}-1]}\right) \times (e^{[2\sigma_{is}(1-\epsilon_{is}/2)^{n_{is}}-\sigma_{is}-\eta_{is}]\tau} - e^{[2\eta_{is}(1-\epsilon_{is}/2)^{n_{is}}-2\eta_{is}]\tau}) \geqslant 1, \tag{5}$$

$$\tau = \tau_{is} + \tau_v, \tag{6}$$

$$\tau_{is} = -\frac{\ln\left(\frac{(1-\epsilon_{is}/2)^{n_{is}}\epsilon_{is}}{[2(1-\epsilon_{is}/2)^{n_{is}}-1](S-1)}\right)}{[2(1-\epsilon_{is}/2)^{n_{is}}-1](\sigma_{is}-\eta_{is})}, \tag{7}$$

and, for a semiconservative viral species,

$$\kappa_v = \left(\frac{\sigma_v\epsilon_v(1-\epsilon_v/2)^{n_v-1}}{(S-1)(\sigma_v-\eta_v)[2(1-\epsilon_v/2)^{n_v}-1]}\right) \times (e^{[2\sigma_v(1-\epsilon_v/2)^{n_v}-\sigma_v-\eta_v]\tau} - e^{[2\eta_v(1-\epsilon_v/2)^{n_v}-2\eta_v]\tau}) \geqslant 1, \tag{8}$$

$$\tau_v = -\frac{\ln\left(\frac{(1-\epsilon_v/2)^{n_v}\epsilon_v}{[2(1-\epsilon_v/2)^{n_v}-1](S-1)}\right)}{[2(1-\epsilon_v/2)^{n_v}-1](\sigma_v-\eta_v)+\delta}. \tag{9}$$

A conservatively replicating virus interacting with a semiconservative immune system will follow the behavior described by the conservative model of Kamp and Bornholdt,

$$\kappa_v = \left(\frac{(e^{(q_v^n\sigma_v-\eta_v)\tau} - e^{(q_v^{n_v}\eta_v-\eta_v)\tau})(1-q_v)\sigma_v}{(S-1)(\sigma_v-\eta_v)q_v}\right), \tag{10}$$

where $q=1-\epsilon$ represents the replicative fidelity per base pair, $\tau=\tau_{is}+\tau_v$ as always, and

$$\tau_v = -\frac{\ln[(1-q_v)/(S-1)]}{q_v^n(\sigma_v-\eta_v)+\delta}. \tag{11}$$

τ_{is} is described by the semiconservative result described by Eq. (7). In Fig. 2 we plot the regions of viability defined by $\kappa_{v,is}\geqslant 1$ with a particular set of parameters for a semiconservative immune receptor interacting with both a conservative and a semiconservative virus. It is immediately clear that the conservative virus can survive under a wider range of conditions than the semiconservative. While the immune system is slightly less robust for the semiconservative virus, this effect is small and lies near the region where the immune system becomes unviable, and hence should have little effect on real systems. Furthermore, in the regions where both species are viable, the conservative virus generally possesses higher values of κ_v [23]. Although this behavior is dependent on the parameters of the model, the qualitative trend was robust for the vast majority of biologically reasonable parameter choices. Further, the increased viability lies both near and far from the optimal immune mutation rate, important since a population of viruses in nature is expected to interact with a range of immune system properties, and the behavior away from the optimum should play an important

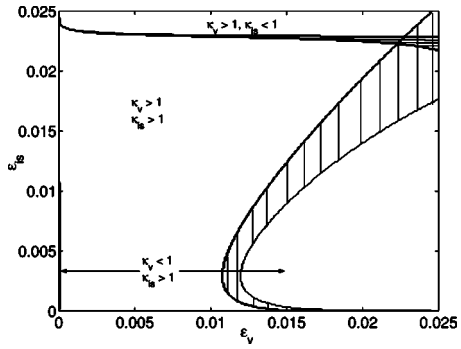

FIG. 2. Regions of viability for a coevolving host-parasite system. The contours for both semiconservative and conservative viruses interacting with a conservative immune system are shown. The vertical lines represent the region where only the conservative virus is viable, while the horizontal lines represent the region where the immune system is only viable for a conservative virus. $n_{is}=n_v=50$, $\sigma_{is}=\sigma_v=100$, $\eta_{is}=\eta_v=1$, $\delta=200$.

role in the evolution of the system. Hence, for any given value of σ, the conservative model generally yields higher values of κ, which measures the viability of the ever-changing fittest sequence on the dynamic landscape. Thus, for the region where both conservative and semiconservative viruses can survive, the conservative virus generally proliferates more rapidly, and a region exists, the vertically hatched region in Fig. 2, where only the conservative virus can survive. This does not mean that any conservative virus will outreplicate any semiconservative. However, it does mean that, for any set of parameters, the conservative virus should outreplicate the semiconservative virus with the same parameters. The key point is that, while other factors, such as the varying costs of replication, may lead to different values for σ in nature (with conservative viruses expected to replicate quicker), we have demonstrated that this fails to tell the whole story. The additional capability to more effectively cope with a dynamic landscape demonstrates a further and fundamental aspect of the dynamics bestowing an additional selective advantage on conservative viruses.

We conclude with a cautionary note. One must always take great care in extracting grand predictions from simplified models such as this one. In nature, a myriad of evolutionary pressures battle for dominance and it is often difficult to pinpoint the selective advantage for a given trait. The relative merits of conservative replication are numerous, ranging from structural aspects to the cost of replication. However, the model presented here likely captures many robust features of the evolution, and suggests a possible explanation for the success of conservative viruses in nature, as well as demonstrating an interesting aspect of host-parasite coevolution.

The authors gratefully acknowledge useful discussions with Emmanuel Tannenbaum and Brian Dominy. This research was supported by NIH.

[1] D. Voet and J. Voet, *Biochemistry* (John Wiley and Sons, New York, 1995).
[2] D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, and M. A. Martin, *Fundamental Virology* (Lippincott, Williams and Wilkins, London, 2001).
[3] L. Marsh, C. Huntley, G. P. Pogue, J. Connell, and T. Hall, Virology **182**, 76 (1991).
[4] D. Goldsby, R. Goldsby, and T. Kindt, *Immunology* (W. H. Freeman Company, New York, 2002).
[5] M. Eigen, Naturwiss. **58**, 465 (1971).
[6] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[7] M. Nilsson and N. Snoad, Phys. Rev. Lett. **84**, 191 (2000).
[8] B. Drossel, Adv. Phys. **50**, 209 (2001).
[9] C. O. Wilke, C. Ronnewinkel, and T. Martinetz, Phys. Rep. **349**, 395 (2001).
[10] P. Tarazona, Phys. Rev. A **45**, 6038 (1992).
[11] S. Galluccio, Phys. Rev. E **56**, 4526 (1997).
[12] P. Schuster and P. F. Stadler, Comput. Chem. (Oxford) **18**, 295 (1994).
[13] L. Peliti, e-print cond-mat/9712027.
[14] S. Altmeyer and J. McCaskill, Phys. Rev. Lett. **86**, 5819 (2001).
[15] P. R. A. Campos and J. F. Fontanari, Phys. Rev. E **58**, 2664 (1998).
[16] D. Alves and J. F. Fontanari, Phys. Rev. E **57**, 7008 (1998).
[17] J. Krug and C. Karl, Physica A **318**, 137 (2003).
[18] S. Crotty, C. E. Cameron, and R. Andino, Proc. Natl. Acad. Sci. U.S.A. **98**, 6895 (2001).
[19] L. A. Loeb, J. M. Essigmann, F. Kazazi, J. Zhang, K. D. Rose, and J. I. Mullins, Proc. Natl. Acad. Sci. U.S.A. **96**, 1492 (1999).
[20] C. Kamp and S. Bornholdt, Phys. Rev. Lett. **88**, 068104 (2002).
[21] C. Kamp, C. O. Wilke, C. Adami, and S. Bornholdt, Complexity **8**, 28 (2003).
[22] E. Tannenbaum, E. J. Deeds, and E. I. Shakhnovich, Phys. Rev. E **69**, 061916 (2004).
[23] Y. Brumer and E. Shakhnovich, Phys. Rev. E **69**, 061909 (2004).